Two-Tier Cache-Aided Full-Duplex Hybrid Satellite–Terrestrial Communication Systems

Quynh T. Ngo, Khoa T. Phan, Wei Xiang, Abdun Mahmood, and Jill Slay

Abstract

Enabling global Internet access is challenging for the Internet of Things (IoT) attributed to the limited range of terrestrial network services. One viable solution is to deploy IoT over satellite systems for coverage extension. However, operating a hybrid satellite-terrestrial IoT system incurs high satellite bandwidth consumption and excessive service latency. This work aims to reduce content delivery delays from the Internet-connected gateway to the remote users, using a two-tier cache-enabled model with full-duplex wireless transmissions where content caches are present at both the satellite and the ground station. A closed-form solution to the successful delivery probability of content files within allocated time-slot under the general caching policy is derived considering the requested content distributions and channel statistics. From the derived results, the successful delivery probability performance under common caching policies can be conveniently obtained. The results are also used to optimize cache placement under caching capacity constraints. Simulation results are presented to demonstrate a superior performance of the proposed system over those of single-tier cache-aided as well as half-duplex transmission systems.

Index Terms

caching, full-duplex, half-duplex, satellite IoT, successful delivery probability

I. INTRODUCTION

Although being around for many years, cellular-based technologies for the Internet of Things (IoT) have not fulfilled the demand of globally connecting everyone and everything. To offer
service across all geographic regions, integrating satellites into IoT networks has been touted. With better service reliability and coverage, satellite communications can provide more effective solutions for IoT than cellular networks. For instance, satellites can provide more than 99.9% availability with proper constellation arrangements [1] and can cover an enormous area of the Earth that is suitable for bringing connectivity to rural or remote areas. Reflecting the utilization of satellite systems as a promising solution complimenting terrestrial networks, many projects have been developed such as the new radio supporting integrated satellite–terrestrial networks by 3GPP [2], the Satellite and Terrestrial Network for 5G (SaT5G) [3], SpaceX’s Starlink satellite project [4].

Operating a hybrid satellite–terrestrial IoT system, however, poses challenges in service delays and satellite bandwidth costs. A content delivered to end-users from the Internet-connected gateway is relayed through satellite(s) and ground station(s), which extends the serving time in addition to very pricey and often limited satellite bandwidth. Excessive content delivery delay can hinder potential use of satellite IoT systems to support delay-sensitive applications. To address these challenges, caching has been proposed as a mechanism to relocate content storage to edge devices that are closer to users [5]–[12]. In general wireless networks, edge caching has been shown to shorten delivery latency and reduce network congestion [13]. In addition, caching contents in satellites or terrestrial stations helps avoid re-transmissions of the same contents and significantly reducing in-network traffic [14].

In this paper, we study a two-tier cache-aided satellite IoT model considering end-to-end communications (i.e., from Internet-connected gateway via satellite and ground station to end-users). Our work focuses on the content delivery time analysis in terms of the successful delivery probability (SDP) considering full-duplex (FD) transmissions at the satellite and ground station, albeit with imperfect self-interference cancellation. Deploying FD communications will potentially shorten the service delivery time compared to the half-duplex (HD) mode providing sufficiently effective self-interference cancellation, and hence increasing the SDP. The main contributions of this paper are summarized as follows.

1) A closed-form SDP expression for the two-tier cache-enabled satellite IoT systems is derived under the FD mode taking into account the requested content characteristics, satellite/terrestrial realistic channel statistics, and caching configurations. Consequently, three common caching policies, namely uniform caching, ground station most popular caching, and satellite most popular caching, are considered;
2) Using the derived results, the SDP performance under different transmission modes (FD with and without FastForward relaying or HD mode) and caching configurations is thoroughly investigated. The network capacity in terms of maximum supportable users given a minimum SDP requirement is also studied;

3) A SDP maximization-based cache placement is designed subject to the caching capacity constraints of the two cache tiers. While the optimal solution is hard to compute, a low-complexity algorithm is developed achieving superior performance as compared with that of common caching configurations.

4) Various simulation scenarios are conducted to illustrate the superior performance of the proposed system over its single-tier cache-aided counterparts as well as HD transmission systems. Results show that the greater self interference cancellation parameter is, the worse performance FD transmission achieves, and the critical point where the FD mode performs worse than the HD mode is also determined. In addition, numerical results show that the proposed system outperforms the single-tier cache systems in [8]–[10] with at least 25% higher average SDP. Other parameters such as the file popularity, file size, the number of supportable ground users and caching capacity of the two layers that affect the system performance are also investigated.

The remainder of this paper is organized as follows. Section II introduces the related work on edge caching in satellite-terrestrial networks. Section III describes the system model and the transmission scheme for full-duplex communications. Section IV analyzes the successful delivery probability of the general caching and special caching policies. Full-duplex without FastForward relaying protocol, half-duplex transmission scheme and single-tier cache system are employed as baselines for the proposed two-tier cache full-duplex system. Cache placement design and numerical results are presented in Sections V and VI, respectively. Finally, conclusions are made in Section VII.

II. RELATED WORK

In this section, the state-of-the-art edge caching techniques in satellite-terrestrial networks (STNs) are reviewed.

Two types of satellites that are commonly used in STNs are the low earth orbit (LEO) and geosynchronous equatorial orbit (GEO) satellites. Employing LEO satellites in STNs can achieve lower transmission latency than using GEO satellites. However, using LEO satellites poses a very challenging problem on dynamic network topology since they are only in contact with stations
on the ground for a brief period due to their traveling speeds.

Taking into account the relative motion of LEO satellites and terrestrial users, [5] proposes a cross-time-slot graph for the time-varying topology file distribution in combination with the back-tracing partition directed on-path caching distribution mechanism (BPDM) to reduce the caching overhead. The proposed BPDM focuses on cache location decision using a modified Dijkstra routing algorithm and cache content placement using the popularity-aware multiple regions cooperative cache strategy. Proposing a time-evolving covering set in-network caching mechanism, [6] exploits the local connectivity of different regions in choosing the intermediate nodes with better connectivity that relay the objective file. The proposed mechanism includes a novel event updating graph that captures the time-varying topology of the network, a minimum time-evolving covering set algorithm for cache location decision, and a query-based minimum delay distribution path algorithm for optimal path finding during file access procedure. Leveraging the broadcast characteristic of satellite to efficiently provide traffic offloading for terrestrial networks, [7] proposes an integrated LEO satellite/terrestrial cooperative caching scheme, in which the terrestrial access point is able to share its cached content with the satellite.

Employing GEO satellites in STNs is suited to latency-non-sensitive services due to the satellite high altitude. STNs with GEO satellites have the advantage of the static network topology and non-interrupting transmission between the satellites and devices on the ground. Incorporating wireless caching into STNs, [8] proposes an amplify-and-forward relaying protocols, where cache is enabled at relays on the ground. The proposal in [8] considers the most popular and uniform content-based cache placement schemes, which shows a substantial improvement in outage probability over the traditional approach without caching. To offload the backhaul of terrestrial networks, [9] proposes using hybrid STN in combination with an off-line edge caching algorithm. Caches are placed at base stations, in which each cache is divided into two parts. The first part is filled with the most local popular files whilst the other part is filled with the most global popular files. On the performance investigation of 5G edge caching over satellite, [10] uses off-line caching approaches to offload terrestrial backhaul networks. The proposal in [10] focuses on the multimode satellite backhaul in two use cases, i.e., dense urban areas and sparsely populated regions. The effectiveness of the system is studied by virtue of the cache hit ratio and cost per bit for satellite transmission.

It should be emphasized that the system models in [5]–[10] are single-tier cache-enabled systems where caching is used for terrestrial devices only. For STNs employing GEO satellites,
[8]–[10] consider dual-hop downlink satellite–relay(s)–users transmission. The satellite uplink communications from the Internet-connected gateway to the satellite is not considered, which is usually the bottleneck in satellite IoT systems due to large bandwidth consumption.

Enabling cache capacity at satellite, [11], [12] propose two-tier caching models, where the first- and second-tier caches are placed in the terrestrial base stations and the satellite, respectively. The two-tier cache system will potentially: (i) reduce service latency since more contents are stored closer to the end-users; (ii) efficiently utilize satellite bandwidth as the contents are prefetched at the satellite during off-peak hours; Furthermore, when same contents are requested by multiple users, having the contents cached at satellite and broadcasting them at once can save satellite bandwidth. The work in [11] focuses on analyzing the satellite bandwidth consumption with an off-line caching scheme where popular contents are cached at each ground station for local services, while the satellite’s cache is used for the most popular contents within its coverage (containing multiple ground stations) to take advantage of the satellite’s broadcast nature to the ground stations. Non-cached contents can be retrieved from the gateway if needed. On the other hand, our current work focus on investigating the file delivery time through the successful delivery probability (SDP) metric considering the realistic channel propagation models. SDP analysis is typically relevant in delay-sensitive applications [15]. In addition, while [11] considers a scenario where a file is fully cached, our work considers more general caching policies allowing partial caching of files. Hence, caching capability can be exploited better for improved performance. Employing LEO satellite into a two-tier caching model, [12] designs a terrestrial link between the core network and base stations to maintain services when quality satellite connection is absent. The design, however, does not consider the handover problem of the link between a satellite and base stations, and the dynamic of the satellite link.

### III. System Model

The satellite-assisted IoT system is composed of a satellite \( S \), an Internet-connected satellite gateway \( G \), a ground station \( G_s \), and a set of \( K \) ground end-users \( U_i, i = 1, \ldots, K \) as depicted in Fig. 1. In this system, \( S \) is a GEO satellite and \( G_s \) is a low power base station equipped with a satellite receiver. Both \( S \) and \( G_s \) are cache enabled. Assuming there is no direct link from user \( U_i \) to satellite \( S \) and from the ground station to the gateway due to weather factors, long distance, and heavy shadowing.
A. Two-tier Caching Model

The caching model consists of two tiers, namely the first tier at the ground station with storage capacity of $C_1$ (bits) and the second tier at the satellite with storage capacity $C_2$ (bits). The gateway is connected to the internet and hosts $N$ files $W_1, \ldots, W_N$, which are assumed to be of equal size of $F$ (bits). This assumption is for the scenario of heterogeneous IoT applications. For other IoT applications, the analysis can be easily extended to the case of unequal file size. Without loss of generality, assume $C_i/F = N_i$, $i = 1, 2$ being integer numbers with $N_1 + N_2 < N$ as it is not possible to cache all the file contents. The generic caching policy considered in this paper is represented by $\mu_1 = [\mu_1^1, \ldots, \mu_1^N]$, $\mu_2 = [\mu_2^1, \ldots, \mu_2^N]$ where $\mu_1^n, \mu_2^n \in [0, 1]$ are the portions of $W_n$ fetched at the first-tier cache and the second-tier cache, respectively. Assuming that the cached portions of the same file do not have overlapping contents. Other conditions of the caching policy are

$$\mu_1^n + \mu_2^n \leq 1, \forall n, \quad \sum_{n=1}^{N} \mu_i^n = N_i, i = 1, 2.$$ 

It is always optimal to utilize all caching capacities.
B. File Popularity Model

The request probability $q_n$ for file $W_n$ follows the Zipf distribution, i.e.,

$$q_n = \frac{n^{-\alpha}}{\sum_{m=1}^{N} m^{-\alpha}}$$

(1)

where $0 < \alpha < 1$ denotes the Zipf skewness factor [16]. A large value of $\alpha$ means the requests on high-popularity files, whereas a small value of $\alpha$ is related to the requests with heavy-tailed popularity. Without loss of generality, it is assumed that files $W_1, \ldots, W_N$ have decreasing popularity.

C. Channel Model

We consider block-based communications, where a transmission section is accomplished within a coherence time $T$ (seconds). The satellite–terrestrial downlink and gateway–satellite uplink channels are similarly modeled; Hence, only the satellite downlink channel model is described.

The satellite–terrestrial links have multipath fading and shadow fading. The shadow fading is composed of the line-of-sight (LOS) shadow fading and multiplicative shadow fading; The multipath fading composes of one LOS and many weak scatter components [17]. Therefore, the channel coefficient $\tilde{h}_S(t)$ for the satellite–terrestrial link can be modeled as

$$\tilde{h}_S(t) = A(t)e^{j\vartheta(t)} + Z(t)e^{j\zeta_0}$$

(2)

where $j^2 = -1$, $A(t)$ is the independent stationary random process representing the amplitude of the scatter components and following the Rayleigh distribution; $Z(t)$ represents the LOS component and follows the Nakagami distribution; $\vartheta(t)$ is the stationary random phase process with a uniform distribution over $[0, 2\pi)$; $\zeta_0$ is the deterministic phase of the LOS component.

The normalized channel power gain $h_S(t) = |\tilde{h}_S(t)|^2/\sigma^2$, where $\sigma^2$ is the additive white Gaussian noise power, has the probability distribution function (PDF) $f_{h_S}(x)$ given by [17], [18]

$$f_{h_S}(x) = \alpha_1 \exp\left(-\frac{x}{b_1}\right) \frac{1}{\sigma_1^2} \frac{\Gamma(m_1, 1, \sigma_1 x)}{\sigma_1^2}$$

(3)

with coefficients

$$\alpha_1 = \left(\frac{b_1 m_1}{b_1 m_1 + \Omega_1}\right)^{m_1}, \quad \frac{1}{\sigma_1} = \frac{\Omega_1}{b_1 (b_1 m_1 + \Omega_1)}.$$
where $b_1 = \mathbb{E}[|A(t)|^2]$ represents the average power of the scatter components; $\Omega_1 = \mathbb{E}[|Z(t)|^2]$ represents the average power of the LOS component; $m_1 = \Omega_1^2 / \text{Var}[|Z(t)|^2]$ is the Nakagami parameter; $\text{1}_F(a; 1; z)$ is the confluent hypergeometric function of the first kind [19, eq.(9.210.1)]:

$$\text{1}_F(a; 1; z) = \sum_{k=0}^{a-1} \frac{(-1)^k \Gamma(1-a+k)z^k e^z}{\Gamma(1-a)(k!)^2}.$$  

Note that we omit the time-dependent index due to the stationarity assumption.

**Lemma 1.** The cumulative distribution function (CDF) $F_{h_S}(x)$ of $h_S(t)$ is

$$F_{h_S}(x) = \sum_{k=0}^{m_1-1} \sum_{q=0}^{k} \frac{\alpha_1 e^{(\sigma_1-1/b_1)x}(-1)^{2k-q}\Gamma(1-m_1+k)\sigma_1^k x^q}{\Gamma(1-m_1)k!q!(\sigma_1-1/b_1)^{k-q+1}}.$$  

\hspace{2cm} (4)

**Proof.** $F_{h_S}(x)$ of $h_S(t)$ can be derived using elementary functions and finite summations

$$F_{h_S}(x) = \int_0^x f_{h_S}(y)dy$$

$$= \sum_{k=0}^{m_1-1} \sum_{q=0}^{k} \alpha_1 e^{(\sigma_1-1/b_1)x}(-1)^{2k-q}\Gamma(1-m_1+k)\sigma_1^k x^q \frac{1}{\Gamma(1-m_1)k!q!(\sigma_1-1/b_1)^{k-q+1}} \int_0^x y^{k-1}e^{(\sigma_1-1/b_1)y}dy.$$  

Using the following result on integration with exponential function:

$$\int y^k e^{cy} dy = e^{cy} \sum_{q=0}^{k} \frac{(-1)^{k-q}k!}{q!c^{k-q+1}} y^q,$$

we can derive (4). 

To validate (4), Fig. 2 depicts the PDF and CDF of $h_S(t)$ with $\{m_1, b_1, \Omega_1\} = \{5, 0.502, 0.279\}$. As can be observed from Fig. 2, as $x$ becomes large, $F_{h_S}(x)$ approaches one.

**The terrestrial link** between the ground station and user $U_i$ is modeled as a Rayleigh fading channel with the (normalized) channel power gain $h_i(t)$ governed by the following distribution functions

$$f_{h_i}(x) = \frac{1}{\bar{h}_i} \exp\left(-\frac{x}{\bar{h}_i}\right), \quad F_{h_i}(x) = 1 - \exp\left(-\frac{x}{\bar{h}_i}\right)$$  

\hspace{2cm} (5)

with $\bar{h}_i$ is the average channel power gain taking into the effects of path-loss, small-scale fading, etc.
Fig. 2: PDF and CDF of $h_S(t)$.

D. Full-duplex Transmission

When caching IoT data, it is essential to maintain data freshness [20]. This model assumes that the users will be served in a time-division multiple access (TDMA) manner which allows each user to be consecutively active in $T/K$ (seconds). The time-slot $T/K$ will be large enough to ensure data freshness in practical IoT applications. Under TDMA, inter-user interference does not exist. Both the satellite and ground station operate in the FD mode. When the channel state information is known at the transmitter, rate adaption is employed and the achievable rates (bps) on the links are

$$R_i = B_{G_s} \log_2 (1 + P_{G_i} h_i), \quad R_S = B_S \log_2 \left(1 + \frac{P_S h_S}{\beta P_{G_s} + 1}\right), \quad R_G = B_G \log_2 \left(1 + \frac{P_G h_G}{\beta P_S + 1}\right)$$

where $B_{G_s}$, $B_S$ and $B_G$ are the bandwidths (Hz) of the satellite uplink, downlink, and the terrestrial downlink respectively; $P_S$, $P_G$, and $P_{G_s}$ denote the transmit powers of the satellite, gateway, and ground station respectively; $\beta P_S$ and $\beta P_{G_s}$ represent the normalized residual self-interference (SI) powers at the satellite and ground station respectively, which are assumed to be proportional to the transmit powers with coefficient $\beta \geq 0$ being the SI cancellation quality parameter [21]. For notational simplicity, denote:

$$\bar{P}_S \triangleq P_S/\left(\beta P_{G_s} + 1\right), \quad \bar{P}_G \triangleq P_G/\left(\beta P_S + 1\right).$$

When a user requests file $W_n$, the gateway needs to transfer the non-cached $(1 - \mu^n_1 - \ldots$
portion to the satellite, which also needs to send accumulated \((1 - \mu_1^n)\) portion to the ground station. By employing the FastForward protocol [22] with the FD mode at the satellite, the completion time is \(\max \{(1 - \mu_1^n - \mu_2^n)F/R_G, (1 - \mu_1^n)F/R_S\}\). Similarly, by deploying the FastForward protocol with the FD mode at the ground station, the (end-to-end) delivery time for transferring the entire file to the user is determined by the time for sending the \((1 - \mu_1^n)\) portion over the satellite downlink to the ground station and the time for sending the whole file over the terrestrial link to the user, i.e.,

\[
t_{n,i} = \max \{F/R_i, (1 - \mu_1^n)F/R_S, (1 - \mu_1^n - \mu_2^n)F/R_G\}.
\] (7)

We will further consider FD transmissions without the FastForward protocol in the ensuing section.

IV. SUCCESSFUL DELIVERY PROBABILITY

A. SDP Analysis

The successful delivery probability (SDP) of serving the requested file \(W_n\) is the probability that a user receives the file within the user’s active time-slot. The following analysis is for file \(W_n\) and user \(U_i\), and is true for all users. The SDP of file \(W_n\) is

\[
\psi_{n,i}(\mu_1^n, \mu_2^n) = \Pr (t_{n,i} \leq T/K).
\] (8)

The probability operator is taken with respect to the channel power gain variables. A closed-form expression for \(\psi_{n,i}\) is given in (9).

\[
\psi_{n,i}(\mu_1^n, \mu_2^n) = \exp \left( -\left(2^{\sigma_1/B_1} - 1\right)/P_G \right) \frac{\Gamma(1 - m_1 + k)}{\Gamma(1 - m_1)} \frac{\Gamma(1 - m_2 + k)}{\Gamma(1 - m_2)} \frac{\left(2^{(1 - \mu_1^n)\tau/B_1} - 1\right)}{P_S} \frac{\left(2^{(1 - \mu_2^n)\tau/B_2} - 1\right)}{P_G} .
\] (9)

Proof. Substituting (7) into (8) gives rise to

\[
\psi_{n,i} = \Pr \left( \max \left\{ \frac{F}{R_i}, \frac{(1 - \mu_1^n)F}{R_S}, \frac{(1 - \mu_1^n - \mu_2^n)F}{R_G} \right\} \leq \frac{T}{K} \right)
= A_1 \times A_2 \times A_3
\]
where we have

\[ A_1 = \Pr \left( \frac{F}{R_i} \leq \frac{T}{K} \right) = 1 - F_{h_i} \left( \frac{2^{\tau/B_{Gs}} - 1}{P_{Gs}} \right), \]

\[ A_2 = \Pr \left( (1 - \mu^n_1) \frac{F}{R_S} \leq \frac{T}{K} \right) = 1 - F_{h_S} \left( \frac{2^{(1-\mu^n_1)\tau/B_S} - 1}{\bar{P}_S} \right), \]

\[ A_3 = \Pr \left( (1 - \mu^n_1 - \mu^n_2) \frac{F}{R_G} \leq \frac{T}{K} \right) = 1 - F_{h_G} \left( \frac{2^{(1-\mu^n_1 - \mu^n_2)\tau/B_G} - 1}{\bar{P}_G} \right), \]

and \( \tau \triangleq \frac{F}{(T/K)} \) is the average file serving rate.

We have the following special cases:

- If \( W_n \) is cached at ground station, we have

  \[ \psi_{n,i}(1, 0) = \exp \left( \frac{(2^{\tau_{n}/B_{Gs}} - 1)}{(P_{Gs}\bar{h}_i)} \right); \]

- If \( W_n \) is cached at the satellite, the SDP is \( \psi_{n,i}(0, 1); \)

- If \( W_n \) is fully retrieved from the gateway, the SDP is \( \psi_{n,i}(0, 0). \)

The SDP of file \( W_n \) under the effect of \( \mu^n_1 \) and \( \mu^n_2 \) is illustrated in Fig. 3. It shows that the SDP can be improved by deploying the second-tier caching at the satellite, and the larger the caching capacity is, the more SDP improvement is. The SDP is the largest or smallest when the file is cached at the ground station or fully retrieved from the gateway, respectively.

![Fig. 3: \( \psi_{n,i} \) under different caching portions.](image)

The average SDP of the system is defined as a weighted sum of the users’ SDPs:

\[ \psi(\mu_1, \mu_2) = \sum_{i=1}^{K} \omega_i \psi_i(\mu_1, \mu_2) = \sum_{i=1}^{K} \omega_i \sum_{n=1}^{N} q_n \psi_{n,i}(\mu^n_1, \mu^n_2) \]  \( (10) \)
with weighting coefficients $\omega_i \in (0, 1)$ and $\sum_i \omega_i = 1$. In the following, without loss of generality, we assume homogeneous users with $\omega_i = 1/K$, and average channel gains $\bar{h}_i = \bar{h}_k$, $\forall i, k$. Thus, $\psi = \psi_1$.

B. SDP Under Different Caching Policies

In this section, we analyze the average SDP under three different caching policies, i.e., uniform caching, ground station most popular caching, and satellite most popular caching.

1) Uniform Caching Policy: The same portions of $N$ files are prefetched in both tiers. The SDP under this caching policy is

$$\psi^{uni} = \sum_{n=1}^{N} q_n \psi_{n,i}(N_1/N, N_2/N).$$

The caching policy does not consider the file popularity distribution. Hence, it tends to be optimal when all files have similar popularity.

2) Ground Station Most Popular Caching: The ground station caches $N_1$ most popular files, and the satellite caches $N_2$ most popular files of the remaining files. The average SDP $\psi^{grnd}$ is

$$\psi^{grnd} = \sum_{n=1}^{N_1} q_n \psi_{n,i}(1, 0) + \sum_{n=N_1+1}^{N_1+N_2} q_n \psi_{n,i}(0, 1) + \sum_{n=N_1+N_2+1}^{N} q_n \psi_{n,i}(0, 0).$$

This caching policy tends to be better if only a number of files are most popular, i.e., the skewness factor $\alpha$ is large. In this case, by caching these files at the ground station, we can increase the SDP.

3) Satellite Most Popular Caching: The satellite caches $N_2$ most popular files, and the ground station caches $N_1$ most popular files of the remaining files. The average SDP $\psi^{sat}$ is

$$\psi^{sat} = \sum_{n=1}^{N_2} q_n \psi_{n,i}(0, 1) + \sum_{n=N_2+1}^{N_2+N_1} q_n \psi_{n,i}(1, 0) + \sum_{n=N_2+N_1+1}^{N} q_n \psi_{n,i}(0, 0).$$

This caching policy is generally more suitable for a two-tier cache-enabled system with multiple ground stations. When most popular files are requested by users associated with different ground stations, the files, when cached at the satellite, are broadcast from the satellite to the ground station only once. In this case, each ground station does not need to separately cache these popular files. For the system under consideration, we have the following result.

**Proposition 1.** Assume $N_1 = N_2$. It is true that $\psi^{grnd} > \psi^{sat}$. 
Proof. We have
\[ \psi^{\text{grnd}} - \psi^{\text{sat}} = \sum_{n=1}^{N_1} q_n (\psi_{n,i}(1, 0) - \psi_{n,i}(0, 1)) - \sum_{n=N_1+1}^{N_1+N_2} q_n (\psi_{n,i}(1, 0) - \psi_{n,i}(0, 1)) \]
\[ = \left( \sum_{n=1}^{N_1} q_n - \sum_{n=N_1+1}^{N_1+N_2} q_n \right) (\psi_{n,i}(1, 0) - \psi_{n,i}(0, 1)) > 0 \]

C. Baselines

1) Single-tier Caching: Our derived results can be used to evaluate the SDP of single-tier caching policies. In the single-tier scenario, only cache \( C_1 \) is enabled at the ground station. There are two cases:

- When the same portion of \( N \) files are cached at the ground station, we have:
  \[ \psi_{1, \text{uni}} = \sum_{n=1}^{N} q_n \psi_{n,i}(N_1/N, 0). \]
- When \( N_1 \) most popular files are cached at the ground station, we have:
  \[ \psi_{1, \text{pop}} = \sum_{n=1}^{N_1} q_n \psi_{n,i}(1, 0) + \sum_{n=N_1+1}^{N} q_n \psi_{n,i}(0, 0). \]

2) Full-duplex Without FastForward Relaying: The delivery time of file \( W_n \) is completed in three slots. In the first slot, the ground station transmits \( \mu_1^n F \) to the user while receiving \( \mu_2^n F \) from the satellite; the satellite transmits \( \mu_2^n F \) to the ground station while receiving \((1 - \mu_1^n - \mu_2^n) F \) from the gateway. In the second slot, the ground station transmits \( \mu_2^n F \) to the user while receiving \((1 - \mu_1^n - \mu_2^n) F \) from the satellite. Finally, in the third slot, the ground station transmits \((1 - \mu_1^n - \mu_2^n) F \) to the user. The delivery time is given by:
\[ \bar{t}_{n,i} = \max \left\{ \mu_1^n F/R_i, \mu_2^n F/R_S, (1 - \mu_1^n - \mu_2^n) F/R_G \right\} \]
\[ + \max \left\{ \mu_2^n F/R_i, (1 - \mu_1^n - \mu_2^n) F/R_S \right\} + (1 - \mu_1^n - \mu_2^n) F/R_i. \]
The SDP of file \( W_n \) is \( \bar{\psi}_{n,i} = \Pr \left( \bar{t}_{n,i} \leq T/K \right) \). We will then evaluate the system SDP \( \bar{\psi} = \sum_n q_n \bar{\psi}_{n,i} \) using Monté Carlo simulation in Section VI.
3) **Half-duplex Transmission:** In this transmission scheme, both the satellite and ground station adopt the HD mode. The achievable rates on the links are

\[ R_{i}^{\text{HD}} = B_{G} \log_{2} (1 + P_{G} h_{i}) , \quad R_{S}^{\text{HD}} = B_{S} \log_{2} (1 + P_{S} h_{S}) , \quad R_{G}^{\text{HD}} = B_{G} \log_{2} (1 + P_{G} h_{G}) . \]

The timeline for delivering file \( W_{n} \) is divided into three slots as shown in Fig. 4. The delivery time for file \( W_{n} \) is

\[ t_{n,i}^{\text{HD}} = \max \{ \mu_{n}^{i} F / R_{i}^{\text{HD}} , (1 - \mu_{n}^{i} - \mu_{n}^{2}) F / R_{G}^{\text{HD}} \} + (1 - \mu_{n}^{i}) F / R_{S}^{\text{HD}} + (1 - \mu_{n}^{i}) F / R_{i}^{\text{HD}} . \]

The SDP of file \( W_{n} \) for the HD mode case is \( \psi_{n,i}^{\text{HD}} = \Pr \left( t_{n,i}^{\text{HD}} \leq T / K \right) \). While \( \psi_{n,i}^{\text{HD}} = \sum_{n} q_{n} \psi_{n,i}^{\text{HD}} \) is too complex to be found in closed form, it is evaluated using Monte Carlo simulation in Section VI.

![Fig. 4: Half-duplex transmission scheme for file \( W_{n} \).](image)

**V. Cache Placement Design**

Cached contents are placed into caches \( C_{1} \) and \( C_{2} \) via satellite uplink and downlink transmissions during off-peak hours. In this section, the cached portions \( \mu_{1}^{n} \) and \( \mu_{2}^{n} \) are optimized with the objective of maximizing the average SDP. Under the generic caching policy, the optimization problem for the prefetched content in the two cache tiers subject to caching capacity constraints is formulated as follows:

\[ \max_{\mu_{1}, \mu_{2}} \psi(\mu_{1}, \mu_{2}) \quad \text{s.t.} \quad \sum_{n=1}^{N} \mu_{i}^{n} \leq N_{i}, \quad i = 1, 2. \quad (11) \]

The problem is a non-linear non-convex optimization one, and hence it is very challenging to find the optimal solution. We resort to sub-optimal caching placement solutions by confining the original \( 2N \)-dimensional search space into two \( N \)-dimensional search spaces. Specifically, we consider that the cached files are prefetched with the same portions, i.e., \( \mu_{i}^{n} = \bar{\mu}_{i}, \quad i = 1, 2 \) if file \( W_{n} \) is cached at the respective tier, and \( \mu_{i}^{n} = 0 \) otherwise. We consider the following two policies.
Generalized uniform caching policy. The caching policy can be described by the vectors \( \mu_1 = [\mu_1, \ldots, \mu_1, 0, \ldots, 0] \) (first \( N_1 \) elements equal to \( \mu_1 \)), where \( N_1 \leq \bar{N}_1 \leq N \) and \( \bar{\mu}_1 \bar{N}_1 = N_1 \); and \( \mu_2 = [\mu_2, \ldots, \mu_2, 0, \ldots, 0] \) (first \( N_2 \) elements equal to \( \mu_2 \)), where \( N_2 \leq \bar{N}_2 \leq N \) and \( \bar{\mu}_2 \bar{N}_2 = N_2 \). It is required that \( \bar{\mu}_1 + \bar{\mu}_2 \leq 1 \). In other words, \( \bar{N}_1 \) most popular files are chosen for caching and the same portions \( \bar{\mu}_1 \) and \( \bar{\mu}_2 \) of \( \bar{N}_1 \) files are cached at the ground station and satellite, respectively. This generalized uniform caching policy is inspired by the fact that uniform caching policy (Section IV-B1) naively caches all files in the library regardless of the file request rate. The cache resource is wasted for prefetching files that are rarely requested. The generalized uniform caching policy becomes the uniform caching policy when \( \bar{N}_i = N, i = 1, 2 \).

Generalized ground station most popular caching policy. The caching policy can be described by the vectors \( \bar{\mu}_1 = [\bar{\mu}_1, \ldots, \bar{\mu}_1, 0, \ldots, 0] \) as above; and \( \bar{\mu}_2 = [0, \ldots, 0, \bar{\mu}_2, \ldots, \bar{\mu}_2, 0, \ldots, 0] \) (first \( \bar{N}_1 \) elements being 0, followed by \( \bar{N}_2 \) elements being \( \bar{\mu}_2 \)), where \( N_2 \leq \bar{N}_2 \leq N \) and \( \bar{\mu}_2 \bar{\bar{N}}_2 = N_2 \). It is required that \( \bar{N}_1 + \bar{N}_2 \leq N \). Similar to the generalized uniform caching policy, this policy only chooses \( \bar{N}_1 \) and \( \bar{N}_2 \) most popular files with higher portion compared to that of the ground station most popular caching policy (Section IV-B2) to cache at the ground station and satellite, respectively. This caching policy degrades to the ground station most popular caching policy when \( \bar{N}_i = N_i, i = 1, 2 \).

Then the resulting problem can be formulated as

\[
\max_{\bar{\mu}_1, \bar{\mu}_2} \psi(\bar{\mu}_1, \bar{\mu}_2).
\]

We can conduct an exhaustive search to find solutions to (12) with a computational complexity of \( O(n^2) \). First, we search for \( \bar{N}_1 \) and \( \bar{N}_2 \) constraining the policy being one of the generalized caching policies. Then, we determine the caching solution with higher SDP.

Remark 1. The caching solution to problem (12) always achieves a higher SDP than the ground station most popular and uniform caching strategies.

With the focus on caching the likely requested files, the generalized caching policies allow more popular contents to be cached closer to users. Therefore, the serving time under generalized caching policies will be shorter than that of the generic uniform and ground station most popular caching policies.
VI. NUMERICAL RESULTS

The values for key parameters [23] used in this section are presented in Table I. Because it requires more than 100 dB of self-interference cancellation for the full-duplex system to achieve the same signal-to-noise-ratio-plus-interference as that of the half-duplex system [24], we choose the self-interference cancellation quality parameter $\beta = 0.0001$ for most of our system configuration under FD transmission. With this value of $\beta$, self-interference cancellation in the system is achieved from 110 dB to 120 dB. The fading states for the satellite-terrestrial links are defined by $\{m_1, b_1, \Omega_1\} = \{m_2, b_2, \Omega_2\} = \{5, 0.502, 0.279\}$ when approaching average shadowing [25].

| Parameter | Value | Parameter | Value |
|-----------|-------|-----------|-------|
| $N$       | 1000 files | $B_G$ | 2 GHz |
| $F$       | 100 Mb  | $B_S$ | 10 GHz |
| $P_S$     | 10 W    | $B_{G_S}$ | 2 GHz |
| $P_G$     | 30 W    | $\sigma^2$ | -120 dBm/Hz |
| $P_{G_S}$ | 2 W     | $m_i/b_i/\Omega_i$ | 5/0.502/0.279 |

A. FD versus HD: Effect of SI Cancellation Parameter $\beta$

We first investigate the average SDP when $\beta$ varies. Consider $C_1 = 50$ Gb and $C_2 = 10$ Gb, and $K = 20$ users under uniform caching policy.

With the HD scheme as a reference, the effect of $\beta$ on the FD scheme SDP is displayed in Fig. 5. As $\beta$ increases (i.e., worse self-interference cancellation in the FD system), the SDP decreases and starts getting worse than that of HD transmission. The FD scheme achieves the same SDP as that of HD with $\beta = 0.02$. The self-interference cancellation in the FD system in this case is 97 dB to 100 dB which is almost the recommended value in [24]. The SDP of FD is 20% worse than that of HD with $\beta = 0.1$ where self-interference cancellation is 85 dB to 90 dB. We also present numerical result for hybrid FD/HD transmission. In this hybrid scheme, the channel realizations are generated, then the delivery time of each file is the minimum delivery time between the FD and HD modes. In practical applications, once knowing the channel condition, devices in the system can dynamically switch between the FD and HD modes for better performance.
B. Two-tier versus Single-tier Caching

The single-tier cache-enabled system in this section is adopted from STN system models in [8]–[10], where content caches are placed at ground stations only. The caching policies for the single-tier cache system correspond to those at the ground station in the two-tier cache system. To compare the performance between two-tier and single-tier caching, we set up both systems with the ground station cache capacity of 80 Gb serving 20 users. The uniform and ground station most popular caching policies are applied to study the average SDP.

Under the uniform caching policy, $N_1/N$ varies and $N_2/N$ is set to be 0.2 in the two-tier caching (meaning the satellite cache capacity is 20 Gb). The result is shown in Fig. 6 with the two-tier scheme outperforming its single-tier counterpart. At best performance with $N_1/N = 0.7$, the achieved SDP in the two-tier scheme is 25% higher than that of the single-tier scheme. Under the ground station most popular caching policy, the number of most popular file cached $N_1$ varies; $N_2$ is fixed at 200 files in the two-tier scheme. The two-tier scheme gives approximately 40% SDP higher than that of the single-tier scheme in best performance scenarios with the ground station caching 600 most popular files.

C. FD: FastForward versus Non-FastForward Relaying Protocols

To compare the performance of the FD protocols, the system is set up with varying values for $C_1$ and $C_2$ of 20 Gb serving 20 users. Full-duplex with and without the FastForward relaying
protocols are examined under the uniform and ground station most popular caching policies. File popularity is set to be $\alpha$ of 0.8 and $\beta$ of 0.0001 for the FD mode. As shown in Fig. 7, the system with the FastForward relaying protocol achieves a higher SDP. This result is expected since the delivery time using the FastForward protocol is shorter than that without using FastForward relaying.
D. Network Capacity

To examine the serving capability, the system is set up with the two layers’ caches $C_1$ of 50 Gb and $C_2$ of 20 Gb. The number of users $K$ varies while other parameters are listed in Table I. Fig. 8 shows the number of supportable users by the two-tier system under two caching policies. A smaller average SDP under uniform caching is achieved when supporting a small number of users; however, it surpasses the SDP of the ground station most popular caching scheme with a larger number of users. The acceptable SDP value depends on practical applications. In this case, with SDP of 50%, the results follow a common understanding that the uniform caching scheme is not as effective as the ground station most popular caching scheme.

![Fig. 8: Number of supportable users under different caching policies.](image)

E. Caching Capacity

In this section, we analyze the two-tier caching capacity when serving 20 users. The system is configured using the parameters specified in Table I. The cache capacities of the two layers are shown in Fig. 9. As expected, the ground station most popular caching policy outperforms uniform caching. The more cache capacity is allocated to the ground station, the higher the SDPs are attainable; and the more content is cached at the satellite, the more the SDPs improve. Under uniform caching, the highest SDP is achieved at 70%, while it is 80% under ground station most popular caching.
Fig. 9: Effect of caching capacity on the average SDP.

F. Effect of File Popularity

In this subsection, to examine the effect of the file popularity on the SDP under the ground station and satellite most popular caching policies, the system is configured with two tier caches $C_1$ and $C_2$ of 50 Gb and 20 Gb, respectively, and serves 20 ground users. Through varying Zipf skewness factor $\alpha$, the resulting SDP is plotted in Fig. 10. Higher SDPs are achieved for both most popular caching schemes with an increase in the Zipf skewness factor. The results also match Proposition 1. This is because having one ground station in the system, the ground station most popular caching policy always gives better SDP results than under satellite most popular caching.

G. Effect of the File Size

To investigate the effect of the file size on SDP, we vary the sizes of $N$ files at the gateway following a uniform distribution with a mean of 100 Mb and a standard deviation of 50 Mb. Two caching schemes are set up as follows: (i) $C_1$ is prefetched with the largest files and then $C_2$ is filled with the largest files among the remained files; (ii) $C_2$ is prefetched with the largest files and then $C_1$ is filled with the largest files among the remained files. The results are shown in Fig. 11. The average SDP is better when the largest files are cached at the ground station, resulting in a gain of approximately 10% to 20% in SDP.
H. Generalized Caching Solutions

To investigate the performances of the generalized caching policies, we find the solutions to the problem in (12) when the caching capacity at satellite is fixed at 20 Gb and at ground station varies up to 80 Gb. The performances are illustrated in Fig. 12. As can be seen from the figure, generalized caching policies yield higher SDPs than the naive uniform and ground station most popular caching policies. The solutions for each case of $C_1$ is detailed in Table II.
Fig. 12: Average SDP under the generalized caching policies.

TABLE II: Solutions of the generalized caching policies.

| $C_1$ (Gb) | $\bar{N}_1$ | $\bar{\mu}_1$ | $\bar{N}_2$ | $\bar{\mu}_2$ | $\bar{N}_3$ | $\bar{\mu}_3$ | $\bar{N}_4$ | $\bar{\mu}_4$ |
|------------|--------------|----------------|--------------|----------------|--------------|----------------|--------------|----------------|
| 10         | 250          | 0.40           | 400          | 0.50           | 220          | 0.45           | 370          | 0.54           |
| 20         | 350          | 0.57           | 500          | 0.40           | 300          | 0.67           | 370          | 0.54           |
| 30         | 450          | 0.66           | 650          | 0.30           | 500          | 0.60           | 330          | 0.60           |
| 40         | 650          | 0.61           | 650          | 0.30           | 570          | 0.70           | 300          | 0.66           |
| 50         | 700          | 0.71           | 800          | 0.25           | 600          | 0.83           | 280          | 0.71           |
| 60         | 800          | 0.75           | 900          | 0.22           | 600          | 1.00           | 200          | 1.00           |
| 70         | 910          | 0.76           | 900          | 0.22           | 700          | 1.00           | 200          | 1.00           |
| 80         | 950          | 0.84           | 1000         | 0.15           | 800          | 1.00           | 200          | 1.00           |

VII. CONCLUSION

In this paper, we studied the SDP of a two-tier cache enabled satellite IoT system with full-duplex communications. A thorough analysis for the SDP under the FD mode with the FastForward relaying protocol was carried out under different caching policies. Cache placement was also designed with the goal of maximizing the average SDP. Using the FD mode without the FastForward protocol, HD mode and single-tier cache system as benchmarks, numerical results were presented when investigating the system performance in terms of SDP, maximum number of supportable users, caching capacity in each tier, and the effect of file popularity on the system performance under different caching policies. For a better understanding on parameters
that affect the proposed system performance, a special case with a variety of file sizes was also investigated.

REFERENCES

[1] S. K. Routray and H. M. Hussein, “Satellite based iot networks for emerging applications,” arXiv:1904.00520, Mar. 2019. [Online]. Available: https://arxiv.org/abs/1904.00520
[2] A. Guidotti, A. Vanelli-Coralli, M. Conti, S. Andrenacci, S. Chatzinotas, N. Maturo, B. Evans, A. Awoseyila, A. Ugolini, T. Foggi, L. Gaudio, N. Alagha, and S. Cioni, “Architectures and key technical challenges for 5g systems incorporating satellites,” IEEE Transactions on Vehicular Technology, vol. 68, no. 3, pp. 2624–2639, Mar. 2019.
[3] European Union. (2018) Satellite and terrestrial network for 5g. [Online]. Available: https://www.sat5g-project.eu/
[4] FCC Report. (2018) Starlink fcc application. [Online]. Available: https://fcc.report/IBFS/SAT-MOD-20181108-00083
[5] Y. Li, Y. Wang, P. Yuan, Q. Zhang, and Z. Yang, “Popularity-aware back-tracing partition cooperative cache distribution for space–terrestrial integrated networks,” IET Communications, vol. 13, no. 17, pp. 2786–2796, Oct. 2019.
[6] Z. Yang, Y. Li, P. Yuan, and Q. Zhang, “Tscc: A novel file distribution strategy in integrated leo satellite-terrestrial networks,” IEEE Transactions on Vehicular Technology, vol. 69, no. 5, pp. 5426–5441, May 2020.
[7] J. Li, K. Xue, D. S. L. Wei, J. Liu, and Y. Zhang, “Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks,” IEEE Transactions on Wireless Communications, vol. 19, no. 4, pp. 2367–2381, Jan. 2020.
[8] K. An, Y. Li, X. Yan, and T. Liang, “On the performance of cache-enabled hybrid satellite-terrestrial relay networks,” IEEE Wireless Communications Letters, vol. 8, no. 5, pp. 1506–1509, Jun. 2019.
[9] 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 (GLOBECOM), Singapore, 2017, pp. 1–6.
[10] T. Vu, Y. Poirier, S. Chatzinotas, and N. Maturo, “Modeling and implementation of 5g edge caching over satellite,” International Journal of Satellite Communications and Networking, vol. 38, no. 5, pp. 395–406, Mar. 2020.
[11] H. Wu, J. Li, H. Lu, and P. Hong, “A two-layer caching model for content delivery services in satellite-terrestrial networks,” in 2016 IEEE Global Communications Conference (GLOBECOM), DC, USA, 2016, pp. 1–6.
[12] L. Liu, J. Zhang, X. Zhang, P. Wang, Y. Wang, and L. Ouyang, “Design and analysis of cooperative multicast-unicast transmission scheme in hybrid satellite-terrestrial networks,” in 2018 IEEE International Conference on Communication Systems (ICCS), Wuxi, China, 2018, pp. 309–314.
[13] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, “The role of caching in future communication systems and networks,” IEEE J. Sel. Areas Commun., vol. 36, no. 6, pp. 1111–1125, Sept. 2018.
[14] Z. Zhang, W. Zhang, and F. Tseng, “Satellite mobile edge computing: Improving qos of high-speed satellite-terrestrial networks using edge computing techniques,” IEEE Network, vol. 33, no. 1, pp. 70–76, Jan. 2019.
[15] T. X. Vu, L. Lei, S. Chatzinotas, B. Ottersten, and A. V. Trinh, “On the successful delivery probability of full-duplex-enabled mobile edge caching,” IEEE Communications Letters, vol. 23, no. 6, pp. 1016–1020, Jun. 2019.
[16] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web caching and Zipf-like distributions: Evidence and implications,” in IEEE Conf. Comput. Commun. Soc. (INFOCOM), NY, USA, 1999, pp. 126 – 134.
[17] A. Abdi, W. C. Lau, M. Alouini, and M. Kaveh, “A new simple model for land mobile satellite channels: first- and second-order statistics,” IEEE Trans. Wireless Commun., vol. 2, no. 3, pp. 519–528, May 2003.
[18] M. R. Bhatnagar and A. M.K., “Performance analysis of af based hybrid satellite-terrestrial cooperative network over generalized fading channels,” IEEE Communications Letters, vol. 17, no. 10, pp. 1912–1915, Sept. 2013.
[19] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, vol. 6th ed., 2000.

[20] S. Vural, P. Navaratnam, N. Wang, C. Wang, L. Dong, and R. Tafazolli, “In-network caching of internet-of-things data,” in *2014 IEEE International Conference on Communications (ICC)*, Sydney, Australia, 2014, pp. 3185–3190.

[21] K. T. Phan and T. Le-Ngoe, “Power allocation for buffer-aided full-duplex relaying with imperfect self-interference cancelation and statistical delay constraint,” *IEEE Access*, vol. 4, pp. 3961–3974, Jul. 2016.

[22] B. Dinesh and K. Sachin, “Fastforward: Fast and constructive full duplex relays,” in *Proceedings of the 2014 ACM Conference on SIGCOMM (SIGCOMM ’14)*, IL, USA, 2014, p. 199–210.

[23] HughesNet. (2017) Echostar xix satellite. [Online]. Available: https://www.hughesnet.com/whats-new/satellite-launch

[24] H. Li, J. Van Kerrebrouck, O. Caytan, H. Rogier, J. Bauwelincx, P. Demeester, and G. Torfs, “Self-interference cancellation enabling high-throughput short-reach wireless full-duplex communication,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6475–6486, Aug. 2018.

[25] N. I. Miridakis, D. D. Vergados, and A. Michalas, “Dual-hop communication over a satellite relay and shadowed rician channels,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4031–4040, Sept. 2015.