On the Exploration and Exploitation Trade-off in Cooperative Caching-enabled Networks

Srikanth Bommaraveni, Sumit Gautam, Thang X. Vu, and Symeon Chatzinotas
Interdisciplinary Centre for Security, Reliability and Trust (SnT), the University of Luxembourg.
Email: {srikanth.bommaraveni, sumit.gautam, thang.vu, symeon.chatzinotas}@uni.lu.

Abstract—Edge-caching is considered as a promising solution to address network congestion and to reduce delivery latency by bringing relevant contents close to users. In this context, the commonly used notion involves the storage of the most popular contents in the cache, while consequently increasing the cache hit ratio (CHR). In the majority of prior works, the content popularity is assumed to be perfectly known and often a priori. However, in reality, the content popularity has to be explored especially for uncertain contents, such as new entrants and items with fast time-varying popularity. In this paper, we develop a framework to analyze the joint exploration and exploitation trade-off by caching both popular and uncertain contents to enable more efficient content caching. Particularly, we formulate an optimization problem to maximize the trade-off between exploration and exploitation subject to the storage capacity, guaranteed CHR and back-haul energy budget constraints. Furthermore, we solve the formulated mixed-integer combinatorial problem by using branch-and-bound method and relaxing the binary to box constraints. The superiority in performance of the proposed method over the state-of-the-art solution is demonstrated in terms of the CHR and back-haul energy on a realistic Movie-lens dataset.

Index Terms—Small cell base station, Cooperative, Cache-enabled Networks, Exploration, Exploitation.

I. INTRODUCTION

The rapid development of smart mobile devices and their hunger for data services like high-resolution video is growing exponentially year-by-year. It is predicted that by 2023, the number of interconnected devices will be more than three times the global population, while the data traffic from these devices is projected to reach 4.8 zettabytes per year [1]. Further, repeated downloads of a few popular contents have been considered as a primary factor for the growth of data traffic, which causes more power consumption that in turn increases carbon emissions. Therefore to address this challenge brought by data traffic over mobile networks and to reduce the back-haul load, content caching at the edge node has been considered as a promising solution [2]–[7]. The main principle of caching is that the storage of most popular contents at the edge node effectively reduces the back-haul load while avoiding duplicate content transmissions from the original content servers [8]. As a result, mobile devices can download data from the edge nodes without utilizing the back-haul link, thereby resulting in the reduction of excess common/popular data transmission. Additionally, the energy consumption of the edge node is reduced by non-essential utilization of back-haul links [9].

Content popularity, which indicates the frequency in the demands, plays an important role in the caching systems [10]. The popularity of different contents is constantly evolving, which consequently influences the Quality-of-Service (QoS). In general, it is observed that only a small number of popular contents are frequently requested [11]. By storing these contents from the content servers placed closer to the end mobile users, downloading of the same content multiple times through the back-haul links can be avoided. Herein, the content popularity is predicted using machine learning techniques in the edge networks [12]–[16]. Most of the existing caching techniques focus on caching in isolated domains, i.e., either at the edge node or at the small cell base stations (SBS). However, in the practical scenarios, content caching is always constrained by the limited storage capacity [12]. An effective solution is to enable the edge nodes to cooperate with each other, resulting in an enlarged set of cached contents at the edge nodes for exploiting the caching diversity [17]. A cost model for content retrieval from the neighboring SBS and the content servers in a decentralized cooperative scenario was proposed in [18]. The authors in [19] studied a distributed caching framework by leveraging user mobility and network coding. In terms of minimizing the system costs like storage, system reconfiguration, content access latency, and content migration, the authors in [20] proposed a cooperate caching system wherein the central offices cooperatively allocate the cloud resources. A graph-assisted cooperative caching based on local content popularity was investigated in [21]. However, all the aforementioned works assume that content popularity is perfectly known a priori, and without any cost constraints.

Recent works on caching are designed to maximize the CHR metric i.e., content served from the cache is a

cache hit otherwise cache miss. Such metrics help design a caching technique to reduce the back-haul load and data transmissions. However, they completely ignore the energy-specific costs related to the transportation of the data. When requested content is cached at the edge node, mobile users download the relevant contents from the cache directly. As a result, the network throughput is improved and the energy consumption of the edge node caching is considered alongside the data transmission costs. When the requested contents are not cached at the edge node, mobile users are required to download contents from the content servers via back-haul link [5]. In this case, the limited back-haul capacity does not only influence the throughput but also affects the energy consumption of networks. Therefore, it is necessary to analyze the energy efficiency performance of cache-enabled networks in conjunction with the development of efficient placement metric designs.

In this paper, we investigate the performance of edge caching wireless networks in which the small cell base stations cooperate by sharing the resources, e.g., energy, back-haul load, power, or financial costs. Specifically, we study the cooperative edge caching wireless networks by storing both the most popular and uncertain contents. The novelty in this work is the formulation of a new objective function using the exploration and exploitation trade-off metric. Further, we formulate a joint optimization problem to maximize the objective function, while taking into consideration the overall back-haul energy cost constraint for the content fetching/placement.

The rest of the paper is organized as follows. Section II describes the system model. In section III, we present the problem formulation. Section IV presents the simulation results and finally, conclusions are drawn in Section V.

Notation: Lower or upper case letters represent scalars, boldface lower case for vectors, $\odot$ is the element-wise product or also called as Hadamard product, $(\cdot)^T$ denotes the transpose operator and $|\cdot|$ represents the cardinality of the set.

II. SYSTEM MODEL

We consider a cache-aided cooperative multi-cell network consisting of N small base stations (SBSs), as shown in Fig. 1. Each SBS $n \in \mathcal{N} = \{1, \ldots, N\}$ is connected to the mobile core network through back-haul links and serves a set of user terminals (UTs), indexed by $\mathcal{U} = \{1, \ldots, K\}$. Further, each SBS is equipped with a cache of size D_n Gigabits (Gbits) to store the fetched content files from the content server (CS). A UT randomly requests the content file f_i from a content file library of F files denoted by $\mathcal{F} = \{f_1, \ldots, f_F\}$, where each content file has size s_F in data unit, and $\mathbf{s} = [s_1, \ldots, s_F]$. Note that each UT is associated with one SBS based on the received signal-to-noise ratio (SNR).
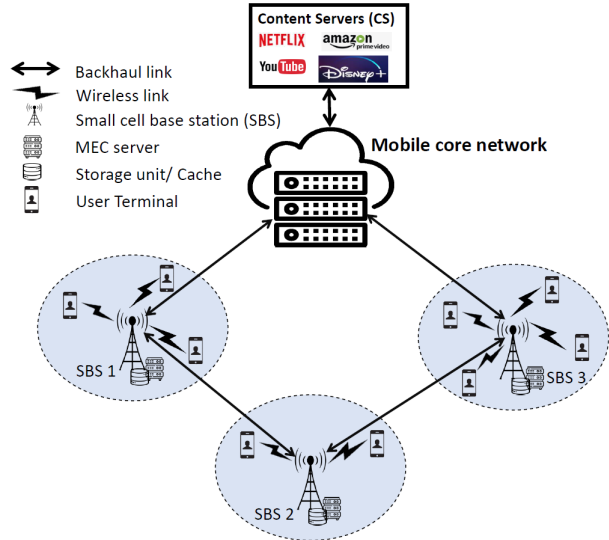

Fig. 1: An illustration of the proposed cache enabled cooperative network with caching performed at the SBS.

Suppose the UT's requested content file is available in the local cache of its associated SBS. In that case, it can directly access the content file from its associated SBS (i.e., avoids the back-haul load). Otherwise, the associated SBS needs to fetch the content file from either the adjacent SBSs or the central content server. The number of contents fetched from the adjacent SBSs or content server to update the cache is defined as the back-haul load, denoted by B_n. However, the incurred cost for retrieving content files from the central content server is higher than the cost incurred for fetching the content files from the adjacent SBSs.

TABLE I: NOTATIONS AND DEFINITIONS

Notation	Definition
$\mathcal{N}$	Set of small cell base stations (SBS)
$\mathcal{U}$	Set of user terminals
$\mathcal{F}$	Set of content files
$\mathbf{s}$	Size of contents
D_n	Size of cache at n^{th} SBS
B_n	Black-haul load of nth SBS
$\mathbf{p}_n$	Popularity vector of nth SBS
$\mathbf{u}_n$	Uncertainty vector of nth SBS
$\mathbf{z}_n$	Placement vector of nth SBS
$\mathbf{x}_n$	Placement vector of popular contents at n^{th} SBS
$\mathbf{y}_n$	Placement vector of uncertain contents at n^{th} SBS
E_n	Energy consumed to refresh cache at n^{th} SBS

At the time interval t, where $t = 1, \ldots, T$, the mobile edge computing server (MEC) monitors the UT request data and keeps the record of the requests received. Let $\mathbf{p}_n(t) = [p_1, \ldots, p_F]$ and $\mathbf{u}_n(t) = [u_1, \ldots, u_F]$ be the popularity and uncertainty vectors of nth SBS at time slot t respectively, which are calculated using the active

learning-based matrix completion[1] as discussed in [6]. Both the popular and uncertain contents are intended to be stored in the SBS's cache. The system leverages this approach to obtain the trade-off between exploration and exploitation. Correspondingly, the exploitation is related to content popularity whereas the exploration is related to the uncertainty of contents.

Let $\mathbf{z}_n(t) \in \{0,1\}^{F \times 1}$ be the placement vector of nth SBS at time slot t. Since both the popular and uncertain contents have to be stored in the cache, we define two binary vectors wherein the elements indicate 1 if the content is stored in the cache, and 0 otherwise. Let $\mathbf{x}_n(t) \in \{0,1\}^{F \times 1}$ and $\mathbf{y}_n(t) \in \{0,1\}^{F \times 1}$ be the binary vectors defining the cache placement of the contents, where $\mathbf{x}_n(t)$ is comprised of the selected popular contents for achieving exploitation, while $\mathbf{y}_n(t)$ consists of the uncertain contents to facilitate the exploration process. The final placement vector is given by

$$\mathbf{z}_n(t) = \mathbf{x}_n(t) + \mathbf{y}_n(t); \mathbf{z}_n(t) \leq 1. \quad (1)$$

Using the defined notations, the popularity of the contents that are used for exploitation are weighted with content size over the inner product of popularity and content size. This is termed as the average CHR metric (Exploitation) of the nth SBS at time slot t, which is mathematically written as

$$A(\mathbf{x}_n(t)) = \frac{\mathbf{p}_n(t)^T (\mathbf{s} \odot \mathbf{x}_n(t))}{\mathbf{p}_n(t)^T \mathbf{s}}. \quad (2)$$

Similarly, we define the uncertainty of contents i.e, the variance of contents that are used for exploration, weighted with content size over the inner product of uncertainty and content size. This is defined as the uncertainty ratio metric (Exploration) of the nth SBS at time slot t, which is mathematically expressed as

$$B(\mathbf{y}_n(t)) = \frac{\mathbf{u}_n(t)^T (\mathbf{s} \odot \mathbf{y}_n(t))}{\mathbf{u}_n(t)^T \mathbf{s}}. \quad (3)$$

The cache is updated based on the popularity and uncertainty measures of the contents at each SBS. Further, the SBS fetches contents either from adjacent SBS or the content sever depending on the backhaul cost. Let the energy consumed to refresh the cache at the nth SBS at time slot t is given by

$$E_n(t) = \Big(\max\big(\mathbf{z}_n(t) - \mathbf{z}_n(t-1), \mathbf{0}\big)\Big)^T \Big(\mathbf{s} \odot \xi_n\big(\mathbf{Z}(t-1)\big)\Big), \quad (4)$$

where $\mathbf{Z}(t-1)$ is the placement matrix at $(t-1)^{\text{th}}$ time slot where the column i indicates the placement vector at i^{th} SBS and $\xi_n(\mathbf{Z}(t-1))$ is the energy consumed per bit transmitted (Joule/bit). It depends on the fixed network

---

[1]It can be replaced with other uncertainty metric without loss of generality.

topology captured in ξ and the previous time slot cached contents captured in $\mathbf{z}_n(t-1)$. Note that, in this work, $\xi_n(\cdot)$ represents the energy consumed, however, it can be generalized to other cost functions such as back-haul load, financial costs, power, etc. To calculate the energy consumed $\xi_n(\mathbf{Z}(t-1))$, we make use of the shortest path algorithm [22]. The links among the SBS's and content servers represent the edges of a graph. The cost to fetch the contents represents the weight on the edge and it is non-negative. More details on the calculation of this cost is provided in the numerical results section based on arbitrary network topology.

In the following, we formulate the optimization problem, to maximize the weighted sum of exploration and exploitation.

III. PROBLEM FORMULATION

The optimization problem is formulated to maximize the objective function (i.e., expressed in terms of average CHR and uncertainty ratio metrics) subject to the binary constraints (i.e., for caching popular and uncertain contents), maximum storage capacity, guaranteed CHR, and back-haul energy budget constraints. The optimization is repeated for each time slot and it requires as inputs the popularity and uncertainty metrics, the energy per bit ξ as they evolve over time. The complete optimization problem is given as

$$\mathcal{P}_1 : \max_{\mathbf{x}_n(t), \mathbf{y}_n(t)} (1-\tau) \sum_{n=1}^{N} A(\mathbf{x}_n(t)) + \tau \sum_{n=1}^{N} B(\mathbf{y}_n(t)) \quad (5)$$

$$\text{s.t.} \quad C_1 : \mathbf{x}_n(t), \mathbf{y}_n(t) \in \{0,1\}^{F \times 1}, \ \forall n = 1 \ldots N,$$

$$C_2 : \mathbf{x}_n(t) + \mathbf{y}_n(t) \leq \mathbf{1}, \ \forall n = 1 \ldots N,$$

$$C_3 : \mathbf{z}_n(t) = \mathbf{x}_n(t) + \mathbf{y}_n(t), \ \forall n = 1 \ldots N,$$

$$C_4 : \mathbf{z}_n(t)^T \mathbf{s} \leq D_n, \ \forall n = 1 \ldots N,$$

$$C_5 : \frac{\max\big(\mathbf{p}_n(t) - \mathbf{u}_n(t), \mathbf{0}\big)^T (\mathbf{s} \odot \mathbf{x}_n(t))}{\mathbf{p}_n(t)^T \mathbf{s}} \geq \theta,$$

$$\forall n = 1 \ldots N,$$

$$C_6 : \sum_{j=1}^{N} E_j(t) \leq \epsilon.$$

The constraints are explained as follows,

- The objective of $\mathcal{P}_1$ consists of the joint weighted sum of uncertainty ratio metric (exploration) and the average CHR (exploitation).
- C_1 ensures that when the cache stores the popular file, the popularity vector's element related to the corresponding content file is 1. Otherwise, it is 0. Likewise, when it stores the uncertain file, the uncertainty vector's element related to the content file is 1. Otherwise, it is 0.
- Constraint C_2 implies that content can be either popular or uncertain, but not both.
- C_3 is the final placement vector which is the sum of popular and uncertain contents placement vectors.

- Constraint C4 specifies the local cache capacity of each SBS.
- C_5 guarantees a lower bound on the CHR, where $\theta \in [0,1]$. This worst case CHR differs from the average CHR in (2) and is calculated by assuming that the actual hits are calculated based on the worst case popularity values i.e. $\max(\mathbf{p}_n(t) - \mathbf{u}_n(t), \mathbf{0})$
- Constraint C_6 ensures that the backhaul link's energy consumption should not be more than the considered threshold value (i.e., ϵ) to update or refresh the cache with new content files.

The formulated optimization problem[2] is a mixed-integer combinatorial problem due to the binary variables/ constraint (C_1), challenging to solve in polynomial time. To get a tractable solution to this problem, we need to relax the binary variables between 0 and 1. After relaxing the binary constraint, the problem becomes a continuous linear problem or convex problem easily solved using the standard branch-and-bound optimizers [23]. Note that herein we utilize the Mosek solver of CVX [24] to obtain an approximate solution.

The computational complexity of the optimization problem is a function of the sum of the number of decision variables and the constraints. So, the problem has $(2N)^3$ decision variables and $(6N + 1)$ constraints. Hence, the computational complexity of $\mathcal{P}_1$ is $\mathcal{O}\left((2N)^3)(6N + 1)\right)$ [25].

IV. NUMERICAL RESULTS

In this section, we present the simulation results to show the performance of the proposed optimization framework in terms of CHR and back-haul energy consumption.

A. Real-world data

We use a real-life, Movie-Lens 1M dataset [26], which consists of ratings on a scale of $\{1, \ldots, 5\}$ for $K = 6040$ UTs, $F = 3952$ movies within the years 2000 to 2003. The dataset is very sparse, i.e., over 96% of UTs gave less than 5 ratings for all the movies, and over 94% of movies have less than 25 ratings for all the UTS. Further, we assumed that each rating corresponds to one request irrespective of the rating number. Moreover, using the zip-codes, we choose only the UTs in New York City, i.e, the zip-codes between 10000 and 10350. Further, we divide these zip-codes into 3 groups where each group corresponds to an SBS. After the preprocessing, each SBS has a total number of contents $|\mathcal{F}| = 146$. Since, the dataset does not contain the content sizes, we assume the normalized content sizes which are uniformly randomly generated between 0.1 and 1 and is shown in Fig. 2. Finally, the popularity and uncertainty vectors are

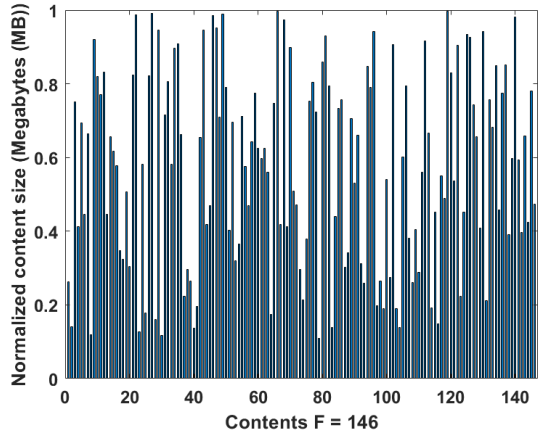[2]We assume that both θ and ϵ are fixed over time without loss of generality.

Fig. 2: Content size distribution

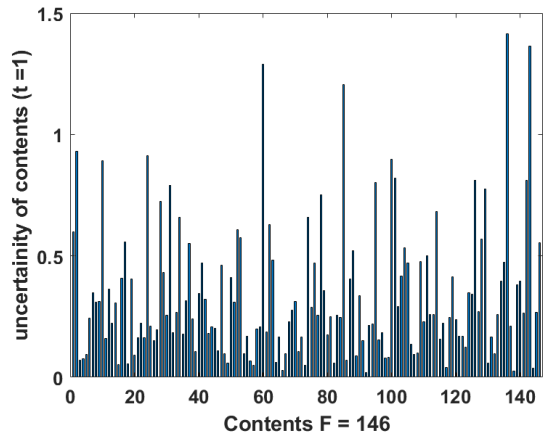Fig. 3: Uncertainty of contents at time interval (t = 1)

calculated using the active learning matrix completion presented in [6]. In order to show the uncertainty of contents, at time slot $t = 1$, we plot the uncertainty of contents versus the contents in Fig. 3.

B. System parameters

The back-haul energy is defined as the energy utilized to fetch the contents from the adjacent SBSs or content servers, which depends on the transport technology. In the literature, the energy per bit transmitted is given by $\delta = 0.5 \times 10^{-8}$ Joule/bit, similar to the values in literature [27]–[30]. Therefore, the back-haul energy of nth SBS at time slot t is calculated as

$$\text{Back-haul energy}_n(t) = B_n(t) \times \delta \quad \text{(Joule/bit)}. \quad (6)$$

By taking into account the true popularity values which lie in the interval $[\mathbf{p} - \mathbf{u}, \mathbf{p} + \mathbf{u}]$, the true CHR of nth SBS at time slot t is calculated as

$$\text{CHR}_{n,\text{true}}(t) = \frac{\mathbf{p}_{n,\text{true}}(t)^T (\mathbf{s} \odot \mathbf{z}_n(t))}{\mathbf{p}_{n,\text{true}}(t)^T \mathbf{s}}. \quad (7)$$
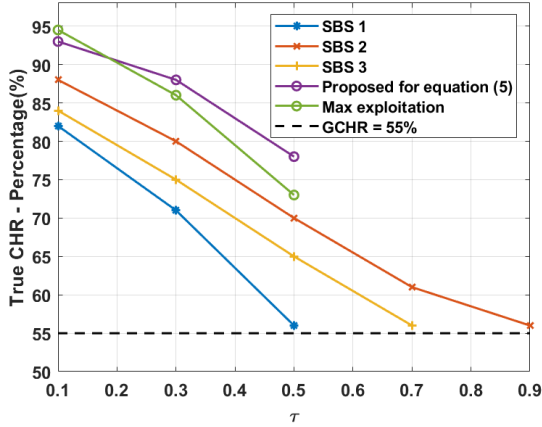
Fig. 4: True CHR vs τ, D = 70 Gbits, θ = 55%, ϵ = 10 Joules
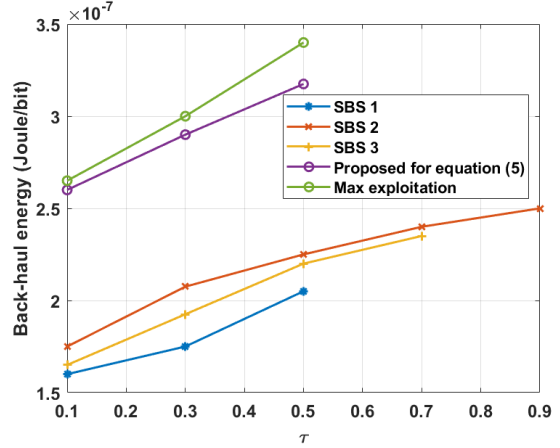

Fig. 5: Back-haul energy vs τ, D = 70 Gbits, θ = 55%, ϵ = 10 Joules

We use the acronym 'Proposed for equation (5)' to refer to the proposed optimization problem of (5) and 'Max exploitation' refers to only exploitation term i.e., the average CHR (2) as the objective function and with the same set of constraints. We assume that the network topology follows the graph properties of adjacency matrix or also called the connection matrix [31]. The adjacency matrix is a symmetric square matrix with entries 1 when there is an edge from vertex and 0 when there is no edge. Moreover, the diagonal elements of the adjacency matrix are all 0, because the edges from a vertex to itself is a loop. The weighted graph using adjacency matrix for the considered network topology is called a cost matrix, which is given as

$$
\mathbf{A} = \begin{array}{c} \\ \text{SBS 1} \\ \text{SBS 2} \\ \text{SBS 3} \\ \text{CS} \end{array} \begin{array}{c} \text{SBS 1} \quad \text{SBS 2} \quad \text{SBS 3} \quad \text{CS} \\ \begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 \\ 2 & 0 & 2 & 0 \end{pmatrix} \end{array}
$$

However, our numerical results can be generalized to any arbitrary network topology. The simulations are performed in this paper for one-time slot of multiple slot simulation. The individual true CHR for each SBS is calculated using the equation (7) and for cooperative and conventional, the true CHR is calculated as the average of all SBS's true CHR.

C. Results and discussions

In Fig. 4, we illustrate the achieved CHR percentages according to the values of τ (i.e., trade-off factor for exploration and exploitation). As can be seen from the results in Fig. 4, we observe that the percentage of CHR reduces by increasing τ. At the higher τ values, the cache mostly stores the uncertain content, leading to the decrement of CHR percentage. In particular, the results illustrate that the system with the conventional scheme achieves higher CHR performance than that with the cooperative scheme at $\tau = 0.1$. This is occurred due to caching of more popular contents at the lower τ values. Also, it is observed from the results that, the proposed cooperative scheme achieves 4% higher CHR than that with the conventional method at $\tau = 0.5$. On the contrary, the results also illustrate that the system experiences an in-feasibility state when it utilizes the cooperative policy or only SBS 1 or SBS 3 for accessing the contents. This is because, for SBS 1, when $\tau = 0.7$ the problem becomes infeasible since the guaranteed CHR is 0.55 which is 55%. But the maximum CHR SBS 1 can achieve is 56% only when $\tau = 0.5$. As a result, the problem becomes infeasible when the τ value is greater than 0.5. A similar observation can be noted for SBS 3. Also, the problem is infeasible for τ greater than 0.5, because in the cooperative and conventional scheme if one SBS has an infeasible solution then the problem becomes infeasible. This result is occurred due to the considered constraint (C5) for the guaranteed CHR. However, using a line search over τ we can find the maximum tolerable exploration while guaranteeing the exploration constraint.

Fig. 5 shows the back-haul energy versus the τ values. From the results in Fig. 4, it is evident that the system utilizes the higher back-haul capacity (i.e., back-haul energy) at the higher τ values. At the higher τ values, the system caches the most unpopular contents leading to the utilization of more back-haul load to fetch the contents from the adjacent SBS or content server. Further, it is clear from the results that the system with the proposed cooperative scheme consumes a lower back-haul energy than that with the conventional method.
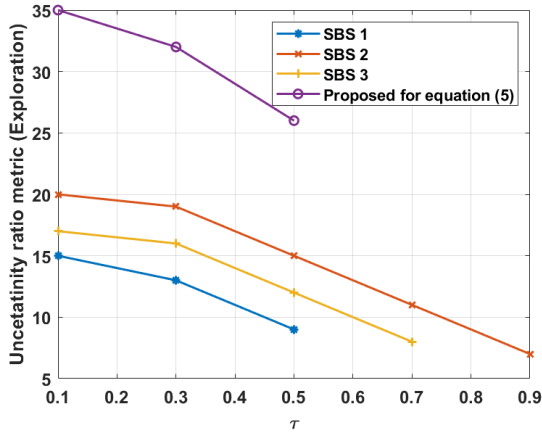
Fig. 6: Exploration objective (uncertainty ratio metric (3)) vs τ, D = 70 Gbits, θ = 55%, ϵ = 10 Joules
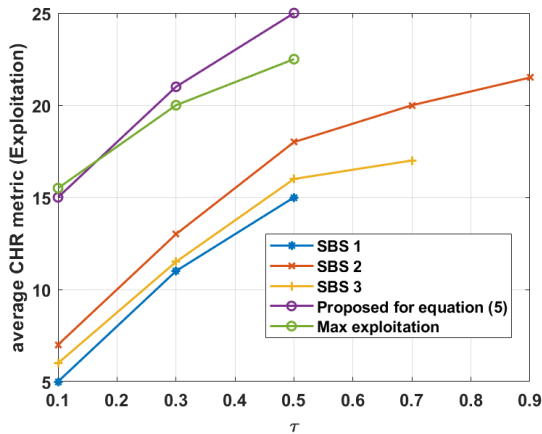

Fig. 7: Exploitation objective (average CHR metric (2)) vs τ, D = 70 Gbits, θ = 55%, ϵ = 10 Joules

In Fig. 6, we show the exploration objective (uncertainty ratio metric (3)) as a function of τ. From the results in Fig. 6, it is observed that as the τ increases, the uncertainty of the contents decreases. This is because more uncertain contents are stored in the cache for exploration as a result the uncertainty of the contents decreases as τ increases.

The exploitation objective (average CHR metric (2)) is shown as a function of τ in Fig. 7. As can be seen from the result in Fig. 7, it is evident that the popularity of contents increases by increasing the τ values. The reason is the uncertainty of the contents decreases by increasing the τ values, leading to the increment of the popularity of the contents.

V. CONCLUSION

In this paper, we investigated the problem of energy-efficient content caching for the cooperative scenario. We developed and analyzed a framework to store both the popular and uncertain contents in the cache to enable more efficient content caching. Based on the estimated popularity and uncertainty of contents, we defined two metrics i.e., the average CHR and uncertainty ratio that was synonymous with the exploitation and exploration, respectively. Further, a joint weighted optimization problem for content caching was formulated, while taking into consideration the aspect of the overall energy consumption and CHR guarantees. The superiority in the performance of the proposed method over the state-of-the-art schemes was established through simulations in terms of CHR and back-haul energy based on the realistic Movielens dataset.

REFERENCES

[1] Cisco, "Cisco Annual Internet Report (2018–2023)," *White paper*, Mar. 2020.
[2] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3098–3130, 2018.
[3] T. X. Vu, E. Baştuğ, S. Chatzinotas, and T. Q. S. Quek (Eds), *Wireless Edge Caching: Modeling, Analysis, and Optimization*. Cambridge University Press, 2021.
[4] K. N. Doan, T. Van Nguyen, T. Q. S. Quek, and H. Shin, "Content-Aware Proactive Caching for Backhaul Offloading in Cellular Network," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3128–3140, 2018.
[5] M. Ji, G. Caire, and A. F. Molisch, "Fundamental Limits of Caching in Wireless D2D Networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, 2016.
[6] S. Bommaraveni, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Active Content Popularity Learning and Caching Optimization With Hit Ratio Guarantees," *IEEE Access*, vol. 8, pp. 151 350–151 359, 2020.
[7] S. Mehrizi, A. Tsakmalis, S. Chatzinotas, and B. Ottersten, "A Bayesian Poisson–Gaussian Process Model for Popularity Learning in Edge-Caching Networks," *IEEE Access*, vol. 7, pp. 92 341–92 354, 2019.
[8] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.
[9] B. Perabathini, E. Baştuğ, M. Kountouris, M. Debbah, and A. Conte, "Caching at the edge: A green perspective for 5G networks," in *IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 2830–2835.
[10] G. S. Paschos, G. Iosifidis, M. Tao, D. Towsley, and G. Caire, "The Role of Caching in Future Communication Systems and Networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1111–1125, 2018.
[11] 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, 2016.
[12] S. Tamoor-ul Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-aho, "Caching in Wireless Small Cell Networks: A Storage-Bandwidth Tradeoff," *IEEE Wireless Commun. Lett.*, vol. 20, no. 6, p. 1175–1178, Jun 2016.

[13] J. Song, M. Sheng, T. Q. S. Quek, C. Xu, and X. Wang, "Learning-Based Content Caching and Sharing for Wireless Networks," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4309–4324, 2017.

[14] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and Scalable Caching for 5G Using Reinforcement Learning of Space-Time Popularities," *IEEE J. Sel. Topics Signal Process*, vol. 12, no. 1, pp. 180–190, 2018.

[15] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A Learning-Based Approach to Caching in Heterogenous Small Cell Networks," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1674–1686, 2016.

[16] S. O. Somuyiwa, A. György, and D. Gündüz, "A Reinforcement-Learning Approach to Proactive Caching in Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1331–1344, 2018.

[17] X. Li, X. Wang, K. Li, Z. Han, and V. C. M. Leung, "Collaborative Multi-Tier Caching in Heterogeneous Networks: Modeling, Analysis, and Design," *IEEE Trans. Commun*, vol. 16, no. 10, pp. 6926–6939, 2017.

[18] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "In-network caching and content placement in cooperative small cell networks," in *International Conference on 5G for Ubiquitous Connectivity*, 2014, pp. 128–133.

[19] K. Poularakis and L. Tassiulas, "Code, Cache and Deliver on the Move: A Novel Caching Paradigm in Hyper-Dense Small-Cell Networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 675–687, 2017.

[20] L. Pu, L. Jiao, X. Chen, L. Wang, Q. Xie, and J. Xu, "Online Resource Allocation, Content Placement and Request Routing for Cost-Efficient Edge Caching in Cloud Radio Access Networks," *IEEE J. Sel. Areas Commun*, vol. 36, no. 8, pp. 1751–1767, 2018.

[21] Y. Jiang, X. Cui, M. Bennis, F. Zheng, B. Fan, and X. You, "Cooperative caching in fog radio access networks: a graph-based approach," *IET Communications*, vol. 13, no. 20, pp. 3519–3528, 2019.

[22] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[23] "Practical guidelines for solving difficult mixed integer linear programs," *Surveys in Operations Research and Management Science*, vol. 18, no. 1, pp. 18 – 32, 2013.

[24] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[25] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali, *LMI control toolbox user's guide*, 05 1995.

[26] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, Dec. 2015. [Online]. Available: https://doi.org/10.1145/2827872

[27] E. Baştuğ, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, 2014, pp. 649–653.

[28] Y. Xu, Y. Li, Z. Wang, T. Lin, G. Zhang, and S. Ci, "Coordinated caching model for minimizing energy consumption in radio access network," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 2406–2411.

[29] J. Llorca, A. M. Tulino, M. Varvello, J. Esteban, and D. Perino, "Energy efficient dynamic content distribution," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2826–2836, 2015.

[30] F. Gabry, V. Bioglio, and I. Land, "On energy-efficient edge caching in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3288–3298, 2016.

[31] Ravindra B. Bapat, *Adjacency Matrix. In: Graphs and Matrices*. Springer, London, 2010.