Modeling and Implementation of 5G Edge Caching Over Satellite

Thang X. Vu[1], Yannick Poirier[2], Symeon Chatzinotas[1], Nicola Maturo[1], Joel Grotz[3], and Frederic Roelens[2]

[1] Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg

[2] Inverto/FTA, Luxembourg

[3] SES Engineering, L-6815 Betzdorf, Luxembourg

Email: thang.vu@uni.lu

Abstract

The fifth generation (5G) wireless networks have to deal with the high data rate and stringent latency requirements due to the massive invasion of connected devices and data-hungry applications. Edge caching is a promising technique to overcome these challenges by prefetching the content closer to the end users at the edge node's local storage. In this paper, we analyze the performance of edge caching 5G networks with the aid of satellite communication systems. Firstly, we investigate the satellite-aided edge caching systems in two promising use cases: a) in dense urban areas, and b) in sparsely populated regions, e.g., rural areas. Secondly, we study the effectiveness of satellite systems via the proposed satellite-aided caching algorithm, which can be used in three configurations: i) mono-beam satellite, ii) multi-beam satellite, and iii) hybrid mode. Thirdly, the proposed caching algorithm is evaluated by using both empirical Zipf-distribution data and the more realistic Movielens dataset. Last but not least, the proposed caching scheme is implemented and tested by our developed demonstrators which allow real-time analysis of the cache hit ratio and cost analysis.

I. Introduction

During the past 40 years, the Internet has followed an extraordinary evolution and has become an integral part of the modern society. However, this evolution has kept momentum and there

Part of this work has been presented in the 36th ICSSC conference [24].

are constantly new services and contents distributed through this global communication network. Based on Cisco's report [1], it is predicted that the mobile data traffic will grow 74% by 2021. Particularly, the mobile video will increase eleven-fold between the mentioned years. The main causes of this traffic growth are the vast availability of mobile devices, e.g., smart phones, tablets, and notebooks, as well as the fast growth of video content on the Internet and their increasing quality. Using these mobile devices, more and more users are immigrating from traditional linear broadcasting services (TV channels) to streaming services, such as YouTube and NetFlix. Another factor that contributes to the traffic is the increasing video quality, i.e., 3D, 4K video, Virtual Reality etc., which can be translated to increased bandwidth requirements for both the core and access networks. This perspective seems very promising for content providers, since they can provide improved services and reap the benefits either through subscriptions or advertising. Nevertheless, looking at this from a telecom operator's point of view, it is obvious that video traffic will become a bottleneck and put excessive strain on current communication infrastructure. On the other hand, video traffic also results in revenue growth. However, telecom operators do not have direct access to the revenue generated through video content delivery and as a result they cannot use it to upgrade their infrastructure. In parallel, spectrum has become scarce and the operators cannot easily access new frequency bands to expand their wireless access and backhaul segments.

These are the reasons why in the last few years, telecom operators started to build their own content delivery networks (CDNs). The aim of a CDN is to serve contents to end-users with high performance by using edge catching. The main benefit of CDNs is represented by the higher degree of cross-optimization between the physical infrastructure and the network service that leads to an improved transmission efficiency. It becomes obvious that closer interaction between operators and content providers will be needed in order to optimize content delivery and overcome the projected bottleneck due to video traffic [3]. Some steps have been made in this direction, for instance Dhiraagu, the Maldives operator, have deployed Google caches successfully in its network [2].

One of the challenges in the edge caching is how to effectively prefetch the popular content to the caches considering the high volume of data [4]. In order to overcome this issue, we propose to use satellite backhauling for cache placement phase to exploit the large coverage of the satellite beams. Satellite systems have the ability to provide high throughput links and to operate in multi/broad-cast modes for immense area coverage.

Due to their multi-hop unicast architecture, the cached content via terrestrial networks has to go through multiple links and has to be transmitted individually towards each base station (BS). On the other hand, with wide area coverage, the satellite backhaul can broadcast content to all BSs or multi-cast contents to multiple groups of BSs. Therefore, bringing these two technologies together can further off-load the network. The main idea is to integrate the satellite and terrestrial telecommunication systems in order to create a hybrid federated content delivery network, which can improve the user experience. The joint deployment of satellite and terrestrial networks can be found in [7], [8]. In this paper, we consider the satellite channel as the only mean for cache placement. The application of satellite communications in feeding several network caches at the same time using broad/multi-cast is investigated in [5], [6], [7]. The work of [6] proposes using the broad/multi-cast ability of the satellite to send the requested contents to the caches located at the user side. Online satellite-assisted caching is studied in [7]. In this work, satellite broadcast is used to help placing the files in the caches located in the proxy servers. Each server uses the local and global file popularity to update the cache.

In this work, a satellite-aided caching algorithm is proposed. We use off-line caching approaches [8], [9], [10], [11], [12] to off-load the backhaul of the terrestrial network. We focus on the role of multimodal satellite backhauling, which provides flexible backhaul's transmission modes, e.g., broadcast and broadband, and its effectiveness on edge caching. The proposed algorithm can be used with three different satellite's configurations: i) broadcast mono-beam, ii) broadband multi-beam where the content of each beam can be selected independently, and iii) hybrid mode that is a combination of the first two modes. Focusing on promising satellite markets, we evaluate the performance of the proposed caching algorithm in two use cases: in dense urban areas and in sparsely populated rural regions. Based on both empirical Zipf-distribution data [13] and the more realistic Movielens dataset [14], we show via numerical results that the multi-beam satellite will outperform the mono-beam system when the demands are less correlated, and that the hybrid achieves a cache hit ratio between the multi-beam and mono-beam schemes. These observations are very much dependent on the popularity of the content. Despite that, the proposed caching algorithm is capable of adapting to different means of content delivery to optimize the system cost function. Furthermore, the proposed caching algorithm is implemented in our developed demonstrators which enable real-time analysis of the cache hit ratio and cost analysis.

The rest of the paper is organized as follows. Section II provides technological enablers for caching over satellite. Section I describes the system parameters and the caching algorithm. Section IV presents the FLUTE protocol which will be used in our demonstrator. Section V provides cost analysis for different transmission modes. Section VI presents numerical results. The implementation and tested results are given in Section VII. Finally, Section VIII concludes the paper.

II. Technological Enabler for Satellite-assisted Edge Caching

A. Hybrid Satellite

The satellite architecture in general can be classified in two traditionally different cases: satellite with a very wide single beam, mainly used for broadcast services, and satellite with many small beams, mainly used for broadband services (see Fig. 1 for example of satellite coverage). While the distinction between these two types of architectures is well established nowadays, in the future satellite system there may be no real distinction between broadcast and broadband payload. Thanks to the adoption of new technologies, like digital transparent payload (DTP) and (semi-)active on-board antennas, it will be possible to have both services sharing the same hardware power and spectrum resources in reconfigurable hybrid broadcast/broadband payloads. Many satellite manufacturers are currently working towards this new type of payload [15] and in Fig. 2 we provide a pictorial representation of this hybrid architecture.

As an example shown in Fig. 2, the satellite is receiving 3 different streams: 2 broadband streams (the light and the dark blue lines) and 1 broadcast stream (the red line). At a beam ports level (the output of the DTP) we still have 3 different streams that are the input of the active antenna. Thanks to the active antenna technology, it will be possible to create overlapping beams of variable granularity and use a single broadcasting system to drive a large broadcast beam and at the same time it will be possible to generate on the same payload separated broadband beams. The flexible payload would be able also to shape the broadcasting beams to concentrate the transmitted power where it is needed. It is interesting to note that in this configuration there isn't a direct relation between the beams and the feed, but basically all the feeds cooperate to create all the beams using digital beamforming (BF). In this way is possible to optimize the power consumption of the high power amplifiers (HPAs) serving the feeds.

Because the creation of the beams is driven by the digital beamforming network (DBFN), the beams design can be very easily reconfigured. This type of flexibility is certainly extremely important for satellite operators, so that they can modify their satellite configuration in order

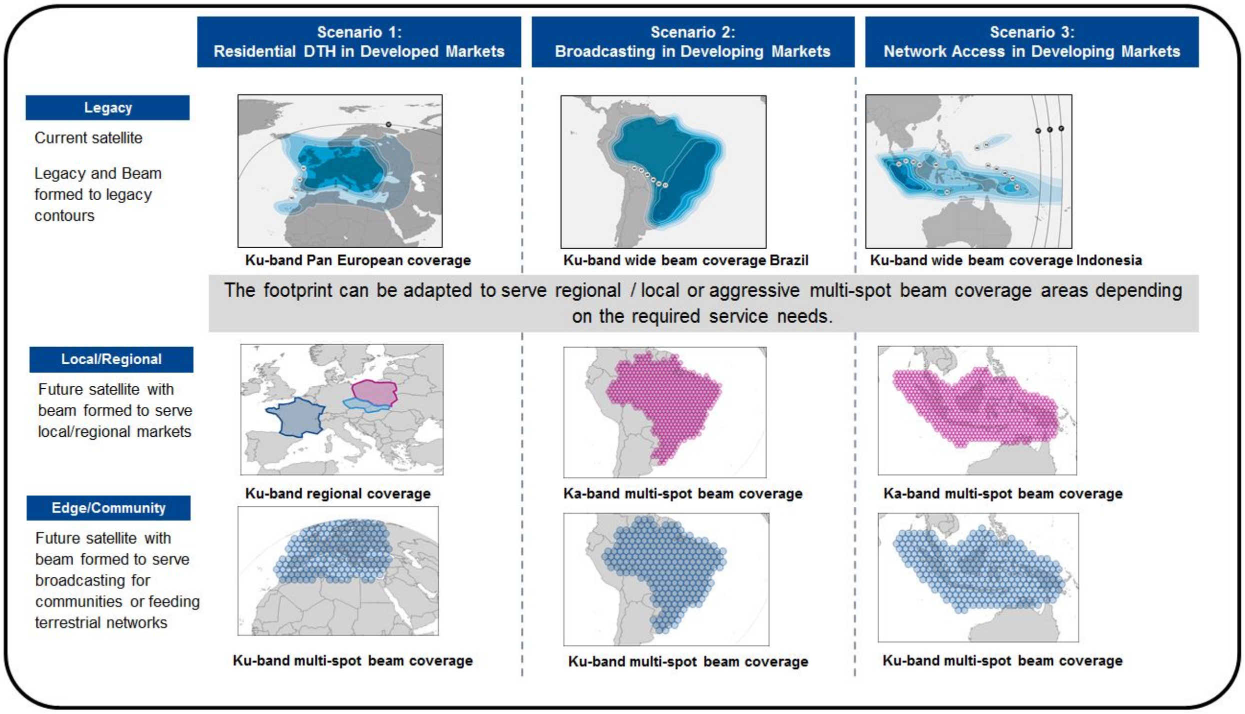

Scenario 1: Residential DTH in Developed Markets
Scenario 2: Broadcasting in Developing Markets
Scenario 3: Network Access in Developing Markets
Legacy
Current satellite
Legacy and Beam formed to legacy contours
Ku-band Pan European coverage
Ku-band wide beam coverage Brazil
Ku-band wide beam coverage Indonesia
The footprint can be adapted to serve regional / local or aggressive multi-spot beam coverage areas depending on the required service needs.
Local/Regional
Future satellite with beam formed to serve local/regional markets
Ku-band regional coverage
Ka-band multi-spot beam coverage
Ka-band multi-spot beam coverage
Edge/Community
Future satellite with beam formed to serve broadcasting for communities or feeding terrestrial networks
Ku-band multi-spot beam coverage
Ku-band multi-spot beam coverage
Ku-band multi-spot beam coverage

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

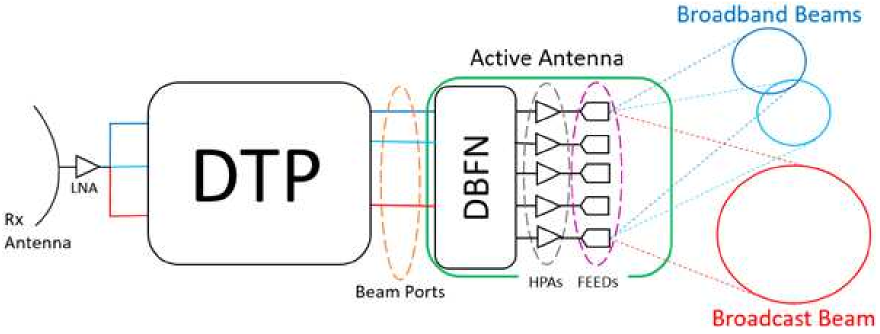


Fig. 2: Hybrid Satellite Architecture

to better accommodate the variation of the traffic demands during the satellite lifetime. In this context we refer in particular to the content-related traffic evolution determined by various factors such as population evolution and changes in the content consumption preferences.

B. Integrated Satellite-Terrestrial System

The classic cellular backhaul comprises of fibre, copper lines and wireless links, which can efficiently transfer the backhaul traffic from point to point. The multimodal architecture enhances the conventional backhaul by overlaying wide satellite beams and narrower broadcasting cells (e.g. HPHT), which are capable of P2MP (point to multi-point) multi/broadcasting. Studies in [16] have shown that a small percentage of extremely popular video files can pose a huge load on current networks that do not support broadcasting on the physical layer. The proposed satellite overlay caching solution can relieve this load by simultaneously and efficiently distributing popular content the edge caches using traditional P2P backhauling with broad/multicasting P2MP backhauling. It should be noted that these multi/broadcast systems can inherently reach a large number of BSs with a single PHY-layer transmission in contrast to the NET-layer multicasting which implies packet replication and multiple individual PHY connections. In addition to above mentioned benefit, this converged solution brings several distinct benefits: 1) additional backhaul capacity based on existing infrastructure, 2) spectrally-efficient physical-layer multi/broadcasting, 3) variable cell sizes for broadcast backhauling (wide coverage for satellite, narrow coverage for high broadcasting towers).

III. System model and caching algorithm

We consider a satellite system serving the users within its coverage via a number of edge nodes, e.g., base stations (BS), which are equipped with a satellite receiver. The satellite is capable to deliver content via both mono-beam and multi-beam transmissions. In order to exploit wide coverage, globally popular contents will be prefetched via satellite links, whereas locally popular contents will be transmitted via terrestrial networks, as depicted in Fig. 3. The considered system can find application in both urban and rural areas. In the former, each BS serves a large number of users, while in the latter, there are less users served by one BS. Each BS has a local storage which can store up to M bits. The users are interested in requesting contents from a library $\mathcal{F} = \{1, \dots, f, \dots, F\}$ consisting of F files. Let Q_f denote the size (in bits) of file f.

We consider offline caching which consists of two consecutive placement phase and delivery phase. The cache placement phase is executed periodically in off-peak hours [8], e.g., from midnight to early morning. Our focus is to design efficient cache placement phase in order to exploit the benefit of the satellite channels. Denote C as the caching capacity of the satellite links, which is the maximum load the satellite channel can deliver during the placement phase. In addition, we consider three operating modes of the satellite channels: mono-beam, multi-beam, and hybrid.

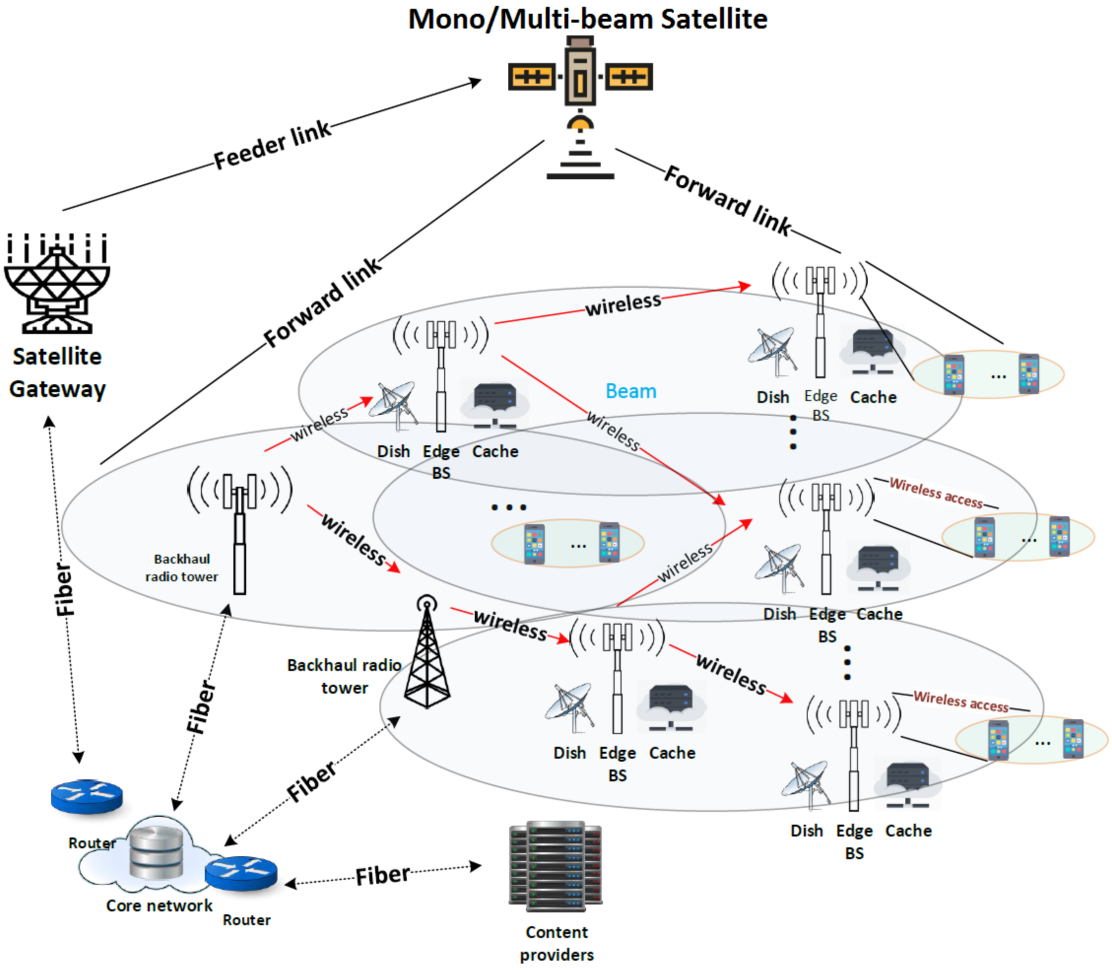


Fig. 3: Hybrid satellite-terrestrial network using a multi-beam satellite for more accurate content placement phase. Globally popular contents are sent via satellite channels, while locally popular contents can be sent via either terrestrial network or multi-beam satellite.

A. Caching via Broadcast mono-beam

With mono-beam satellite mode, both broadcast and multicast are possible. However, the efficiency of the multicast in this mode is in general lower than in multibeam systems due to the wide beam. This is the reason why in our study, we assume that the monobeam is used for broadcasting. The cache placement is done based on the global popularity.

Based on the user preference sent to the network controller via the return link, a local user preference (content popularity) is obtained, which is then used to determine caching strategy. The local popularity is estimated based on the user requests sent to the corresponding BS in the previous phase. Denote $n_{k,f}$ is the number of requests for file f from BS k. The local content popularity at BS k is computed as $\mathbf{p}_k = [p_{k,1}, \dots, p_{k,F}]$, where

$$p_{k,f} = \frac{n_{k,f}}{\sum_{f=1}^{F} n_{k,f}}.$$

The global content popularity is computed from local popularities, $\mathbf{p}_k$, given as $\mathbf{p}_G = \frac{1}{K}\sum_{k=1}^{K} \mathbf{p}_k$, where K is the number of BSs. The gateway then use $\mathbf{p}_G$ to determine the most popular contents to be sent via the satellite channel as follows. First, the gateway obtains $\tilde{\mathbf{p}}_G = \pi(\mathbf{p}_G)$, $\pi()$ denotes the sorting operator. Denote $\Psi(m; \tilde{\mathbf{q}}_G) = \sum_{f=1}^{m} Q_f \mathbf{1}_{\tilde{q}_{G,f}}$ as the total file size (in bits) of the first m contents in $\tilde{\mathbf{p}}_G$, where $\mathbf{1}_x \triangleq \mathtt{sgn}(x)$. Then the gateway chooses

$$m_G = \arg\max_m \{\Psi(m; \tilde{\mathbf{q}}_G) \mid \Psi(m; \tilde{\mathbf{q}}_G) \leq C\}$$

first files in $\tilde{\mathbf{p}}_G$ to broadcast over the satellite channel. Each BS k will choose the first m_k files of $\tilde{\mathbf{p}}_k = \pi(\mathbf{p}_k)$ to be stored in the kth BS's cache, where $m_k = \arg\max_m \{\Psi(m; \tilde{\mathbf{q}}_k) \mid \Psi(m; \tilde{\mathbf{q}}_k) \leq M\}$. Obviously, if $C > \sum_{f=1}^{F} Q_f$, all the BSs can cache the most locally popular files in their cache. However, when $C < \sum_{f=1}^{F} Q_f$, some BSs might not be able to cache the most popular files (locally) if $\{\mathbf{p}_k\}$ are uncorrelated, hence the CHR during the delivery phase is degraded.

The cache hit ratio (CHR) is the main performance indicator to be considered. It is defined as the ratio between the number of requests served by the local cache, divided by the total number of requests:

$$CHR = \frac{\sum_{k=1}^{K} \bar{n}_k}{\sum_{k=1}^{K} n_k}, \tag{1}$$

where $\bar{n}_k$ is the number of requested files available in the kth local cache, and n_k is the total number of requests at the kth BS.

B. Caching via Broadband multi-beam

The CHR performance can potentially be improved by using multi-beam satellite as the content deliver is done per beam and not for the whole widebeam. For a satellite with multi-beam and flexible beam coverage, the caching can target, in fact, a subset of beams. It is often expected that the narrower beam can reach a higher spectral efficiency, compared to the global beam, while utilizing similar satellite resources in terms of power and bandwidth. However, the use of multi-beam system may introduce additional inference due to the reuse of the spectrum. The level of co-channel interference will depend on the isolation among different beams, the beam size, etc. In our analyses, the efficiency of the link is taken into account as a configurable parameter.

TABLE I: CACHING ALGORITHM VIA SATELLITE

```
Input: $C$ - caching capacity, $M_k, \forall k$ - cache size at the base station, $n_{k,f}$ - number of requests for file $f$ from BS $k$, $L$ - number of beams.
1. BS $k$ estimates $\mathbf{p}_k = [p_{k,1}, \dots, p_{k,F}]$, where $p_{k,f} = \frac{n_{k,f}}{\sum_{f=1}^{F} n_{k,f}}$, then forward it to the gateway
2. if $(\texttt{mono} - \texttt{beam})$
   2.1. The global cache manager (GCM) at the gateway computes the global popularity $\mathbf{p}_G = \frac{1}{K}\sum_{k=1}^{K} \mathbf{p}_k$
   2.2. The GCM sorts $\mathbf{p}_G$ in the decreasing order to obtain $\tilde{\mathbf{p}}_G$
   2.3. The GCM chooses the first $m_G = \arg\max_m \{\Psi(m; \tilde{\mathbf{p}}_G) \mid \Psi(m; \tilde{\mathbf{p}}_G) \leq C\}$ files in $\tilde{\mathbf{p}}_G$ to broadcast for
        caching, where $\Psi(m; \tilde{\mathbf{p}}_G)$ denotes the total volume of the first $m$ files in $\tilde{\mathbf{p}}_G$.
   2.4. Each BS $k$ sorts $\mathbf{p}_k$ in the decreasing order, and chooses the first $m_k =$ files in the sorted local popularity
        for caching, where $m_k = \arg\max_m \{\Psi(m; \tilde{\mathbf{p}}_k) \mid \Psi(m; \tilde{\mathbf{p}}_k) \leq M\}$.
        A file is cached at the local cache of BS $k$ if it is in the list and is sent by the GCM
3. else_if $(\texttt{multi} - \texttt{beam})$
   3.1. The GCM constructs $L$ beams. $\mathcal{K}_l$ denotes the set of BSs in the $l$-th beam.
   3.2. For each beam, the GCM applies the caching policy in step 2.
4. else_if $(\texttt{hybrid} - \texttt{mode})$
   4.1. Determine $C_{\text{mono}}, C_{\text{mul}}$ as the caching capacity in mono-beam and multi-beam satellites, respectively.
   4.2. The GCM applies step 2 to with the caching capacity $C_{\text{mono}}$ to determines the files being broadcasted
        via mono-beam satellite, denoted by $\mathcal{F}_{\text{mono}}$
   4.3. The GCM excludes files in $\mathcal{F}_{\text{mono}}$ from the requests $\{n_{k,f}\}_{\forall k,f}$, then applies step 3 with the caching
        capacity $C_{\text{mul}}$.
```

Let L denote the number of beams. We define the number of files that each beam is capable of delivering in the placement phase as the caching capacity C bits. Denote $\mathcal{K}_l$ as the set of BSs in the l-th beam. The gateway will calculate the popularity of the user requests within each beam as follows: $\mathbf{p}_G^l = \frac{1}{K_l}\prod_{k \in \mathcal{K}_l} \mathbf{p}_k$. Then the l-th beam will apply the same caching technique as in the previous subsection to broadcast the most popular contents this beam.

C. Caching via Hybrid design

In the hybrid setting, some contents can be prefetched via the mono-beam mode, while the rest are sent via the multi-beam satellite. A similar approach was proposed for the terrestrial Centralized-RAN (CRAN) architecture in [19]. Denote $C_{\text{mono}}, C_{\text{mul}}$ as the caching capacity of the mono-beam satellite and multi-beam satellite, respectively. First, the mono-beam caching algorithm in Section III-A is applied subject to the caching capacity C_{mono}. The cached files after this phase are removed from the requests. Next, the multi-beam caching algorithm in

Section III-B with the caching capacity C_{mul} bits is used on the remaining requested files. The proposed caching algorithm is summarized in Table I.

IV. FLUTE PROTOCOL FOR CONTENT DELIVERY OVER SATELLITE

In this section, we briefly introduce the file delivery over unidirectional transport protocol (FLUTE), which will be used to implement the caching algorithm over satellite channels. The principle of FLUTE enables scalability and realizability which are suitable for broadcast networks [20]. Since FLUTE is built on top of asynchronous layered coding protocol (ALC), it permits to transfer binary objects with multiple rate congestion control and application-level forward erasure correction (AL-FEC) to an unlimited number of concurrent receivers from a single sender.

A complete FLUTE-based file transfer protocol consists of a FLUTE sender and multiple FLUTE receivers. The former is responsible for encoding data with proper code rates to guarantee some given reliability and sending the coded files over the network. The FLUTE sender is to able send a collection of files in the form of packets. In particular, multiple files are sliced into blocks, which are then sliced into packets, as depicted in Fig. 4. Blocks and packets are numbered by SND and ESI, respectively. It is worthy noting that the FLUTE packets are compatible with UDP/IP protocol. The FLUTE receivers receives UDP/IP packets which contain ALC payload. The received packets are then reordered and reconstructed into blocks, as shown in . Fig. 5. Since employing a Al-FEC decoder, some missing packets can be reliably recovered.

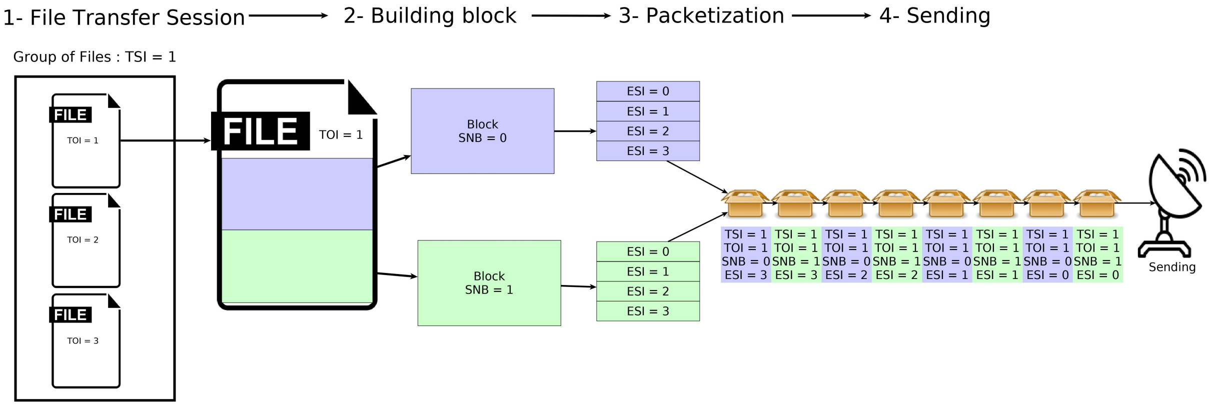

1- File Transfer Session
2- Building block
3- Packetization
4- Sending
Group of Files : TSI = 1
FILE
TOI = 1
FILE
TOI = 2
FILE
TOI = 3
FILE
TOI = 1
Block
SNB = 0
Block
SNB = 1
ESI = 0
ESI = 1
ESI = 2
ESI = 3
ESI = 0
ESI = 1
ESI = 2
ESI = 3
TSI = 1 TOI = 1 SNB = 0 ESI = 3
TSI = 1 TOI = 1 SNB = 1 ESI = 3
TSI = 1 TOI = 1 SNB = 0 ESI = 2
TSI = 1 TOI = 1 SNB = 1 ESI = 2
TSI = 1 TOI = 1 SNB = 0 ESI = 1
TSI = 1 TOI = 1 SNB = 1 ESI = 1
TSI = 1 TOI = 1 SNB = 0 ESI = 0
TSI = 1 TOI = 1 SNB = 1 ESI = 0
Sending

Fig. 4: FLUTE sender.

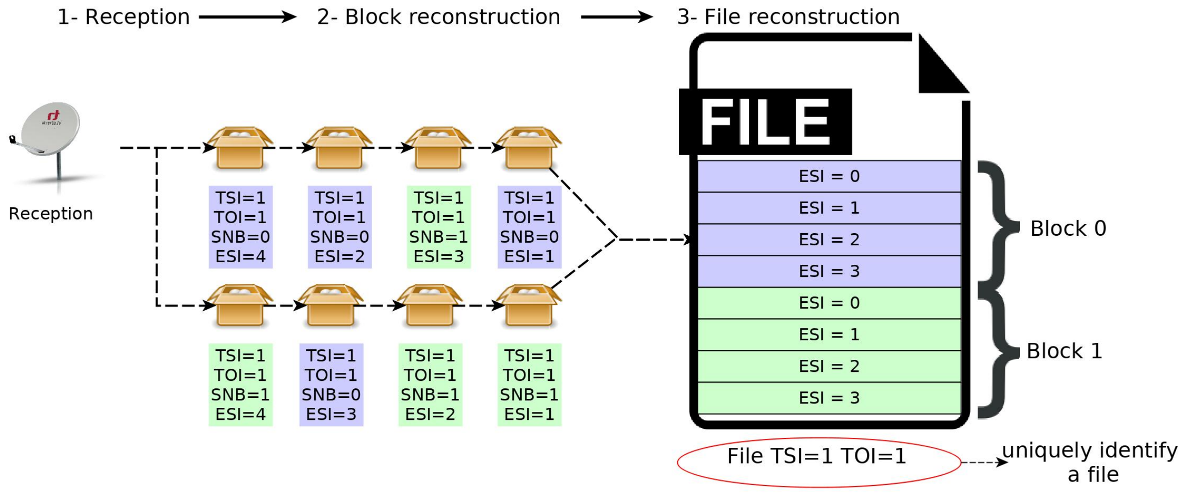


Fig. 5: FLUTE receiver.

A. FLUTE/DASH protocol

Media contents from movie database are encoded using dynamic adaptive streaming over HTTP format (DASH). DASH specifies XML and binary formats that enable the delivery of media content from standard HTTP servers to HTTP clients [21]. DASH permits to encode multiple representations of the content with different bit rates or resolutions. The media content is sliced in small pieces of files called segments. Multiple segments will be transferred via FLUTE as binary files. DASH contains the media presentation description (MPD), an XML document that describes how media segments are relative to each other. It contains metadata about the content and information permitting to select video, audio, caption components for the clients.

A content in DASH format is transferred over multiple FLUTE sessions. MDP and service-based transport session instance description (S-TSID) are transferred over a FLUTE session with a transport session identifier (TSI) of 0. Then each media representation is transferred over its own FLUTE session. [22]. FLUTE receivers must listen to FLUTE session with TSI equal to 0 and acquire the MPD and S-TSID. Once these two files are received, the S-TSID is used to determine the mapping between a DASH media representation and a FLUTE session. Details of the FLUTE/DASH protocol is depicted in Fig. 6.

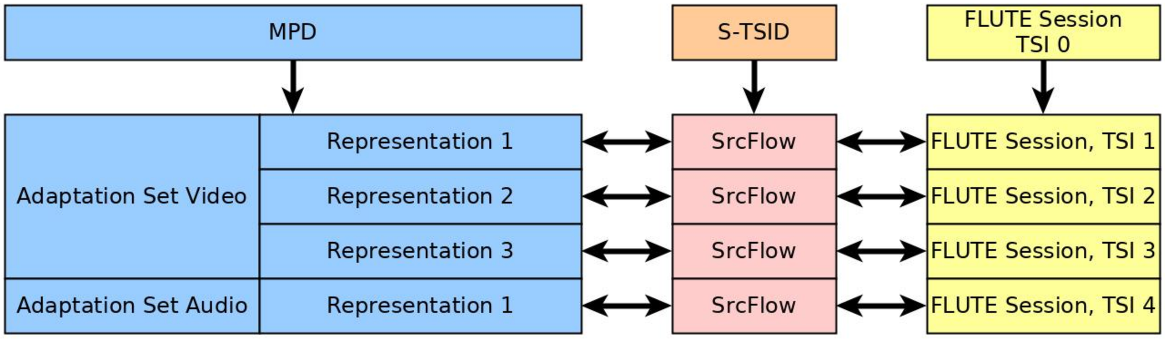


Fig. 6: Mapping between DAHS and FLUTE.

B. NACK protocol

Since negative acknowledgment protocol (NACK) is unavailable in the standard FLUTE, we have implemented NACK on top of the FLUTE protocol to improve the file transfer reliability as it allows to request and retransmit lost packets at an expense of a return channel. NACK protocol is based on the file repair procedure defined in TS 126 346 [23, chapter 9]. File repair capability is signaled by the FLUTE sender in an associated procedure description instance (APD) file. The APD is an XML file that is transmitted over FLUTE in a carousel. The APD contains file repair properties needed by the FLUTE receivers in order to initialize the file repair procedure.

In this work, we implement two file transfer protocols. The first one uses HTTP client/server to send NACK and receive responses. The second one is based on UDP. NACK requests are encapsulated in a single UDP packet and then the repair packets can be transmitted over UDP via unicast/multicast links.

V. Cost per bit calculation

In order to define the cost per bit for the satellite transmission, different aspects need to be considered, such as the entire service costs of the transmission and the capacity of the satellite system for all the possible transmissions of services.

For contents distribution service, we assume that the transmission is organized in a multicast manner allowing many terminals to receive the same content. In the following, we will so compare the cost per transferred bit between a widebeam multicarrier mode and a multibeam single carrier mode. For this we need to consider the required resources of the satellite and take into account the link budget results to assess the overall picture. This is a first order assessment of the cost per bit compared to what the total capacity of the satellite is, so it does not consider eventual opportunity costs resulting from specific operational contexts or second order effects that one service might have on alternative usage of the capacity. This is a comparison on the basis of the total satellite cost per transmitted information bit.

We assume multi-cast link budget assessment results indicating the required E_S/N_0 values for both the considered transmission modes. From this assessment, we derive the transmission efficiency, taking into account all overhead parameters required.

The power and bandwidth used in the link budgets to assess the transmission efficiency have to be considered. If it is multiplexed with other traffic, the proportional DC power resources and bandwidth utilization resources is taken in to account. The DC power equivalent is deduced from the assessment of the power assessment taking into account the amplifier efficiency and output backoff (OBO) of the amplifiers required. The total bandwidth available over all beams is considered as bandwidth reference for the satellite. Here the total bandwidth of the simultaneously transmitted traffic has to be considered for all possible beams which are active at the same time.

The total cost of the transmission can be estimated based on the reference of same satellite services,which takes into account CAPEX (and OPEX in a refined model, but not required in first order approximation) costs with related assumptions on fill rates per lifetime. This is deduced as cost per Mbps assessment for the computed traffic load. We may consider different levels of costs to take into account the fact that an underutilized time period could be used for the data transfer. The total is a $COST_{LINK}$ result as $RESOURCE_{PERCENT}$ * $COST_{SAT}$.

Then the total file transfer cost results in:

$$COST_{file} = Size * COST_{LINK}/Efficiency * Bw.$$

The relative cost comparison between mono-beam and multi-beam used in the simulation is given in Table II.

VI. Performance Evaluation via Numerical Results

In this section, we evaluate the performance of the proposed caching algorithms in two forms of numerical results and file transfer implementation.

In this subsection, the cache hit ratio (CHR) performance of the proposed caching algorithms is evaluated via numerical results. The user's requests follow Zipf distribution with the skewness factor equal to 0.8. The library consists of $F = 500$ files of equal size of Q bits. Since the

TABLE II: RELATIVE COMPARISON FOR COST PER BIT

	Widebeam	**Multibeam**	**Unit**
	Multicarrier	**Singlecarrier**	
E_S/N_0 (worst case)	8.0	7.5	dB
Efficiency	2.4	2.1	bps/Hz
Total Sat RF Power	8556	5398	W
Total Sat Bandwidth	4320	80000	MHz
Beams per coverage	1	10	#
Power per beam	120	60	W
Bandwidth per beam	60	500	MHz
Power Ratio	1.4%	11.1%	
Bandwidth Ratio	1.4%	6.3%	
$RESOURCE_{PERCENT}$	1.4%	11.1%	
$COST_{LINK}$	0.014	0.111	relative
Size	5000	5000	Mbyte
$COST_{file}$	3.89	4.19	
Cost w.r.t. widebeam	1	1.09	relative

files have the same size, the storage capacity is normalized by the file size for simplicity. All the BSs are equipped with a cache of size M (files). To make a fair comparison, the caching capacity in mono-beam and multi-beam settings are chosen such that both schemes have same total placement cost. This means that we fix the caching capacity of the monobeam mode and we scale the caching capacity of the multibeam mode in accordance with the results of Table II. In particular, let C_{mono} and C_{mul} denote the caching capacity in the mono-beam and multi-beam modes, respectively. The caching capacity of the hybrid mode is calculated in accordance with the usage percentage of the monobeam, $C_{\text{hybrid},1}$, and the multibeam mode, $C_{\text{hybrid},2}$. In order to meet the same placement cost, we have $C_{\text{mono}} = 1.09C_{\text{mul}}$ = $C_{\text{hybrid},1} + 1.09C_{\text{hybrid},2}$. The satellite provides service for an area consisting of 1000 BSs, each is serving 2000 users in use case 1 (e.g., urban area) and 400 users in use case 2 (e.g., rural area).

Fig. 7 presents the CHR for the two considered use cases. The total caching capacity is equal to 200 files in use case 1, and 100 files in use case 2. We assume that the user requests form four geographical regions which are weakly correlated. More specifically, 10% of the files is globally popular across all regions. The popularity of other files are randomly assigned for each region. The multi-beam mode sends the data via 4 beams. It is observed that the multi-

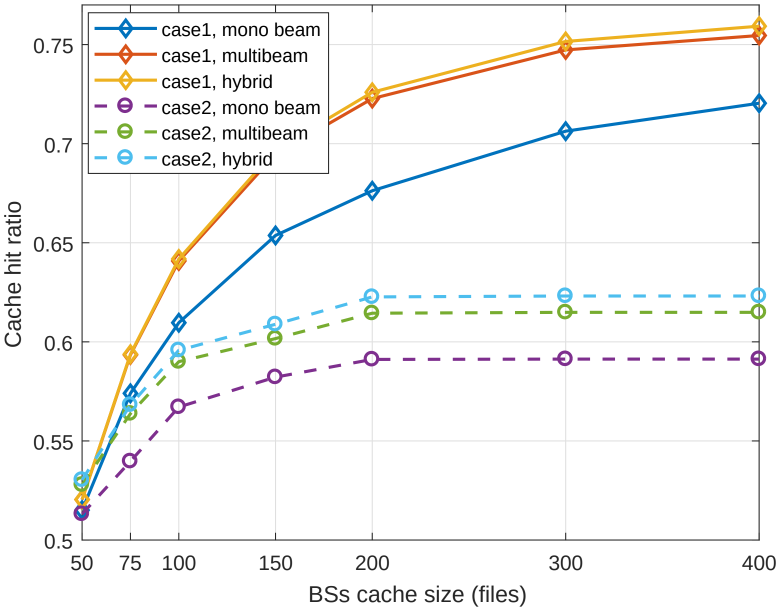


Fig. 7: CHR performance of the proposed caching algorithm in two use cases. For each case, the caching algorithm is designed for three different beam settings: mono beam only, multi-beam only, and hybrid mono beam and multi beam.

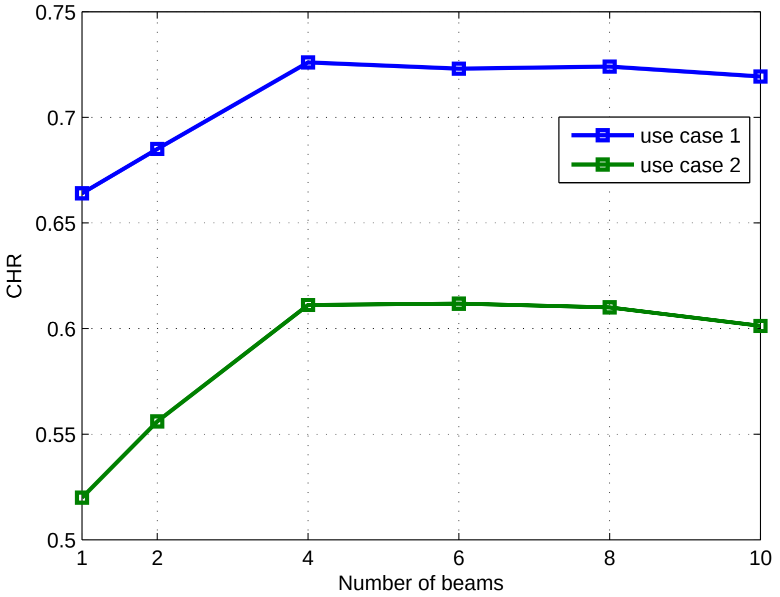


Fig. 8: CHR performance of the multi-beam scheme v.s. number of beams. The user requests form 4 weakly correlated regions.

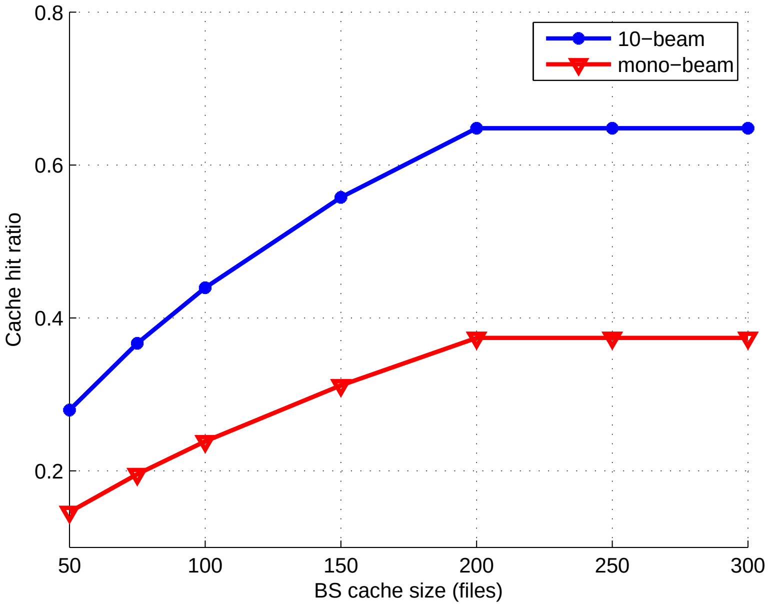


Fig. 9: CHR performance based on Movielens dataset.

beam satellite achieves a CHR significantly larger than the mono-beam scheme. This is because the multi-beam satellite can better match the local popularity with the user demands. More importantly, the proposed hybrid mode, which prefetches the cache via both the mono-beam and the multi-beam satellite, can further improve the CHR performance. The usage percentage of the mono-beam mode in the hybrid scheme is optimized and equal to the most popular files in all regions, e.g., $C_{\text{hybrid},1} = 50$ and $C_{\text{hybrid},2} = 138$ in use case 1. It is also shown that usecase 1 outperforms usecase 2 due to larger caching capacity. Indeed, this initial result shows one of the major advantages of the hybrid and multi-beam caching with respect to the mono-beam caching, i.e. the access to the geographical diversity of the content. This way, the hybrid and multibeam modes can cache the less correlated content popularity in a more efficient manner.

Fig. 8 shows the CHR of the two use cases as a function of number of beams or cluster of beams according to the fact that we are considering the new hybrid satellite architecture, presented in Section II-A, or a more conventional and less flexible payload. It is shown that the number of beams equal to 4 gives the largest CHR for both scenarios. This is because there the requests form 4 geographical areas which are weakly correlated. In this case, using four beams is sufficiently efficient. It is noted that using more beams than the regions may degrade the CHR since the cost per bit of multi beam mode is larger than the that in the mono beam mode.

Fig. 9 presents the CHR of usecase 1 with Movielens demand [14]. The system parameters are similar in the previous section. A significant gain is observed for multi beam satellite (4 beams) compared with the mono beam counterpart. In particular, a CHR gain of 30% is achieved by the 10-beam setup compared with the mono beam. This is because the user demands across the beams are weakly correlated. Therefore, using the multi-beam satellite can serve the local demands better than the mono beam. Indeed, multi-beam satellites bring the CHR in this case to a level which is much more beneficial. It is also observed that the CHR in both setups saturates due to the caching capacity limit. This is due to the fact that the caching capacity of the transmission scheme is lower than the storage capacity of the terminals. This would be not a limitation if we had run the simulation for a longer period of time, where the storage capacity would have been filled in the caching period of several consecutive days.

VII. Testing setup and implementation results

In this section, we present the testing results executed in our developed testbed, which is capable of demonstrating live satellite file transfer to edge caches using FLUTE protocol. The test bed is highly configurable and permits to cover different use-cases (residential direct-to-home broadcasting in developed markets, broadcasting in developing markets, feeding video to terrestrial networks) with different technologies, e.g., widebeam, multispot beams and hybrid beams. The test bed is able to emulate a maximum of 20000 users per local cache. It is possible to run up to 6 local caches on single device. The gateway embeds a hard drive that contains 3738 movies encoded in mpeg-dash. The size of the local cache storage adjustable. The current local storage is able to store a maximum of 300 movies.

The testbed allows to analyse the profitability of caching mechanisms. Metrics for the file transfer cost have been defined with the caching algorithms to determine the cost of transmission, and establish a comparison with traditional unicast file transfer. Cost ratio between satellite and terrestrial file transfer has been used with various use-case of distribution -wide beam, multi spot beam- in order to determine the profitability of satellite based file distribution. The testbed can access such performance indicators for different transfer cost assumptions, using MovieLens [14] and Netflix Prize datasets for user behavior emulation. Fig. 10 depicts the testing setup, which has been trialed by the SES iCast demonstration over satellite using the developed technology.

The robustness of the FLUTE implementation enables the demonstrator to be capable for further trials and integration into a first product prototyping.

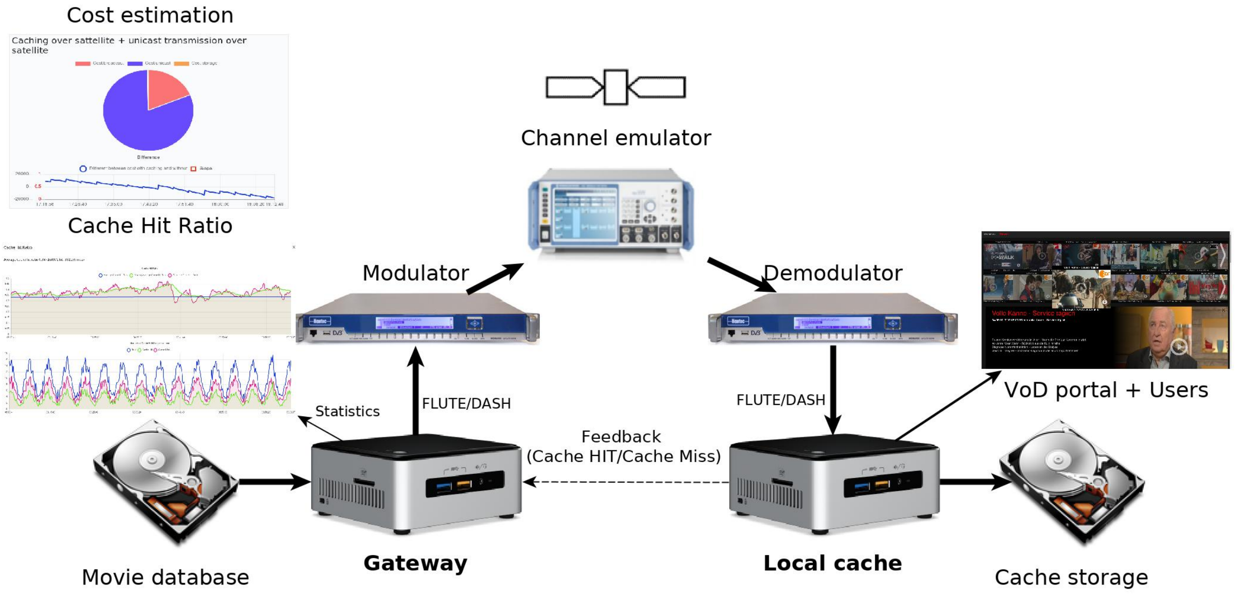
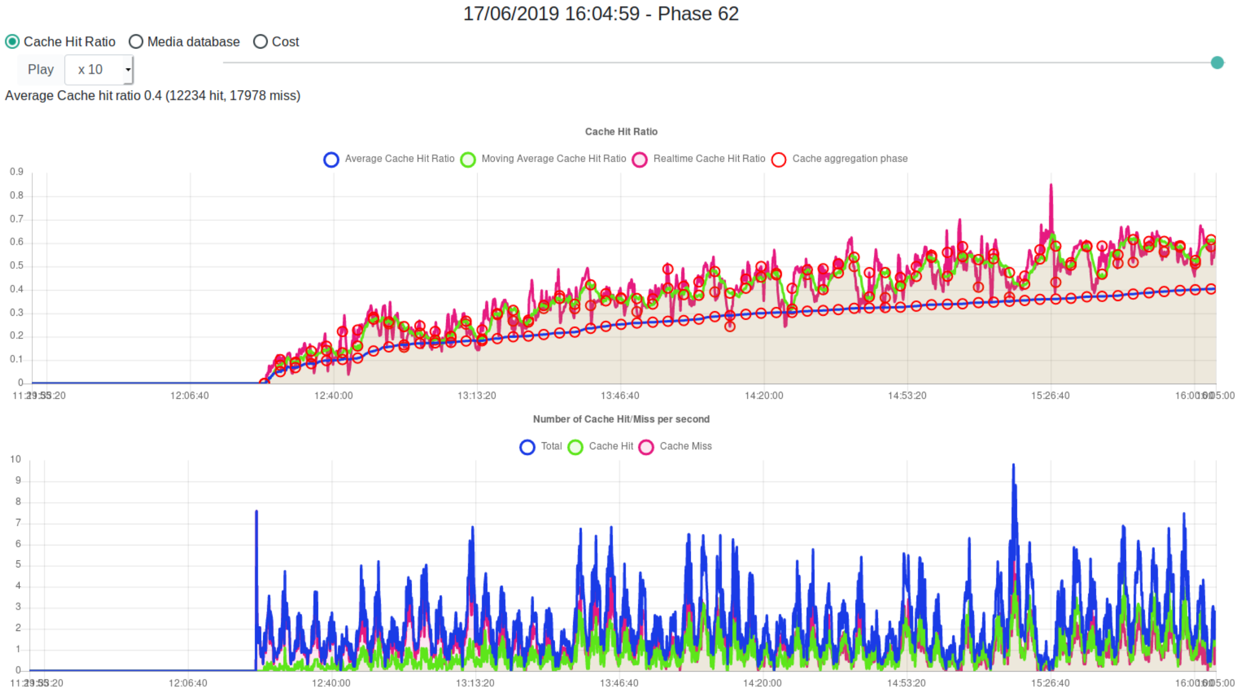


Fig. 10: Testbed permitting to analyse cache performance.

A. *CHR performance*

In this subsection, we run a test to evaluate the performance of the proposed caching algorithm. It is assume that all the local caches are empty at the beginning of the test. During the test, the local cache manager will aggregate the user requests to estimate the content popularity, which is then used to determine the contents to be cached.

Fig. 11 (top) presents a snapshot of the CHR performance for both real-time and average CHR values. The real time CHR measures the instantaneous CHR while the latter accounts for all requests from the beginning of the test. It is observed that the average CHR increases over time. This is because at the beginning of the test, CHR is equal to zero since all the local caches are empty. And due to the limited backhaul capacity, it takes time to implement the cache placement phase. This is different from common theoretical analysis results that usually ignore the cost (time and bandwidth) to fill the cache. In Fig. 11 bottom, we also presents the cache miss probability - the percentage of requests that are not served by the local cache, in addition to the CHR. As the time comes, the CHR increases while the cache miss probability decreases.



17/06/2019 16:04:59 - Phase 62

● Cache Hit Ratio ○ Media database ○ Cost

Average Cache hit ratio 0.4 (12234 hit, 17978 miss)

Fig. 11: Testbed permitting to analyse cache performance.

This expected observation results from the fact that over time the local caches better estimate the file popularity, making the caching algorithm more efficient.

B. *Cost for serving user requests*

This subsection presents cost analysis to understand the benefits of the proposed caching algorithm. The cost is calculated as the product of the cost per bit (see Section. V) with the total file size. In particular, the implemented demonstrator allows to find the cost parameters and simulation duration for which the cost of the caching scheme becomes less than the cost of the solution without caching. Fig. 12 plots the relative cost of the caching scheme compared with the solution without caching. A positive value of the different cost means that the caching scheme costs more that without caching. And the vice versus, a negative value indicates the using the caching solution is more cost efficient (the black y-axis). It is observed that the the relative cost decreases over time and become negative eventually, which confirms the effectiveness of the proposed caching algorithm.

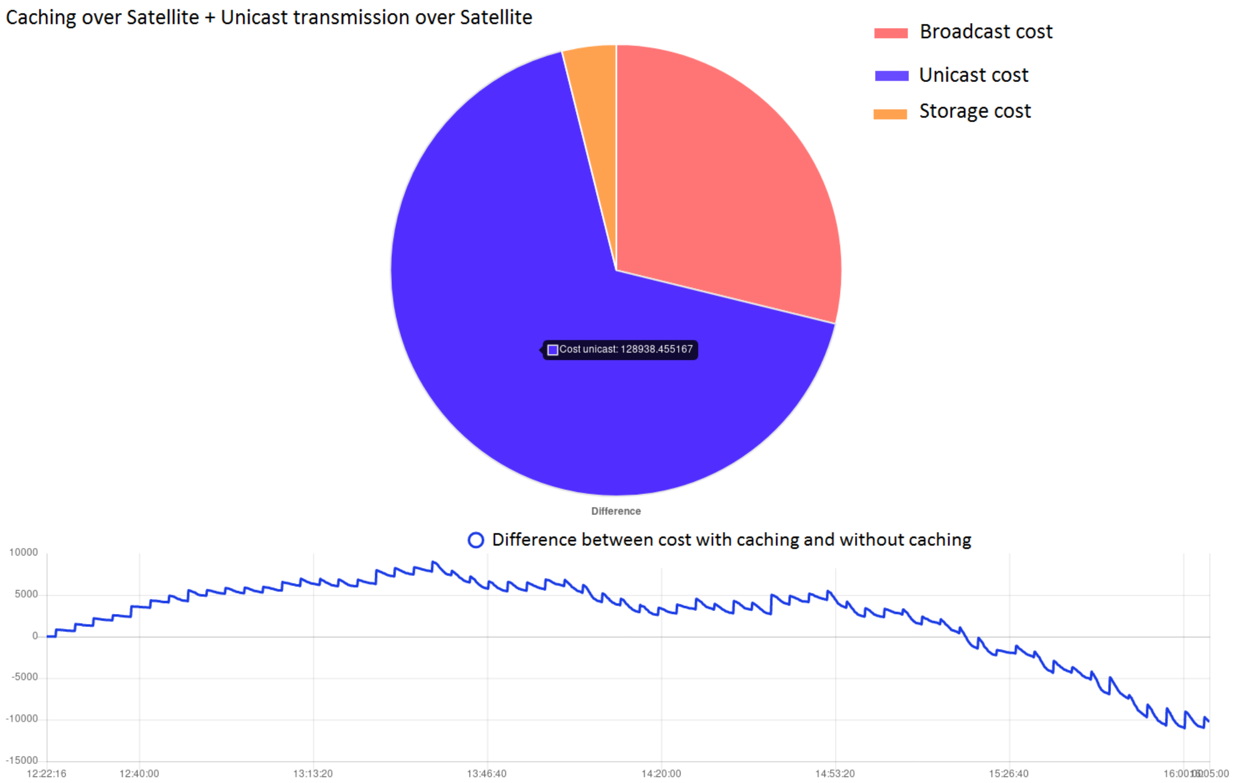


Fig. 12: Testbed permitting to analyse cache performance.

VIII. Conclusions and Future Works

We have demonstrated the effectiveness of the multimodal satellite backhauling on edge caching systems in the presence of highly uncorrelated content which is likely to be the trend of the future content consumption. The proposed offline caching algorithm is shown to be capable for flexible deployments of the satellite channels: mono beam, multi beam and hybrid. We have shown that the multi-beam and the hybrid modes become useful with respect to widebeam in caching as the geographical distribution of content popularities becomes uncorrelated. In particular, the future flexible multi-beam payloads with their adjustable coverage will be able to better distinguish between the clusters of beams with less popularity correlation and hence further improve the CHR. As future development, a cost based optimization of the rates of the different modes of the hybrid setup, in order to better fit the traffic distribution, can be realized. In addition, these results suggest a promising extension to on-line caching strategies [17], [18], where the user preference periodically changes time to time.

ACKNOWLEDGMENTS

This work is partially supported by the Luxembourg National Research Fund under the project FNR CORE ProCAST, and by the European Space Agency, ESA ESTEC, Noordwijk, The Netherlands, under the project title "High Throughput Digital Broadcasting Satellite Systems" (Contract No. 4000120692/17/NL/EM). Opinions, interpretations, recommendations and conclusions expressed herein are those of the authors and are not necessarily endorsed by the European Space Agency.

REFERENCES

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

[2] "Content delivery networks 3.0," CTOiC White paper. Online: https://www.slideshare.net/BenSchwarz1/content-deliverynetworks30.

[3] 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.

[4] S. Ramesh, I. Rhee, and K. Guo, "Multicast with cache (Mcache): an adaptive zero-delay video-on-demand service," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 440–456, Mar. 2001.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] T. X. Vu, S. Chatzinotas, B. Ottersten, and T. Q. Duong, "Energy Minimization for Cache-Assisted Content Delivery Networks With Wireless Backhaul," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 332-335, Jun. 2018.

[10] T. X. Vu, S. Chatzinotas, and B. Ottersten, "Edge-Caching Wireless Networks: Performance Analysis and Optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2827-2839, Apr. 2018.

[11] T. X. Vu, S. Chatzinotas, and B. Ottersten,, "Coded caching and storage allocation in heterogeneous networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, San Francisco, CA, 2017, pp. 1–5.

[12] 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.

[13] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *Annual Joint Conf. of the IEEE Computer and Commun. Societies*, vol. 1, Mar. 1999, pp. 126–134.

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

[15] Glyn Thomas. "Enabling Technologies for flexible HTS payloads", 33rd AIAA International Communications Satellite Systems Conference and Exhibition, International Communications Satellite Systems Conferences (ICSSC), (AIAA 2015-4348)

[16] D. Ciullo, V. Martina, M. Garetto and E. Leonardi, “How Much Can Large-Scale Video-on-Demand Benefit From Users’ Cooperation?,” in *IEEE/ACM Transactions on Networking*, vol. 23, no. 6, pp. 1846-1861, Dec. 2015.

[17] 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.

[18] 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.

[19] Christopoulos D., Chatzinotas S., Ottersten B., “Cellular-Broadcast Service Convergence through Caching for CoMP Cloud RANs,” in *IEEE Symposium on Communications and Vehicular Technology in the Benelux*, SCVT 2015, Luxembourg, November 2015.

[20] “FLUTE - File Delivery over Unidirectional Transport”, *RFC 6726*, Nov. 2012 https://tools.ietf.org/html/rfc6726.

[21] “Dynamic Adaptive Streaming over HTTP (DASH) part 1, Media presentation description and segment formats”. *ISO/IEC 23009-1*

[22] “Signaling, Delivery, Synchronization and Error Protection.” *A/331*, 21 June 2016.

[23] “Universal Mobile Telelecommunications System (UMTS); LTE; Multimedia Broadadcast/Multicast Service (MBMS); Protocols and codecs.” *3GPP TS 26.3.346.*

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