Cache-enabled Device-to-Device Communications: Offloading Gain and Energy Cost

Binqiang Chen, Chenyang Yang and Andreas F. Molisch

Abstract

By caching files at users, content delivery traffic can be offloaded via device-to-device (D2D) links if a helper user is willing to transmit the cached file to the user who requests the file. In practice, the user device has limited battery capacity, and may terminate the D2D connection when its battery has little energy left. Thus, taking the battery consumption allowed by the helper users to support D2D into account introduces a reduction in the possible amount of offloading. In this paper, we investigate the relationship between offloading gain of the system and energy cost of each helper user. To this end, we introduce a user-centric protocol to control the energy cost for a helper user to transmit the file. Then, we optimize the proactive caching policy to maximize the offloading opportunity, and optimize the transmit power at each helper to maximize the offloading probability. Finally, we evaluate the overall amount of traffic offloaded to D2D links and evaluate the average energy consumption at each helper, with the optimized caching policy and transmit power. Simulations show that a significant amount of traffic can be offloaded even when the energy cost is kept low.

Index Terms

Caching, D2D, Traffic offloading, Energy cost.

I. INTRODUCTION

Device-to-device (D2D) communications boosts the throughput of cellular networks by offloading traffic [1-4], and thus is a promising way to achieve the goal of 5th generation (5G) mobile networks. Traditional D2D communication, which does not cache content locally, can

Binqiang Chen and Chenyang Yang are with the School of Electronics and Information Engineering, Beihang University, Beijing, China, Emails: chenbq@buaa.edu.cn, cyyang@buaa.edu.cn. Andreas F. Molisch is with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA. Email: molisch@usc.edu.
only offload peer-to-peer (P2P) traffic from cellular networks if source and destination are in proximity at the time they wish to communicate, such as gaming and relaying [1-5]. However, the lion’s share of cellular traffic is video dissemination, a kind of client/server (C/S) services, which will generate more than 2/3 of mobile data traffic by 2019 [6].

Motivated by the fact that a large amount of content requests are asynchronous but redundant, i.e., the same content is requested at different times, caching at the wireless edge has become a trend for content delivery, which improves the throughput and energy efficiency of the network and the quality of experience (QoE) of the users [7-13].

Recent work [14, 15] has shown that caching at the user devices enables offloading also of C/S traffic, in particular video, to D2D connections. Without caching at the devices, the users need to fetch their requested video via base station (BS) from a remote server. By pre-downloading popular files to users during the off-peak time, say at night, the file requested by a user can be transmitted via D2D links by other users in proximity that have cached the file. Such a proactive caching policy largely alleviates the burden to the BSs during the peak time, yielding high offloading gain [14-18]. To improve the performance of cache-enabled D2D communications, proactive caching policies were optimized in [17, 18], and a distributed reactive caching mechanism was designed in [16].

When D2D communications are used for supporting P2P services, the users acting as transmitters are by definition willing to send messages to the destination users. However, offloading content delivery traffic by cache-enabled D2D communications needs the help of other users who are not obligated to help. Due to the limited battery capacity, a natural question from a helper user in such a network is: “why should I spend energy of my battery to provide you with faster video download? [14]” This makes the energy consumption of a helper user a big concern in cache-enabled D2D communications. In practice, a helper user may only be willing to use a fraction of its battery for transmitting files to other users, if properly rewarded by the operator. It is thus important to quantify the offloading gain when the helper users’ allowed battery consumption is taken into account, and to evaluate the average energy consumed by a helper user to deliver the files to others.

In previous research efforts for cache-enabled D2D communications [14-18], the energy of the battery is implicitly assumed infinite and the energy costs at helper users are never considered. Consequently, (i): maximal transmit power is used by all D2D transmitters to deliver the files,
and (ii): once a D2D link is established, the file is assumed to be able to be delivered completely without considering whether there is still energy in the battery or whether a helper is willing to contribute more energy.

In this paper, we quantify the offloading gain of a cache-enabled D2D communication system by taking maximal permissible battery consumption into account, and evaluate the energy cost for a user to transmit the file. With the allowed battery consumption, a helper user may only transmit part of a file to the user requesting the file. To control the energy spent by the helper user for transmitting a file, we consider a user-centric caching and transmission strategy, where only the users within a collaboration distance $r_c$ of the requesting user can serve as helpers. When the collaboration distance is large, the probability that the users can fetch their desired contents via D2D links is high, and thus more traffic can be offloaded. However, since the possible D2D link distance increases, the energy cost of a helper user also grows and then more files cannot be conveyed completely via D2D links.

Aimed to find the maximal offloading gain, we first introduce a user-centric probabilistic caching policy, where the users proactively cache files according to a $r_c$-dependent caching distribution. We optimize the policy to maximize the amount of traffic that can be possibly offloaded with a given collaboration distance and the user demands statistics. In [17], a cluster-centric caching policy was proposed, which was optimized to maximize the same objective with given cluster size and demands statistics, but is not optimal under the user-centric framework. Then, we optimize the transmit power at each helper to maximize the probability that a requested file can be found in adjacent users and transmitted completely via a D2D link, considering two extreme cases in terms of interference level. Finally, we quantify the total offloaded amount of traffic by taking complete and partial transmission into account, evaluate the average energy consumption for each D2D transmitter with optimized caching policy and transmit power, and characterize the relationship between offloading gain and energy cost.

The contributions of this paper are summarized as follows:

- We analyze the offloading gain when the user only allows partial energy in its battery to be consumed. To the best of the authors’ knowledge, this is the first paper to characterize the offloading gain given limited battery consumption in cache-enabled D2D communications.
- We investigate the relationship between the offloading gain of the system and the energy costs of the helper user, and show the impact of the allowed battery consumption.
The rest of the paper is organized as follows. Section II presents the system model. Section III optimizes the caching policy. Section IV optimizes the transmit power, and evaluates the offloading gain and energy cost. Section V shows simulations. Section VI concludes the paper.

II. SYSTEM MODEL

Consider a cell where users’ locations follow a Poisson Point Process (PPP) with density $\lambda$. Each single-antenna user has local cache to store files, and can act as a helper to transmit but with only a fraction of its battery capacity. For simplicity of notation, assume that each user only stores one file in its local cache as in [15, 18], though generalization to storage of multiple files is straightforward.

When a helper transmits a file in the local cache via D2D link to a user requesting the file, i.e., a D2D receiver (DR), the helper becomes a D2D transmitter (DT). To control the energy spent by a DT for transmitting to a DR, we introduce a user-centric protocol. A DT will send a cached file to the DR only if their distance is smaller than a given value $r_c$, called collaboration distance. The users with distance $r$ less than $r_c$ are called adjacent users. Assume that a fixed bandwidth is assigned to the D2D links to avoid the interference between D2D and cellular links [5], and all DTs transmit with same transmit power. The BS is aware of the files cached at the users and coordinates the D2D communications.

A. Content Popularity and Caching Placement

We consider a static content catalog consisting of $N_f$ files that all users in the cell may request, which are indexed in descending order of popularity, i.e., the 1st file is the most popular file. Each file has size of $F$ bits, but the analysis can be easily extended to general cases by dividing each file into chunks of equal size. The probability that the $i$th file is requested follows a Zipf distribution

$$p_r(i) = \frac{i^{-\beta}}{\sum_{k=1}^{N_f} k^{-\beta}},$$  

where $\sum_{i=1}^{N_f} p_r(i) = 1$, and the parameter $\beta$ reflects how skewed the popularity distribution is, with large $\beta$ meaning that a few files are responsible for the majority of requests [19].

Since deterministic caching policy designed for wired networks with fixed topology is not applicable for a wireless scenario with user locations that are unknown a priori, we consider a probabilistic caching policy. Specifically, each user caches a file according to a $r_c$-dependent
caching distribution, which is the probability that the $i$th file is cached at users, $i = 1, \cdots, N_f$. All users in the cell that have cached with the $i$th file constitute a user set, called the $i$th helper set, as shown in Fig. 1.

**B. User Allowed Battery Consumption and Content Delivery**

The content popularity usually changes at a much slower speed than the traffic variation of cellular networks (e.g., one week for movies [14]), which is often regarded as invariant over a period. Consequently, the files can be proactively downloaded by the BS during the off-peak time, without the need to be updated frequently. The energy consumed at users during content placement is negligible since users will usually be connected to the AC power during the download time (say at night).

Assume that each user requests one file from the catalog independently. If a user can find its requested file in the local caches of its adjacent users, a D2D link is established between the user and its nearest adjacent user cached with the file to convey the file. Assume that each user device has the same battery capacity of $Q$ (mAh), and only a fraction $\rho$ of each DT’s battery capacity can be consumed for transmitting a file to the DR. Denote the operating voltage of the user device as $V_0$. When a DT has consumed $\rho Q V_0$ energy to transmit a file to a DR, the DT
interrupts the D2D link, and the DR has to receive the remaining data of the file from the BS. In fact, another helper in the adjacent of the DT can take over the transmission. We do not consider the hand over among DTs due to the following reason. The distances between the DR and other not-busy helpers are always longer than the distance between the DR and its first-established DT, and hence the corresponding channel conditions are worse in high probability (e.g., when \( r_c = 100 \text{ m} \) and \( \beta = 1 \), this probability is 97\%). As a result, the handover will introduce higher energy cost for other DTs and more signaling overhead for the BS to coordinate. Therefore, two cases may occur for the established D2D links depending on their channel conditions.

- **Complete transmission:** A DR can receive a complete file via D2D link, which is called a *satisfied* DR.
- **Partial transmission:** A DR only receives a fraction of the file from a DT, which needs to access to the BS to fetch the remaining file.

If a user cannot find its requested file in the local caches of its *adjacent users*, the user fetches the file from the BS. If a user can find the desired file in its own local cache, such a self-serve can offload traffic without establishing D2D link. Since we focus on the energy cost of a DT in cache-enabled D2D communications, we ignore self-serve (also called self-offloading in literature) in the forthcoming analysis (similar to \([15,18]\)), but we will evaluate its impact via simulations in section V.

We consider two metrics regarding offloading by the cache-enabled D2D communications.

- **Offloading probability:** This is the probability that a DR enjoys complete transmission via D2D links, which reflects the percentage of the satisfied users.
- **Offloading ratio:** This is the ratio of the amount of data offloaded by both complete and partial transmission via D2D links to the total amount of data in the cell, which reflects the offloading gain of the system.

To focus on the energy cost issue, we assume that the distance between DT and DR remains fixed

\[ \rho \] can reflect the user incentive in terms of battery consumption to serve as a helper. We assume that all users are initially with full battery and hence each user allows to employ the same amount of energy to help others. In practice, the user devices may have different battery capacities. Moreover, a helper may be requested more than once over several hours before recharging its battery, especially when the file is very popular. When a DT serves the second request, the remaining energy in its battery may be less than \( \rho Q V_0 \). For analysis simplicity, we assume that one user only sends one request, and hence each DT is only requested once. The impact that a DT serves multiple requests will be shown via simulation later.
during transmission (again following most of previous works \[14-18\]), although user mobility is one of the key factors that affects the offloading gain of cache-enabled D2D communications.

III. OPTIMAL CACHING POLICY

To optimize the probabilistic caching policy with known user demand statistics, we need to find the optimal caching distribution. Because the contents are proactively placed at users before they initiate requests, we optimize the caching distribution to maximize *offloading opportunity* as in \[14-17\], defined as the probability that the desired file of a user can be found in adjacent users. Such an opportunity reflects how much traffic can be *possibly* offloaded by D2D communications for a given collaboration distance under the assumption of infinite battery capacity.

Denote the probability that the \(i\)th file is cached at a user as \(p_c(i)\). Then, the set \(\{p_c(i)\} = [p_c(1), p_c(2), \ldots, p_c(N_f)]\) constitutes the caching distribution. The locations of the users who belong to the \(i\)th helper set follow a PPP with density \(\lambda_i = \lambda p_c(i)\) according to the thinning property of PPP \[20\]. Thus, the probability that a user requesting the \(i\)th file can find its desired file in the cache of any user within the collaboration distance \(r_c\) is \(p_f(i) = 1 - e^{-\lambda_i \pi r_c^2}\). Then, the offloading opportunity with given caching distribution and \(r_c\) can be derived as

\[
p_o = \sum_{i=1}^{N_f} p_r(i) p_f(i) = \sum_{i=1}^{N_f} p_r(i) (1 - e^{-\lambda p_c(i) \pi r_c^2}).
\]  

(2)

The optimal caching distribution that maximizes the offloading opportunity can be found from the following problem

\[
\max_{p_c(i)} p_o \text{ s.t. } \sum_{i=1}^{N_f} p_c(i) = 1, \quad p_c(i) \geq 0, \quad i = 1, \ldots, N_f.
\]  

(3)

Because the objective function is the sum of \(N_f\) exponential functions and the constraints are linear, this problem is convex \[21\]. It is not hard to show from its Karush-Kuhn-Tucker (KKT) conditions that the optimal caching distribution should satisfy the following conditions

\[
p_c^*(i) = \left[ \frac{1}{\lambda \pi r_c^2} \ln(p_r(i)) - \frac{1}{\lambda \pi r_c^2} \ln(\frac{\mu}{\pi \lambda r_c^2}) \right]^+,
\]  

(4)

where \(1 \leq i \leq N_f, \sum_{i=1}^{N_f} p_c^*(i) = 1\), \(p_r(i)\) is the Zipf distribution in \[1\], and \([x]^+ = \max(x, 0)\).

**Proposition 1:** If \(\frac{(N_f)^{N_f}}{N_f!} < e^{-\frac{\lambda \pi r_c^2}{\pi}}\), then the optimal caching distribution is

\[
p_c^*(i) = \frac{1}{N_f} \left(1 + \frac{\beta}{\lambda \pi r_c^2} \sum_{j=1}^{N_f} \ln(i_j)\right).
\]  

(5)
Otherwise, the optimal caching distribution is
\[
p^*_c(i) = \begin{cases} 
\frac{1}{i^*} \left( 1 + \frac{\beta}{\lambda \pi r_c^2} \sum_{j=1}^{i^*} \ln\left(\frac{j}{i^*}\right) \right), & i \leq i^*, \\
0, & i^* < i \leq N_f,
\end{cases}
\] (6)
where \( i^* \) is upper and lower bounded as
\[
\frac{\lambda \pi r_c^2}{\beta} - 1 \leq i^* \leq \frac{\lambda \pi r_c^2}{\beta} + \ln\left(\sqrt{2\pi N_f}\right) + 1.
\]

Proof: See Appendix A. \qed

The gap between the upper and lower bounds of \( i^* \) in (6) is \( \ln\left(\sqrt{2\pi N_f}\right) + 2 \), which is small. For example, when \( N_f = 1000 \), the gap equals to 4.4. This suggests that \( i^* \) and hence the optimal caching distribution \( p^*_c(i), i = 1, \ldots, N_f \) can be obtained efficiently. \( p^*_c(i) \) depends on the collaboration distance \( r_c \), user density \( \lambda \), as well as content statistics \( N_f \) and \( \beta \).

When \( r_c \to \infty \), \( \frac{(N_f)^{N_f}}{N_f!} < e^{\frac{\lambda \pi r_c^2}{\beta}} \) holds, and according to (5) \( p^*_c(i) = \frac{1}{N_f} \). In this case, the optimal caching distribution is a uniform distribution, i.e., each user can randomly choose a file to cache, because the number of adjacent users for any user trends to infinity.

By using the conditions below (6) and setting \( i^* = N_f \), it is not hard to show that when \( r_c \leq \sqrt{\frac{(N_f+1)\beta}{\pi \lambda}}, \frac{(N_f)^{N_f}}{N_f!} \geq e^{-\frac{\lambda \pi r_c^2}{\beta}} \). In this case, \( p^*_c(i) \) is computed with (6), and the less popular files with indices larger than \( i^* \) are never cached at the users. Because the number of adjacent users are limited when \( r_c \) is small, only the files with high popularity are cached. When \( r_c \to 0 \), \( p^*_c(1) = 1 \) and \( p^*_c(i) = 0, 1 < i \leq N_f \), i.e., only the most popular file is cached at each user.

IV. OFFLOADING GAIN AND AVERAGE ENERGY COSTS

In this section, we investigate the offloading gain of the system and the energy cost at each DT. To this end, we first optimize the transmit power of each DT to maximize the offloading probability, which yields maximal user satisfaction rate and hence high offloading gain. Then, we evaluate the offloading ratio and the average energy consumed at each DT to transmit a file via D2D links with the optimized transmit power and optimized caching policy.

Considering that the interference among D2D links has large impact both on the offloading gain and the energy cost, for mathematical tractability we analyze two extreme cases in terms of interference level: full reuse and time division multi-access (TDMA). With full reuse, all DTs in a cell simultaneously transmit over the time and frequency resources are assigned for D2D communications without any interference coordination. With TDMA, only one DT in the whole cell transmits at a time, and the DTs are scheduled according to round robin (or random)
scheduling with equal time slot duration. While further improvements could be achieved through scheduling, it is known that optimal scheduling in D2D networks is NP-hard. On the other hand, cluster-based scheduling as in [15] is not aligned with the user-centric transmission strategy that forms the basis for our model.

A. Case 1: Full Reuse

Once a D2D link is established, the DT can transmit its cached file to the DR that requests the file. In the full reuse case, each DR treats the interference among the D2D links as noise when decoding the desired signal. The signal to interference plus noise ratio (SINR) at the DR requesting the $i$th file from its corresponding DT is

$$\gamma_{1}(i, r) = \frac{P_{t} h r^{-\alpha}}{\sum_{j \neq i} P_{t} h_{j} r_{j}^{-\alpha} + \sigma^{2}} = \frac{h r^{-\alpha}}{I_{i, r} + \sigma_{0}^{2}},$$

(7)

where $P_{t}$ is the transmit power at each DT, $h$ is the channel power gain that follows an exponential distribution with unit mean for Rayleigh fading, $r$ is the distance between the DT and the DR, $\alpha$ is the path loss exponent, $I_{i, r} = \sum_{j \neq i} h_{j} r_{j}^{-\alpha}$ is the total interference from other DTs normalized by $P_{t}$, $\sigma^{2}$ is the variance of white Gaussian noise, and $\sigma_{0}^{2} = \sigma^{2} / P_{t}$ Then, the data rate is $R_{1}(i, r) = W \log_{2} \left(1 + \frac{h r^{-\alpha}}{I_{i, r} + \sigma_{0}^{2}}\right)$, where $W$ is the bandwidth assigned to D2D links.

To evaluate the energy cost of each DT, we consider both circuit power and transmit power. Then, the energy consumed to transmit the $i$th file via a D2D link with distance $r$ is

$$E_{1}(i, r) = \frac{F}{W \log_{2} \left(1 + \frac{h r^{-\alpha}}{I_{i, r} + \sigma_{0}^{2}}\right)} \left(\frac{1}{\eta} P_{t} + P_{c}\right),$$

(8)

where $\eta$ is the power amplifier efficiency and $P_{c}$ is the circuit power at the DT.

Because only a fraction $\rho$ of the battery capacity is permitted to be used at each DT to help a DR, a DT can transmit the $i$th file completely only if $E_{1}(i, r) \leq \rho V_{0} Q$.

1) Optimal Transmit Power: Because the files not completely delivered via D2D links need to be fetched from the BS, which not only introduces extra signaling overhead but also may degrade the user experience, we optimize the transmit power at a DT to maximize the user satisfaction rate. In other word, we maximize the offloading probability for a given collaboration

Note that this model neglects shadowing and incorporating shadowing would lead to a change of the exponential channel gain distribution to an approximate lognormal distribution [22]. We neglect shadowing, in line with most works in D2D literature.
distance $r_c$, which is the probability that a requested file can be found in adjacent users and transmitted completely via a D2D link.

**Proposition 2:** The offloading probability in the full reuse case is

$$p_1(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_{0}^{r_c} f_i(r) e^{-\phi_i(\Gamma_1, r)} \, dr,$$

where $\Gamma_1 = e^{F(P_t+\eta P_c) \ln 2} - 1$, $f_i(r) = 2\pi r \lambda_i e^{-\lambda_i \pi r^2}$ is the probability density function (pdf) of the D2D link distance, $\phi_i(x, y) = x y^\alpha \sigma_0^2 + \pi(\lambda_i \xi_1 - \lambda_i^d \xi_2) y^2 x^{2\alpha}$, $\lambda_i = \sum_{i=1}^{N_f} \lambda_i^d$ is the density of all DTs and $\lambda_i^d = \lambda_i \left(1 - \left(1 + \frac{\lambda_i r_c}{3.5}\right)^{-3.5} \theta_i\right)$ is the density of DTs cached with the $i$th file, $\theta_i = \frac{\Gamma(3.5, 0) - \Gamma(3.5, 3.5 \lambda_i + \lambda_i r_c, \pi r_c^2)}{\Gamma(3.5, 0) - \Gamma(3.5, 3.5 \lambda_i, \pi r_c^2)}$, $\Gamma(s, x) = \int_{x}^{\infty} t^{s-1} e^{-t} dt$ is the upper incomplete gamma function [23], $\xi_1 = \int_{0}^{+\infty} \frac{1}{1+t^{\alpha/2}} dt$, and $\xi_2 = \int_{0}^{x} \frac{1}{1+t^{\alpha/2}} dt$.

**Proof:** See Appendix B

The expression in (9) depends on the values of $\lambda$, $r_c$, $\rho$ and $P_t$, but not on the user’s location and channel. To maximize the offloading probability for the cache-enabled D2D communications with given values of $\lambda$, $r_c$ and $\rho$, the transmit power at each DT can be optimized as

$$\max_{P_t} p_1(P_t, \rho) \quad \text{s.t.} \quad 0 < P_t \leq P_{\text{max}},$$

where $P_{\text{max}}$ is the maximal transmit power of a DT.

Due to the complicated expression of $p_1(P_t, \rho)$, in general the optimal solution $P_t^*$ can only be found by using similar method as in [15]. When $r_c$ is small, all D2D links experience a line of sight (LOS) environment [17], i.e., $\alpha = 2$. In such a special case, both closed-form expressions of $p_1(P_t, \rho)$ and $P_t^*$ can be obtained.

**Proposition 3:** When $\alpha = 2$, the offloading probability can be approximated as

$$p_1(P_t, \rho) \approx \sum_{i=1}^{N_f} \frac{p_r(i) \pi \lambda_i}{\varphi_i(P_t)} \left(1 - e^{-\varphi_i(P_t) r_c^2}\right),$$

which first increases and then decreases with $P_t$, where $\varphi_i(P_t)$ is defined in (C.2).

**Proof:** See Appendix C

The approximation is accurate when the file catalog size $N_f$ is large. As shown in Appendix C, the closed-form solution of $P_t^*$ can be obtained by solving a cubic equation, which is not provided herein for conciseness.
2) **Offloading Gain:** To evaluate the offloading gain provided by cache-enabled D2D communications, which is characterized by the offloading ratio, both complete transmission and partial transmission should be taken into account.

**Proposition 4:** The offloading ratio in the full reuse case is

\[
P^a_1(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} \frac{f_i(r)}{\ln(1+\Gamma_1)} \int_0^{\Gamma_1} \frac{e^{-\phi_i(t,x)}}{1+t} \, dt \, dr,
\]  

and\( p_1(P_t, \rho) \leq p^a_1(P_t, \rho) \leq p_o \), both equalities will hold if \( \rho \to \infty \) or if \( r^o\sigma^2_0 \to 0 \) and \( \lambda_I \to 0 \).

**Proof:** See Appendix [D]

The first condition \( \rho \to \infty \) means that all helpers have infinite battery capacity, which is the scenario where the user devices are charging when acting as the DTs. The second condition \( r^o\sigma^2_0 \to 0 \) and \( \lambda_I \to 0 \) indicate that all interference are eliminated and the SNR is infinite, because \( r^o\sigma^2_0 = 1/(P_t r - \alpha \sigma^2_0) \) is the inverse of the receive signal to noise ratio (SNR) at the DR averaged over fading. In this case, although battery is limited, the data rate can be extremely high to complete all transmission via D2D links. In either condition, the offloading probability, offloading ratio and offloading opportunity are equal.

3) **Energy costs:** In what follows, we derive the energy cost of a DT for a given transmit power and caching policy, with which we can evaluate the energy cost of a DT with the optimized transmit power and caching policy.

**Proposition 5:** The average energy consumed at a DT for a complete transmission is

\[
\bar{E}_1 = \rho V_0 Q - \rho V_0 Q \sum_{i=1}^{N_f} \frac{p_r(i)}{p_1(P_t, \rho)} \int_0^{r_c} f_i(r) \ln(1 + \Gamma_1) \int_0^{\Gamma_1} \frac{1}{1+t} e^{-\phi_i(t,x)} \, dt \, dr
\]  

\[= \rho V_0 Q - \rho V_0 Q \sum_{i=1}^{N_f} \frac{p_r(i)}{p_1(P_t, \rho)} \int_0^{r_c} f_i(r) \ln(1 + \Gamma_1) \int_0^{\Gamma_1} \frac{1}{1+t} e^{-\phi_i(t,x)} \, dt \, dr,
\]  

where (a) is obtained by substituting (2), (12) and (13).

To reflect how much energy consumed at a DT by serving as a helper occupies the battery capacity, we define the energy cost as \( \bar{e}_1 = \frac{\bar{E}_1}{V_0 Q} \).
B. Case 2: TDMA

By using TDMA, the DT of a randomly scheduled D2D link transmits the requested file to its corresponding DR, while other DTs stay mute. The data rate of each DR is given by

\[ R_2(r) = \frac{W}{N_a} \log_2\left(1 + \frac{P_{th} r^{-\alpha}}{\sigma^2}\right), \]

(15)

where \( N_a = p_o \lambda S \) is the average number of DRs in a cell and \( S \) is the area of the cell. The muting DTs can turn off some circuits to save energy. We call the circuit power consumed by a muting DT as idle power, denoted as \( P_{cI} \), which ranges from a few to tens of mW [24]. Then, the energy consumed at a DT to transmit a file via the D2D link can be obtained as,

\[ E_2(r) = F \frac{E_{\text{r}}(r)}{N_a R_2(r)} \left( \frac{1}{\eta} P_t + P_{cI} \right) + \left( \frac{E_{\text{r}}(r)}{N_a R_2(r)} \right) P_{cI} = F \frac{E_{\text{r}}(r)}{W \log_2(1 + \frac{P_{th} r^{-\alpha}}{\sigma^2})} \left( \frac{1}{\eta} P_t + P_{cI}^T \right), \]

(16)

where \( P_{cI}^T \triangleq P_{cI} + (N_a - 1) P_{cI} \). Note that \( E_2(r) \) is not related to the file index \( i \) since the received signals are only corrupted by noise, which is different from \( E_1(r, i) \).

1) Optimal Transmit Power: From the definition of the offloading probability and (16), it is easy to obtain \( p_2(P_t, \rho) \) by letting \( \lambda_f \xi_1 - \lambda_f^d \xi_2 = 0 \) in (9) as

\[ p_2(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^c f_i(r) e^{-\Gamma_2 r^2 \sigma^2} dr, \]

(17)

where \( \Gamma_2 = e^{\frac{E(P_t + \eta P_{cI})}{W \rho N_0 \sigma^2}} - 1 \). With the growth of both the number of DRs \( N_a \) and idle power \( P_{cI} \), the circuit power \( P_{cI}^T \) and hence \( \Gamma_2 \) increase, which results in the reduction of \( p_2(P_t, \rho) \).

To maximize the offloading probability for the cache-enabled D2D communications, the transmit power at each DT can be optimized as follows

\[ \max_{P_t} p_2(P_t, \rho) \]

\[ s.t. \quad 0 < P_t \leq P_{\text{max}}. \]

(18)

Again, the closed-form expression of \( p_2(P_t, \rho) \) is hard to obtain in general. When \( \alpha = 2 \), by using the similar way to derive (11), we can approximate the offloading probability as

\[ p_2(P_t, \rho) \approx \sum_{i=1}^{N_f} p_r(i) \pi \lambda_i \frac{1 - e^{-\sigma^2 (P_{\text{th}} + \pi \lambda_i)^2}}{\sigma^2 (P_{\text{th}} + \pi \lambda_i)^2}. \]

(19)

Despite that the offloading probability has complicated expression in general cases, the closed-form solution of the optimal transmit power for all values of \( \alpha \) can be found as follows.
Proposition 6: The optimal transmit power is
\[
P_t^* = \begin{cases} 
P_{\text{max}}, & P_{\text{max}} < \eta P_T^c \left(\sqrt{\frac{1}{a\eta P_T^c}} + \frac{1}{4} - \frac{1}{2}\right), \\
\eta P_T^c \left(\sqrt{\frac{1}{a\eta P_T^c}} - \frac{1}{2}\right), & \text{otherwise}
\end{cases}
\]  \tag{20}

where \( a = \frac{E_{\ln 2}}{W\rho Q V_0}. \)

Proof: See Appendix F

2) Offloading Gain: By considering both the complete and partial transmission, we can obtain the offloading ratio by using a similar method as for the proof of Proposition 4 as
\[
p_2^o(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} \frac{f_i(r)}{\ln(1+\Gamma_2)} f_i(r) \ln(1+\Gamma_2) e^{-r^2} \sigma_0^2 dt dr
\]
\[
= \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} \frac{f_i(r)}{\ln(1+\Gamma_2)} e^{-r^2} \sigma_0^2 \left(\text{Ei}\left(-r^2 \sigma_0^2 (\Gamma_2 + 1)\right) - \text{Ei}\left(-r^2 \sigma_0^2\right)\right) dr,
\]
where \( \text{Ei}(x) = \int_{-x}^{\infty} e^{-t^2} dt \) is a frequently-used special function.

3) Energy costs: By using the similar derivation as for Proposition 5 we can obtain the average energy consumed at a DT for a complete transmission with given transmit power and caching policy as
\[
\bar{E}_2 = \rho V_0 Q - \rho V_0 Q \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \ln(1+\Gamma_2) e^{-r^2} \sigma_0^2 dt dr.
\]  \tag{21}

Since a percentage \( \frac{p_2(P_t, \rho)}{p_o} \) of the requested files can be completely conveyed via D2D transmissions, and the DT only consumes energy \( \rho V_0 Q \) to help the DR for a partial transmission, by considering both complete and partial transmission, the average energy consumption for a DT can be obtained as
\[
\bar{E}_2^a = \frac{p_2(P_t, \rho)}{p_o} \bar{E}_2 + \left(1 - \frac{p_2(P_t, \rho)}{p_o}\right) \rho V_0 Q
\]
\[
= \rho V_0 Q - \rho V_0 Q \sum_{i=1}^{N_f} \frac{p_r(i)}{p_o} \int_0^{r_c} f_i(r) \ln(1+\Gamma_2) e^{-r^2} \sigma_0^2 dt dr,
\]  \tag{22}
where (a) is obtained by substituting (2), (17) and (21).

Then, the energy cost for a DT to transmit a file is \( \bar{v}_2 = \frac{\bar{E}_2^a}{V_0 Q}. \)

V. Simulations

In this section, we validate previous analytical results, and evaluate the offloading gain of the system and the energy cost at each DT via simulations.
We consider a square cell with side length 500 m. The users’ locations follow a PPP with $\lambda = 0.01$, so that in average there is one user in a $10 \text{ m} \times 10 \text{ m}$ area. The path-loss model is $37.6 + 36.8 \log_{10}(r)$, where $r$ is the distance of the D2D link [17]. $W = 20 \text{ MHz}$ and $\sigma^2 = -100 \text{ dBm}$, the maximal transmit power of each DT is $P_{\text{max}} = 200 \text{ mW}$ (23 dBm), the circuit power for an active DT is $P_c = 115.9 \text{ mW}$, and the power amplifier efficiency is $\eta = 0.5$ [4]. The typical idle power for a muting DT with TDMA is $P_{\text{cI}} = 25 \text{ mW}$ [24]. The operating voltage at a user device is $V_0 = 4V$ and the battery capacity is $Q = 1800 \text{ mAh}$ (typical for current-generation smartphones). The file catalog contains $N_f = 1000$ files, where each file has a size of 30 Mbytes (a typical video size on the Youtube website [14]). The parameter of the Zipf distribution $\beta = 1$. This setup is used in the sequel unless otherwise specified.

Besides the optimized caching policy in Proposition 1 (with legend “Optimal-x”), we also provide the results for uniform caching policy (i.e., each user selects a file from the catalog uniformly, with legend “Uniform-x”) and popularity based caching policy (i.e., each user selects a file from the catalog according to the content popularity, with legend “Popularity-x”) as the baseline caching policies, where “x” is the number of files cached at each user.

A. Optimal Caching Distribution and Offloading Opportunity

In Fig. 2(a), we show the optimal caching distribution for different collaboration distance $r_c$, Zipf parameter $\beta$, and user density $\lambda$. With the increase of $\beta$ and $\lambda$ or decrease of $r_c$, the
probability of caching popular files increases, which makes the distribution more “skewed”, and vice versa. When $r_c$ is large enough, say $r_c = 500$ m, the caching distribution reduces to a uniform distribution. When $r_c$ is very small, say $r_c = 20$ m, the caching distribution makes the probability for caching most popular files very high, which agrees with Proposition 1.

In Fig. 2(b), we show the simulated offloading opportunity versus the collaboration distance, where each user allows to cache more files. When each user has cached one or two files, the optimized caching policy has the potential to offload more traffic than the popularity based caching policy and even more than the uniform caching policy. When each user caches more files, the offloading opportunity is improved for all policies, as expected. For large value of $r_c$, say 100 m, the offloading opportunities of all caching policies can achieve nearly 0.8. This indicates that when $r_c$ is large and a user is willing to cache more files, uniform caching policy can also achieve high traffic offloading, despite that it is not good for D2D throughput in general [15]. Nonetheless, the offloading opportunity for caching one file and more files exhibits the same trend, which indicates that caching more files offers essentially the same insight with caching one file, which justifies our assumption in previous analysis.

B. Validation of Analytical Results

In Fig. 3 we compare the numerical and simulation results for the offloading probability $p_1(P_t, \rho)$, $p_2(P_t, \rho)$ and offloading ratio $p_{a1}(P_t, \rho)$, $p_{a2}(P_t, \rho)$ respectively for the full reuse and
TDMA cases in Fig. 3(a) and energy cost $\bar{e}_1$ and $\bar{e}_2$ for the two cases in Fig. 3(b) versus $P_t$. We can see that the numerical results almost overlap with simulation results, which validates our analysis. Moreover, the trend for $p_1(P_t, \rho)$ and $p_2(P_t, \rho)$ changing with $P_t$ are the same as $p_{1a}(P_t, \rho)$ and $p_{2a}(P_t, \rho)$, respectively. This suggests that the optimized transmit power to maximize the offloading probability can also maximize the offloading gain. For the full reuse case, $p_{1a}(P_t, \rho)$ first increases to achieve the maximal value and then decreases, and $\bar{e}_1$ first decreases and then increases. This is due to the severe interference and the allowed battery consumption. Comparing Fig. 3(a) and Fig. 3(b), we can observe that the optimal value of $P_t$ to maximize the offloading probability $p_{1a}(P_t, \rho)$ can nearly minimize the energy cost $\bar{e}_1$. This is because to maximize the offloading ratio the transmit power should be reduced in an interference environment and then more DTs consume less power than the allowed battery consumption. For the TDMA case in the considered setting, increasing $P_t$ can always improve $p_{2a}(P_t, \rho)$. Moreover, $\bar{e}_2$ always decreases, because increasing $P_t$ can shorten the duration of transmission and hence can reduce the circuit power consumption.

In Fig. 4 we compare the numerical and simulation results for the offloading probability under special channel models. By using the same approach as for deriving the closed form solution for the LOS channel, we can also derive the closed-form expressions for the offloading probability under $\alpha = 4$, which are not shown for conciseness. We can see that the numerical results almost
overlap with the simulation results, which indicates that the approximations in (11) and (19) are accurate. When $\alpha = 2$ and $\alpha = 4$, the offloading probability for the full reuse case first increases and then decreases with $P_t$, and the offloading probability for TDMA case always grows with $P_t$ in the considered setting, which are the same as Fig. 3. It implies that the optimal solution of $P_t$ can be found by bisection searching efficiently in general channels, because $2 \leq \alpha \leq 4$ in practical channels among D2D links [17].

C. Impacts of Key Parameters and Self Offloading

In what follows, we analyze the impact of the file size, idle power, content popularity, energy consumption allowed by user device, as well as the self offloading on the offloading gain and energy cost for full reuse and TDMA cases with numerical results. The optimized transmit power and optimized caching policy are used, unless otherwise specified.

In Fig. 5, we show the impact of file size. We can see that for TDMA, $P_t^* \neq P_{\text{max}}$ only when $\rho = 0.01$ and $F > 2.2 \text{ GBytes}$. In all other cases, transmitting with $P_{\text{max}}$ is optimal for TDMA. We can also observe that with the growth of $F$, the offloading probability for both full reuse and TDMA cases decreases. This implies that cache-enabled D2D communications is more applicable for offloading traffic of delivering small size files.

In Fig. 6, we show the impact of the idle power $P_{c_I}$. We can see that the offloading ratio of TDMA is always larger than that of the full reuse. The increase of $P_{c_I}$ directly leads to higher
energy cost at each DT with TDMA. When only 1% battery capacity can be used and $P_{c_I} > 40$ mW, the energy cost at a DT with TDMA is larger than full reuse. When $\rho = 30\%$ (not shown in the figure), the energy cost of full reuse is around 12%, which is much larger than the energy cost of TDMA that changes from 0 to 2.5 % with the increase of $P_{c_I}$.

In Fig. [7] we show the impact of the Zipf parameter $\beta$ and self offloading. As expected, the offloading ratio increases rapidly with $\beta$ due to the high cache hit rate. The energy cost increases with $\beta$ for TDMA, but first increase and finally decreases with $\beta$ for full reuse. This can be
explained as follows. On one hand, larger $\beta$ leads to smaller D2D link distance, which reduces the energy cost. On the other hand, larger $\beta$ leads to more DTs, which generates more severe interference for full reuse and longer muting time for TDMA, both of which increase the energy cost. Since the reduction in D2D link distance is dominant for full reuse, the energy cost finally decreases. Besides, the offloading ratio including both cache-assisted D2D communications and self offloading is larger than that only contributed by the cache-enabled D2D, and the energy cost including both is less than that only considering D2D. However, the contribution of self offloading on the performance is marginal.

In Fig. 8 we show the impact of the allowed battery consumption $\rho$. We can see that the offloading ratio first increases rapidly and then slowly with $\rho$, whereas the energy cost increases with $\rho$ but is always much less than the allowed battery consumption. This is because for D2D links with better channel state, the DTs can transmit complete files to corresponding DRs with less than $\rho Q V_0$ of energy. The results suggest that choosing a proper $\rho$ is important for operators to balance benefits (e.g., the offloading gain) and costs (e.g., rewarding users for a larger value of $\rho$). We can also observe that the energy cost for the full reuse case grows more rapidly than the TDMA case, because there are more partial transmission links in the full reuse case than the TDMA case, whose energy consumed by each DT equals the allowed battery consumption.

![Graphs showing offloading ratio and energy cost](image_url)

(a) Offloading ratio  
(b) Energy cost

Fig. 8. Impact of the allowed fraction of battery consumption $\rho$. 
D. Relationships Between Offloading Gain and Energy Cost

In the sequel, we show the relation between offloading ratio and energy cost with the optimal transmit power and optimal caching policy.

In Fig. 9(a), the offloading ratio is adjusted by changing the collaboration distance $r_c$ from 10 m to 400 m, where popularity based policy is also simulated for comparison. We can observe an optimal $r_c$ to maximize the offloading ratio for a given $\rho$, which are 350 m, 300 m, 100 m, 80 m, respectively for “TDMA, $\rho = 0.3$”, “TDMA, $\rho = 0.1$”, “Full Reuse, $\rho = 0.3$”, and “Full Reuse, $\rho = 0.1” with the optimal caching policy. This is because the full reuse scheme is interference limited and the TDMA scheme is transmit power limited. With the growth of $r_c$, the average D2D communication distance increases, and hence the energy cost increases, whereas the very limited battery consumption allowed for helping others makes the offloading ratio decrease. Compared with popularity based caching policy, the optimal caching policy can improve the offloading ratio and reduce energy cost.

In Fig. 9(b), the offloading ratio is adjusted by changing $\rho$ from 0 to 1, $r_c = r^\ast_c$. To show what happens if a helper serves multiple requests, here each user sends $N_r$ requests sequentially according to the Zipf distribution. As a result, the helper that cached the most popular files may be requested multiple times and serve as a DT for multiple users. When $N_r = 1$, there exists a tradeoff between offloading gain and energy cost. When $N_r > 1$, a large energy cost may not yield a
high offloading gain. This is because with larger $\rho$, a DT will consume more energy before interrupting the transmission for a D2D link with bad channel condition, and will soon run out of battery for serving subsequent requests. Consequently, each DT can serve fewer requests, which leads to the reduction of the offloading gain. Nonetheless, it is interesting to observe that the energy cost to support high offloading ratio is low. Even when $N_r = 10$, to offload around 80% of traffic, the average energy consumption at each DT with TDMA only consumes around 10% battery capacity. This suggests that cache-enabled D2D communications is cost-efficient for offloading by optimizing the collaboration distance and selecting a proper transmission scheme.

In the following, we provide a brief summary of the simulation results.

- **Caching policy**: When the collaboration distance is small or only one file is cached, optimizing caching policy can improve offloading gain and reduce energy cost.
- **Transmission scheme**: When the file size is not large, TDMA is superior to full reuse with typical value of idle power. The optimization of transmit power to maximize the offloading probability also helps increase the offloading gain and reduce the energy cost. For the TDMA case, the DT can simply transmit with $P_{max}$ to maximize the offloading gain if the file size is not too large. For the full reuse case, optimizing transmit power is important.
- **Parameter setting**: There exists an optimal value of $r_c$ that maximizes the offloading gain for a given $\rho$. When each DT only serves one request, both offloading gain and energy cost increase with $\rho$. When each DT could serve multiple requests, a large value of $\rho$ not only causes large energy cost but also reduces the offloading gain.
- **Gain and costs**: When the file size is not very large, a high offloading gain can be achieved by a low energy cost if the collaboration distance, transmission scheme and caching policy are judiciously designed and the value of $\rho$ is properly selected.

VI. CONCLUSION

In this paper, we quantified the offloading gain of cache-enabled D2D communications after taking the user allowed battery consumption into account and evaluated the energy consumed at a helper user. We considered a user-centric caching and transmission protocol, where the energy consumed for transmission can be controlled by a collaboration distance. We first optimized a proactive caching policy with given collaboration distance, with which the offloading opportunity can be maximized. For either full reuse or TDMA (round-robin) scheduling, we then optimized
the transmit power to deliver a file via D2D link, where the percentage of satisfied users is maximized. With the optimized probabilistic caching policy and optimized transmit power, we evaluated the offloading gain of the system and the energy cost of a D2D transmitter, and investigated their relationship. Simulation results showed that high offloading gain can be obtained in practice by cache-enabled D2D with low energy cost at each help user, if the collaboration distance, transmission scheme and caching policy are optimized and the allowed battery consumed by each D2D transmitter for conveying one file is properly set.

**APPENDIX A**

**PROOF OF PROPOSITION 1**

Denote $x_i \triangleq \frac{\ln(p_r(i))}{\lambda \pi r_c^2}$ and $v \triangleq \frac{1}{\lambda \pi r_c^2} \ln\left(\frac{-\mu}{\pi \lambda r_c^2}\right)$. Then, considering $\sum_{i=1}^{N_f} p_c(i) = 1$ and from (4) we have

$$\sum_{i=1}^{N_f} [x_i - v]^+ = 1. \quad (A.1)$$

Since problem (3) is convex, the solution of $v$ found from this necessary condition is globally optimal, and with it the optimal caching distribution can be obtained.

As shown in (4), $p_c^*(i)$ decreases when $p_r(i)$ decreases. As shown in (1), $p_r(i)$ is a decreasing function of file index $i$. This indicates that $p_c^*(i)$ is a decreasing function of $i$. Thus, there exists a unique file index $i^* \leq N_f$, with which $p_c^*(i) > 0$ if $i \leq i^*$, $p_c^*(i) = 0$ otherwise. As a result, finding the solution of $v$ from (A.1) is equivalent to finding the index $i^*$ from $\sum_{i=1}^{i^*} (x_i - v) = 1$. Once $i^*$ is found, the solution of (A.1) can be obtained as

$$v^* = \frac{\sum_{i=1}^{i^*} x_i - 1}{i^*}. \quad (A.2)$$

**Case 1:** When $i^* = N_f$, from (4) and $p_c^*(i) \geq 0$ we have $p_c^*(N_f) = x_{N_f} - v > 0$, which can be rewritten as $\sum_{i=1}^{N_f} (x_i - x_{N_f}) < 1$ after substituting $v$ in (A.2), then

$$\sum_{i=1}^{N_f} (x_i - x_{N_f}) = \sum_{i=1}^{N_f} \frac{\ln(p_r(i)) - \ln(p_r(N_f))}{\lambda \pi r_c^2} = \frac{\beta}{\lambda \pi r_c^2} \sum_{i=1}^{N_f} \ln(N_f/i) = \frac{\beta}{\lambda \pi r_c^2} \ln\left(\frac{N_f^{N_f}}{N_f!}\right) < 1, \quad (A.3)$$

which can be rewritten as $\frac{(N_f)^{N_f}}{N_f!} < e^{\frac{\lambda \pi r_c^2}{\beta}}$. By substituting $v^*$ in (A.2) into (4), the optimal caching distribution can be derived as

$$p_c^*(i) = \frac{\beta}{\lambda \pi r_c^2 N_f} \sum_{j=1}^{N_f} \ln\left(\frac{j}{i}\right) + \frac{1}{N_f}. \quad (A.4)$$
Case 2: When \( i^* < N_f \), \( p_c^i(i^*) = x_{i^*} - v > 0 \) and \( x_{i^*+1} - v \leq 0 \). By substituting \( v \) in (A.2) into these two inequalities, we have \( \sum_{i=1}^{i^*} (x_i - x_{i+1}) \geq 1 \) and \( \sum_{i=1}^{i^*} (x_i - x_{i+1}) < 1 \), which can be further derived by substituting \( p_r(i) \) in (1) and \( x_i \) as
\[
\frac{\beta}{\pi r^2} \ln\left(\frac{(i^*+1)x^*}{i^*+1}\right) \geq 1, \quad \frac{\beta}{\pi r^2} \ln\left(\frac{(i)x^*}{i+1}\right) < 1. \tag{A.5}
\]

Then, \( i^* \) satisfies \( \frac{(i^*+1)x^*}{i^*+1} \geq e^{\frac{\lambda \pi r^2}{\beta}} \) and \( \frac{(i)x^*}{i+1} < e^{\frac{\lambda \pi r^2}{\beta}} \). With Stirling formula \[25\], \( \sqrt{2\pi n} \left(\frac{n}{e}\right)^n < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \), (A.5) can be further derived as \( i^* - \ln(\sqrt{2\pi i^*} e^{\frac{1}{12n}}) < \frac{\lambda \pi r^2}{\beta} \), \( i^* + 1 - \ln(\sqrt{2\pi i^*}) > \frac{\lambda \pi r^2}{\beta} \). Since \( i^* < N_f \), \( \ln(\sqrt{2\pi i^*}) < \ln(\sqrt{2\pi N_f}) \), we can obtain the range of \( i^* \) as \( \frac{\lambda \pi r^2}{\beta} - 1 \leq i^* < \frac{\lambda \pi r^2}{\beta} + \ln(\sqrt{2\pi N_f}) + 1 \).

By substituting \( v \) in (A.2) into (4), we obtain
\[
p_c^i(i) = \begin{cases} \frac{\beta}{\pi r^2} \sum_{j=1}^{i^*} \ln\left(\frac{j}{i^*+1}\right) + \frac{1}{i^*}, & i \leq i^* \\ 0, & i^* < i \leq N_f. \end{cases} \tag{A.6}
\]

Finally, we prove Proposition 1.

**APPENDIX B**

**PROOF OF PROPOSITION 2**

The cumulative distribution function (cdf) of the distance between a DR requesting the \( i \)th file and the nearest helper in the \( i \)th helper set can be obtained as \( F_i(r) = 1 - e^{-\lambda \pi r^2} \) [20]. Then, the pdf of the D2D link distance can be obtained as \( f_i(r) = \frac{dF_i(r)}{dr} = 2\pi r \lambda_i e^{-\lambda_i \pi r^2} \).

According to the definition, the offloading probability can be obtained as
\[
p_1(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P} \left[ E_i(i, r) \leq \rho V_0 Q \right] dr, \tag{B.1}
\]
where \( \mathbb{P}[\cdot] \) is the probability of an event.

Denoting \( \Gamma_1 = e^{\frac{F_i(P_t + \eta P_t)}{W_0 Q V_0}} - 1 \), the energy constraints \( E_i(i, r) \leq \rho V_0 Q \) can be rewritten as \( \frac{h r^{-\alpha}}{I_i, r + \sigma_0^2} \geq \Gamma_1 \), and (B.1) can be further written as
\[
p_1(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P} \left[ \frac{h r^{-\alpha}}{I_i, r + \sigma_0^2} \geq \Gamma_1 \right] dr. \tag{B.2}
\]

Since \( h \) follows an exponential distribution with unit mean, \( \mathbb{P} \left[ \frac{h r^{-\alpha}}{I_i, r + \sigma_0^2} \geq \Gamma_1 \right] \) can be obtained as
\[
\mathbb{P} [h \geq \Gamma_1 r^{-\alpha} (\sigma_0^2 + I_i, r)] = \mathbb{E} \left[ \exp \left(-r^{-\alpha} \Gamma_1 (\sigma_0^2 + I_i, r)\right) \right] = e^{-\Gamma_1 r^{-\alpha} \sigma_0^2 \mathcal{L}_{I_i, r} (r^{-\alpha} \Gamma_1)}, \tag{B.3}
\]

June 10, 2016 DRAFT
Then, for the DR requesting the $i$th file, we can be closer than the desired DT.

\[ \sum_{i=1}^{N_I} p_c(i) p_s(i) \]

where (a) is obtained using the same method as in [26], and $L_{I_s}(s)$ is the Laplace transform of the random variable $I_{s,r}$. Note that if some DTs interrupt during transmission, the interference should be lower. However, we ignore this situation to make the analysis tractable to reflect the essential insights.

To derive $L_{I_s} (\rho \alpha \Gamma_1)$, we need to obtain the probability that a helper acts as a DT, called active probability.

As mentioned in Section III, the users in the $i$th helper set follow a PPP with density $\lambda_i$. Since it is hard to directly derive the active probability, we derive its complementary probability. We first derive the probability that a helper cached with the $i$th file does not act as a DT, denoted as $p_s(i)$. Considering that the coverage area of each helper can not exceed $\pi r_c^2$ and the probability that no DRs requesting the $i$th file are located in the coverage with area $x$ of a helper cached with the $i$th file is $e^{-\lambda p_c(i) x}$, $p_s(i)$ can be derived as

\[ p_s(i) = \frac{\text{B.5}}{\text{B.4}} \]

where $\theta_i \triangleq \frac{\Gamma(3.5, 0) - \Gamma(3.5, 3.5\lambda_i + \lambda p_c(i) \pi r_c^2)}{\Gamma(3.5, 0) - \Gamma(3.5, 3.5\lambda_i \pi r_c^2)}$, $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt$ is the upper incomplete Gamma function [23], $g_s(x, \lambda_i)$ is the pdf of the coverage area for a typical Voronoi cell in a Poisson random tessellation, which can be fitted as

\[ g_s(x, \lambda_i) = \frac{3.5^{3.5 \lambda_i^5} x^{2.5} e^{-3.5 \lambda_i x}}{\Gamma(3.5, 0)^{3.5 \lambda_i^5}} \]

Then, for the probabilistic caching policy, the active probability can be obtained as

\[ p_a = 1 - \sum_{i=1}^{N_I} p_c(i) p_s(i). \]

The DTs cached with the $i$th file follow a PPP distribution with density $\lambda_i d = (1 - p_s(i)) p_c(i) \lambda = (1 - p_s(i)) \lambda_i$ [28], and the density of all DTs is $\lambda_I = \sum_{i=1}^{N_I} \lambda_i d = \sum_{i=1}^{N_I} (1 - p_s(i)) p_c(i) \lambda = p_a \lambda$.

Recall that the DR establishes a D2D link with the nearest helper cached with the desired file. Then, for the DR requesting the $i$th file, the DTs cached with other files with density $\lambda_I - \lambda_i d$ can be closer than the desired DT.
Using Theorem 2 in [26], when $\alpha > 2$, $\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1)$ can be derived as

$$
\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1) = \exp \left( -2\pi \left( \lambda_I - \lambda_d^d \right) \int_0^\infty \frac{1}{\Gamma_1 + (v/r)^{\alpha}} \cdot v du - 2\pi \lambda_d^d \int_r^\infty \frac{1}{\Gamma_1 + (v/r)^{\alpha}} \cdot v du \right)
$$

$$
= \exp \left( -2\pi \lambda_I \int_0^\infty \frac{1}{\Gamma_1 + (v/r)^{\alpha}} \cdot v du + 2\pi \lambda_d^d \int_r^\infty \frac{1}{\Gamma_1 + (v/r)^{\alpha}} \cdot v du \right)
$$

$$
= \exp \left( -\pi \lambda_I r^{2\alpha} \int_0^\infty \frac{1}{1+u^{\alpha/2}} \cdot du + \pi \lambda_d^d r^{2\alpha} \int_{r}^{\infty} \frac{1}{1+u^{\alpha/2}} \cdot du \right) = \exp \left( -\pi r^{2\alpha} (\lambda_I \xi_1 - \lambda_d^d \xi_2) \right),
$$

where $\xi_1 = \int_0^\infty \frac{1}{1+u^{\alpha/2}} \cdot du$ and $\xi_2 = \int_{r}^{\infty} \frac{1}{1+u^{\alpha/2}} \cdot du$.

By substituting (B.3) and (B.6) into (B.2), we prove Proposition 2.

**APPENDIX C**

**PROOF OF PROPOSITION 3**

Since (B.6) holds only when $\alpha > 2$, in case of $\alpha = 2$, we need to re-derive the Laplace transform $\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1)$. By ignoring the interference generated by the DTs with distance larger than $r_{\max}$ as in [29], we can approximate $\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1)$ as

$$
\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1) \approx \exp \left( -2\pi \left( \lambda_I - \lambda_d^d \right) \int_0^{r_{\max}} \frac{1}{\Gamma_1 + (v/r)^{2}} \cdot v du - 2\pi \lambda_d^d \int_{r}^{r_{\max}} \frac{1}{\Gamma_1 + (v/r)^{2}} \cdot v du \right)
$$

$$
\overset{(a)}{\approx} \exp \left( -\pi \lambda_I r^{2} \int_0^{r_{\max}} \frac{1}{1+u^{\alpha/2}} \cdot du \right) = \exp \left( -\pi r^{2} \lambda_I \ln(1 + r_{\max}) \right),
$$

where (a) is obtained by using $\lambda_I \gg \lambda_d^d$ when the file catalog size $N_f$ is large.

Then, the offloading probability can be obtained by substituting $\mathcal{L}_{I_{r,s}}(r^{\alpha} \Gamma_1)$ into (B.3) as

$$
p_1(P_t, \rho) \approx \sum_{i=1}^{N_f} p_r(i) \int_0^{r_{\max}} e^{-\rho \lambda_I r^2} e^{-\Gamma r^2 \sigma_0^2 - \pi \lambda_I \ln(1 + r_{\max})} r^2 \Gamma_1 \cdot dr = \sum_{i=1}^{N_f} \frac{p_r(i) \rho \lambda_I}{\varphi_i(P_t)} (1 - e^{-\varphi_i(P_t) r^2}),
$$

where $\varphi_i(P_t) = \sigma_0^2 \Gamma_1 + \pi \lambda_I \xi_s \Gamma_1 + \pi \lambda_i$, and $\xi_s = \ln(1 + r_{\max})$.

By denoting $g_i(P_t) = \frac{1-e^{-\varphi_i(P_t) r^2}}{\varphi_i(P_t)}$, and taking the derivative of $g_i(P_t)$ with respect to $P_t$, we can obtain

$$
g_i'(P_t) = \frac{\kappa(\varphi_i(P_t) r^2)}{(\varphi_i(P_t))^2} \cdot \left( 1 - e^{-\varphi_i(P_t) r^2} \right) = \frac{(\varphi_i(P_t) r^2 - 1)}{(\varphi_i(P_t))^2} \cdot \kappa'(P_t),
$$

where $\kappa(t) \triangleq (1 + t)e^{-t}$, $t \geq 0$. It is not hard to show that $\kappa'(t) = -te^{-t} \leq 0$, so $\kappa(t)$ is a decreasing function of $t$ and $\kappa(0) = 1$ is the maximal value of $\kappa(t)$. Therefore, $\kappa(t) \leq 1$ and the equality holds when $t = 0$. Because $\varphi_i(P_t) r^2 > 0$, $\kappa(\varphi_i(P_t) r^2) < 1$ always holds. Then, part (I) in (C.3) $\frac{(\kappa(\varphi_i(P_t) r^2)-1)}{(\varphi_i(P_t))^2} < 0$ always holds.
By changing variable $P_t \rightarrow x$ and denoting $a = \frac{F \ln 2}{W \rho Q V_0 \eta}$, we have $\varphi_i(x) = \sigma^2 \frac{(e^{a(x+\eta P_c)}-1)}{x} + \pi \lambda f \xi_s (e^{a(x+\eta P_c)} - 1) + \pi \lambda i$, whose first-order derivative can be obtained as

$$
\varphi'_i(x) = \frac{\sigma^2}{a^2} \left( \frac{e^{a(x+\eta P_c)} a x - e^{a(x+\eta P_c)} + 1}{u_1(x)} + \frac{a \pi \lambda f \xi_s e^{a(x+\eta P_c)}}{u_2(x)} \right).
$$

(C.4)

The first-order derivative of $u_1(x)$ can be derived as

$$
u'_1(x) = \frac{1}{x^2} \left( a^2 e^{a(x+\eta P_c)} x^2 - 2 a e^{a(x+\eta P_c)} x + 2 e^{a(x+\eta P_c)} - 2 \right).
$$

(C.5)

It is not hard to obtain that $v'(x) = a^3 e^{a(x+\eta P_c)} x^2 \geq 0$. Therefore, $v(x)$ is an increasing function of $x$ and $v(0) = 0$. Because $x > 0$, we know that $v(x) \geq 0$ always holds, i.e., $u'_1(x) \geq 0$, and hence $u_1(x)$ an increasing function.

By using the same approach, we can show that $u_2(x) = a \pi \lambda f \xi_s e^{a(x+\eta P_c)}$ in (C.4) is an increasing function. Then, $\varphi'_i(x)$ is an increasing function. Besides, when $x \rightarrow 0$, $\lim_{x \rightarrow 0} \varphi'_i(x) = \lim_{x \rightarrow 0} (a \pi \lambda f \xi_s e^{a(x+\eta P_c)} - \frac{(e^{a(x+\eta P_c)}-1) \sigma^2}{x^2}) = -\infty$.

When $x = P_{max}$, $\varphi'_1(P_{max}) = \frac{\sigma^2}{P_{max}} \left( \frac{e^{a(P_{max}+\eta P_c)} a P_{max} - e^{a(P_{max}+\eta P_c)} + 1}{a \pi \lambda f \xi_s e^{a(P_{max}+\eta P_c)}} \right) > \frac{\sigma^2}{P_{max}^2} (e^{a(P_{max}+\eta P_c)} + 1) + a \pi \lambda f \xi_s e^{a(P_{max}+\eta P_c)}$, because $\frac{\sigma^2}{P_{max}} \ll 1$ in practice. Further considering that part (I) in (C.3) is negative, $g'_i(P_t)$ is first greater than zero and then less than zero, and hence $g_i(P_t)$ first increases and then decreases with $P_t$. Therefore, the global optimal transmit power $P^*_t$ to maximize $g_i(P_t)$ can be obtained by solving $\varphi'_i(P_t) = 0$.

It is worthy to note that although $\varphi_i(x)$ depends on $i$, its first-order derivative does not. Therefore, the optimal transmit power $P^*_t$ to maximize $p_t(P_t, \rho)$ in (C.2) is the same as that to maximize $g_i(P_t)$.

Considering that today’s device battery capacity is usually large, the file size is not very large (typically less than 3 GBytes), and $\frac{\rho Q V_0 \eta}{P_{max} + \eta P_c}$ is the maximal time that a DT can transmit with $P_{max}$, we have $a(P_{max} + \eta P_c) = \frac{F \ln 2(P_{max} + \eta P_c)}{W \rho Q V_0 \eta} = F \ln 2 \left( \frac{1}{W} \frac{P_{max} + \eta P_c}{\rho Q V_0 \eta} \right) \ll 1$. By using the approximation $e^t \approx 1 + t$, when $t \ll 1$, $\varphi'_1(x) = \frac{\sigma^2}{x^2} \left( e^{a(x+\eta P_c)} a x - e^{a(x+\eta P_c)} + 1 \right) + a \pi \lambda f \xi_s e^{a(x+\eta P_c)} \approx \frac{\sigma^2}{x^2} \left( a(x+\eta P_c) + 1 \right) a x - \left( a(x+\eta P_c) + 1 \right) + a \pi \lambda f \xi_s (a(x+\eta P_c) + 1)$.

Then, the optimal $x^*$ satisfying $\varphi'_1(x) = 0$ can be obtained by solving the cubic equation $a^2 \mu x^3 + (a(\sigma^2 + \mu \eta P_c) + \mu) a x^2 + a^2 \eta P_c \sigma^2 x - a \eta P_c \sigma^2 = 0$, where $\mu = \pi \lambda f \xi_s$. From the equation, we can obtain the closed-form of $P^*_t$.

This proves Proposition 3.
Appendix D

Proof of Proposition 4

Denote $\delta_1(i, r)$ as the ratio of the data conveyed via D2D links to the file size $F$, which can be obtained as $\delta_1(i, r) = \min \left( R_1(i, r) \frac{\rho V_0 Q}{F(\frac{1}{2} P_t + P_c)}, 1 \right) = \min \left( \log_2 (1 + \gamma_1(i, r)) \frac{W \rho V_0 Q}{F(\frac{1}{2} P_t + P_c)}, 1 \right)$. From the definition, the offloading ratio can be obtained as

$$ p_i^O(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{E}_h[\delta_1(i, r)] \, dr $$

$$ = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[\delta_1(i, r) = 1] \, dr + \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[\delta_1(i, r) < 1] \mathbb{E}_h[\delta_1(i, r) \mid \delta_1(i, r) < 1] \, dr $$

(a)$$= \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[\delta_1(i, r) < 1] \mathbb{E}_h[\delta_1(i, r) \mid \delta_1(i, r) < 1] \, dr $$

$$ = p_1(P_t, \rho) + \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[\delta_1(i, r) < 1] \mathbb{E}_h[\delta_1(i, r) \mid \delta_1(i, r) < 1] \, dr, $$

(D.1)

where (a) comes from the fact that $\mathbb{P}[\delta_1(i, r) = 1] = \mathbb{P}[E_1(i, r) \leq \rho V_0 Q]$.

From the expression of $\delta_1(i, r)$ and $R_1(i, r)$, we have

$$ \mathbb{E}_h[\delta_1(i, r) \mid \delta_1(i, r) < 1] = \mathbb{E}_h \left[ \ln(1 + \gamma_1(i, r)) \frac{W \rho V_0 Q}{F(\frac{1}{2} P_t + P_c)} \mid \ln(1 + \gamma_1(i, r)) \frac{W \rho V_0 Q}{F(\frac{1}{2} P_t + P_c)} < 1 \right] $$

(a)$$= \mathbb{E}_h \left[ \frac{\ln(1 + \gamma_1(i, r))}{\ln(1 + \gamma_1(i, r))} \left\{ \ln(1 + \gamma_1(i, r)) \right\} < 1 \right] = \frac{1}{\ln(1 + \Gamma_1)} \mathbb{E}_h \left[ \ln(1 + \gamma_1(i, r)) \mid \ln(1 + \gamma_1(i, r)) < \ln(1 + \Gamma_1) \right], $$

where (a) is obtained by substituting $\ln(1 + \Gamma_1) = \frac{F(P_t + \eta P_c) \ln 2}{W \rho_2 V_0 n}$, and $\Gamma_1 = e^{-\frac{F(P_t + \eta P_c) \ln 2}{W \rho_2 V_0 n}} - 1$.

For a positive random variable $x$ with cdf $F(x)$ and pdf $f(x)$, we have

$$ \mathbb{E} [x \mid x < X_0] = \int_0^{X_0} x f(x) \frac{F(x)}{F(X_0)} \, dx = X_0 - \frac{1}{F(X_0)} \int_0^{X_0} \mathbb{P} [x < t] \, dt, $$

(D.2)

where (a) comes from the fact that the conditional pdf of $x$ given $x < X_0$ is $f(x) \frac{F(x)}{F(X_0)}$.

Then, we have

$$ \mathbb{E}_h \left[ \ln(1 + \gamma_1(i, r)) \mid \ln(1 + \gamma_1(i, r)) < \ln(1 + \Gamma_1) \right] $$

$$= \ln(1 + \Gamma_1) - \frac{1}{\mathbb{P}[h \leq \Gamma_1(\sigma_0^2 + 4 t \sigma_0^2 r^a) \geq 1]} \int_0^{\ln(1 + \Gamma_1)} \mathbb{P} [h < (e^t - 1)(\sigma_0^2 + I_t r^a)] \, dt $$

(a)$$= \ln(1 + \Gamma_1) - \frac{1}{\ln(1 + \Gamma_1)} \int_0^{\ln(1 + \Gamma_1)} 1 - e^{-\phi_i(e^t - 1, r)} \, dt, $$

(b)$$= \phi_i(x, y) = x y^\alpha - \pi (\lambda_1 \xi_1 - \lambda_2 \xi_2) y^2 x^{2/\alpha}.$
Therefore, we have
\[ \mathbb{E}_h [\delta_1(i, r) \mid \delta_1(i, r) < 1] = 1 - \frac{1}{1 - e^{-\phi_i(1, r)}} \int_0^{\ln(1 + \Gamma_1)} \frac{1 - e^{-\phi_i(e^{1-r} - 1, r)}}{\ln(1 + \Gamma_1)} \, dt. \] (D.4)

On the other hand, we can obtain
\[ \mathbb{P} [\delta_1(i, r) < 1] = \mathbb{P} [E_1(i, r) > \rho V_0 Q] = 1 - \mathbb{P} [E_1(i, r) \leq \rho V_0 Q] \overset{(a)}{=} 1 - e^{-\phi_i(\Gamma_1, r)}, \] (D.5)
where (a) is obtained according to (B.3). By substituting (D.5) and (D.4) into (D.1), we can obtain the expression of \( p^o_1(P_t, \rho) \) in Proposition 4.

From (D.1), we can show that \( p_1(P_t, \rho) \leq p^o_1(P_t, \rho) \). Considering \( \delta_1(i, r) \leq 1 \), we can obtain
\[ p^o_1(P_t, \rho) = \sum_{i=1}^{N_f} p^o_r(i) \int_0^{r_c} f_i(r) \mathbb{E}_h [\delta_1(i, r)] \, dr \leq \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \, dr = p_o. \] Then, \( p_1(P_t, \rho) \leq p^o_1(P_t, \rho) \leq p_o. \) When \( \phi_i(\Gamma_1, r) = 0 \), according to Proposition 2, \( p_1(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) e^0 \, dr = \sum_{i=1}^{N_f} p_r(i)(1 - e^{\lambda_i r^2}) \, dr = p_o \) and both equalities hold, where the conditions that lead to \( \phi_i(\Gamma_1, r) = 0 \) can be further derived as follows.

From the expression of \( \Gamma_1 \), we have \( \lim_{\rho \to \infty} \Gamma_1 = \lim_{\rho \to \infty} e^{\frac{\rho F(P_t + \rho \rho_v)^{1/2}}{V_0^2}} - 1 = 0 \). Hence, \( \lim_{\rho \to \infty} \phi_i(\Gamma_1, r) = \lim_{\rho \to \infty} \Gamma_1 r^2 \sigma_i^2 + \pi(\lambda_i \xi_1 - \lambda_i^d \xi_2)^2 r^2 \Gamma_1^{2/a} = 0. \) Since \( 0 \leq \lambda_i^d \leq \lambda_i, \lambda_i \to 0 \) leads to \( \lambda_i^d \to 0 \). Further considering that \( \text{SNR} = P_t r^{-\alpha} / \sigma^2 = 1 / r^a \sigma_0^2 \), we have \( \lim_{\text{SNR} \to \infty \lambda_i \to 0} \phi_i(\Gamma_1, r) = 0. \) Therefore, the upper bound of \( p^o_1(P_t, \rho) \) can be achieved when \( \rho \to \infty \), or when \( \text{SNR} \to \infty \) and \( \lambda_i \to 0 \).

**APPENDIX E**

**PROOF OF PROPOSITION 5**

The average energy consumed at a DT for complete transmission can be obtained as
\[ \bar{E}_1 = \sum_{i=1}^{N_f} p'_r(i) \int_0^{r_c} f_i(r) \mathbb{E}_h [E_1(i, r) \mid E_1(i, r) < \rho V_0 Q] \, dr, \] (E.1)
where \( p'_r(i) \) is the probability that the \( i \)th file is requested by a satisfied DR, \( f'_i(r) \) is the pdf of the D2D link distance for a DR that requests the \( i \)th file and is satisfied.

We can obtain \( p'_r(i) \) as \( p'_r(i) = \frac{p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q | \rho V_0 Q] \, dr}{p_r(P_t, \rho)} \overset{(a)}{=} \frac{p_r(i) \int_0^{r_c} f_i(r) e^{-\phi_i(\Gamma_1, r)} \, dr}{p_r(P_t, \rho)} \), where \( p_r(i) \int_0^{r_c} f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q] \, dr \) is the probability that a DR requests the \( i \)th file and can be satisfied, \( p_r(P_t, \rho) \) is the probability that a DR requesting any file can be satisfied, and (a) is obtained analogously to deriving (B.1). The cdf of the D2D distance for a satisfied DR that requests the \( i \)th file can be obtained as \( F'_i(R) = \mathbb{P}[r \leq R] = \frac{\int_0^R f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q] \, dr}{\int_0^{r_c} f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q] \, dr} = \frac{\int_0^R f_i(r) e^{-\phi_i(\Gamma_1, r)} \, dr}{\int_0^{r_c} f_i(r) e^{-\phi_i(\Gamma_1, r)} \, dr} \), where \( \int_0^{r_c} f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q] \, dr \) is the probability that a DR that desires the \( i \)th file can be
satisfied with a D2D transmission distance smaller than \(R\), and \(\int_0^{r_c} f_i(r) \mathbb{P}[E_1(i, r) \leq \rho V_0 Q] \, dr\) is the probability that a DR desiring the \(i\)th file can be satisfied. Then, the pdf \(f'_i(r)\) can be obtained as \(f'_i(r) = \frac{dF'_i(r)}{dr} = \frac{f_i(r)e^{-\phi_i(r)}}{f_0(r)e^{-\phi_i(0)}}\).

Considering that \(E_1(i, r) = \frac{1}{\ln(1 + \Gamma_1(r))}\), we have
\[
\mathbb{E}_h [E_1(i, r) | E_1(i, r) < \rho V_0 Q] = \ln(1 + \Gamma_1) \rho V_0 Q \mathbb{E}_h \left[ \frac{1}{\ln(1 + \gamma_1(i, r))} \middle| \frac{1}{\ln(1 + \gamma_1(i, r))} < \Gamma_1' \right],
\]
where \(\Gamma_1' = \frac{1}{\ln(1 + \Gamma_1)}\).

Moreover, the expectation in (E.2) can be derived as
\[
\mathbb{E}_h \left[ \frac{1}{\ln(1 + \gamma_1(i, r))} \middle| \frac{1}{\ln(1 + \gamma_1(i, r))} < \Gamma_1' \right] = \Gamma_1' - \int_{\Gamma_1'}^{\Gamma_1} \mathbb{P} \left[ \frac{1}{\ln(1 + \gamma_1(i, r))} < t \right] dt
\]
\[
= \frac{1}{\mathbb{P}[h > \Gamma_1'(\sigma_0^2 + I_{r,i})^{2\alpha}]} \int_{\Gamma_1'}^{\Gamma_1} \mathbb{P} \left[ h > (e^{\frac{t}{\alpha}} - 1)(\sigma_0^2 + I_{r,i})^{2\alpha} \right] dt
\]
\[
= \frac{1}{\mathbb{E}_h \left[ \frac{1}{\ln(1 + \gamma_1(i, r))} \middle| \frac{1}{\ln(1 + \gamma_1(i, r))} < \Gamma_1' \right]} \int_{\Gamma_1'}^{\Gamma_1} \mathbb{E}_h \left[ \frac{1}{\ln(1 + \gamma_1(i, r))} \middle| \frac{1}{\ln(1 + \gamma_1(i, r))} < t \right] dt,
\]
where (a) is obtained according to (D.2), (b) is obtained by substituting the definition of \(\gamma_1(i, r)\) in (7), (c) is because \(h\) follows an exponential distribution with unit mean, and (d) is because \(\phi_i(x, y) = xy^\alpha \sigma_0^2 - \pi(\lambda_1 \xi_1 - \lambda_1^d \xi_2) y^2 x^{2/\alpha}\).

By substituting \(p'_i(i), f'_i(r)\) and (E.3) into (E.2) and (E.1) and after some further manipulations, Proposition 5 follows.

**APPENDIX F**

**PROOF OF PROPOSITION 6**

By denoting \(A = r^\alpha \sigma_0^2\), \(a = \frac{F \ln 2}{W \rho Q V_0 \eta}\), and \(g(P_l) = \frac{A P_l}{P_t}\), the offloading probability can be expressed as
\[
p_2(P_t, \rho) = \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) e^{-g(P_l)} \, dr,
\]
where \(\Gamma_2 = e^{a(\rho + \eta P_t^\alpha)} - 1\).

To simplify the notation, we change the variable \(P_t \rightarrow x\). We can obtain the first-order derivative of \(g(x)\) as \(g'(x) = \frac{A}{x^2} (\frac{\partial P_t}{\partial x} x - \Gamma_2)\). Then, the first-order derivative of \(p_2(x, \rho)\) can be obtained as
\[
p_2'(x, \rho) = -\sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) g'(x) e^{-g(x)} \, dr = -g_1(x) \sum_{i=1}^{N_f} p_r(i) \int_0^{r_c} f_i(r) \frac{A}{x^2} e^{-g(x)} \, dr,
\]
where \( g_1(x) = \frac{d^2}{dx^2} x - \Gamma_2 \). The first-order derivative of \( g_1(x) \) with respect to \( x \) can be obtained as \( g'_1(x) = \frac{d}{dx} x = a^2 e^{a(x+\eta P_c^T)} x \geq 0 \).

Therefore, \( g_1(x) \) is an increasing function, where \( g_1(0) = 1 - e^{a\eta P_c^T} < 0 \) and \( g_1(+\infty) \to +\infty \). Then, from (F.2), we can see that \( p_2(P_t, \rho) \) first increases and then decreases, and the optimal \( x \) to maximize \( p_2(x, \rho) \) can be obtained by solving \( g_1(x) = 0 \). Again, considering that \( a(P_{\text{max}} + \eta P_c^T) \ll 1 \) and using the approximation \( e^t \approx 1 + t \) that is accurate when \( t \ll 1 \), \( g_1(x) \) can be derived from (F.2) as

\[
g_1(x) = e^{a(x+\eta P_c^T)}(ax-1) + 1 \approx (a(x+\eta P_c^T) + 1)(ax-1) + 1 = a^2(x^2 + \eta P_c^T(x-\frac{1}{a})). \tag{F.3}
\]

Then, from \( g_1(x) = 0 \) and \( 0 < x \leq P_{\text{max}} \), the optimal transmit power can be obtained as

\[
P^*_t = \begin{cases} P_{\text{max}}, & P_{\text{max}} < \eta P_c^T \left(\sqrt{\frac{1}{a\eta P_c^T} + \frac{1}{4}} - \frac{1}{2}\right), \\ \eta P_c^T \left(\sqrt{\frac{1}{a\eta P_c^T} + \frac{1}{4}} - \frac{1}{2}\right), & \text{otherwise} \end{cases} \tag{F.4}
\]

This proves Proposition 6.

REFERENCES

[1] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, “Device-to-device communication as an underlay to LTE-advanced networks,” IEEE Commun. Mag., vol. 47, no. 12, pp. 42–49, 2009.

[2] X. Lin, J. Andrews, A. Ghosh, and R. Ratasuk, “An overview of 3GPP device-to-device proximity services,” IEEE Commun. Mag., vol. 52, no. 4, pp. 40–48, 2014.

[3] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, “Exploring social ties for enhanced device-to-device communications in wireless networks,” IEEE GLOBECOM, 2013.

[4] S. Andreev, O. Galinina, A. Pyattaev, K. Johnsson, and Y. Koucheryavy, “Analyzing assisted offloading of cellular user sessions onto D2D links in unlicensed bands,” IEEE J. Sel. Areas Commun., vol. 33, no. 1, pp. 67–80, 2015.

[5] A. Asadi, Q. Wang, and V. Mancuso, “A survey on device-to-device communication in cellular networks,” IEEE Commun. Surveys Tuts., vol. 16, no. 4, pp. 1801–1819, 2014.

[6] Cisco Visual Networking, “Global mobile data traffic forecast update 2014-2019,” White Papers, 2015.

[7] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, “Cache in the air: exploiting content caching and delivery techniques for 5G systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 131–139, 2014.

[8] E. Bastug, M. Bennis, and M. Debbah, “Living on the edge: The role of proactive caching in 5G wireless networks,” IEEE Commun. Mag., vol. 52, no. 8, pp. 82–89, Aug. 2014.

[9] M. A. Maddah-Ali and U. Niesen, “Fundamental limits of caching,” IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2856–2867, 2014.

[10] D. Liu and C. Yang, “Energy efficiency of downlink networks with caching at base stations,” IEEE J. Sel. Areas Commun., to appear.

[11] B. D. Higgins, J. Flinn, T. J. Giuliani, B. Noble, C. Peplin, and D. Watson, “Informed mobile prefetching,” ACM MobiSys, 2012.
[12] K. Wang, Z. Chen, and H. Liu, “Push-based wireless converged networks for massive multimedia content delivery,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2894–2905, 2014.

[13] B. Chen and C. Yang, “Performance gain of precaching at users in small cell networks,” *IEEE PIMRC*, 2015.

[14] 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 Commun. Mag.*, vol. 51, no. 4, pp. 142–149, 2013.

[15] N. Golrezaei, P. Mansourifard, A. Molisch, and A. 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, 2014.

[16] J. Jiang, S. Zhang, B. Li, and B. Li, “Maximized cellular traffic offloading via device-to-device content sharing,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 82–91, 2016.

[17] M. Ji, G. Caire, and A. Molisch, “Wireless device-to-device caching networks: Basic principles and system performance,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, 2016.

[18] D. Malak and M. Al-Shalash, “Optimal caching for device-to-device content distribution in 5G networks,” *IEEE GLOBECOM*, 2014.

[19] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web caching and Zipf-like distributions: Evidence and implications,” *IEEE INFOCOM*, 1999.

[20] D. Stoyan, W. S. Kendall, J. Mecke, and L. Ruschendorf, *Stochastic geometry and its applications*. Wiley New York, 1987, vol. 2.

[21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[22] N. B. Mehta, J. Wu, A. F. Molisch, and J. Zhang, “Approximating a sum of random variables with a lognormal,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2690–2699, 2007.

[23] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, vol. 55.

[24] H. Kim and G. de Veciana, “Leveraging dynamic spare capacity in wireless systems to conserve mobile terminals’ energy,” *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 802–815, 2010.

Stirling formula, *Encyclopaedia of Mathematics*. [Online]. Available: [http://www.encyclopediaofmath.org/index.php?title=Stirling_formula&oldid=13618](http://www.encyclopediaofmath.org/index.php?title=Stirling_formula&oldid=13618)

[25] J. Andrews, F. Baccelli, and R. Ganti, “A tractable approach to coverage and rate in cellular networks,” *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.

[26] J.-S. Ferenc and Z. Néda, “On the size distribution of poisson voronoi cells,” *Physica A: Stat. Mechanics and its App.*, vol. 385, no. 2, pp. 518–526, 2007.

[27] S. Lee and K. Huang, “Coverage and economy of cellular networks with many base stations,” *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1038–1040, 2012.

[28] H. S. Lichte, S. Valentin, and H. Karl, “Expected interference in wireless networks with geometric path loss: a closed-form approximation,” *IEEE Commun. Lett.*, vol. 14, no. 2, pp. 130–132, 2010.