Context-Aware Caching With Social Behavior in MEC-enabled Wireless Cellular Networks

Xinwei Liu, Chuanhao Sun, Xing Zhang

Wireless Signal Processing and Network Laboratory
Key Laboratory of Universal Wireless Communication, Ministry of Education
Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876, P.R. China
Email: lxwalyw@gmail.com

***Abstract*—Recent advances in big data processing and mobile edge computing (MEC) have demonstrated the feasibility of further performance improvement on wireless caching via exploiting contextual information, like mobility-aware caching. However, most related works either focus on the utilization of single contextual information or ignore the impact of social network on reshaping a user's behavior when integrating various context data. In this work, a context-aware caching scheme with social behavior that integrates various contextual information is proposed for MEC-enabled small cell networks (SCNs), with a novel model provided to characterize how social networks influence and reshape behaviors. Based on a multi-armed bandit algorithm, a user's behavior patterns such as its mobility pattern, preference towards different files and the amount of traffic it may consume in each time slot are learned under the impact of social relationships. Numerical results and comparisons show that the proposed caching scheme outperforms normal context-aware caching schemes by more than 30% in terms of backhaul data offloading ratio.**

I. Introduction

The unprecedented global growth of mobile data traffic has become the performance bottleneck of 5G networks and it is predicted by [1] that such an explosion of data traffic will be continuous and more fierce in the future. Traditional network-centric methods like network densification and frequency division multiplexing are not the economical and efficient ways to address the issue because of the limited capability of backhaul links and the high deployment costs.

Caching at the edge has been widely recognized as a promising method to alleviate the burden of backhaul links and tackle the aforementioned challenge. Equipped with storage, small base stations (SBSs) can serve their subscribers without backhaul congestion if it has already cached the requested files. Exploiting user's request pattern, [2] studied caching scheme by modeling content popularity follows Zipf distribution. The authors of [3] proposed their caching schemes for heterogeneous cellular networks considering the interference among BSs on the basis of stochastic geometry theory. In [4], the tradeoff between transmission diversity and content diversity in cluster-centric cooperative SCNs was discussed. All these works showed that, integrated with fundamental user context information (e.g., the overall user preferences), wireless caching can achieve a higher performance.

In recent years, with the progress of big data processing and MEC, mobile edge networks are not just cache-enabled, but also able to analyze the huge amount of collected Radio Access Network (RAN) context data through machine learning tools in real time, which paves the ways for a more intelligent caching decision. In [5], mobility-aware caching scheme was proposed for MEC-enabled SCNs on the basis of an inter-contact model that characterizes user mobility by the length of sojourn time. The authors of [6] proposed a social-aware distributed caching strategy for device-to-device (D2D) networks, where the closeness between users are measured by both their physical distance and the social relationships. Although these recent works have provided valuable insights into context-aware caching schemes in in cellular networks and D2D systems, they only focus on single contextual information. It was showed by [7] that popularity and mobility aware caching can achieve a high performance by integrating more RAN context data simultaneously. However, [7] actually studied different user contextual information separately and assumed these contexts to be independent, which is unreasonable under the influence of social networks.

Motivated by above insights and shortcomings, in this paper, we propose a context-aware caching scheme with social behavior for a MEC-enabled SCN, where various contextual information such as, a user's mobility pattern, preference towards files, the amount of its consumed traffic and its social relationships will be jointly considered. Furthermore, a novel model that characterizes how social networks influence and reshape an individual's behavior patterns is provided for a more reasonable analysis and integration of various contexts. Based on a multi-armed bandit algorithm[7], a SBS periodically updates its cache in each time slot according to the mappings between user types and the corresponding behavior patterns unknown a priori. The main contributions of this paper are summarized as follows:

- We propose a context-aware caching scheme with social behavior for MEC-enabled SCNs, aiming to maximize offloaded backhaul data.
- A novel model is provided to characterize how social networks influence and reshape an individual's behavior

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

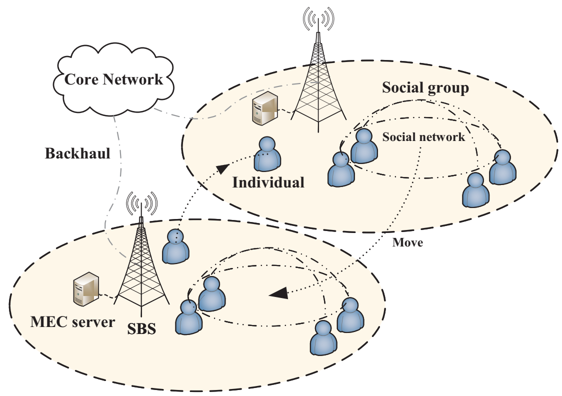

Fig. 1: An illustration of a one-tiered SCN, where some users move in a group and others move alone.

patterns on the basis of the intensity of the social ties between users.

- Numerical results are obtained under different settings. Comparisons demonstrate that the proposed caching scheme outperforms normal context-aware caching schemes in terms of cumulative offloaded backhaul data, traffic offloading ratio and storage efficiency.

The remainder of this paper is organized as follows. Section II describes the system model, including network model, user behavior patterns model and a novel model characterizing the impact of social networks on a user's behavior patterns. In Section III, the investigated problem is formulated, with the optimal caching decisions obtained through a fundamental reinforcement learning algorithm. In Section IV, numerical results are illustrated and performance comparisons are conducted. Finally, conclusions are summarized in Section V.

II. System Model

In this section, we first present the network model and then introduce the user behavior patterns model. Finally, a novel model is provided to characterize the impact of social networks on a user's behavior patterns.

A. Network Model

As shown in Fig. 1, we consider a one-tiered SCN, where each SBS is equipped with a MEC server to collect various RAN context information and cache files more intelligently. The MEC server can store up to M files out of the whole content library that consists of F files. Let $\mathcal{F} = \{1, 2, \ldots, F\}$ denotes the set of file indices. Therefor, the following equation is derived:

$$M = \theta F, \tag{1}$$

where $0 < \theta < 1$ indicates the storage capacity of a MEC server and we henceforth call θ the *storage capability coefficient*. Without loss of generality, we assume that there are N_S SBSs deployed in the area with U active users, and the coverage areas of the these SBSs are non-overlapping. The set of user indices is denoted by $\mathcal{U} = \{1, 2, \ldots, U\}$. A user will be served by a specific SBS if he or she is in its coverage. When the serving SBS has already cached the file that the user is requesting, no additional traffic load is put on the backhaul and the user can enjoy the high transmission rate. Otherwise, the SBS has to download the file from the core network, which results in a heavy burden on backhaul links.

B. User Behavior Patterns Model

According to [7], users can be divided into different types based on personal characteristics which may, for example, be demographic factors (e.g., age, gender), personality, job and so on. Furthermore, [7] assumed that each user type maintains a specific mapping to content popularity that is unknown a priori. In this work, we extend the aforementioned assumption and suppose that there are mapping relationships, which are unknown a priori and needed to be learned, between user types and behavior patterns. Additionally, we also assume that users are divided into K types based on their personal information, and let $\mathcal{K} = \{1, 2, \ldots, K\}$ indicates the set of type indices.

We describe a user's behavior patterns by its mobility pattern, preferences towards files and the amount of consumed data traffic in each time slot. In terms of mobility pattern modeling, similar to that of [8], location-based model is applied to capture a user's moving preference. Specifically, we use a matrix $\mathbf{P}$ of size $N_S \times N_S \times K$ to denote the set of the transition probability, where $0 \leq p_{i,j}^k \leq 1$ indicates the probability of a user who belongs to k-th type leaves SBS i for SBS j in each time slot. In particular, $i = j$, i.e., $p_{i,i}^k$ means the probability of a user who belongs to k-th type stay in SBS i in each time slot.

As for the file preferences, in the same vein, a matrix $\mathbf{Q}$ of size $F \times K$ is provided to denote the set of the request probability of each file, where $0 \leq q_f^k \leq 1$ indicates the the request probability of a user of k-th type for the file f. In terms of the amount of consumed data traffic in each time slot, let a vector $\mathbf{D} = [d_1, d_2, ..., d_K]$ denotes the set of the amount of consumed data traffic of each user type, where d_k indicates the amount of consumed data traffic of a user who is k-th type. To sum up, there exist a specific type $k_u \in \mathcal{K}$ for a user u and therefore the corresponding behavior patterns can be described by $p_{i,j}^{k_u}$, $q_f^{k_u}$ and d_{k_u}. We henceforth call 4-tuple $(k_u, p_{i,j}^{k_u}, q_f^{k_u}, d_{k_u})$ the *tuple of behavior patterns* for a user u.

C. Impact of Social Networks on Behavior Patterns

We use a directed, weighted complete graph $\mathcal{G} = (V, E)$ to express a social network, where V is the set of users, with $|V| = U$. The edges in E represent the relationships between users and the corresponding edge weights indicate the intensity of the social ties between users. Formally, a matrix $\mathbf{W}$ of size $N_S \times N_S$ is used to denote the set of edge weights, where $0 \leq w_{i,j} \leq 1$ reflects the degree of how friendly is user i to user j, or in other words, the personal influence of user j on user i. The value of $\mathbf{W}$ can actually be quantified through the analysis on the data trace collected from social platforms, such as Facebook, Twitter, WeChat and so on [9]. However, the detailed data processing is beyond the scope of this paper, and the $\mathbf{W}$ is thus assumed to be known a priori. In particular, for $i \in \mathcal{U}$, $w_{i,i}$ is set to be 1.

Based on empirical knowledge, we suppose that users under the coverage of the same SBS will act together if the degree of the relationships between them are high enough, which is in accordance with the core idea of [6] and [10] that take both physical distance and social relationships among users into account. Specifically, for user *i* and user *j*, if both $w_{i,j} \geq \sigma$ and $w_{j,i} \geq \sigma$, then they will form a *social group* as shown in Fig. 1. $0 \leq \sigma \leq 1$ is a threshold controlling the tendency of the social group formation, the value of which we assume is determined geographically. A small σ means people, for example, in a business street or a college, tend to act together while a large σ indicates people, for example, in library or residential area, are likely inclined to enjoy their personal time. Let σ_i denotes the threshold of the coverage area of SBS *i*. We thus formally use ψ denote a social group under the coverage of SBS *i*, the definition of which is provided as follows.

Definition 1(Social group): A social group under the coverage of SBS *i* is:

$$\psi = \{x, y | x, y \in \mathcal{U}, w_{x,y} \geq \sigma_i, w_{y,x} \geq \sigma_i\}. \tag{2}$$

Without loss of generality, we assume that users in the same social group ψ act together in an influence-oriented manner. In other words, for each user $x \in \psi$, its individual behavior patterns, i.e., $(k_x, p_{i,j}^{k_x}, q_f^{k_x}, d_{k_x})$ will be distorted under the influence of the internal relationships of ψ. We henceforth regard a social group ψ as a *hyper user* Φ who can also be described like an individual by a 4-tuple. On the basis of the aforementioned assumption, the user type of such a hyper user Φ is determined by the power of influence of all the members in ψ. Let I_x indicates the personal influence of a member *x*, which is defined as follows.

Definition 2(Personal influence): The personal influence of a member *x* in ψ is:

$$I_x = \sum_{u \in \psi, u \neq x} w_{u,x}, \ \forall x \in \psi. \tag{3}$$

According to the above definition, let θ_x denotes the comprehensive influence coefficient of a member $x \in \psi$, which reflects its comprehensive influence in ψ and is defined as follow.

Definition 3(Comprehensive influence coefficient): For a member $x \in \psi$, its comprehensive influence coefficient is:

$$\theta_x = \frac{I_x}{\sum_{u \in \psi} I_u}, \ \forall x \in \psi, \tag{4}$$

where $\theta_x \in [0, 1]$. A large θ_x means the member *x* is of great influence among users in ψ, and vice versa. Combing the aforementioned definitions and assumptions, we thus derived the model that characterizes how social networks influence and reshape an individual's behavior patterns in an influence-oriented manner. The user type of a hyper user Φ is modeled as follows:

$$k_\Phi = \min_{k \in \mathcal{K}} \left| k - \sum_{x \in \psi} \theta_x k_x \right|, \tag{5}$$

where $\sum_{x \in \psi} \theta_x k_x$ indicates the actual user type of a hyper user Φ, which is determined by the internal internal relationships of ψ. Intuitively, a member of great influence will make the type of such a hyper user more similar to that of its via a large θ_x, and vice versa. For an unified analysis on both individuals and hyper users, $\sum_{x \in \psi} \theta_x k_x$ is approximated by the nearest $k \in \mathcal{K}$. Therefore, no matter whether *u* is an individual or a hyper user, the tuple of its behavior patterns can be denoted by $(k_u, p_{i,j}^{k_u}, q_f^{k_u}, d_{k_u})$ in a unified way with d_{k_u} is derived as follows:

$$d_{k_u} = \sum_{i \in \psi} d_{k_i}, \tag{6}$$

where d_{k_i} indicates the amount of the consumed traffic of a member $i \in \psi$ in each time slot. In particular, if *u* is an individual, then $|\psi| = 1$, which means an individual can be considered as a special social group.

III. Problem Formulation and Solution

A. Problem Formulation

The goal of a SBS *x* is to maximize the expected cumulative amount of offloading data traffic up to the finite time horizon *T*. We henceforth regard an individual or a hyper user unifiedly as a *generalized user* denoted by δ. In addition, let $\mathcal{U}_{x,t}$ indicates the set of generalized users who are under the coverage of SBS *x* in time slot *t*. Then, the problem of cache content placement can be formally written as follows:

$$\begin{aligned} \max & \sum_{t=1}^{T} \sum_{f \in \mathcal{F}} y_{t,f} \sum_{i=1}^{N_S} \sum_{\delta \in \mathcal{U}_{i,t}} p_{i,x}^{k_\delta} q_f^{k_\delta} d_{k_\delta}, \\ \text{s.t.} \quad & k_\delta \in \mathcal{K}, \\ & M < |\mathcal{F}|, \\ & \textstyle\sum_{f \in \mathcal{F}} y_{t,f} \leq M, \ t = 1, \ldots, T, \\ & y_{t,f} \in 0, 1, \ t = 1, \ldots, T, \ f \in \mathcal{F}, \end{aligned} \tag{7}$$

where the binary variable $y_{t,f}$ indicates the caching decision whether file *f* is cached in time slot *t*. $y_{t,f} = 1$ means file *f* should be cached in time slot *t*, and vice versa. The aforementioned problem is an integer linear programming problem, which can be decoupled into *T* independent sub-problems. Each sub-problem is a special case of the knapsack problem [7], where the knapsack capacity is *M* and each item is of unit weight. Therefore, in time slot *t*, ranking the files in descending order according to their value measured on the basis of $\mathcal{K}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{D}$ and $\mathbf{W}$, the optimal caching decisions can be derived by choosing the *M* top of all *F* files, which is denoted by Φ_t^*. Let $v_{f,t}^x$ indicates the value of a file *f* in time slot *t* if it is cached in SBS *x*, then its expression is:

$$v_{f,t}^x = \sum_{i=1}^{N_S} \sum_{\delta \in \mathcal{U}_{i,t}} p_{i,x}^{k_\delta} q_f^{k_\delta} d_{k_\delta}. \tag{8}$$

The above equation indicates that if a generalize user δ who is a heavy consumer of data traffic (i.e., with a large d_{k_δ}) will leave SBS *i* for SBS *x* and request file *f* with high probability in time slot *t* (i.e., both $p_{i,x}^{k_\delta}$ and $q_f^{k_\delta}$ are large), then it is

TABLE I: Simulation parameters

Parameter	Value	Parameter	Value
N_S	5	U	100
M(default)	40	F	200
K	64	T	6000

reasonable and desirable for SBS x to cache file f in time slot t. Formally, we denote the file with the i-th file in Φ_t^* by $f_{i,t}^*$ and thus derived the following equations:

$$f_{i,t}^* = \begin{cases} \arg\max\limits_{f\in\mathcal{F}} v_{f,t}^x, & i=1 \\ \arg\max\limits_{f\in\mathcal{F}\setminus(\cup_{k=1}^{i-1}\{f_k^*\})}, & 1<i\leq M \end{cases}. \quad (9)$$

In time slot t, given $\mathcal{K}$, **P**, **Q**, **D** and **W**, the set of files cached in SBS x $\Phi_t^* = \cup_{i=1}^{M}\{f_{i,t}^*\}$ can be derived by plugging (8) into (9).

B. *Context-Aware Caching With Social Behavior Algorithm*

In practice, **P**, **Q** and **D** are unknown a priori and therefore a SBS x has to learn the corresponding expected values [7]. To address the issue, an algorithm call *Context-Aware Caching With Social Behavior (CCSB) Algorithm* built on a multi-armed bandit algorithm is provided to derived the estimated values of **P**, **Q** and **D** , denoted by $\hat{\mathbf{P}}$, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{D}}$ respectively, via renewing the sample means of their elements in each time slot. The pseudocode is presented in Algorithm 1, and the general procedure of which is as follows. At the the begin of a time slot t, the SBS x first collects all the subscriber's personal information with the help of the distributedly deployed MEC servers and thus obtains each subscriber's user type. Furthermore, based on **W** that contains the intensity of social relationships and the threshold σ of each SBS, generalized users are obtained. Therefore, for each SBS i, $\mathcal{U}_{i,t}$ is derived. Then, Φ_t^* is obtained through (8) and (9) on the basis of $\hat{\mathbf{P}}$, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{D}}$, the sample means of which are recently renewed at the end of time slot t-1. In order to balance the tradeoff between caching files about which little information is available (*exploration*) and files of which it believes that they will yield the highest values (*exploitation*)[7], a ϵ-Gready algorithm is applied. Specifically, the SBS x has a probability of ϵ to cache a file randomly from $\mathcal{F}\setminus\Phi_t^*$, otherwise cache a file from Φ_t^* in an order of $f_1^*, f_2^*, \ldots, f_M^*$. Finally, at the end of time slot t, according to the observed behavior patterns of each generalized user u, i.e., $(k_u, p_{i,j}^{k_u}, q_f^{k_u}, d_{k_u})$, renew the corresponding sample means $\hat{p}_{i,j}^{k_u}$, $\hat{q}_f^{k_u}$ and $\hat{d}_{k_u}$.

IV. Numerical Results and Analysis

In this section, we numerically evaluate the proposed caching scheme by comparing its performance with other widely known caching schemes. The simulation parameter settings are listed in TABLE 1. In our experiments, synthetic datasets are used to model the behavior patterns of an individual or a super user. Specifically, for each user type $k\in\mathcal{K}$, the $p_{i,j}^k$, q_f^k and d_k are generated based on some specific distributions respectively (e.g., uniform distribution) [11].

Algorithm 1 CCSB Algorithm

```
Input: T, {σ_i | i = 1, 2, ..., N_S} and W
Output: y_{t,f}
Initialize P̂, Q̂ and D̂
for t=1,2,...T do
    for i=1,2,...N_S do
        Observe generalized users based on W and σ_i
        Obtain U_{i,t}, and k_u of a generalized user u ∈ U_{i,t}
    end for
    Derive Φ*_t on the basis of (8) and (9)
    for f=1,2,...M do
        rand ← a random number ∈ [0, 1]
        if rand < ϵ then
            Cache a file randomly from F \ Φ*_t
        else
            Cache a file from Φ*_t in an order of f*_1, f*_2, ..., f*_M
        end if
    end for
    for i=1,2,...N_S do
        Observe (k_u, p^{k_u}_{i,j}, q^{k_u}_f, d_{k_u}) of a generalized user u ∈ U_{i,t}
    end for
    Renew the sample means of p^{k_u}_{i,j}, q^{k_u}_f and d_{k_u}
end for
```

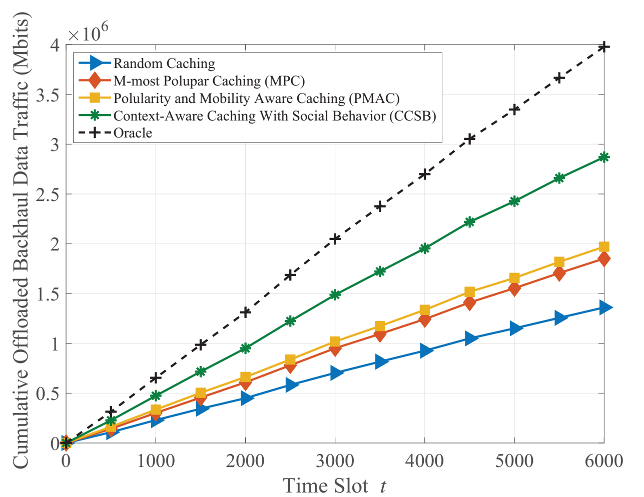

Fig. 2: Cumulative offloaded backhaul data traffic as a function of time slots

Comparisons among the proposed caching scheme and other caching schemes are plotted in Fig. 2 to evaluate the performance of CCSB, with the storage capability coefficient θ set to be 0.2. The black curve labeled by Oracle indicates the performance of CCSB when the **P**, **Q** and **D** are known a priori, which is also the upper bound of the performance of CCSB. Therein, the performance of the random caching scheme is the worst as it just simply fills the storage of a SBS. M-most popular caching (MPC) [3] that only focuses on file popularity achieves a higher cumulative offloaded hackhaul

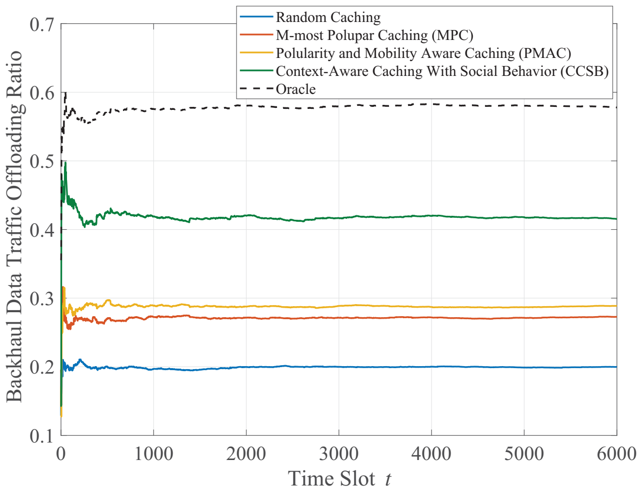

Fig. 3: Backhaul data traffic offloading ratio as a function of time slots

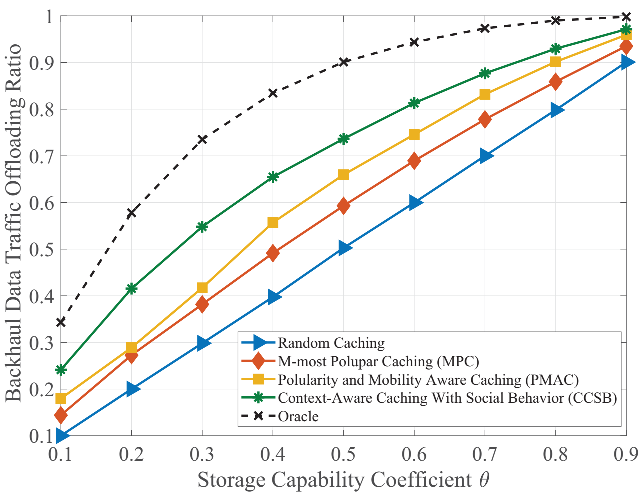

Fig. 4: Backhaul data traffic offloading ratio vs. storage capability coefficient

data traffic than that of random caching, while popularity and mobility aware caching (PMAC) [7] is better due to the additional integration of the contextual information of user mobility. Compared with these caching schemes, the proposed CCSB not only exploits various contextual information but also takes the influence of social networks into account, and thus achieves an obvious performance improvement. Besides, such an advantage is much more remarkable over time.

In order to quantify the performance differences among the aforementioned caching schemes, a relative performance measure called the *backhaul data traffic offloading ratio* [10] is utilized, which is defined as the ratio of the amount of the offloaded backhaul data traffic compared to the overall backhaul data traffic. As shown in Fig. 3, our proposed CCSB outperforms the PMAC [7] in terms of backhaul data offloading ratio by more than 30%.

In Fig. 4, performance comparisons are conducted under different storage capabilities reflected by θ which is defined as the ratio of the number of files a SBS can cache compared to the total number of files. Although all the performances of these caching schemes show an upward trend and finally tend to 100% with the increase of θ, their the storage utilization efficiencies are different. Intuitively, the proposed CCSB exploits storage in a higher efficiency comparing with other caching schemes, especially when storage resources are poor. For instance, when $\theta = 0.2$, CCSB can make more than 40% backhaul data traffic offloaded while the best one of the other three, i.e., PMAC, can only achieve around 30%.

V. Conclusion

In this paper, we propose a context aware caching scheme with social behavior for MEC-enabled SCNs, aiming to maximize the offloaded backhaul data. Various user contextual information such as mobility pattern, preference towards files and the amount of consumed traffic in each time slot are integrated. Furthermore, the impact of social networks is considered, with a novel model characterizing how an individual' behavior patterns are distorted by social relationships. Based on a multi-armed bandit algorithm, a generalized user's behavior patterns are learnt under the impact of social networks. By comparing with other widely known caching schemes, the superiority of the proposed caching scheme is demonstrated in terms of cumulative offloaded backhaul data, traffic offloading ratio and storage utilization efficiency.

Acknowledgment

This work is supported by the National Science Foundation of China (NSFC) under grant 61571054, 61771065 and 61631005.

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021," *White Paper*, Feb 2015.

[2] L. Qiu and G. Cao, "Popularity-aware caching increases the capacity of wireless networks," in *Proc. IEEE GLOBECOM*, 2017, pp. 1-9.

[3] J. Zhang, X. Zhang, M. A. Imran, et al, "Energy Efficiency Analysis of Heterogeneous Cache-enabled 5G Hyper Cellular Networks," in *Proc. IEEE GLOBECOM*, 2016, pp. 1-6.

[4] Z. Chen, J. Lee, T. Q. S. Quek, et al, "Cooperative Caching and Transmission Design in Cluster-Centric Small Cell Networks," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 5, pp. 3401-3415, 2017.

[5] X. Liu, J. Zhang, X. Zhang and W. Wang, "Mobility-Aware Coded Probabilistic Caching Scheme for MEC-Enabled Small Cell Networks," *IEEE Access*, vol. 5, pp. 17824-17833, 2017.

[6] C. Ma et al., "Socially Aware Caching Strategy in Device-to-Device Communication Networks," *IEEE Trans. on Vehicular Technology*, vol. 67, no. 5, pp. 4615-4629, May 2018.

[7] S. Müller, O. Atan, M. van der Schaar and A. Klein, "Context-Aware Proactive Content Caching With Service Differentiation in Wireless Networks," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 2, pp. 1024-1036, Feb. 2017.

[8] K. Poularakis, L. Tassiulas, "Exploiting user mobility for wireless content delivery," in *Proc. IEEE ISIT*, 2013, pp. 1017 -1021.

[9] P. Netrapalli and S. Sanghavi, "Learning the graph of epidemic cascades," in *Proc. 12th ACM Joint Intl. Conf. on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2012.

[10] Y. Wang, M. Ding, Z. Chen and L. Luo, "Caching Placement with Recommendation Systems for Cache-Enabled Mobile Social Networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2266-2269, Oct. 2017.

[11] L. E. Chatzieleftheriou, M. Karaliopoulos and I. Koutsopoulos, "Caching-aware recommendations: Nudging user preferences towards better caching performance," in *Proc. IEEE INFOCOM*, 2017, pp. 1-9.