Individual Preference Aware Caching Policy Design for Energy-Efficient Wireless D2D Communications

Ming-Chun Lee, *Student Member, IEEE*, and Andreas F. Molisch, *Fellow, IEEE*
Ming Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA, USA
Email: mingchul@usc.edu, molisch@usc.edu

***Abstract*—Video caching at the wireless edge is a promising approach to improve throughput and reliability of video delivery. While most existing literature considers homogeneous user preference modeling when designing caching policies, this paper investigates how to exploit *statistical* information on *individual* popularity preferences in the design of caching policies in base station assissted wireless device-to-device networks. Specifically, we formulate the caching policy design problem aiming to minimize average energy consumption and propose two approaches that solve the problem under different conditions: users can or cannot coordinatedly design their policies. By simulating systems using practical individual preference models, we show that the proposed designs are effective, and the performance gain brought by exploiting individual preferences can be clearly observed.**

I. Introduction

Wireless data traffic driven by the demand for video content has increased dramatically over the past several years and is predicted to continue its steady increase for at least the next five years [1]. This motivates the recent urgency for proposing new network-improving approaches, such as, cell densification, large-scale antenna systems, and millimeter-wave communications [2]. Different from those approaches that tend to improve wireless networks without regard to the type of data that are to be transmitted, video caching at the wireless edge exploits the unique video accessing behavior of typical consumers and the cheap storage resources to support the demand for video content with low cost. The potential of video caching renders it widely discussed in recent years [3]–[13].

There are several utilizations of video caching in wireless networks [5]–[13]. By exploiting storage resources in helper nodes and base stations (BSs) to cache video content, femtocaching and BS caching can offer immediate service to users even in the absence of a fast backhaul [5], [6], [9], [10]. By caching files in mobile devices, users can access video content directly from their own storage without energy or spectrum cost and/or from other users by using low cost device-to-device (D2D) communications [8], [11]–[13]. The collaborations between users lead to a denser and lower cost caching network compared to femto- and BS- caching [11]–[13]. Combining caching with multicast coding, coded caching has also been widely investigated [7], [12].

Most of the previous literature investigating caching networks uses the homogeneous popularity model which assumes all users have the same file preference distribution - in other words, each user requests files independently and randomly, according to the *same* popularity distribution. However, this model is counter-intuitive because different users indeed have different taste and preferences. Therefore the homogeneous popularity model can only be an approximation of the true user behavior, and designs based on this model are restricted by their lack of considering individual user preferences [14]–[17]. We note that, by replacing homogeneous modeling with heterogeneous modeling, significant performance improvement has been demonstrated both in [14], [16]–[18] and this work.

Most recently, researchers have started to consider individual preferences and shown the capablility of using this information to improve wireless caching networks [14]–[18]. In [14], individual preferences were studied, and a machine learning approach was used to learn from user preferences and decide which video to preload onto a local device cache purely by consumption of the particular user. While this kind of approach (also known as the ”Netflix challenge”) is very important for recommendation systems and preloading on individual devices, it does not provide a tractable statistical model or network. Our recent work [15] derived a *statistical* model for individual user preferences based on extensive real-world data. In [16] and [17], individual preferences for files were exploited to categorize users into groups and design caching policies for different group to maximize the successful file discovery probability. In [18], a content push strategy was designed and optimized to maximize D2D offlaoding gain by jointly considering the influences of the user preference and sharing willingness.

In this work, we consider BS-assissted D2D caching networks in which users can reach desired video files through the BS link, D2D communications, and self-cache with different costs. Motivated by the potential benefit of exploiting individual preferences, we model the caching design problem exploiting such individual preferences and propose caching policy designs. The design goal is to minimize the average energy consumption of the system.

We consider both the situations that users can or cannot design caching policies by coordinating with each other. For the non-coordinated caching policy design, optimization can be formulated as a convex problem, and we propose an efficient solving algorithm that obtains the optimal solution by using

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

Lagrange multiplier and duality [10], [19]. For the coordinated design, the problem can be cast into a linear program by optimizing the caching policy of a user at a time. However, due to the large scale of the linear program, we propose an iterative algorithm that can provide an effective solution with low complexity. A performance bound for energy consumption of the system is proposed to serve as a benchmark. We use simulations with real-world-data based individual preference modeling [15] to evaluate the performance. The results show that the proposed designs are effective, and that significant performance gain can be brought by exploiting information of individual preferences.

The remaining paper is organized as follows. Section II elaborates the models adopted in this work. The caching policy design problem is formulated in Section III. The proposed caching policy designs are described in Section IV. We provide performance evaluations in Section V. Section VI concludes this work.

II. System and Individual Preference Models

We consider BS-assissted cache-enabled wireless D2D networks with a single BS and K users within a D2D communication radius. Users can use retrieval from their own caches, D2D communications, or BS links to access desired content. We assume the distances between users are small so that the outage for D2D communications caused by poor channel conditions is ignored. The file library consists of M files, and all files have the same size. Each user is able to cache S files in the device storage. Note that we consider $S < M$ because the number of files in a library is generally larger than the number of files that can be stored in a device, and the caching policy design problem is trivial as $S \geq M$. A user prefers to use a file copy stored in its own cache since it can then obtain the content with no cost. Otherwise, the user checks whether other users in the D2D network store the desired content. If so, D2D communications is used to access the desired content with low cost; otherwise, the user needs to use the BS link to access the content with higher cost. It is assumed that the BS has an unlimited backhaul to repositories that store all files in the library, which guarantees that any request from a user always can be satisfied, though potentially at high cost. To concentrate on the influence of individual user preferences, we consider orthogonal accesses of users[1]. The possible interferences between different users will be subject of our future work.

In this work random caching policies are adopted, and different users can have different caching policies. Denoting b_m^k as the probability for user k to cache file m, we have the caching policy of user k given by $\{b_m^k\}_1^M$, where $\sum_{m=1}^M b_m^k \leq S$. Individual user preferences are considered and represented via probabilities. We denote the preference probability of user k for file m, i.e., probability that user k wants file m in the future as a_m^k, in which $0 \leq a_m^k \leq 1, \forall m, k$, and $\sum_{m=1}^M a_m^k = 1, \forall k$. Thus, different preference probabilities of different users for the same file indicate that different users have different preferences.

In this work, a user could access desired files from its own cache, caches of other users, and BS. We assume different amounts of energy are consumed when using different types of accessing approaches. We denote the energy consumption for accessing a file using BS transmission as E_B (Joule/bit); the energy consumption for accessing a file using D2D communications as E_D (Joule/bit); and consider $E_B > E_D$. Zero energy consumption is assumed if the user can access the desired file from its own cache. Intuitively, a user with a selfish caching strategy, i.e., caching purely to maximize the probability of finding a desired video file in its own cache, minimizes the energy consumption for its own use, but possibly decreases the probability that it can provide a file to other users via D2D communication. If all users act in such a selfish way, then the overall energy consumption might be high. Conversely, if each user caches purely for D2D communication, it might miss out on energy savings that are associated with finding the file in its own cache. The optimum energy saving strategy will thus strike a balance between those two extremes.

III. Proposed Caching Policy Design Problem

Our goal in this work is to design caching policies that minimize the system energy consumption by exploiting the knowledge of individual preferences. In this section, the access probabilities of different accessing approaches for a user are firstly derived. These are then used to derive the average energy consumption at each access, and thus formulate the caching policy design problem aiming to minimize average energy consumption.

Considering the system model in Sec. II and given caching policies $\{b_m^k\}_1^M, \forall k$ of users, the probability that user k accesses the desired file through the BS is

$$P_B^k = \sum_{m=1}^M a_m^k \Big[\prod_{l=1}^K (1 - b_m^l)\Big], \tag{1}$$

where $\prod_{l=1}^K (1 - b_m^l)$ is the probability that file m is not in the caches of user devices, and therefore $a_m^k \prod_{l=1}^K (1 - b_m^l)$ is the probability that the user wants file m but file m is not in the caches of user devices. We then define the hit probability as the complement, $P_h^k = 1 - P_B^k$. We define the self-access probability of user k as the probability that the desired file of user k is in its own cache, expressed as

$$p_S^k = \sum_{m=1}^M a_m^k \cdot b_m^k. \tag{2}$$

By using p_h^k and p_S^k, the probability that user k can reach the desired file via D2D commmunications is given by

$$P_D^k = P_h^k - p_S^k = 1 - \sum_{m=1}^M a_m^k \Big[\prod_{l=1}^K (1 - b_m^l)\Big] - \sum_{m=1}^M a_m^k \cdot b_m^k. \tag{3}$$

[1]This can be achieved, e.g., by having "clusters" of users communicate with each other, while different clusters use different frequency bands with "spatial reuse", compare [8]. The "communication radius" we henceforth reference thus corresponds to the dimension of such a cluster, not the cell radius of the macro BS.

Now, we derive the average energy consumption of the system. By using (1), (3), and the energy consumption model in Sec. II, the expected energy consumption of user k to access a desired file can be computed as

$$EC_k = E_D \cdot P_D^k + E_B \cdot P_B^k \quad \text{(Joule/bit)}, \tag{4}$$

and the average over the users is

$$EC = \sum_{k=1}^{K} \frac{E_D P_D^k + E_B P_B^k}{K} \quad \text{(Joule/bit)}. \tag{5}$$

Thus the caching policy design problem aiming to minimize the average energy consumption is given as

$$\begin{aligned} \min_{b_m^k, \forall k, m} \quad & EC = \textstyle\sum_{k=1}^{K} \frac{E_D P_D^k + E_B P_B^k}{K} \\ \text{subject to} \quad & \textstyle\sum_{m=1}^{M} b_m^k \leq S, \forall k, \\ & 0 \leq b_m^k \leq 1, \forall k, m. \end{aligned} \tag{6}$$

There are two important features of (6) that are frequently used in the following sections. First, by using (1) and (3), EC in (6) can be reformulated as

$$EC = \sum_{k=1}^{K} \frac{E_D P_D^k + E_B P_B^k}{K} = E_D + \frac{E_B - E_D}{K}\Big[\sum_{m=1}^{M} S_m \prod_{k=1}^{K}(1 - b_m^k)\Big] - \frac{E_D}{K}\sum_{m=1}^{M}\sum_{k=1}^{K} a_m^k b_m^k, \tag{7}$$

where $S_m = \sum_{k=1}^{K} a_m^k$. The derivation of (7) is given in (8) at the top of next page. Second, we have the following Proposition.

Proposition 1: The optimal solution of (6) must be tight at the equality of the sum constraint, i.e., for the optimal solution $(b_m^k)^*, \forall k, m$, we have

$$\sum_{m=1}^{M} (b_m^k)^* = S, \forall k \tag{9}$$

Proof: See Appendix A.

IV. Proposed Caching Policy Designs

In this section, we propose two caching policy designs used in different system scenarios. Specifically, we first propose the caching policy design that is used in the situation that users cannot coordinatedly design caching policies, i.e., a user has no knowledge about other users' caching policies. Then the coordinated caching policy design is proposed.

A. Proposed Non-Coordinated Caching Policy Design

Here we consider the situation that users cannot coordinatedly design their caching policies. Before providing the detail, we stress that, although we will be able to obtain the optimal solution of the formulated problem, this does not mean the proposed design is optimal under different assumptions or constraints. Our design is only optimal when all users adopt the same caching policy and their sum preference is known.

By assuming that users design their caching policies using the same approach and know the sum preference probabilities of the demanding user set when designing caching policies, we propose the non-coordinated caching policy design problem as

$$\begin{aligned} \min_{b_m, \forall m=1,..,M} \quad & EC_{nc} \\ \text{subject to} \quad & \textstyle\sum_{m=1}^{M} b_m = S, \\ & 0 \leq b_m \leq 1, \forall m, \end{aligned} \tag{10}$$

where $\{b_m\}_1^M$ is the caching policy adopted by any user and

$$EC_{nc} = E_D + \frac{(E_B - E_D)}{K}\sum_{m=1}^{M} S_m (1 - b_m)^K - \frac{E_D}{K}\sum_{m=1}^{M} S_m b_m. \tag{11}$$

We note that (10) is simply the special degeneration of (6) as we restrict users to adopting an identical policy and use Proposition 1. Besides, every user designs the caching policy by independently solving (10). Furthermore, since the global (system) popularity can be regarded as the average individual preference probability over a sufficiently large number of users, $\frac{S_m}{K}$ could gradually converge to the system popularity with inceasing K. In this context, the performance of the proposed non-coordinated caching policy design using individual preferences will converge to its counterpart that uses the system popularity distribution, i.e., the homogeneous popularity model, when K is sufficiently large. However, since K is generally not a large number for a D2D network, we should distinguish between these two cases, and performance improvements can be observed for systems exploiting user group preferences as shown in Sec. V. Note, however, that such advantages are only possible if the characteristics of the group of users that will communicate are known during the caching process. Finally, we offer the following proposition:

Proposition 2: The optimization problem in (10) is convex.

Proof: See Appendix B.

To solve this problem[2], we exploits the Lagrange multiplier and duality [10], [19]. Considering the Lagrange multiplier μ that relates the sum constraint to the Lagrangian, the Lagrangian is given as

$$L_{nc}(\mu, b_1, ..., b_M) = \mu\Big(\Big[\sum_{m=1}^{M} b_m\Big] - S\Big) + E_D + \frac{(E_B - E_D)}{K}\sum_{m=1}^{M} S_m (1 - b_m)^K - \frac{E_D}{K}\sum_{m=1}^{M} S_m b_m. \tag{12}$$

The feasible domain of the function is described by $0 \leq b_m \leq 1, \forall m$. Then since the duality gap of a convex program is zero, the optimal solution of

$$\max_{\mu} \min_{b_1,..,b_M} L_{nc}(\mu, b_1, ..., b_M), \quad s.t. \quad 0 \leq b_m \leq 1, \forall m, \tag{13}$$

is equivalent to the optimal solution of (10).

[2] Although this problem can be solved by any convex solver [19], we herein propose a solving algorithm that is specifically designed for the problem.

$$
\begin{aligned}
\sum_{k=1}^{K}\frac{E_DP_D^k+E_BP_B^k}{K} &= \frac{1}{K}\Big\{\sum_{k=1}^{K}E_D(1-\sum_{m=1}^{M}a_m^k\Big[\prod_{l=1}^{K}(1-b_m^l)\Big]-\sum_{m=1}^{M}a_m^kb_m^k)+E_B\sum_{m=1}^{M}a_m^k\Big[\prod_{l=1}^{K}(1-b_m^l)\Big]\Big\}\\
&= E_D+\frac{(E_B-E_D)}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}a_m^k\Big[\prod_{l=1}^{K}(1-b_m^l)\Big]-\frac{E_D}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}a_m^kb_m^k\\
&= E_D+\frac{(E_B-E_D)}{K}\sum_{m=1}^{M}\Big[\prod_{l=1}^{K}(1-b_m^l)\Big]\underbrace{\sum_{k=1}^{K}a_m^k}_{S_m}-\frac{E_D}{K}\sum_{m=1}^{M}\sum_{k=1}^{K}a_m^kb_m^k\\
&= E_D+\frac{E_B-E_D}{K}\Big[\sum_{m=1}^{M}S_m\prod_{k=1}^{K}(1-b_m^k)\Big]-\frac{E_D}{K}\sum_{m=1}^{M}\sum_{k=1}^{K}a_m^kb_m^k.
\end{aligned}
\tag{8}
$$

To solve (13), the standard approach is using the bisection search for the optimal μ^* and $b_m^*, \forall m$, such that $b_m^* = b_m(\mu^*)$ has the expression

$$
b_m(\mu^*) = \begin{cases} f(\mu^*, S_m), & K\mu^* - S_mE_D \geq 0 \\ 1, & K\mu^* - S_mE_D < 0 \end{cases} \tag{14}
$$

and $\sum_{m=1}^{M} b_m(\mu^*) = S$, where

$$
f(\mu^*, S_m) = \max\Big[1-\Big(\frac{K\mu^*-S_mE_D}{KS_m(E_B-E_D)}\Big)^{\frac{1}{K-1}}, 0\Big]. \tag{15}
$$

The update rule of the ith step of the bisection algorithm is then given by

$$
\mu_{min}^{(i+1)} = \frac{\mu_{min}^{(i)}+\mu_{max}^{(i)}}{2}, \mu_{max}^{(i+1)} = \mu_{max}^{(i)} \tag{16}
$$

if $\sum_{m=1}^{M} b_m(\frac{\mu_{min}^{(i)}+\mu_{max}^{(i)}}{2}) > S$; and

$$
\mu_{min}^{(i+1)} = \mu_{min}^{(i)}, \mu_{max}^{(i+1)} = \frac{\mu_{min}^{(i)}+\mu_{max}^{(i)}}{2} \tag{17}
$$

if $\sum_{m=1}^{M} b_m(\frac{\mu_{min}^{(i)}+\mu_{max}^{(i)}}{2}) < S$. The initial conditions are suggested as $\mu_{min}^{(0)} = \min_m \frac{E_DS_m}{K}$ and $\mu_{max}^{(0)} = \max_m \frac{E_DS_m}{K} + (E_B-E_D)S_m$.

B. Proposed Coordinated Caching Policy Design

Now, we consider the case that users can jointly design their caching policies. Therefore the orginal caching policy design problem in (6) is adopted. Note that we are here jointly designing the caching *policies and probabilities*, not a joint deterministic caching of particular files which would require a more detailed, and more frequent, interaction between the users.

To solve the problem, we propose an approach that iteratively optimizes the caching policy of a user at a time. Specifically, we optimize the caching policy of user k while assuming caching policies of other users fixed, resulting in the sub-problem

$$
\begin{aligned}
&\min_{b_m^k, \forall m} && EC_{LP} && (18a)\\
&\text{subject to} && \textstyle\sum_{m=1}^{M} b_m = S, && (18b)\\
& && 0 \leq b_m \leq 1, \forall m, && (18c)
\end{aligned}
$$

where

$$
\begin{aligned}
EC_{LP} = E_D + &\frac{E_B-E_D}{K}\Big[\sum_{m=1}^{M}S_mC_m^k(1-b_m^k)\Big]\\
&-\frac{E_D}{K}\sum_{m=1}^{M}(D_m^k+a_m^kb_m^k),
\end{aligned}
\tag{19}
$$

$C_m^k = \prod_{l=1, l\neq k}^{K}(1-b_m^l)$, and $D_m^k = \sum_{l=1,l\neq k}^{K} a_m^l b_m^l$. We remark that (19) is simply a reformulation of (7). Besides, (18) is a linear program because C_m^k and D_m^k are constants as we concentracte only on optimizing the caching policy of user k. Furthermore, we consider only equality constraint in (18b) because of Proposition 1.

To solve (18), general linear program solvers are feasible. However, to provide a more insightful and efficient approach, a solving approach with analytical closed-form expressions are proposed. Since minimizing EC_{LP} is equivalent to maximizing

$$
\sum_{m=1}^{M} b_m^k\Big((E_B-E_D)S_mC_m^k+E_Da_m^k\Big), \tag{20}
$$

(18a) can be equivalently expressed as

$$
\max_{b_m^k, \forall m} \quad \textstyle\sum_{m=1}^{M} b_m^k\Big((E_B-E_D)S_mC_m^k+E_Da_m^k\Big). \tag{21}
$$

Then by observing that the optimal solution of (21) subject to constraints (18b) and (18c) is achieved by allocating cache space to the terms that can offer larger payoffs, the optimal solution of (18) is epressed as

$$
(b_m^k)^* = \begin{cases} 1, & m \in \Phi_k, \\ 0, & \text{otherwise}, \end{cases} \tag{22}
$$

where $\Phi_k = \{m : f_m^k$ is among the S largest of all $f_m^k\}$ and $f_m^k = (E_B-E_D)S_mC_m^k + E_Da_m^k$. By iteratively solving (18) via using (22) for different k until convergence, the caching policy design problem in (6) can be effectively solved. We note that the iteration will definitely converge, because each iteration can provide an improved solution for (6) by solving (18) and there is always a lower bound for the energy consumption. We provide the lower bound in the next subsection.

C. Proposed Energy Comsumption Lower Bound

Here we provide a loose lower bound for EC. This lower bound can help in characterizing the performance of the proposed caching policy designs as well as the caching network. Here we consider $a_1^k \geq a_2^k \geq ... \geq a_M^k, \forall k$, without loss of generality. Then without regard to the other users' energy consumptions, the best caching strategy to minimize the energy consumption of user k is: user k caches files $1, 2, ..., S$; other users in the network cache files $S+1, ..., S_{max}$, where $S_{max} = \min(M, KS)$. By this golden strategy, the expected energy consumption for user k is lower bounded by

$$EC_k \geq E_D \sum_{m=S+1}^{S_{max}} a_m^k + E_B \sum_{m=S_{max}+1}^{M} a_m^k, \forall k. \quad (23)$$

It follows that the lower bound of the average energy consumption EC is

$$EC \geq \frac{E_D}{K} \sum_{k=1}^{K} \sum_{m=S+1}^{S_{max}} a_m^k + \frac{E_B}{K} \sum_{k=1}^{K} \sum_{m=S_{max}+1}^{M} a_m^k. \quad (24)$$

We note that this bound might not be achievable due to the following two reasons: (I) it requires every user to adopt their golden strategies which could be infeasible; (II) it constitutes a deterministic coordinated caching strategy where the users coordinate not only caching policy (i.e., user preferences), but the particular files that they download. Note that a similar bound was used in our earlier work [8].

V. Performance Evaluations

In this section, we use computer simulations and real-world-data based individual preference modeling to evaluate the proposed caching policy designs. We consider users randomly distributed in a circle with radius equal to 5 meter (distances between users are within 10 meters), which is the range over which D2D communication is possible (it is *not* the cell radius of the macro BS). The transmission power of the BS and users are 42 dBm and 20 dBm, respectively. System bandwidth, carrier frequency, and noise power density are 20 MHz, 2 GHz, and -174 dBm/Hz, respectively. We adopt the path loss plus shadowing model given as

$$20 \log_{10} \frac{4\pi d_b}{\lambda_c} + 10\alpha \log_{10}(\frac{d}{d_b}) + X_{\sigma^2}, \quad (25)$$

where λ_c is the wavelength of the carrier frequency, $\alpha = 3.5$ is the pathloss exponent, $d_b = 1$ m is the break point distance, and X_{σ^2} is a random variable with log-normal distribution and variance 8 dB. The small-fading is modeled by Rayleigh fading.

We consider users being served in a round-robin manner, and each user accesses the desired file with a rate of 100 Mbits/sec. If the channel conditions of D2D communications cannot support the transmission rate, users are forced to use BS link to access desired files even if the desired files can be found in the D2D network. Note that this is a more general setting than assumed in the theoretical derivations, where it was assumed that the D2D link can always sustain the required rate. To generate practical individual preference probabilities for users, the generation approach and parameters in Sec. V of [15] are adopted. The exception is that we specify different values for the numbers of files within each genre $M_g, \forall g$ in different figures. Note that individual preference probabilties in [15] are modeled by a hierachical structure in which the preference probability of a user for a file is modeled as the probability that a user wants a certain video genre, and then the conditional probability a user wants a file within the genre. Therefore, each file in the model can be categorized into a genre, and we have $M = \sum_{g=1}^{G} M_g$, where G is the total number of genres in the library. Instead of evaluating the performance of proposed designs by using energy consumption, the more popular energy efficiency (bits/Joule) is used in the following figures. By definition, the conversion between the average energy consumption (EC) and energy efficiency (EE) is simply given by $EE \propto \frac{1}{EC}$.

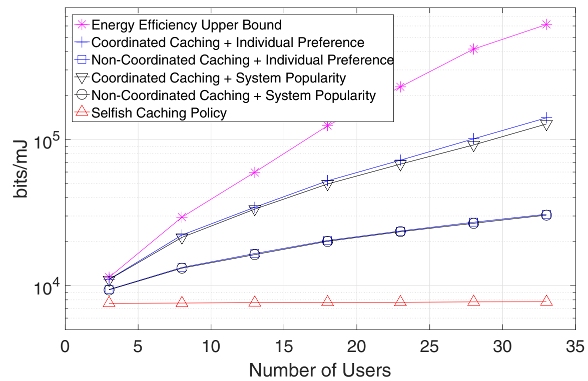

Fig. 1: Evaluation of proposed designs for systems with $S = 5$, $G = 94$, $M_g = 20, \forall g$, and $M = 1880$.

In the following figures, we compare the performance between the proposed non-coordinated caching policy design, coordinated caching policy design, selfish caching policy, and the EE upper bound (the inverse of the EC lower bound of Sec. V.C). Both proposed non-coordinated and coordinated caching policy designs can be implemented either by using knowledge of individual preference probabilities as in Sec. IV or simply by using the system-wide popularity distribution. When implementing using system popularity distribution, the individual preference probabilities of users in (6) are replaced by the system popularities. We note that the system popularities are computed by averaging the individual preference probabilities of 10000 users generated by the same generation model, i.e., the system popularities are statistics constructed from 10000 individuals. For the selfish policy, users cache files according to their own preferences without considering other users. Hence the performance of the upper bound and the selfish caching serves as benchmarks for the evaluation.

In Fig. 1, we consider $S = 5$, $G = 94$, and $M_g = 20, \forall g$. Therefore we have $M = 1880$. From the figure, it can be observed that the coordinated caching policy designs can outperform their corresponding non-coordinated counterparts. Besides, the benefits of exploiting individual preferences are

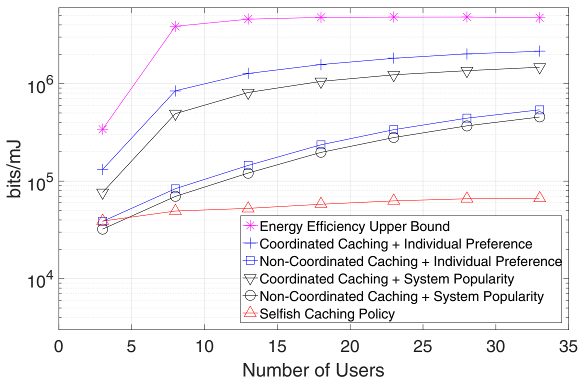

Fig. 2: Evaluation of proposed designs for systems with $S = 40$, $G = 94$, $M_g = 20, \forall g$, and $M = 1880$.

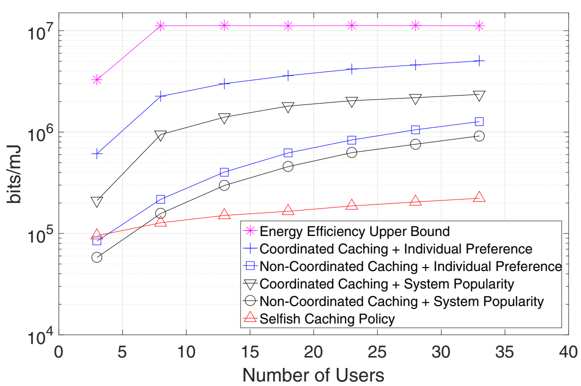

Fig. 3: Evaluation of proposed designs for systems with $S = 35$, $G = 94$, $M_g = 10, \forall g$, and $M = 940$.

small because S is much smaller than M so that users can only cache a tiny portion of files in the library. Specifically, since the cache size of each user is small, to have a low cost transmission, having an effective D2D network is important. In order to enable the effective D2D, caching files that can benefit more users is more important than caching files that only benefit a single user. Thus the strategy that users cache files that are popular globally can provide a fairly good performance even if users have different preferences. However, a different result will be observed below as S is large, i.e., users can cache a large portion of files in the library. In Fig. 2, $S = 40$, $G = 94$, and $M_g = 20, \forall g$. We can observe again that the coordinated caching policy designs outperform the non-coordinated ones. Besides, in contrast to Fig. 1, the performance gain due to considering individual preferences can be readily observed. In Fig. 3, systems with $S = 35$, $G = 94$, and $M_g = 10, \forall g$, are considered. From the figure, we can observe a similar result as in Fig. 2, i.e., the performance gain due to considering individual preferences is significant.

From the simulation results, we conclude that improvements can indeed be achieved by exploiting individual preferences. Besides, by comparing different figures, the conditions offering large performance gain are empirically demonstrated. We note that, although not being shown here, the performance gain of considering individual preferences increase as S increases when fixing the library size M. We also note that, since the models proposed in [15] is based on frequent users, more empirical results are needed for identifying conditions in which the individual preference aware policies can offer large performance gain. As a final remark, from all figures, it can be observed that the performance of the selfish caching is much worse compared to all other policies and the gaps between the coordinated and non-coordinated policies are significant. This indicates the importance for users to coordinate when deciding their caching strategies. We note that, while introducing more cooperations and exploiting information of individual preferences can readily improve the system performance, it also increases the overhead. The investigations of caching policy designs that provide effective trade-off are considered as future work. Future work will also investigate the robustness of the caching strategies to node mobility or uncertainty, i.e., that the users with which D2D communication are partly different from those with which the coordination was done or the users cannot know the exact user set with which the D2D network is constructed.

VI. Conclusions

In contrast to most literature that uses homogeneous user preference modeling, this work considers heterogeneous individual preference modeling and proposes energy efficient caching policy designs in BS assissted wireless D2D networks. The caching policy design problems are formulated and considered from non-coordinated and coordinated perspective, respectively, and effective designs are proposed for both cases. Performance of proposed designs is evaluated through simulations with preference distributions based on real-world data. Results show that the proposed designs are effective, and as compared with designs adopting homogeneous user preference modeling, the performace is readily improved by leveraging information of individual preferences.

Appendix A
Proof of Proposition 1

By (7), the first order partial derivatives of EC is:

$$\frac{\partial EC}{\partial b_m^k} = -\frac{E_B - E_D}{K} S_m \prod_{l=1, l\neq k}^{K} (1 - b_m^l) - \frac{E_D a_m^k}{K}, \forall k, m \tag{26}$$

Then since $E_B > E_D$ and $0 \leq b_m^k \leq 1, \forall k, m$, we have $\frac{\partial EC}{\partial b_m^k} \leq 0, \forall k, m$. Therefore EC is non-increasing respective to $b_m^k, \forall k, m$. This indicates the optimal solution of (6) must be tight at the equality of the sum constraint.

Appendix B
Proof of Proposition 2

Here we prove that the problem in (10) is convex. Note that the problem degenerates to a linear program as $K = 1$. Therefore we generally consider situations with $K \geq 2$.

The first and second order partial derivatives of EC_{nc} in (10) are given by

$$\frac{\partial EC_{nc}}{\partial b_m} = -(E_B - E_D)S_m(1-b_m)^{K-1} - \frac{E_D S_m}{K} \quad (27)$$

and

$$\frac{\partial EC_{nc}}{\partial b_m b_n} = \begin{cases} (K-1)(E_B - E_D)S_m(1-b_m)^{K-2}, & m = n, \\ 0, & m \neq n, \end{cases} \quad (28)$$

respectively. Since $(K-1)(E_B - E_D)S_m(1-b_m)^{K-2} \geq 0$ as $0 \leq b_m \leq 1$, the Hessian matrix of EC_{nc} is a diagonal matrix with entries greater and equal to zero for all feasible solutions. Therefore the Hessian matrix of EC_{nc} is positive semidefinite considering the feasible set of (10). This completes the proof.

ACKNOWLEDGEMENT

Part of this work was supported by the National Science Foundation (NSF).

REFERENCES

[1] "Cisco Virtual Networking Index: Global Mobile Data Traffic Forcast Updatae, 2016-2021," San Jose, CA, USA.

[2] J. G. Andrews, S. Buzzi, W. Choi, and *et. al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065-1082, Jun. 2014.

[3] A. F. Molisch, G. Caire, D. Ott, J. R. Foerster, D. Bethanabhotla, and M. Ji, Caching eliminates the wireless bottleneck in video aware wireless networks, Adv. Elect. Eng., vol. 2014, Nov. 2014, Art. ID 261390.

[4] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commmun. Mag.*, vol. 54, no. 9, pp. 22-28, Sep. 2016.

[5] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commmun. Mag.*, vol. 51, no. 4, pp. 142-149, Apr. 2013.

[6] K. Shanmugam, N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402-8413, Dec. 2013.

[7] M. A. Maddah-Ali and U. Niessen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856-2867, May 2014.

[8] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-Station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665-3676, Jul. 2014.

[9] A. Liu and V. K. N. Lau, "Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming," *IEEE Trans. Sig. Process.*, vol. 63, no. 1, pp. 57-69, Jan. 2015.

[10] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," *IEEE ICC*, Jun. 2015.

[11] M. Ji, G. Caire, and A. F. Molisch "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Area Commun.*, vol. 34, no. 1, pp. 176-189, Jan. 2016.

[12] 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. 849-864, Feb. 2016.

[13] Z. Chen, N. Pappas, and M. Kountouris, "Probabilistic caching in wireless D2D networks: cache hit optimal vs. throughput optimal," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 584-587, Mar. 2017.

[14] D. Karamshuk, N. Sastry, M. Al-Bassam, A. Secker, and J. Chandaria, "Take-Away TV: Recharging wok commutes with predictive preloading of catch-up TV content," *IEEE J. Sel Commun.*, vol. 34, no. 8, pp. 2091-2101, Aug. 2016.

[15] M. C. Lee, A. F. Molisch, N. Sastry, and A. Raman, "Individual preference probability modeling for video content in wireless caching networks," *IEEE GLOBECOM*, Dec. 2017.

[16] Y. Guo, L. Duan, and R. Zhang, "Cooperative local caching and file sharing under heterogeneous file preferences," *IEEE ICC*, May 2016.

[17] Y. Guo, L. Duan, and R. Zhang, "Cooperative local caching under heterogeneous file preferences," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 444-457, Jan. 2017.

[18] Y, Pan, C, Pan, H, Zhu, and *et. al.*, "On consideration of content preference and sharing willingness in D2D assisted offloading," *arXiv preprint*, arXiv:1702.00209, Feb. 2017.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.