Dynamic Bandwidth Allocation and Edge Caching Optimization for Nonlinear Content Delivery through Flexible Multibeam Satellites

Thang X. Vu[1], Nicola Maturo[1], Symeon Chatzinotas[1], Joel Grotz[2], Tom Christophory[2], and Björn Ottersten[1]

[1]Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg,
L-1855 Luxembourg. E-mail: {thang.vu, nicola.maturo, symeon.chatzinotas, bjorn.ottersten}@uni.lu
[2]SES Engineering, L-6815 Betzdorf, Luxembourg. Email: {joel.grotz, tom.christophory}@ses.com

Abstract—The next generation multibeam satellites open up a new way to design satellite communication channels with the full flexibility in bandwidth, transmit power and beam coverage management. In this paper, we exploit the flexible multibeam satellite capabilities and the geographical distribution of users to improve the performance of satellite-assisted edge caching systems. Our aim is to jointly optimize the bandwidth allocation in multibeam and caching decisions at the edge nodes to address two important problems: i) cache feeding time minimization and ii) cache hits maximization. To tackle the non-convexity of the joint optimization problem, we transform the original problem into a difference-of-convex (DC) form, which is then solved by the proposed iterative algorithm whose convergence to at least a local optimum is theoretically guaranteed. Furthermore, the effectiveness of the proposed design is evaluated under the realistic beams coverage of the satellite SES-14 and Movielens data set. Numerical results show that our proposed joint design can reduce the caching feeding time by 50% and increase the cache hit ratio (CHR) by 10% to 20% compared to existing solutions. Furthermore, we examine the impact of multispot beams and multicarrier wide-beam on the joint design and discuss potential research directions.

Index Terms—Edge caching, flexible multibeam satellite, resource allocation, successive convex approximation.

I. INTRODUCTION

Efficient data distribution is one of the most important goals in 5G and beyond networks due to not only the proliferation of connected devices but also the increasing demand for data-hungry applications. The major contributor to these extraordinary evolution of the network traffics is the mobile data, which will grow 74% by 2023 according to recent Cisco's forecasts [1]. This traffic growth results from the increase of mobile handsets, e.g., tablets and smart phones, along with the booming of streaming services, such as YouTube and NetFlix. In addition, the popularity of new applications that require real-time interactions and the increasing video quality, i.e., Tactile internet, Virtual Reality and 4K videos etc., also contribute to the network traffics, which in turns puts more pressure on both the core and access segments. From the content provider perspective, this trend is economically interesting as they can achieve more benefits either through subscriptions or advertising. On the other hand, the telecom operators seem not getting sufficient benefit for upgrading their infrastructure to serve the newly data-hungry applications. Furthermore,

they also have difficulty in accessing new frequency bands to improve their wireless access and backhaul networks as the spectrum resource is scarce. These ignite a new trend for the telecom operators to create data distribution services through their own content delivery networks (CDNs), in which efficient edge caching will be implemented to serve contents to end-users. Since the CDN belongs to the telecom operator, efficient cross-layer optimization can be easily performed between the physical infrastructure and the network service to improve the system resource usage and the user quality of experience [2].

Edge caching through satellites has recently been considered as an efficient way to send popular data to many CDNs over wide areas of interest. Unlike caching via terrestrial networks, in which the cached data must go through multiple hops and be sent to each CDN individually, caching via satellite backhauling can reach many caching nodes due to the large beam coverage of the satellite channels [3]. In addition, the very large bandwidth allows the satellite to transport very high-volume data in short time. The joint consideration of terrestrial and satellite systems for edge caching is studied in [4], [5], which employs satellite backhauling as the complementary solution to the terrestrial network. The large coverage of the satellite beams are shown to be beneficial in improving the caching performance of the proxy servers. A similar setup is considered in [6], in which caching update rules based on the estimated local and global content popularity are executed via multicasting-based satellite links. The authors of [7] propose a broad/multi-cast based satellite to improve the caching performance of the caches at the user end. Hybrid configuration of monobeam and multibeam satellites is studied in [8], [9] in which the whole frequency bandwith is shared between the wide-beam and multi-beam transmission modes. It is noted that the above-mentioned works consider the traditional satellite payload in which the mono/multi-beam configurations are fixed, and hence unable to exploit the flexibility of the next generation multibeam satellites.

Contributions: In this work, we consider the flexible multibeam satellite for fetching popular contents to the edge nodes and focus on the off-line placement phase to optimize the average caching performance [4], [5], [7], [10]. Our contributions

are as follows:

- We formulate the joint bandwidth allocation and caching decision problem to fully exploit the potentials of future flexible multi-beam satellites and the geographical distribution of users towards efficient edge caching algorithm. The proposed design jointly optimizes the frequency bandwidth allocated to each beam and the content indexes to be stored at each edge nodes. We consider two important metrics of the caching systems: i) cache feeding time minimization and ii) cache hits maximization.

- To tackle the non-convexity of the formulated joint optimization problem, we propose an iterative algorithm using the successive convex approximiation (SCA) method. The convergence to at least a local optimum of the proposed iterative algorithm is guaranteed.

- We evaluate the proposed joint design under realistic Movielens data set [16] and the SES-14's coverage [17] over the US East Coast. Numerical results show that the proposed joint design significantly improves both the caching time and cache hit ratio (CHR) performances compared with existing reference schemes. In addition, we discuss potentials issues in flexible satellite systems to further boost the caching performance.

The rest of the paper is organized as follows. Section II provides technological enablers for caching over satellite. Section III describes the system parameters and the proposed joint design. Section IV presents numerical results. Finally, Section V provides discussions and concludes the paper.

II. FLEXIBLE MULTIBEAM SATELLITES: TECHNOLOGICAL ENABLER FOR SATELLITE-ASSISTED EDGE CACHING

Traditional satellite architectures usually consist of two distinct configurations: a wide beam with very large coverage which is mainly used for broadcasting applications, or multiple small beams which are designed for broadband services. These application-oriented configurations are best suitable for dedicated services, however reduce the agility and flexibility in coping with dynamic and complex situations. The main reason of such low flexible reconfiguration is the high cost and delay for changing the payload operation in the conventional architecture. With the development of novel payload technology, digital transparent payload (DTP) [11] and active on-board antennas are practically feasible which allows concurrent services sharing the same radio resources (power, bandwidth) in an efficient and reconfigurable hybrid broadcast/broadband modes. The recent successful launch of the satellite SES-17 with the fully digital payload actualizes the deployment of fully flexible multibeam satellites.

Fig. 1 depicts a pictorial example of flexible payloads sending three data streams: one broadcast stream and two broadband streams. By employing the active antenna technology, the three distinct data streams at the outputs of the DTP are simultaneously guided to the broadcast beam and two broadband beams from the same payload. More importantly, the flexible payload also enables arbitrary shape of the multibeams and concentrates power in the broadcasting beam to efficiently deal with non-uniform distribution of users. It
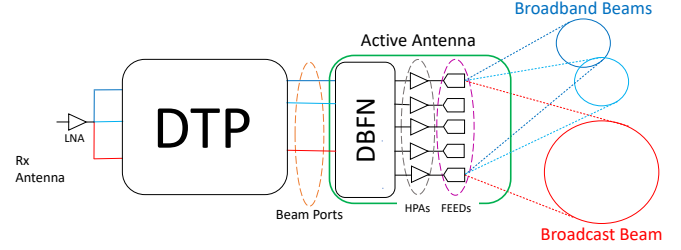

Fig. 1. Illustration of the reference satellite coverage and architectures

is worth to note that this configuration does not require a direct connection between the beams and the feed. Instead, digital beamforming (BF) is employed by all the feeds to cooperatively generate the beams. This special architecture allows to optimize the energy consumption of the high power amplifiers (HPAs) serving the feeds. Furthermore, since the beams are generated by the digital beamforming network (DBFN), they are easy to modified and reconfigure. Such flexibility brings an extremely important feature to satellite operators to reduce controlling overhead in order to better tackle the variation of the traffic demands during the satellite lifetime. This is particularly crucial to edge caching systems, where the user preference (hence content popularity) largely varies cross geographical areas and over time.

III. JOINT CACHING AND BANDWIDTH ALLOCATION FOR FLEXIBLE MULTIBEAM SATELLITES

In this section, we develop a joint resource allocation and caching algorithm, which takes a further step to fully exploit the usage of the wide beam and multi-spot beam that share spectral and power resources through a DTP. In particular, the developed algorithm will jointly determine the caching decisions (files to go through which mode) as well as the required link rate of each mode, subjected to DTP resource constraints. Since our target is to optimize bandwidth to be used by both the multi-beam and the wide-beam modes, we calculate the link budget using the nominal bandwidth and power provided by SES. In addition, we assume that the power spectral density is constant, so that changing the bandwidth we are able to keep constant the signal-to-noise ratio (SNR). In this way the caching algorithm can optimize the bandwidth relying on a fixed SNR value (equivalently the spectral efficiency) for the location of interest.

The considered satellite-assisted edge caching system consists of N CDN caches, which receive a selection of contents via both wide-beam and multi-spot beam coverages of the considered flexible satellite. Each cache serves the users within its coverage who are interested in contents from the common database of F content files. We introduce variable $\boldsymbol{x} \in \{0,1\}^{F \times 1}$ presenting the indexes of files sent through the wide beam and variable $\boldsymbol{y}_n \in \{0,1\}^{F \times 1}, n = 1, \ldots, N$, presenting the indexes of files sent to the n-th cache through the multi-beam channels. In addition, toward intelligent caching we introduce variable $\boldsymbol{x}_n \in \{0,1\}^{F \times 1}, n = 1, \ldots, N$ indicating

the stored file indexes at the n-th cache from the wide beam coverage. Let $\boldsymbol{l}_n \in \mathbb{N}^{F \times 1}, n = 1, \ldots, N$ denote the demands vector, in which the f-th component $\boldsymbol{l}_{n,f}$ is the average number of requests for the f-th content at the n-th CDN. In this off-line setting, the demand vectors are assumed to be obtained from the intelligent unit that predicts the content requests using, e.g., regression model or machine learning techniques[1]. In simulation section, we obtain the demand vectors from the realistic Movielens data set. Therefore, the total number of hits at the CDN n is calculated as $\boldsymbol{l}_n^T(\boldsymbol{x}_n + \boldsymbol{y}_n), \forall n$. Denote $\boldsymbol{q} = [q_1, \ldots, q_f, \ldots, q_F]^T$ as the vector of file sizes, in which q_f is the size of the f-th content file.

A. Minimization of the cache placement time

In this subsection, we aim at minimizing the time duration, τ, that the satellite spends for content fetching while satisfying the minimum cache hit requirements. The joint optimization of bandwidth allocation, cache feeding time, and caching decisions to every CDN is formulated as follows:

$$\underset{\boldsymbol{x},\boldsymbol{x}_n,\boldsymbol{y}_n,\boldsymbol{w},\tau}{\text{Minimize}} \quad \tau, \tag{1}$$

$$\text{s.t.} \quad \boldsymbol{l}_n^T(\boldsymbol{x}_n + \boldsymbol{y}_n) \geq \eta_n, \forall n, \tag{1a}$$

$$\boldsymbol{q}^T \boldsymbol{x} \leq w_0 \gamma_0 \tau, \tag{1b}$$

$$\boldsymbol{q}^T \boldsymbol{y}_n \leq w_n \gamma_n \tau, \forall n, \tag{1c}$$

$$\boldsymbol{q}^T(\boldsymbol{x}_n + \boldsymbol{y}_n) \leq M_n, \forall n, \tag{1d}$$

$$\boldsymbol{x}_n + \boldsymbol{y}_n \leq \mathbf{1}, \forall n \tag{1e}$$

$$\boldsymbol{x}_n \leq \boldsymbol{x}, \forall n, \tag{1f}$$

$$\sum_{n=0}^{N} w_n \leq W_{tot}, \tag{1g}$$

$$\boldsymbol{x}, \boldsymbol{x}_n, \boldsymbol{y}_n \in \{0,1\}^{F \times 1}, \forall n, \tag{1h}$$

where $\boldsymbol{w} = \{w_n\}_{n=0}^N$ is the short-hand notation for the subcarriers, M_n is the cache size of the n-th CDN, and W_{tot} is the total available bandwidth. In the above optimization problem, constraint (1a) is to guarantee the minimum cache hits requirement η_n at every CDN; constraints (1b) and (1c) states that the total amount of sent data can not exceed the caching capacity of the wide-beam and multi-beam during the caching duration, respectively; (1d) sets the maximum of cached data at each CDN not exceeding the cache size M_n; constraint (1e) guarantees efficient caching and avoids one file being sent through both wide-beam and multi-beam; constraint (1f) enables stored files from the wide-beam only if it is sent through the wide-beam; constraint (1g) states that the total bandwidth of the wide-beam and multi beams must not exceed the total bandwidth; finally the last constraint sets binary values of corresponding decision variables.

Although having a linear objective function, solving problem (1) is challenging because of constraints (1b) and (1c). In particular, the presence of products $w_0\tau$ and $w_n\tau$ make constraints (1b) and (1c) non-convex. To tackle this difficulty, we will reformulate these two constraints into a preferable form which can be efficiently handled. By using the equivalent

[1]The joint design for on-line caching setting is left for future work and discussed in Section V

form $w_0\tau = \frac{1}{2}((w_0 + \tau)^2 - w_0^2 - \tau^2)$, constraints (1b) is equivalent to following constraint:

$$(1b) \Leftrightarrow 2\frac{\boldsymbol{q}^T\boldsymbol{x}}{\gamma_0} + w_0^2 + \tau^2 \leq (w_0 + \tau)^2. \tag{2}$$

Because both sides of (2) are convex functions with respective variables, it suggests an successive approximation method to solve (1) iteratively. In particular, let $\bar{w}_0, \bar{\tau}$ are feasible values of constraint (2) at iteration i-th. In the $(i+1)$-th iteration, the first-order approximation of the right-hand-side (RHS) of (2) at $\bar{w}_0, \bar{\tau}$ will be used, which is stated as follows:

$$2\frac{\boldsymbol{q}^T\boldsymbol{x}}{\gamma_0} + w_0^2 + \tau^2 \leq 2(w_0 + \tau)(\bar{w}_0 + \bar{\tau}) - (\bar{w}_0 + \bar{\tau})^2, \tag{3}$$

which is convex because the left-hand-side (LFS) is a convex function and the RHS is linear. Similarly, we can approximate constraint (1c) by the following convex constraint:

$$2\frac{\boldsymbol{q}^T\boldsymbol{y}_n}{\gamma_n} + w_n^2 + \tau^2 \leq 2(w_n + \tau)(\bar{w}_n + \bar{\tau}) - (\bar{w}_n + \bar{\tau})^2. \tag{4}$$

By using (3) and (4) as the inner approximations of constraints (1b) and (1c), respectively, and relaxing the binary constraint (1h), problem (1) can be approximated by a convex optimization problem (5). Because of the nature of the approximation, the optimal solution of (5) depends on the approximated points using in (2) and (4). Therefore, to improve the optimal solution of the approximated problem (5), we propose an iterative Algorithm 1, which consists of a sequence of solving a convex optimization problem. The central principle of the proposed iterative algorithm is to have better initial values $\bar{\tau}, \bar{w}_n$ through iterations, which eventually reduces the gap between the approximated optimal solution and the original problem's solution. It is worthy mentioning that Algorithm 1 outputs continuous values of the caching decision variables $\boldsymbol{x}, \boldsymbol{y}_n, \boldsymbol{x}_n$. Therefore, a binary-recovery step will be performed at the end of Algorithm 1. Furthermore, to actively force the caching variables to be binary, a penalty function should be added to the objective function of (5). More details of penalty function can be found in [12].

Algorithm 1 ITERATIVE ALGORITHM TO SOLVE (1)

1: Initialize $\bar{\tau}, \bar{w}_n, \forall n, \epsilon, T_{old}$ and `error`.
2: **while** `error` $> \epsilon$ **do**
3:      Solve problem (5) to obtain $\tau_k^\star, \boldsymbol{x}^\star, \boldsymbol{x}_n^\star, \boldsymbol{y}_n^\star, w_n, \forall n$
4:      Compute `error` $= |\tau^\star - T_{old}|$
5:      Update $T_{old} \leftarrow \tau^\star; \ \bar{\tau} \leftarrow \tau^\star; \ \bar{w}_n \leftarrow w_n^\star, \forall n$

$$\underset{\boldsymbol{x},\boldsymbol{x}_n,\boldsymbol{y}_n,\boldsymbol{w},\tau}{\text{Minimize}} \quad \tau, \tag{5}$$

$$\text{s.t.} \quad (1a), (1d) - (1g), (3), (4).$$

Proposition 1: The sequence of the objective values generated by Algorithm 1 in solving problem (5) is non-decreasing. The proof of Proposition 1 can be found by employing a technique similar to the one presented in [15, Theorem 3.1], noting that both sides of constraints (3) and (2) are continuously differential. Proposition 1 guarantees the convergence of Algorithm 1 to at least a (local) optimum. Although not guaranteeing the global optimality, it provides justification for the iterative algorithm.

B. Cache Hits Maximization

Although CHR is the popular metric of caching systems, it does not efficiently capture the non-uniform distribution of requests over different geographical areas. Therefore, we choose the cache hits as the performance metric of interest. We aim at jointly optimizing the caching decision and bandwidth allocation of the flexible multi-beam satellites to maximize the total number of cache hits at the CDNs. The joint optimization is formulated as follows:

$$\underset{\boldsymbol{x}, \boldsymbol{x}_n, \boldsymbol{y}_n, \boldsymbol{w}}{\text{Maximize}} \quad \sum_{n=1}^{N} \boldsymbol{l}_n^T (\boldsymbol{x}_n + \boldsymbol{y}_n), \qquad (6)$$
$$\text{s.t.} \quad (1b) - (1h),$$

where the purpose of the constraints are presented in Section III-A. Since both the objective function and constraints are linear, problem (6) is a mixed binary linear problem (MBIP), which can be solved by either standard MBIP tools, e.g., Gurobi and Mosek, or relaxing the binary constraint with penalty function [12].

IV. PERFORMANCE EVALUATION ON REALISTIC SYSTEM PARAMETERS

In this section, we provide performance evaluation of the proposed joint caching and bandwidth allocation design, which exploits the flexible multi-beam satellite. The CHR performance shown in the figures are obtained by the cache hits obtained in Section III-A divided by the total requests.

A. Satellites Coverage

In order to demonstrate the flexible frequency allocation between wide-beam and multi-beam modes, we employ the coverage of the satellite SES-14 [17] on the US east coast as it provides both Ku-band HTS multi-beam footprint and single Ku-beam. Out of all the beams of the HTS system, we down select the 4 beams that cover more or less the same region of the US East monobeam, which are shown in Fig. 2. We used the power flux density information included in these beam patterns, and the per beam power information for calculating the signal to noise ratio (SNR) in any location within the coverage. The SNR at the CDNs' locations varies from 4.08dB to 9.22dB for the monobeam and from 2.48dB to 9.33dB for the multibeam.

B. Movielens-based Content Popularity

To obtain realistic content popularity, we use the 1M data set made available by Grouplens [16], wherein 1 million movie ratings are provided. The data set provides user IDs, user locations (ZIP code), movie IDs, movie genres and the rating time, from which we can compute the distribution of requests (ratings) in arbitrary time duration. Then we use the Zip code information to exact geographical distributions of the requests, by mapping the Zip codes with corresponding Latitude and Longitude components. Because the users in the 1M dataset locate all over the US territory, while the actual satellite's coverage are only the east coast, we only count the users within the satellite coverage when calculating local popularity l_n.
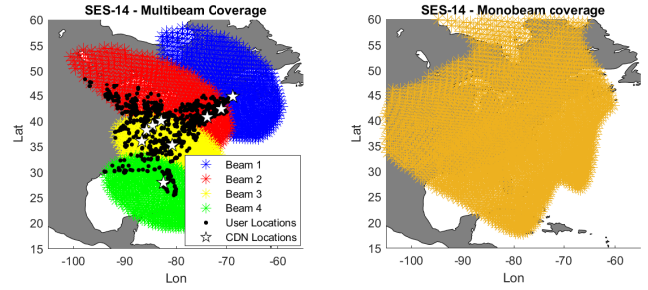

Fig. 2. SES-14 HTS KU multibeams (left) and monobeam (a) coverage on the US east coast. White starts are the CDNs' locations which are in major cities. Black dots are users' location, which are obtained from ZIP code information in the Movielens data set.
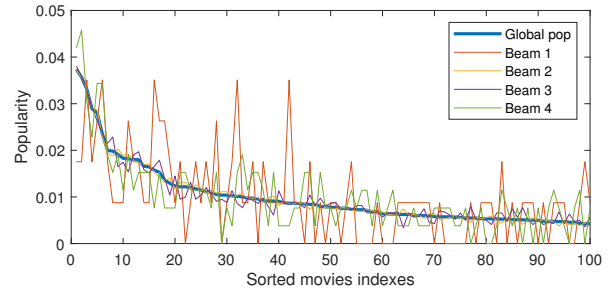

Fig. 3. Content popularity of the considered files on different beams.

Fig. 3 shows the global and local (on each beam) popularities of 100 randomly selected movies. It is observed that the local popularity largely varies over different geographical areas.

C. Performance Evaluations

We conduct numerical results with 9 CDNs which are randomly selected from the satellite coverage. The global content popularity and local popularity on each beam are calculated based on the requests (user ratings in Movielens data set) from 100 movies which are randomly selected and have at least 100 ratings. The movie size is randomly varying between 0.5 GBs and 1 GBs. The cache size at each CDN varies from 5 GBs to 30 GBs.

The proposed joint design is compared with three reference schemes where applicable: i) *Reference 1* which is proposed in [6] and use multi-beam to multicast the cached files; ii) *Reference 2* which uses the whole frequency bandwidth to broadcast the globally popular files via the mono-beam; and iii) *Reference 3* which is a hybrid mono/multi-beam based method [5], [8] in which part of the bandwidth is used for broadcasting the globally popular files and the other part is used for multicasting the locally popular files to each CDN. We remind that for a fair comparison, the total satellite bandwidth of all the schemes remain same.

1) Cache feeding time performance: Fig. 4a shows the time for caching as a function of the required CHR. We note that comparison with the Reference 3 is not available. In general, higher target CHR requires more cache fetching time

as more files need to be sent to the caches. The advantage of the proposed joint design is clearly shown via the smallest caching time of the joint design compared to the references. In particular, at the target CHR of 0.5, the proposed joint design reduces 70% and 50% the caching time compared to Reference 1 and Reference 2, respectively. An interesting observation is that Reference 2 (only uses the mono-beam) is more efficient than Reference 1 (only uses the multi-beam). This is because the caching time of Reference 1 is determined by the worst CDN while the resource allocated to each CDN is not optimized. Fig. 4b shows how the transferred data in the proposed joint design is distributed over different modes, i.e., via wide-beam and multi-beam. As the target CHR increases, more files are sent through the wide-beam because the most popular files are strongly correlated cross different geographical areas.

2) Cache hits performance: In Fig. 5a, we compare the CHR performance of the proposed joint design with three reference schemes. It is observed that the proposed joint design achieves higher CHR compared to the references. This CHR gain comes from the efficient deployment of the multimodal capability of the satellite channel. In particular, two main factors determine the efficiency of the joint design: i) dynamic bandwidth allocation and ii) multimodal satellite channel. The first factor allows efficient bandwidth allocation to the beams with more demands. The second factor better serves some contents which are locally popular in some CDNs. Fig. 5b presents the cached data volumes allocation between the multi-beam and wide-beam modes of the proposed joint design as a function of the CDN's cache size. When the cache size is small in relation to the database size, it is better to utilize both wide-beam and multi-beam modes. When the cache size is large, it can store most of contents sent by the satellites. In this case, using the wide-beam is sufficient.

3) Multi-spot beam vs multicarrier widebeam: In this subsection, we study different configurations of the satellite channel for implementing the multicasting mode in our joint design, namely the multi-spot beam based and the multicarrier wide-beam multicasting. In particular, due to different frequency coloring, frequency reuse can be obtained by the multi-spot beam mode. On the other hand, the multicarrier wide-beam is unable to reuse the frequency. We compare the CHR performance of the proposed joint design under both multi-spot beam and multicarrier wide-beam settings. There are four CDNs in this case, each CDN is served by one beam coverage. We assume that the frequency reuse factor is 2, which means that beam 1 and beam 3 share the same frequency sub-band, and beam 2 and beam 4 operate in the other sub-band. The cache size of the CDN is 30 GBs. The advantage of using the multi-spot beam is clearly shown in the Fig. 6, which outperforms the multicarrier when the cache feeding duration is limited. This gain results from the fact that due to frequency reuse, the multi-spot beam has more bandwidth for multicasting locally popular contents to the CDNs than the multicarrier wide-beam. When the caching time increases, the gain diminishes. This result is expected because a large cache feeding duration can deliver most of popular movies to the CDNs. In this case, additional locally popular movies
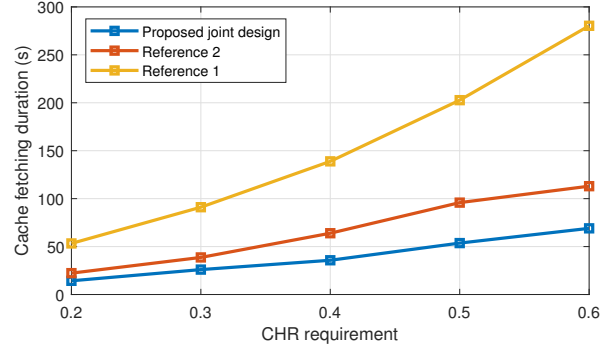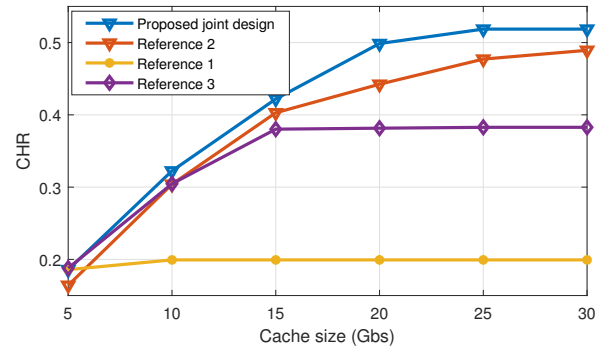
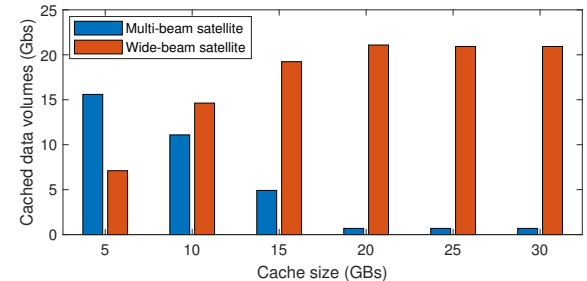

(a) Cache feeding time performance
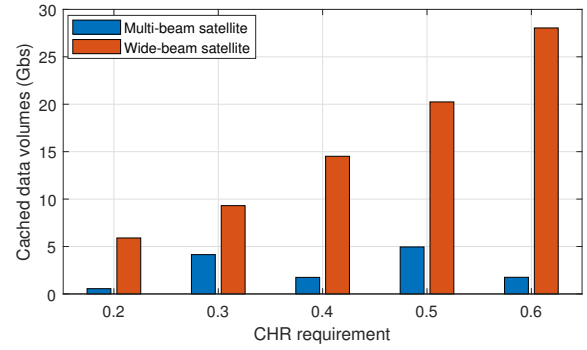
(b) Distribution of cached data

Fig. 4. Caching feeding time comparison of the proposed design with Reference 1 and 2. The Reference 3 is not applicable in this case. The cache size is 30 GBs.

(a) Cache hit ratio comparison

(b) Distribution of cached data

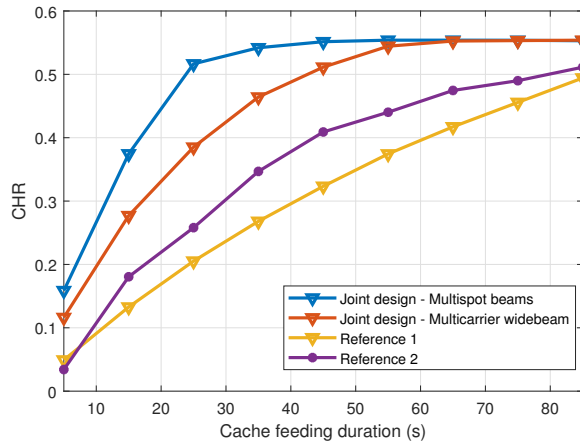Fig. 5. CHR performance comparison. The caching time is equal to 100s.

Fig. 6. CHR performance of the joint design under multicarrier and multi-spot beam configurations. The cache size is 30 GBs.

brought by the multi-spot beam have negligible contribution to the overall CHR.

We would highlight that the considered use case is the worst-case scenario and the performance gap is expected to increase with the multibeam system size. More specifically, the larger the available number of beams is, the more efficient it will become to use the multi-spot beams compared to the multicarrier framework. This can be intuitively explained by the fact that a wide-beam carrier renders the spectrum non-reusable for the entire coverage area, where a spot beam carrier can be reused multiple times depending on the number of supported spotbeams, e.g., the frequency reuse is equal to the number of beams divided by 4 for a 4C system.

V. CONCLUSION AND DISCUSSION

In this paper, we have developed a joint caching and dynamic resource allocation that exploits the flexible multi-beam payload of the next generation multibeam satellites. We optimized the cached contents and the bandwidth allocation between the wide-beam and multi-beam modes in order to maximize the system cache hits and minimize the cache fetching duration. We use the realistic Movielens dataset to extract the content requests as well as the geographical distribution of requests. We demonstrated that the proposed solution achieves about 10% to 20% higher cache hits compared to the reference under realistic request inputs. Furthermore, we demonstrated the benefit of using the multi-spot beam for multicasting to exploit the frequency reuse, compared with the multicarrier wide-beam setting. In addition, we provided a designing guidance for optimizing the system resource to maximize the CHR.

The current work assumes fixed power spectrum density, hence only the frequency bandwidth being optimized. Joint optimization of bandwidth, transmit power and beam coverage is expected to bring further performance gain at an expense of higher computation complexity. Another promising problem is to consider temporary content popularity. In this case, machine-learning based user preference prediction [13], [14]

will be jointly design the dynamic radio resource management of the flexible multibeam satellite.

REFERENCES

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2016-2021," 2017, white paper.

[2] W. Han, A. Liu, and V. K. N. Lau, "PHY-caching in 5G wireless networks: design and analysis," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 30–36, Aug. 2016.

[3] O. Kodheli et al., "Satellite Communications in the New Space Era: A Survey and Future Challenges," *IEEE Communications Surveys& Tutorials*, vol. 23, no. 1, pp. 70-109, Q1. 2021.

[4] B. Evans, M. Werner, E. Lutz, M. Bousquet, G. E. Corazza, G. Maral, and R. Rumeau, "Integration of satellite and terrestrial systems in future multimedia communications," *IEEE Wireless Commun.*, vol. 12, no. 5, pp. 72–80, Oct. 2005.

[5] A. Kalantari, M. Fittipaldi, S. Chatzinotas, T. X. Vu, and B. Ottersten, "Cache-Assisted Hybrid Satellite-Terrestrial Backhauling for 5G Cellular Networks," in *Proc. IEEE Global Commun. Conf.*, Singapore, 2017, pp. 1-6.

[6] C. G. 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. 31st AIAA Int. Commun. Satellite Systems Conf.*, 2013-5664.

[7] H. Linder, H. D. Clausen, and B. Collini-Nocker, "Satellite internet services using DVB/MPEG-2 and multicast web caching," *IEEE Commun. Mag.*, vol. 38, no. 6, pp. 156–161, Jun. 2000.

[8] T. X. Vu et al., "Efficient 5G Edge Caching Over Satellite", in *Proc. Int. Commun.Satellite Syst. Conf.*, Niagara, 2018, pp. 1-5.

[9] T. X. Vu, Y. Poirier, S. Chatzinotas, N. Maturo, J. Grotz, and F. Roelens, "Modeling and Implementation of 5G Edge Caching over Satellite," *Int. J. of Satellite Commun. and Netw.*, Jan. 2020.

[10] T. X. Vu, S. Chatzinotas, and B. Ottersten, "Energy-efficient design for edge-caching wireless networks: When is coded-caching beneficial?" in *Proc. IEEE Int. Workshop Signal Process. Wireless Commun.*, Sapporo, 2017, pp. 1–5.

[11] Glyn Thomas, "Enabling Technologies for flexible HTS payloads," in *Proc. 33th AIAA Int. Commun. Satellite Systems Conf.*, AIAA 2015-4348.

[12] S. Mehrizi, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Trend-Aware Proactive Caching via Tensor Train Decomposition: A Bayesian Viewpoint," in *IEEE OJ-COMS*, vol. 2, pp. 975-989, 2021.

[13] L. Lei, L. You, G. Dai, T. X. Vu, D. Yuan and S. Chatzinotas, "A deep learning approach for optimizing content delivering in cache-enabled HetNet," in *Proc. Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Bologna, 2017, pp. 449-453.

[14] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A learning-based approach to caching in heterogenous small cell networks," *IEEE Transactions on Communications*, vol. 64, Apr. 2016.

[15] H. A. Le Thi, V. N. Huynh, and T. P. Dinh, "DC programming and DCA for general DC programs," in *Advanced Computational Methods for Knowledge Engineering*. Cham: Springer International Publishing, 2014, pp. 15–35.

[16] https://grouplens.org/datasets/movielens/1m/

[17] SES-14 Satellite, https://www.ses.com/our-coverage#/explore/satellite/369