

# Caching at the Edge with Fountain Codes

Estefanía Recayte, Francisco Lázaro, Gianluigi Liva

\*Institute of Communications and Navigation of DLR (German Aerospace Center),  
Wessling, Germany. Email: {Estefania.Recayte, Francisco.LazaroBlasco,Gianluigi.Liva}@dlr.de

**Abstract**—We address the use of linear random fountain code caching schemes in a heterogeneous satellite network. We consider a system composed of multiple hubs and a geostationary Earth orbit satellite. Coded content is memorized in hubs' caches in order to serve immediately the user requests and reduce the usage of the satellite backhaul link. We derive the analytical expression of the average backhaul rate, as well as a tight upper bound to it with a simple expression. Furthermore, we derive the optimal caching strategy which minimizes the average backhaul rate and compare the performance of the linear random fountain code scheme to that of a scheme using maximum distance separable codes. Our simulation results indicate that the performance obtained using fountain codes is similar to that of maximum distance separable codes.

## I. INTRODUCTION

Cache-aided delivery protocols represent a promising solution to counteract the dramatic increase in demand for multimedia content in wireless networks. Caching techniques have been widely studied in literature with the aim of reducing the backhaul congestion, the energy consumption and the latency. In cache-enabled networks content is pre-fetched close to the user during network off-peak periods in order to directly serve the users when the network is congested. In [1] Maddah-Ali *et al.* aim at reducing the transmission rate in a network where each user has an individual cache memory. In that work, the idea of *coded caching* is introduced, so that the cache memory not only provides direct local access to the content but also generates coded multicasting opportunities among users requesting different files.

In [2] and [3], maximum distance separable (MDS) codes are proposed for minimizing the use of the backhaul link during the delivery phase in networks with caches at the transmitter side only. In [4] a delayed offloading scheme based on MDS codes is proposed to spare backhaul link resources in a network with mobile users. Caching schemes leveraging on MDS codes have also been proposed for device to device communication in order to reduce the latency [5].

MDS codes, such as Reed-Solomon codes, are optimal in the sense that they achieve the Singleton bound. The drawback of such codes is their limited flexibility in the choice of the code parameters (e.g. the block length) once the finite field is fixed, and the fact that the rate of the code is set before the encoding takes place. Unlike MDS codes, fountain codes [6] are *rateless*, i.e., their rate can be adapted on-the-fly. This has

the advantage of adding flexibility to the network, allowing a dynamic resource management.

Extensive studies regarding caching for terrestrial applications can be found in literature while limited work is available in the context of heterogeneous satellite networks [7]–[10]. In [10] a off-line caching approach over a hybrid satellite-terrestrial network is proposed for reducing the traffic of terrestrial network. However, spare backhaul resources is of particular importance not only for terrestrial networks but also in satellite systems.

To the best of the authors knowledge, the application of linear random fountain codes (LRFCs) for caching content in satellite networks has not been proposed yet. In this paper we study and optimize the performance of fountain codes for caching-enabled networks with satellite backhauling. We derive the average backhaul rate<sup>1</sup> for such system and optimize the cache placement. Our results show that the performance of the caching system using LRFCs is close to that of a system based on MDS codes already for a field size  $q = 4$ .

The rest of the paper is organized as follows. Section II introduces the system model while in Section III some preliminaries on LRFC are presented. The achievable backhaul rate is presented in Section IV while the optimization of the number of coded symbol to be memorized at each cache is presented in Section V. In Section VI the numerical results are presented. Finally, Section VII contains the conclusions.

## II. SYSTEM MODEL

We consider a heterogeneous network composed of a single geostationary Earth orbit (GEO) satellite and a number of hubs (e.g. terrestrial repeaters or high altitude platform station (HAPS)) with cache capabilities, as shown in Fig. 1. Each hub is connected to the GEO satellite through a backhaul link. Users are assumed to be fixed and to have a limited antenna gain so that a direct connection to the GEO satellite is not possible. Depending on their location users may be connected to one or multiple hubs. We denote by  $\gamma_h$  the probability that a user is connected to  $h$  hubs.

The GEO has access to a library of  $n$  equally long files  $\mathcal{F} = \{f_1, \dots, f_n\}$ . We assume that users request files from the library independently at random. Furthermore, we assume that the probability of file  $f_j$  being requested,  $\theta_j$ , follows a Zipf distribution with parameter  $\alpha$  leading to

$$\theta_j = \frac{1/j^\alpha}{\sum_i 1/i^\alpha} \quad \forall j = 1, \dots, n.$$

<sup>1</sup>We define the average backhaul rate as the average number of coded packets (output symbols) that the GEO needs to send through the backhaul link during the delivery phase to serve the request of a user.

This work has been accepted for publication at IEEE ASMS 2018.

©2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting /republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

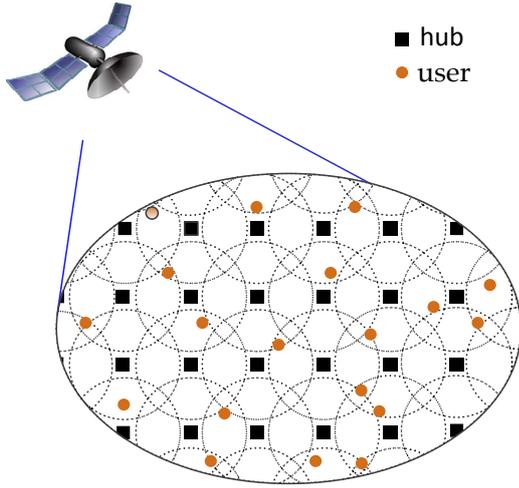


Fig. 1. System model

The assumption is made that all files are fragmented into  $k$  *input symbols* (packets). The GEO satellite encodes each file independently, using a fountain code. Each hub has storage capability for  $M$  files, i.e., for  $kM$  packets.

We shall assume that the coded caching scheme is split into two phases. During the placement phase the GEO sends a number  $w_j$  of output symbols from file  $f_j$  to each hub, which are cached by each hub. Note that for the same file each hub caches the same number of *output symbols* (encoded packets), however, the sets of output symbols cached at different hubs are different. The placement phase is assumed to be carried out offline. During the delivery phase, a user requests a file  $f_j$  at random. In a first stage, the user downloads output symbols of  $f_j$  cached in the  $h$  hubs he is connected to. If the number of symbols is not enough for decoding  $f_j$  successfully, then the GEO satellite generates additional output symbols. The GEO sends the additional output symbols to the user via one of the  $h$  hubs he is connected to. For simplicity we assume that all transmissions are error-free.

### III. LINEAR RANDOM FOUNTAIN CODES

In this work we consider the use of LRFCs for the delivery of the different files in the library. Each file is fragmented into  $k$  input symbols,  $(u_1, u_2, \dots, u_k)$ . For simplicity, we assume in the following  $u_i \in \mathbb{F}_q$ . The case of  $u_i \in \mathbb{F}_q^m$ , i.e. the case in which packets are  $m$  symbols long, can be addressed as a straightforward extension. The LRFC encoder generates a sequence of output symbols  $\mathbf{c} = (c_1, c_2, \dots, c_\ell)$ , where the number of outputs symbols  $\ell$  can grow indefinitely. In particular, the  $i$ -th output symbols is generated as

$$c_i = \sum_{a=1}^k g_{a,i} u_a,$$

where the coefficients  $g_{a,i}$  are picked independently at random with uniform probability in  $\mathbb{F}_q$ . For fixed  $\ell$ , LRFC encoding can be expressed as a vector matrix multiplication

$$\mathbf{c} = \mathbf{u}\mathbf{G}$$

where  $\mathbf{u}$  is the vector of input symbols and  $\mathbf{G}$  is a  $k \times \ell$  matrix, whose elements  $g_{a,i}$  are selected independently and uniformly at random in  $\mathbb{F}_q$ .

In order to download a file, a user must collect a set of  $m \geq k$  output symbols  $\mathbf{y} = (y_1, y_2, \dots, y_m)$ . If we denote by  $\mathcal{I} = (i_1, i_2, \dots, i_m)$  the set of indices corresponding to the  $m$  output symbols collected by the receiver we have

$$y_r = c_{i_r}.$$

The user attempts decoding by solving the system of equations

$$\mathbf{y} = \mathbf{u}\tilde{\mathbf{G}}$$

where  $\tilde{\mathbf{G}}$  is a matrix corresponding to the  $m$  columns of  $\mathbf{G}$  associated to the collected output symbols i.e., the columns of  $\mathbf{G}$  with indexes in  $\mathcal{I}$ . If the system of equations admits a unique solution (i.e., if  $\tilde{\mathbf{G}}$  is full rank), decoding is declared successful after recovering  $\mathbf{u}$ , for example by means of Gaussian elimination. If  $\tilde{\mathbf{G}}$  is rank deficient, a decoding failure is declared. In the latter case the receiver reattempts decoding after collecting one or more additional output symbols.

Let us define  $\delta$  as the receiver overhead  $\delta = m - k$ , that is, the number of output symbols in excess to  $k$  that the receiver has collected. Given  $k$ ,  $\delta$  and  $q$ , the probability of decoding failure of an LRFC is given by

$$P_f(k, \delta, q) = 1 - \prod_{i=1}^k \left( 1 - \frac{q^{i-1}}{q^{k+\delta}} \right)$$

and can be tightly lower and upper bounded as [11]

$$l(\delta, q) \leq P_f(k, \delta, q) < u(\delta, q) \quad (1)$$

where

$$l(\delta, q) := q^{-\delta-1}$$

and

$$u(\delta, q) := \frac{1}{q-1} q^{-\delta}.$$

Note that the bounds are independent from the number of input symbols  $k$  and become tighter for increasing  $q$ .

For notational convenience, in the remaining of the paper we shall use the probability of decoding success rather than the probability of decoding failure, which is simply defined as

$$P_s(k, \delta, q) := 1 - P_f(k, \delta, q).$$

### IV. AVERAGE BACKHAUL RATE

We define the average backhaul rate as the average number output coded symbols that the GEO has to send during the delivery phase in order to fulfill a user request.

In this section, we derive the overhead decoding probability  $\sigma_\delta$ , that is, the probability that a user needs exactly  $k+\delta$  coded symbols to successfully decode the requested file. Then, we calculate the average backhaul transmission rate for a LRFC coded caching scheme.

### A. Overhead Decoding Probability

Let us denote  $S_\delta$  the event that the matrix  $\tilde{\mathbf{G}}$  is full rank when  $m = k + \delta$  output symbols have been collected, where  $\Pr\{S_\delta\} = P_s(k, \delta, q)$ . Let us denote the complementary event, i.e. the rank of  $\tilde{\mathbf{G}}$  is smaller than  $k$ , with  $\bar{S}_\delta$ . We are interested in deriving

$$\sigma_\delta := \Pr\{S_\delta \mid \bar{S}_{\delta-1}\}.$$

Starting from

$$\begin{aligned} \Pr\{S_\delta\} &= \Pr\{S_\delta \mid S_{\delta-1}\} \Pr\{S_{\delta-1}\} \\ &\quad + \Pr\{S_\delta \mid \bar{S}_{\delta-1}\} \Pr\{\bar{S}_{\delta-1}\} \\ &= \Pr\{S_{\delta-1}\} + \Pr\{S_\delta \mid \bar{S}_{\delta-1}\} \Pr\{\bar{S}_{\delta-1}\} \end{aligned}$$

we have that

$$\begin{aligned} \Pr\{S_\delta \mid \bar{S}_{\delta-1}\} &= \frac{\Pr\{S_\delta\} - \Pr\{S_{\delta-1}\}}{\Pr\{\bar{S}_{\delta-1}\}} \\ &= \frac{1 - \Pr\{\bar{S}_\delta\} - (1 - \Pr\{\bar{S}_{\delta-1}\})}{\Pr\{\bar{S}_{\delta-1}\}} \\ &= 1 - \frac{\Pr\{\bar{S}_\delta\}}{\Pr\{\bar{S}_{\delta-1}\}} \end{aligned}$$

which can be rewritten as

$$\sigma_\delta = 1 - \frac{1 - P_s(k, \delta, q)}{1 - P_s(k, \delta - 1, q)}. \quad (2)$$

The expression in (IV-A) holds for  $\delta \geq 0$ , whereas for  $\delta < 0$  we have  $\sigma_\delta = 0$ .

Bounds on equation (IV-A) can be obtained from (III). In particular, for  $\delta \geq 0$  we can write

$$1 - \frac{u(\delta, q)}{l(\delta - 1, q)} < \sigma_\delta < 1 - \frac{l(\delta, q)}{u(\delta - 1, q)}$$

yielding

$$1 - \frac{1}{q-1} < \sigma_\delta < 1 - \frac{q-1}{q^2}. \quad (3)$$

Note that the bounds are independent of the overhead  $\delta$  (for non negative  $\delta$ ) and become tight as  $q$  grows. Note also that for  $q = 2$  the lower bound becomes 0 and, hence, it loses significance.

### B. Overhead Average

Let us denote as  $\Delta$  the random variable associated to the average number of symbols in excess to  $k$  that a user needs to recover the requested content and let us also denote as  $\delta$  its realization. We can calculate the average overhead as follows

$$\mathbb{E}[\Delta] = \sum_{\delta=0}^{\infty} \delta \left[ \prod_{i=0}^{\delta-1} (1 - \sigma_i) \right] \sigma_\delta. \quad (4)$$

By using (IV-A) in (IV-B),  $\mathbb{E}[\Delta]$  can be upper bounded as

$$\mathbb{E}[\Delta] < \frac{q-1}{(q-2)^2} \left( 1 - \frac{q-1}{q^2} \right) := \delta_u. \quad (5)$$

### C. Backhaul Rate

Let  $Z$  be the random variable associated to the number of output symbols for a file requested by a user in the hubs he is connected to, and let  $z$  be its realization. Let  $H$  be the random variable associated to the number of hubs a user is connected to, being  $h$  its realization. Finally, let  $J$  be the random variable associated to the index of the file requested by a user, being  $j$  its realization. We have

$$P_{Z|J,H}(z|j, h) = \begin{cases} 1 & \text{if } z = w_j h \\ 0 & \text{otherwise} \end{cases}$$

where we recall that  $w_j$  stands for the number of coded symbols from file  $j$  stored in every hub. The probability mass function of  $Z$  is

$$\begin{aligned} P_Z(z) &= \sum_j \sum_h P_{Z|J,H}(z|j, h) P_J(j) P_H(h) \\ &= \sum_j \sum_h \theta_j \gamma_h P_{Z|J,H}(z|j, h). \end{aligned}$$

We are interested in deriving the distribution of the backhaul rate, i.e. the number of output symbols which have to be sent over the backhaul link to serve the request of a user, which we denote by random variable  $T$ . If we condition  $T$  to  $Z$ , it is easy to see how the probability of  $T = t$  corresponds to the probability that decoding succeeds when the user has received exactly  $t$  output symbols from the backhaul link in excess to the  $z$  output symbols it received from the hubs through local links, that is, when  $m = z + t$ .

In order to derive  $P_{T|Z}(t|z)$  we shall distinguish two cases.

If  $z \leq k$ , then

$$P_{T|Z}(t|z) = \begin{cases} \prod_{i=0}^{z-k+t-1} (1 - \sigma_i) \sigma_{z-k+t} & \text{if } t \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

If  $z > k$ , then

$$P_{T|Z}(t|z) = \begin{cases} \sum_{j=0}^{z-k} \left( \prod_{i=0}^{j-1} (1 - \sigma_i) \sigma_j \right) & \text{if } t = 0 \\ \prod_{i=0}^{z-k+t-1} (1 - \sigma_i) \sigma_{z-k+t} & \text{if } t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We are after the expectation of  $T$ , which is obtained as

$$\begin{aligned} \mathbb{E}[T] &= \sum_t t P(t) \\ &= \sum_t t \sum_{z=0}^{\infty} P_{T|Z}(t|z) P_Z(z) \\ &= \sum_t t \left( \sum_{z=0}^k P_{T|Z}(t|z) P_Z(z) + \sum_{z=k+1}^{\infty} P_{T|Z}(t|z) P_Z(z) \right) \end{aligned} \quad (6)$$

where in the last equality we distinguished two different cases,  $z \leq k$  and  $z > k$ . Let us define  $\bar{T}_1$  and  $\bar{T}_2$  as

$$\begin{aligned}\bar{T}_1 &:= \sum_t t \sum_{z=0}^k P_{T|Z}(t|z) P_Z(z) \\ \bar{T}_2 &:= \sum_t t \sum_{z=k+1}^{\infty} P_{T|Z}(t|z) P_Z(z)\end{aligned}$$

so that

$$\mathbb{E}[T] = \bar{T}_1 + \bar{T}_2. \quad (7)$$

If we introduce the variable change  $\delta = z - k + t$  in the expression of  $\bar{T}_1$ , we obtain

$$\begin{aligned}\bar{T}_1 &= \sum_{z=0}^k P_Z(z) \sum_{\delta=z-k}^{\infty} (\delta - z + k) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \\ &\stackrel{(a)}{=} \sum_{z=0}^k P_Z(z) \sum_{\delta=0}^{\infty} (\delta - z + k) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \\ &= \sum_{z=0}^k P_Z(z) \left( \sum_{\delta=0}^{\infty} \delta \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right. \\ &\quad \left. + (k - z) \sum_{\delta=0}^{\infty} \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right) \\ &\stackrel{(b)}{=} \sum_{z=0}^k P_Z(z) (\mathbb{E}[\Delta] + k - z) \\ &= (\mathbb{E}[\Delta] + k) \Pr\{Z \leq k\} - \sum_{z=0}^k z P_Z(z)\end{aligned} \quad (8)$$

where equality (a) is due to  $\sigma_\delta = 0$  for  $\delta < 0$  and equality (b) is due to

$$\sum_{\delta=0}^{\infty} \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta = 1.$$

Introducing the same variable change in the expression of  $\bar{T}_2$  we have

$$\begin{aligned}\bar{T}_2 &= \sum_{z=k+1}^{\infty} P_Z(z) \sum_{\delta=z-k}^{\infty} (\delta - z + k) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \\ &= \sum_{z=k+1}^{\infty} P_Z(z) \left( \sum_{\delta=z-k}^{\infty} \delta \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right. \\ &\quad \left. + \sum_{\delta=z-k}^{\infty} (k - z) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right) \\ &\leq \sum_{z=k+1}^{\infty} P_Z(z) \mathbb{E}[\Delta] = \mathbb{E}[\Delta] \Pr\{Z > k\}\end{aligned} \quad (9)$$

where the inequality is due to

$$\sum_{\delta=z-k}^{\infty} \delta \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \leq \mathbb{E}[\Delta]$$

and to

$$\sum_{\delta=z-k}^{\infty} (k - z) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \leq 0$$

since  $z > k$ .

The expression of the average backhaul rate given by

$$\begin{aligned}\mathbb{E}[T] &= (\mathbb{E}[\Delta] + k) \Pr\{Z \leq k\} - \sum_{z=0}^k z P_Z(z) \\ &\quad + \sum_{z=k+1}^{\infty} P_Z(z) \left( \sum_{\delta=z-k}^{\infty} \delta \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right. \\ &\quad \left. + \sum_{\delta=z-k}^{\infty} (k - z) \prod_{i=0}^{\delta-1} (1 - \sigma_i) \sigma_\delta \right).\end{aligned}$$

Finally, we can upper bound the average backhaul rate by use of (IV-B), (IV-C) and (IV-C) in (IV-C) as

$$\mathbb{E}[T] \leq \delta_u + k \Pr\{Z \leq k\} - \sum_{z=0}^k z P_Z(z). \quad (10)$$

## V. LRFC PLACEMENT OPTIMIZATION

The LRFC placement problem calls for minimizing the average backhaul rate during the delivery phase. To this end, we want to optimize the number of coded symbols per file that each hub has to cache. We present in this section the placement optimization problem adapted to a LRFC cached scheme based on the optimization problem proposed for MDS codes in [2].

The optimization problem can be written as

$$\begin{aligned}\min_{w_1, \dots, w_n} \quad & \mathbb{E}[T] \\ \text{s.t.} \quad & \sum_{j=1}^n w_j = Mk, \quad w_j \in \mathbb{N}_0.\end{aligned} \quad (11)$$

The first constraint specifies that the total number of stored coded symbols should be equal to the size cache. The second constraints accounts for the discrete nature of the optimization variable.

Solving exactly the optimization problem requires evaluating (IV-C), which can be computationally complex when  $k$  and  $z$  are large. Hence, as an alternative to minimizing the average backhaul rate, we propose minimizing its upper bound in (IV-C), which leads to the following optimization problem

$$\begin{aligned}\min_{w_1, \dots, w_n} \quad & \left[ \delta_u + k \Pr\{Z \leq k\} - \sum_{z=0}^k z P_Z(z) \right] \\ \text{s.t.} \quad & \sum_{j=1}^n w_j = Mk \quad w_j \in \mathbb{N}_0.\end{aligned} \quad (12)$$

Since the upper bound on  $\mathbb{E}[T]$  in (IV-C) relies on the upper and lower bounds in (III), which are tight, we expect the result of the optimization problem in (V) to be close to the result of the optimization problem in (V).

## VI. RESULTS

In this section, we numerically evaluate the normalized average backhaul rate, which we define as  $\mathbb{E}[T]/k$ .

In all the setups, we consider that users are uniformly distributed within the coverage area of the satellite and border effects are neglected. We consider that each hub covers a circular area of radius  $r$  centered around the hub. For simplicity, we

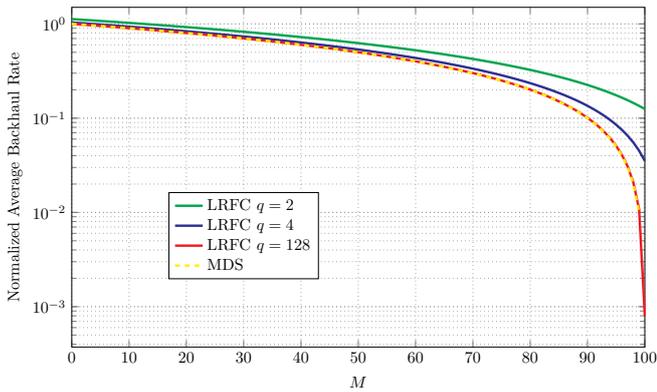


Fig. 2. Normalized average backhaul rate as a function of memory size  $M$  for MDS and LRFC codes over  $\mathbb{F}_q$  for  $q = 2, 4, 128$  given  $n = 100, k = 10, \alpha = 0$  and  $\gamma_1 = 1$ . Solid curves represent LRFC schemes while the dashed curve represents the MDS scheme.

TABLE I  
AVERAGE OVERHEAD REQUIRED FOR SUCCESSFUL DECODING FOR A LRFC WITH  $k = 10$

$q$	$\mathbb{E}[\Delta]$	upper bound (IV-B)
2	1.1981	-
4	0.3490	0.6094
8	0.1447	0.1792
16	0.0669	0.0720
32	0.0323	0.0334
64	0.0159	0.0161
128	0.0079	0.0079

assume that the hubs are arranged according to a uniform two dimensional grid, with spacing  $d$ . Unless otherwise specified, we assume  $r = 60$  km and  $d = 45$  km. Thus, the coverage areas of different hubs partially overlap, as it can be observed in Fig. 1. After some simple geometrical calculations we obtain that the following connectivity distribution

$$\begin{aligned} \gamma_1 &= 0.2907, \gamma_2 = 0.6591, \\ \gamma_3 &= 0.0430, \gamma_4 = 0.0072. \end{aligned} \quad (13)$$

We first evaluate the tightness of the upper bound (IV-B) on average overhead. Table I shows  $\mathbb{E}[\Delta]$  for different values of  $q$  are shown. The values in the second column were numerically derived from equation (IV-B) while values in the third column were derived from the bound in equation (IV-B). We can see that the bound becomes tighter for increasing  $q$ .

In the first scenario, we study how the cache size  $M$  impacts in the average backhaul rate. In this setup, we assume all user are connected to exactly one hub, i.e.  $\gamma_1 = 1$ , furthermore, we consider that file popularity is uniformly distributed (i.e Zipf distributed with parameter  $\alpha = 0$ ). Moreover, we assume a library size  $n = 100$  and we assume that each file is fragmented into  $k = 10$  input symbols. We optimized the number of LRFC coded symbols  $w_j$  cached at each hub by solving the problem (V) for  $q = 2, q = 4$  and  $q = 128$  and we calculated the average backhaul rate the fountain coding caching scheme. As a benchmark we used the MDS caching scheme from [2].

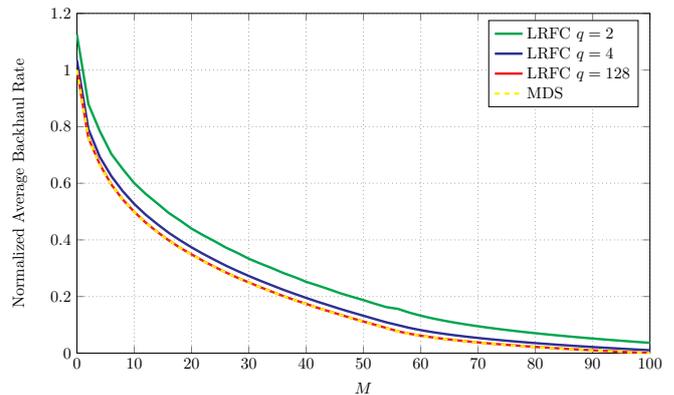


Fig. 3. Normalized average backhaul rate as a function of memory size  $M$  for MDS and LRFC codes over  $\mathbb{F}_q$  for  $q = 2, 4, 128$  given  $n = 100, k = 10, \alpha = 0.8$  and  $\gamma_1 = 0.2907, \gamma_2 = 0.6591, \gamma_3 = 0.0430, \gamma_4 = 0.0072$ .

In Fig. 2 the normalized average backhaul rate is shown as a function of the memory size  $M$ . We can observe how the penalty on the average rate for using LRFC with respect to a MDS code becomes smaller for increasing  $q$  and already for  $q = 128$  is almost negligible. We remark that for  $M = 100$  the cache size coincides with the library size, hence, the backhaul rate for the MDS scheme becomes zero, whereas for the LRFC schemes the average backhaul rate coincides with the average overhead.

In our second setup, we assume that users can be connected to multiple hubs. We consider the connectivity distribution given in (VI) and file popularity distribution with parameter  $\alpha = 0.8$ . The library size is set to  $n = 100$  and the number of input symbols is  $k = 10$ , the same as in the previous scenario. The optimal cache placing is computed for each LRFC scheme as well as for the MDS.

In Fig. 3 we show the impact of memory size  $M$  on the normalized average backhaul rate for different code caching schemes when users can be connected to multiple hubs. Similarly to the previous scenario, we see that for sufficiently large  $q$  the performance of the LRFC caching scheme approaches that of the MDS code. Note that since a MDS code achieves the best possible performance, this result shows implicitly that solving the optimization problem in (V) yields a solution that is close to that of solving the optimization problem in (V). We further observe that LRFC caching with storage capabilities equal to 10% of the library size can reduce the average backhaul rate for at least 40% with respect to a system with no caching ( $M = 0$ ).

For the same connectivity distribution, a fixed memory size  $M = 10$  and library size  $n = 100$ , we investigate how the file distribution impacts on the average backhaul rate. In Fig. 4 the normalized average backhaul rate is shown as a function of the file parameter distribution  $\alpha$ . As expected, when  $\alpha$  increases caching schemes become more efficient since the majority of the requests is concentrated in a small number of files. Looking at the figure we can observe how for  $\alpha = 0.2$ , a LRFC in  $\mathbb{F}_2$  requires roughly 12% more transmissions over backhaul link than a LRFC in  $\mathbb{F}_{128}$ . For  $\alpha = 1.5$  the LRFC of order  $q = 2$  requires only 4.7% more than in  $q = 128$ .

## VII. CONCLUSIONS

In this work we analyzed fountain code schemes for caching content at the edge. We considered a heterogeneous satellite network, composed of a GEO satellite and a number of hubs which can cache content. We focus on reducing the average rate in the backhaul link which connects the GEO to the hubs. For this setting, we derived the analytical expression of the average backhaul rate as well as an upper bound of it. Making use of this upper bound, we formulated the cache placement optimization problem for a fountain coding caching scheme using linear random fountain codes (LRFCs). Finally, we presented simulation results where we compared the performance of the LRFC scheme with a MDS scheme. Our simulation results indicate that the performance of the LRFC caching scheme built over a finite field moderate order approaches that of the MDS caching scheme.

## REFERENCES

- [1] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *IEEE Int. Symposium on Info. Theory (ISIT)*, Hong Kong, Jun. 2015.
- [2] V. Bioglio, F. Gabry, and I. Land, "Optimizing MDS codes for caching at the edge," in *Proc. IEEE Globecom*, San Diego, U.S.A., Dec. 2015.
- [3] J. Liao, K. K. Wong, M. R. A. Khandaker, and Z. Zheng, "Optimizing cache placement for heterogeneous small cell networks," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 120–123, Jan. 2017.
- [4] E. Ozfatura and D. Gündüz, "Mobility and popularity-aware coded small-cell caching," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 288–291, Feb. 2018.
- [5] A. Piemontese and A. G. i. Amat, "MDS-coded distributed storage for low delay wireless content delivery," in *Int. Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, Brest, France, Sep. 2016.
- [6] D. J. C. MacKay, "Fountain codes," *IEEE Proc.-Commun.*, vol. 152, pp. 1062–1068, 2005.
- [7] T. de Cola, G. Gonzalez, and V. E. V. Mujica, "Applicability of ICN-based network architectures to satellite-assisted emergency communications," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016.
- [8] H. Wu, J. Li, H. Lu, and P. Hong, "A two-layer caching model for content delivery services in satellite-terrestrial networks," in *Proc. IEEE Globecom*, Washington, DC, USA, Dec. 2016.
- [9] S. Liu, X. Hu, Y. Wang, G. Cui, and W. Wang, "Distributed Caching Based on Matching Game in LEO Satellite Constellation Networks," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 300–303, Feb. 2018.
- [10] H. Kalantari, M. Fittipaldi, S. Chatzinotas, T. X. Vu, and B. Ottersten, "Cache-assisted hybrid satellite-terrestrial backhauling for 5G cellular networks," in *Proc. IEEE Globecom*, Singapore, Singapore, Dec. 2017.
- [11] G. Liva, E. Paolini, and M. Chiani, "Performance versus overhead for fountain codes over  $\mathbb{F}_q$ ," *IEEE Comm. Letters.*, vol. 14, no. 2, pp. 178–180, 2010.

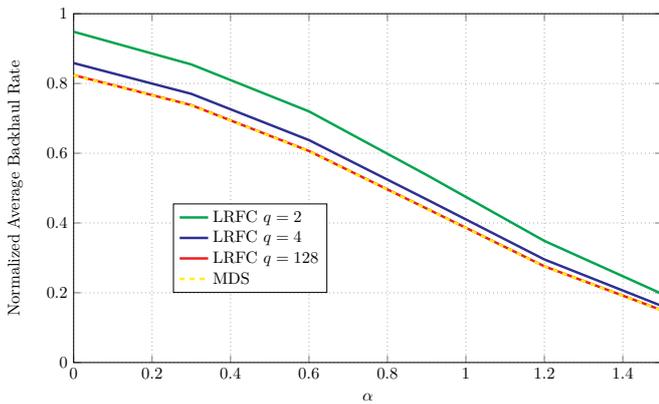


Fig. 4. Normalized average backhaul rate as a function of the file parameter distribution  $\alpha$  for MDS and LRFC codes over  $\mathbb{F}_q$  for  $q = 2, 4, 128$  given  $n = 100$ ,  $k = 10$ ,  $M = 10$  and  $\gamma_1 = 0.2907$ ,  $\gamma_2 = 0.6591$ ,  $\gamma_3 = 0.0430$ ,  $\gamma_4 = 0.0072$ .

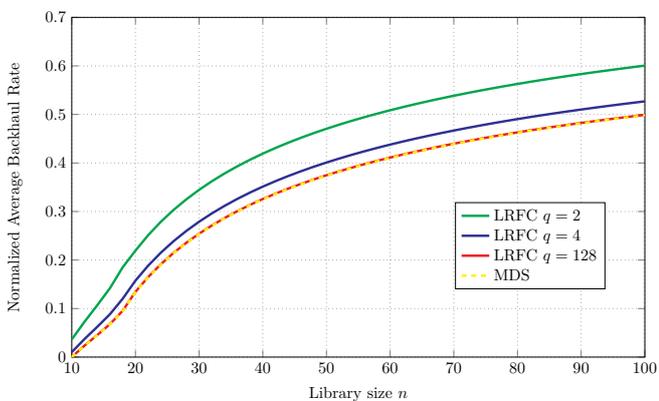


Fig. 5. Normalized average backhaul rate as a function of the library size  $n$  for MDS and LRFC codes over  $\mathbb{F}_q$  for  $q = 2, 4, 128$  given  $\alpha = 0.8$ ,  $k = 10$ ,  $M = 10$  and  $\gamma_1 = 0.2907$ ,  $\gamma_2 = 0.6591$ ,  $\gamma_3 = 0.0430$ ,  $\gamma_4 = 0.0072$ .

In our last setup we consider  $\alpha = 0.8$ ,  $M = 10$ ,  $k = 10$  and the distribution given in (VI). We evaluate the average backhaul rate for different cardinalities of the library. In Fig. 5 the normalized average backhaul rate is shown as a function of the library size. For a fixed memory size the average backhaul rate increases as the library size increases. As it can be observed, also in this case the proposed LRFC caching scheme performs similarly to a MDS scheme.