Mobility-Aware Coded Storage and Delivery

Emre Ozfatura and Deniz Gündüz
Information Processing and Communications Lab
Department of Electrical and Electronic Engineering
Imperial College London
Email: {m.ozfatura, d.gunduz}@imperial.ac.uk

***Abstract*—Content caching at small-cell base stations (SBSs) is a promising method to mitigate the excessive backhaul load and delay, particularly for on-demand video streaming applications. A cache-enabled heterogeneous cellular network architecture is considered in this paper, where mobile users connect to multiple SBSs during a video downloading session, and the SBSs request files, or fragments of files, from the macro-cell base station (MBS) according to the user requests they receive. A novel coded storage and delivery scheme is introduced to reduce the load on the backhaul link from the MBS to the SBSs. The achievable backhaul delivery rate as well as the number of sub-files required to achieve this rate are studied for the proposed coded delivery scheme, and it is shown that the proposed scheme provides significant reduction in the number of sub-files required, making it more viable for practical applications.**

I. Introduction

There are two prominent approaches used extensively in the literature to reduce the backhaul load, namely coded storage and coded delivery. In a broad sense, coded storage is designed from the perspective of users, and allows users to efficiently receive a file from multiple access points without worrying about overlapping bits. Maximum distance separable (MDS) coded storage has been studied extensively, for example, in multi-access downlink scenarios, such as a static user downloading a content from multiple SBSs [3]–[5], mobile users (MUs) connecting to different SBSs sequentially to download content [6]–[8], or MUs utilizing device-to-device (D2D) communication opportunities [9].

Coded delivery, on the other hand, is designed from the point of the server, which utilizes the caches of the users to seek multi-casting opportunities [1], [2]. Coded delivery scheme consists of two phases. In the *placement phase*, files are divided into sub-files, and each user stores a certain subset of the sub-files. In the *delivery phase*, the server carefully constructs the multicast messages as XORed combinations of the requested sub-files. Each user recovers its request from the multicasted messages using its cache contents. We note that the multicast gain increases with the number of users, which implies that a large number of users is an advantage, allowing lower delivery rates. However, at the same time the number of sub-files increases exponentially with the number of users, which is one of the main challenges in front of practical implementations of this scheme [10].

This work was supported in part by the Marie Sklodowska-Curie Action SCAVENGE (grant agreement no. 675891), and by the European Research Council (ERC) Starting Grant BEACON (grant agreement no. 725731).

In this paper, we introduce a novel coded storage and delivery scheme, which is designed taking into account the random mobility patterns of the users. The proposed scheme divides the SBSs into smaller groups according to the mobility patterns of the users, and applies coded delivery to each group of SBSs independently. MDS-coded storage is used to make sure that the MUs can collect useful information from any of the SBSs they connect to until they completely download their requested file. We also introduce an efficient grouping strategy via utilizing the analogous well known frequency reuse pattern problem [11].

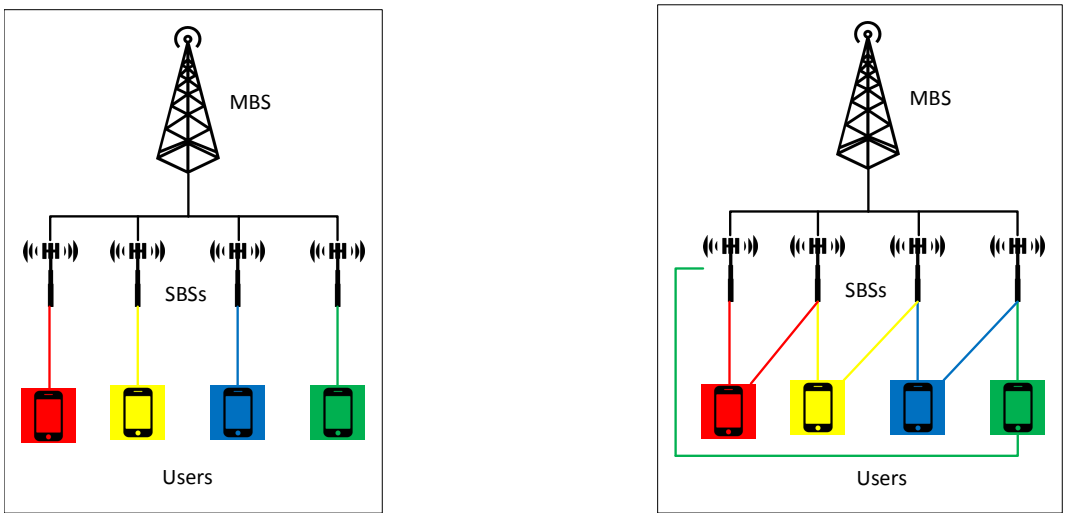
The rest of the paper is organized as follows. The system model is introduced in Section II and the proposed coded storage and delivery scheme is explained in Section III. Performance of the proposed coded storage and delivery scheme in comparison to coded delivery scheme in [1] is numerically analyzed in Section IV. Finally, in Section V we conclude the paper with summary of our contributions and future directions.

Notations. Throughout the paper, for positive integer N, the set $\{1,\dots,N\}$ is denoted by $[N]$. We use $\bigoplus$ to denote the bit-wise XOR operation, while $\binom{j}{i}$ represents the binomial coefficient.

II. System Model

A. Network model

We consider a cellular network architecture that consists of one MBS and K SBSs, i.e., $SBS_1,\dots,SBS_K$. The MBS has a fixed access to a content library of N files, $W_1,\dots,W_N$, over a high capacity transmission link (e.g., fiber link). We assume that all the files have the equal size of F bits. Each SBS_k is equipped with a cache memory of MF bits. The SBSs are connected to the MBS over a shared wireless backhaul link. Hence, when a user requests a content from a SBS, the SBS first checks its content cache. If the requested content is fully cached, then the SBS directly delivers the corresponding file. If the requested content is not cached at all, or partially cached, then the remaining parts of the content are first transferred from the MBS to the SBS over the shared wireless backhaul link. In this paper, we assume that all the files in the library are requested by the users with the equal probability, i.e., file W_n, is requested by a MU with probability $1/N$. Under this assumption, if N is large, which is the case in realistic scenarios, then it is safe to assume that all the users request a different content. For instance, when $K = 30$ and $N = 10000$ the probability of each user requesting a different



(a) Single access model.

(b) Multi-access model with uniform access pattern.

Fig. 1: Static user access models studied in [1] and [2], respectively.

file is 0.975. This assumption has been widely accepted in the coded caching literature as it also represents the worst case scenario. We also assume that the number of users in the network is limited by the number of SBSs; hence, there are K users in the network, i.e., $U_1, \dots, U_K$, and the requests of the users are denoted by the demand vector $\mathbf{d} \triangleq (d_1, \dots, d_K)$. The placement phase, i.e., caching at the SBSs takes place before the demand vector $\mathbf{d}$ is revealed. The required delivery rate for the backhaul link, $R(M, \mathbf{d})$, is defined as the minimum number of bits that must be transmitted over the shared link, normalized by the file size, for a given normalized cache capacity M, in order to satisfy all the user demands. Since, we are interested in the case in which each user requests a different file, we simply use $R(M)$ to denote the worst case delivery rate, instead of $R(M, \mathbf{d})$.

The delivery rate over the backhaul link is directly related to the access model of the users. In this paper, we are particularly interested in a single access model with mobility, in which a MU is connected to exactly one SBS at a particular time instant; however, due to mobility, it connects to multiple SBSs over time. To better motivate and explain our model and results, we will first explain the previously studied access models in the literature, and then provide a detailed explanation of the considered single access model with mobility.

B. User access models

In this subsection, we will first briefly explain the two previously studied access models in order to highlight the fundamental differences of the model considered in this paper.

1) Static single access model: In this model, it is assumed that each user has access to exactly one SBS, as illustrated in Figure 1a. Hence, under this assumption the backhaul delivery rate problem is identical to the shared link problem introduced in [1]. The caching and coded delivery method introduced in [1], for this model, works as follows. In the placement phase, for $t \triangleq \frac{MK}{N}$, file W_n, $n \in [N]$, is cached at *level* t, which means that it is divided into $\binom{K}{t}$ non-overlapping sub-files of equal size, and each sub-file is cached by a distinct subset of t SBSs. Then, each sub-file can be identified by a subset $\mathcal{I}$, where $\mathcal{I} \subseteq [K]$ and $|\mathcal{I}| = t$, such that sub-file $W_{n,\mathcal{I}}$ is cached by SBSs $k \in \mathcal{I}$. Following a placement phase in which all the files are cached at level t, in the delivery phase, for each subset $\mathcal{S} \subseteq [K]$, $|\mathcal{S}| = t + 1$, all the requests of the SBSs in $\mathcal{S}$ can be served simultaneously by MBS via multicasting

$$\bigoplus_{s \in \mathcal{S}} W_{d_s, \mathcal{S}\setminus\{s\}}. \tag{1}$$

Thus, with a single multicast message the MBS can deliver $t + 1$ sub-files, and achieve a *multicasting gain* of $t + 1$. Accordingly, the achievable delivery rate for the backhaul link is $R(M) = \frac{K-t}{t+1}$, where $t = MK/N$ and integer valued.

2) Static multi-access model: In this model, users are allowed to access multiple SBSs, where each user is connected to an equal number of SBSs, and remains connected to the same SBSs during the whole video downloading process; thus, we call this model the static multi-access model. A particular case of this problem is studied in [2], where each user connects to L SBSs following a certain pattern, such that user U_k is connected to $SBS_k, \dots, SBS_{k+L-1 \mod K}$. The case of $L = 2$ is illustrated in Figure 1b. In [2], the authors divide the SBSs into L groups according to their index, such that the lth group is formed by $\mathcal{G}_l \triangleq \{SBS_k : k \mod L = l\}$. Then, they modify the coded delivery scheme in [1] as follows. In the placement phase each file is first divided into L disjoint fragments, i.e., W_n^l is the lth fragment of file W_n. Then, for each $l \in [L]$, all the fragments in $\mathcal{W}^l \triangleq \{W_1^l, \dots, W_N^l\}$ are cached by the SBSs in $\mathcal{G}_l$. For the placement of a particular group $\mathcal{G}_l$, we use the same caching scheme as in the static single access model with $\acute{K} = K/L$ SBSs and cache sizes

$\acute{M} = ML$ [1]. Therefore, each fragment of each file is cached at level $t \triangleq \frac{\acute{K}\acute{M}}{N} = MK/N$ i.e., sub-file $W^l_{n,\mathcal{I}}$, where $\mathcal{I} \subseteq \{k : k \mod L = l\}$ and $|\mathcal{I}| = t$, is cached by SBS_k, $k \in \mathcal{I}$. Similarly, the coded delivery phase is executed for each $\mathcal{G}_l$, $l \in [L]$, separately. The coded delivery algorithm for this model is given in Algorithm 1, and the corresponding backhaul delivery rate is found as $R(M) = \frac{K-Lt}{1+t}$. We note that the delivery rate decreases with L, the number of SBSs each user connects to.

Algorithm 1: Delivery

```
for l = 1 : L do
    for ĺ = 0 : L − 1 do
        for S ∈ {k : k mod L = l}, |S| = t do
            ⊕_{s∈S} W^l_{d_{(s−ĺ) mod K}, S\{s}}
        end
    end
end
```

3) Single access model with mobility: In this model, MUs may connect to different SBSs during the video downloading process, but unlike in the previous model, each MU is connected only to the nearest SBS at any time instant. We consider equal-length time slots, whose duration corresponds to the minimum time duration a MU remains connected to the same SBS. We assume that each SBS is capable of transmitting B bits to a MU within one time slot. Hence, a file of size F bits can be downloaded in $T = \frac{F}{B}$ slots. We define the mobility path of a user as the sequence of small-cells visited during these T time slots. For instance, for $K = 7$ and $T = 3$, SBS_2, SBS_3, SBS_4 is a possible mobility path. We assume that during the video downloading session of T time slots each MU is connected to exactly T different SBSs, which we call as the *high mobility assumption*.

C. Problem definition

Our aim is to minimize the normalized delivery rate over the backhaul link under the high mobility assumption for MUs. We note that, the single access model with mobility can be easily treated as a static single access model in the following way: each file is divided into T disjoint fragments, and each fragment is considered as a separate file so that the size of the file library and the size of the caches are scaled as NT and MT, respectively. Then the placement phase is executed as in the coded delivery scheme in [1] according to caching level $t = \frac{KMT}{NT} = KM/N$ and at each time slot the delivery phases can be executed as in the coded delivery scheme in [1]; hence, $R(M) = \frac{K-t}{t+1}$, $t = KM/N$, is still achievable. At this point we remark that, in addition to reducing the normalized backhaul delivery rate, it is equally important to bound the number of sub-files used in the delivery phase in order to obtain a practically viable caching strategy.

The approach introduced in [2] is an efficient method to reduce both the number of sub-packets and the normalized delivery rate of the backhaul link, when the users connect to the SBSs in a uniform manner, as described in the previous section. However, this method is not applicable when the users do not follow the prescribed access patterns.

[1]Since the size of a fragment is $1/L$ of the original file.

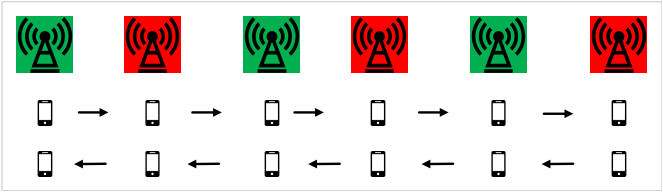


Fig. 2: Deterministic mobility path along a line

On the other hand, MDS-coded caching can be employed when the users are mobile, or access to the SBSs with non-uniform patterns [3], [6], [7], [12]. The key advantage of MDS-coded caching at the SBSs is to reduce the amount of data that need to be cached at each SBS for each file. Consider the following simple example with $K = 4$ SBSs, where a MU can connect to any 3 of them. In this case, each file is divided into 3 fragments, and they are encoded into 4 fragments through a (3,4) MDS code. Then, each SBS caches a different fragment so that a MU that connects to any 3 of the SBSs can recover the file. In this example each SBS needs to cache only one fragment for each file, equivalently, $1/3$ of the original file. Accordingly, under high mobility assumption, for a given $T = F/B$, it is sufficient to store only $1/T$ portion of each file at each SBS. Hence, when $M \geq N/T$, via MDS coded caching, the normalized delivery rate of the backhaul link can be reduced to zero which means that all the user requests can be delivered locally. Otherwise, only MT files can be cached and delivered locally using MDS coded caching. The main drawback of MDS-coded caching is that, coded delivery techniques can not be applied directly to MDS-coded files since the multicast gain of coded delivery stems from the overlaps among the cached sub-files at different SBSs.

III. Solution Approach

In this section, we will introduce a new method that utilizes both the MDS-coded caching and coded delivery techniques, and analyze its performance under the high mobility assumption. For the sake of exposition, we first consider a special class of mobility patterns, for which the coded delivery technique in [2] can be applied directly.

A. Special case: Deterministic paths

In this special case, we consider a particular mobility scenario, in which a user's mobility path is completely determined by its direction and the first SBS that the MU connects to. This can model, for example, MUs on a train connecting to SBSs located by the rail tracks in a known order. In this special case, MUs can be considered as moving on a line as illustrated in Figure 2.

Although a MU is connected only to the nearest SBS at any time instant, the coded delivery technique introduced in [2] can be applied in this special case. For given file size F and SBS transmission rate B, each file is divided into $T = F/B$

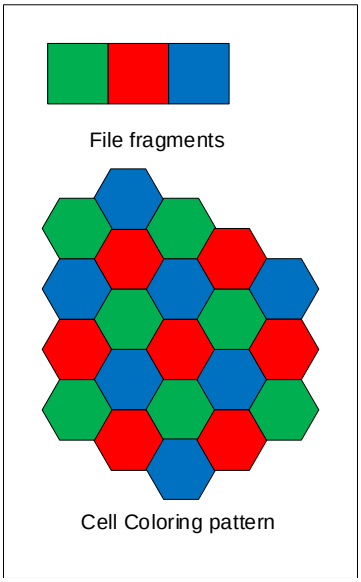


Fig. 3: File fragmentation and associated cell coloring.

fragments. Similarly, the set of all SBSs are also divided into T disjoint groups, where SBSs in each group caches only one fragment of each file as in the placement phase in [1]. These groups of SBSs should be constructed in a way that, through any mobility path a MU must connect exactly to one SBS from each group in order to receive all the fragments of the requested files. Grouping of the SBSs can be considered as a coloring where the SBSs are colored using T different colors such that any adjacent T of them have different colors. An example for $T = 2$ is illustrated in Figure 2, where any two neighboring SBSs have different colors. Coded delivery phase is executed at each time slot separately for each group of SBSs $\{\mathcal{G}_l\}$, $l \in [T]$. Hence, for the special case of deterministic paths, the achievable delivery rate for the backhaul link is $R(M) = \frac{K-tT}{1+t}$, where $t = KM/N$, and integer valued as before.

B. General Case: Random paths

In this subsection, we consider a general path model in which the users move on a 2D grid, and each SBS covers a disjoint, equal size area with hexagonal shape as illustrated in Figure 3. The fundamental difference between the deterministic path model and the random path model is that; in the random path model it may not be possible to group all the SBSs using only T colors while ensuring that in any path of length T a MU connects to exactly one SBS from each group.

Assume that for given path length T, there is a coloring $\mathcal{C}_T$ of the SBSs such that over any mobility path of length T a MU connects to T SBSs with different colors. Let L be the number of colors used by $\mathcal{C}_T$. The following theorem states the achievable delivery rate over the backhaul link for this network.

Theorem 1: For given values of the variables N, M, K, T and L, where $L \geq T$, the following delivery rate over the backhaul link is achievable:

$$R(M) = \frac{K - tL}{1 + t}, \tag{2}$$

where $t \triangleq \frac{KMT}{NL}$, and integer valued.

When $L \geq T$, we use the following strategy for the placement phase. First, each file is divided into T disjoint fragments with equal size. These are then encoded into L fragments using a (T,L) MDS code. Hence, any T fragments out of total L is sufficient to decode the original file. Consequently, each group of SBSs (SBSs with the same color) cache a different fragment using the placement phase in [1].

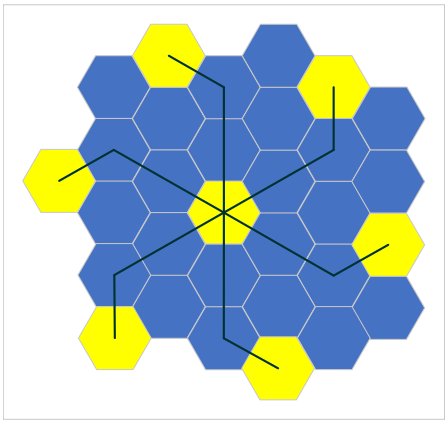


Fig. 4: Frequency reuse pattern i=2, j=1

C. Cell coloring

A simple example for $T = 2$ is illustrated in Figure 3. As one can observe from Figure 3, three colors are sufficient to group the SBSs to ensure that in any mobility path of length two a MU always connects to two SBSs with different colors. Hence, in the placement phase of the given example, each file is initially divided into two fragments which are labeled with green and red colors. These fragments are then coded into additional parity fragment using a (2,3) MDS code which is labeled with blue. Eventually, all the SBSs in the same group (i.e., those with the same color) cache the fragment that has been assigned the same color. Then, at each time slot, the coded delivery phase is executed for each group of SBSs independently.

Assume that the number of SBSs in each group is equal to $\acute{K} = K/L$. If we consider the coded delivery phase of a particular group at a particular time slot, this is identical to the single access model with $\acute{K}$ SBSs each with a cache memory of size $\acute{M} = MT$ files; and hence the corresponding delivery rate is $\frac{\acute{K}-\acute{K}\acute{M}/N}{1+\acute{K}\acute{M}/N}\frac{1}{T} = \frac{K/L-t}{t+1}\frac{1}{T}$, where $t \triangleq \frac{KMT}{NL}$. Accordingly, the overall delivery rate for the backhaul link is found as $R(M) = \frac{K-tL}{1+t}$. We note that this delivery rate is achievable when the group level multicasting gain $t = \frac{KMT}{NL}$ is an integer. For non-integer t values the following lemma can be used to calculate the corresponding achievable delivery rate.

Storage capacity (M/N)	Coded delivery method and network scenario	Number of sub-files	Normalized Delivery rate
$\frac{1}{8}$	Coded delivery [1], for $K=24$	4048	5.25
	Mobility-aware coded delivery for $K=24$	56	6
$\frac{1}{4}$	Coded delivery [1], for $K=24$	2.69×10^5	2.57
	Mobility-aware coded delivery for $K=24$	140	2.4
$\frac{1}{8}$	Coded delivery [1], for $K=48$	2.45×10^7	6
	Coded delivery for $K=48$ with clustering	4048	10.5
	Mobility-aware coded delivery for $K=48$	3640	7.2
$\frac{1}{4}$	Coded delivery [1], for $K=48$	1.39×10^{11}	2.77
	Coded delivery for $K=48$ with clustering	2.69×10^5	5.14
	Mobility-aware coded delivery for $K=48$	2.57×10^4	2.66

TABLE I: Comparison of the proposed mobility-aware coded storage and delivery scheme with conventional coded delivery scheme of [1] in terms of the number of required sub-files and the achieved normalized delivery rate.

Lemma 1: If $t=\frac{KMT}{NL}$ is not an integer, then the following rate is achievable by *memory sharing*

$$R(M)=\left(\gamma\frac{\frac{K}{L}-\lfloor t\rfloor}{\lfloor t\rfloor+1}+(1-\gamma)\frac{\frac{K}{L}-\lceil t\rceil}{\lceil t\rceil+1}\right)L, \qquad (3)$$

where $\gamma\triangleq\lceil t\rceil-t$.

We note that the delivery rate achieved for a random path is higher than the one achieved for a deterministic path, if $L>T$; and it increases further with the number of colors used for coloring cells. Hence, in the single access model mobility the main objective is to find the optimal coloring $\mathcal{C}_T$ that minimizes the number of colors L used for grouping the SBSs. In the following section we study the optimal coloring method for the SBSs and the corresponding backhaul delivery rate.

Recall that, for a given mobility path of length T, our objective is to color the cells using the minimum number colors while ensuring that, in any mobility path each color is encountered at most once. This problem is analogous to the well known *frequency reuse pattern* problem in cellular networks [11], in which co-channel cell (cells serving in the same frequency) locations are determined according to the given distance constraint (the distance between the center of two co-channel cells). In [11], frequency reuse pattern (co-channel cell pattern) is defined via the integer valued shift parameters i and j in the following way: starting from a cell, "move i cells along any chain of hexagons; turn counter-clockwise 60 degrees; move j cells along the chain that lies on this new heading". A frequency reuse pattern example with $i=2$ and $j=1$ is illustrated in Figure 4. It is shown that using a reuse pattern with shift parameters i and j, the two nearest co-channels are separated with a distance $D=\sqrt{3C}$ (scaled with the cell diameter), where $C=i^2+j^2+ij$ is the cluster size (the total number of different frequencies used in the network).

We remark that when the nearest co-channel cells are separated with a distance $D=\sqrt{3C}$ according to the reuse pattern with shift parameters i and j, a user in a particular cell should visit at least $i+j$ (including the current cell) cells to reach the nearest co-channel cell. Therefore, the frequency reuse pattern problem is analogous to our problem, where the length of the mobility path T is equivalent to $i+j$, and the cluster size C is equivalent to the number of colors L. In our problem, we want to minimize the number of colors $L=T^2-ij$ for given mobility length $T=i+j$. Hence, we use the reuse pattern (i,j), with $i=\lceil\frac{T}{2}\rceil$ and $j=\lfloor\frac{T}{2}\rfloor$ to minimize the number of colors L.

Corollary 1: For a given path length T, in order to ensure that in an any path a color is encountered at most once, the minimum number of colors should be

$$L_{min}=\begin{cases}3n^2, & \text{if } T=2n,\\ 3n^2+3n+1, & \text{if } T=2n+1,\end{cases} \qquad (4)$$

for some positive integer n.

IV. Numerical Results

For our simulation results, we consider two heterogeneous cellular network topologies with $K=24$ and $K=48$ SBSs, respectively. We also consider a mobility path of length $T=2$. Hence, the cells are colored according to the reuse pattern $(i=1,j=1)$ with a total of $L=3$ colors as in Figure 3.

We compare the performance of our mobility-aware coded delivery scheme with the coded delivery scheme [1] in terms of two metrics: the number of required sub-files and the normalized backhaul delivery rate. Although the coded delivery scheme is originally designed to reduce the backhaul delivery rate, the number of sub-files (i.e., the sub-packetization level) is also an important metric since the excessive sub-packetization is one of the main obstacles for practical implementations of coded delivery in real networks [13]. For each topology, we analyze the performance of these schemes for two different storage capacities of $M/N=1/4$ and $M/N=1/8$, respectively. The numerical results are presented in Table I.

In the first group of simulations we analyze the network topology with $K=24$ SBSs and observe that our mobility-aware coded delivery scheme reduces the number of sub-files dramatically. When, $M/N=1/8$ our mobility-aware coded delivery scheme induces 12.5% extra delivery rate while reducing the number of sub-files by approximately $1/72$. The more interesting results are observed when the storage capability is higher, i.e., $M/N=1/4$, where the proposed mobility-aware coded delivery scheme outperforms the coded delivery scheme in both performance metrics. At the first

glance this might be counterintuitive since there is a trade-off between the delivery rate and the number of sub-files [13]. However, mobility-aware approach not only utilizes the multicasting gain, but it also utilize the multi-access gain which is clearly visible in the deterministic path scenario. In this network setting the number of required sub-files goes from 269000 down to 140.

In the second group of simulations, we analyze the network topology with $K = 48$ SBSs. When $M/N = 1/8$ our mobility-aware coded delivery scheme results in a 20% increase in the delivery rate, while reducing the number of sub-files by approximately four orders of magnitude.

At this point, one can argue that the number of sub-files could also be reduced by simply clustering the SBSs to obtain two sub-networks with $K/2$ SBSs and then applying the coded delivery scheme to each sub-network independently. Indeed, the clustering approach could reduce the number of sub-files significantly; however, it leads to a further increase in the backhaul delivery rate. The results with the clustering approach, assuming two clusters, each consisting of $K/2 = 24$ SBSs, are also included in Table I. We note that when there are two clusters, the corresponding delivery rate is simply the sum of the delivery rates corresponding to each cluster. Hence, the coded delivery scheme with two clusters uses the same number of sub-files as the coded delivery scheme for the network topology with $K = 24$ SBSs; however, the delivery rate is doubled. One can easily observe that for both $M/N = 1/8$ and $M/N = 1/4$ our mobility-aware coded delivery scheme outperforms the coded delivery scheme with two clusters in terms of both performance metrics. We also observe that the mobility-aware coded delivery approach becomes more efficient compared to to the other two schemes, particularly when the storage capacity is higher. To highlight this fact, for the given path length $T = 2$, consider the extreme point $M/N = 1/2$. In this case the backhaul delivery rate reduces to zero, while the number of subfiles is only two.

V. Conclusions

We have introduced a novel MDS-coded storage and coded delivery scheme that adopts its caching strategy to the mobility patterns of the users. Our scheme exploits a coloring scheme for the SBSs, inspired by frequency reuse patterns in cellular networks, that have been extensively studied in the past to reduce interference. The files in the library are then MDS-coded and stored in the SBS caches, allowing users to satisfy their demands from multiple SBSs on their path under a high mobility assumption. We have showed that the proposed strategy achieves a significant reduction in the number of sub-files, while sometimes also reducing the delivery rate, particularly for high cache capacity scenarios. We are currently extending our analysis to the case of non-uniform file popularity.

References

[1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, May 2014.

[2] J. Hachem, N. Karamchandani, and S. N. Diggavi, "Coded caching for multi-level popularity and access," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, May 2017.

[3] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, Dec. 2013.

[4] X. Xu and M. Tao, "Modeling, analysis, and optimization of coded caching in small-cell networks," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3415–3428, Aug 2017.

[5] J. Liao, K. K. Wong, Y. Zhang, Z. Zheng, and K. Yang, "Coding, multicast, and cooperation for cache- enabled heterogeneous small cell networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6838–6853, Oct 2017.

[6] K. Poularakis and L. Tassiulas, "Code, cache and deliver on the move: A novel caching paradigm in hyper-dense small-cell networks," *IEEE Transactions on Mobile Computing*, vol. 16, March 2017.

[7] E. Ozfatura and D. Gündüz, "Mobility and popularity-aware coded small-cell caching," *IEEE Communications Letters*, vol. 22, no. 2, pp. 288–291, Feb 2018.

[8] T. Liu, S. Zhou, and Z. Niu, "Mobility-aware coded-caching scheme for small cell network," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[9] M. Chen, Y. Hao, L. Hu, K. Huang, and V. K. N. Lau, "Green and mobility-aware caching in 5G networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8347–8361, Dec 2017.

[10] G. Paschos, E. Bastug, I. Land, G. Caire, and M. Debbah, "Wireless caching: Technical misconceptions and business barriers," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 16–22, August 2016.

[11] V. H. M. Donald, "Advanced mobile phone service: The cellular concept," *The Bell System Technical Journal*, vol. 58, no. 1, pp. 15–41, Jan 1979.

[12] N. Mital, D. Gündüz, and C. Ling, "Coded caching in a multi-server system with random topology," 2017.

[13] L. Tang and A. Ramamoorthy, "Low subpacketization schemes for coded caching," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 2790–2794.