Throughput-Outage Scaling Laws for Wireless Single-Hop D2D Caching Networks with Physical Models

Ming-Chun Lee, *Member, IEEE*, Andreas F. Molisch, *Fellow, IEEE*, and Mingyue Ji, *Member, IEEE*

***Abstract*—Throughput-Outage scaling laws for single-hop cache-aided device-to-device (D2D) communications have been extensively investigated under the assumption of the protocol model. However, the corresponding performance under physical models has not been explored; in particular it remains unclear whether link-level power control and scheduling can improve the asymptotic performance. This paper thus investigates the asymptotic throughput-outage tradeoff and derives its outer bound for cache-aided D2D networks under two common physical models. The results show that the asymptotic performance of the network under physical models is identical to that under the protocol model when requests are served with equal quality. This indicates that the throughput-outage performance cannot be improved asymptotically by using link-level power control and scheduling.**

I. Introduction

As video traffic demand has rapidly increased, satisfying such demand has been one of the major challenges of wireless networks [1]. One of the most promising approaches, caching at the wireless edge, exploits the unique video accessing pattern of users and cheap storage to trade off bandwidth against memory [1] and has drawn significant attention [2].

Due to the ability to improve video services without the need of newly installed infrastructure, cache-aided wireless D2D networks has been one of the most popular implementations of caching at the wireless edge, and drawn significant attentions [1]. While many papers investigate the designs and implementations of cache-aided D2D networks, another set of investigations concentrates on the scaling laws of the networks, which characterize the asymptotic behaviors as number of users N tend to infinity in order to understand the fundamental limits and benefits of such networks. For example, scaling laws of cache-aided multi-hop D2D networks with and without hierarchical cooperation were investigated in [3]–[6]. Furthermore, scaling laws of single-hop cache-aided networks were studied in [7]–[9]. Specifically, [7] investigated the maximum expected throughput scaling law without considering the outage probability. Considering the outage probability, [8] analyzed the throughput-outage performance of the networks and provided both the achievable performance and outer bound under protocol model [10] and the Zipf popularity distribution. Following [8], [9] derived the throughput-outage analysis for the more general Mandelbrot-Zipf (MZipf) popularity distribution.

The throughput-outage scaling law analyses for single-hop cache-aided D2D networks have been conducted mostly for the protocol model. However, this model may be oversimplified, since it does not incorporate the link-level power control and scheduling into the analysis. A more realistic model is the physical (interference) model [10], under which suitable scheduling and power control algorithms (usually for a given caching distribution) have been designed, and performance in finite-size networks has been investigated, usually by simulation, [11], [12]. However, to the best of our knowledge, the scaling laws for physical models have not been explored yet, and it is unclear whether the link-level power control and scheduling can further improve the scaling laws derived under protocol model. This paper thus aims to address this question.

This paper considers single-hop cache-aided D2D networks with MZipf popularity distribution whose Zipf factor is smaller than 1 and conducts the throughput-outage scaling law analysis for two common physical models. Specifically, by exploring the relationship between the communication distances of users and the throughput-outage performance, we derive the throughput-outage outer bounds of the networks for both physical models in the regime that the outage probability is negligibly small or converging to zero when the numbers of users in the network and files in the library goes to infinity, which corresponds to the practical consideration that the outage probability should be small. The results show that when the user powers are finite, the throughput-outage performance under physical models is asymptotically identical to those derived in [8] and [9] under the protocol model. This indicates that the performance of single-hop cache-aided D2D networks cannot be asymptotically improved by using link-level power control and scheduling when requests are served with equal quality, though the constant factor might be improved significantly in practice.

II. Network Model

We consider a random dense network where users are placed according to a binomial point process (BPP) within a unit square-shaped area $[0,1] \times [0,1]$. Accordingly, we assume that the number of users in the network is N, and users

M.-C. Lee is with Institute of Communications Engineering, National Yang Ming Chiao Tung University (National Chiao Tung University), Hsinchu 30010, Taiwan (email: mingchunlee@nctu.edu.tw).

A. F. Molisch is with Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, USA (email: molisch@usc.edu).

M. Ji is with Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112, USA (email: mingyue.ji@utah.edu).

This work was supported in part by the National Science Foundation (NSF) under grants CNS-1816699, CCF-1817154 and SpecEES- 1824558, and by the Ministry of Science and Technology (MOST) of Taiwan under grant MOST 109-2218-E-009-002.

are distributed uniformly at random within the network. We assume each device in the network can cache S files and each file has a equal size. We consider a library consisting of M files. We assume that users request the files from the library independently according to a request distribution modeled by the MZipf distribution [13]:

$$P_r(f) = \frac{(f+q)^{-\gamma}}{\sum_{m=1}^{M}(m+q)^{-\gamma}}, \quad (1)$$

where γ is the Zipf factor and q is the plateau factor of the distribution. We can see that the MZipf distribution degenerates to a Zipf distribution when $q = 0$. We consider the decentralized random caching policy for all users [14], in which users cache files independently according to the same caching policy. Denoting $P_c(f)$ as the probability that a user caches file f, the caching policy is fully described by $P_c(1), P_c(2), ..., P_c(M)$, where $0 \leq P_c(f) \leq 1, \forall f$; thus users cache files according to the caching policy $\{P_c(f)\}_{f=1}^{M}$. To satisfy the cache space constraint, we have $\sum_{f=1}^{M} P_c(f) = S$. Note that by using the random caching policy in [14], such constraint can result in the use of the cache space being exactly S. In this paper, we assume that S and γ are some constants.

We consider the asymptotic analysis in this paper, in which we assume that $N \to \infty$, $M \to \infty$, and $q \to \infty$ [9]. We consider $\gamma < 1$ and restrict to $M = o(N)$ and $q = \mathcal{O}(M)$. The main reason for restricting to $M = o(N)$ is to render the users of the network the sufficient ability to cache the whole library. Besides, we assume $q \to \infty$ because q is asymptotically influential only if $q \to \infty$. Furthermore, when q goes to infinity, it is sufficient to consider $q = \mathcal{O}(M)$. This is because the MZipf distribution would behave like a uniform distribution asymptotically as $q = \omega(M)$ which is less interesting.

We consider single-hop D2D delivery for the network. Users can obtain their desired files through only single-hop D2D delivery. Note that since S is a constant, the probability that a user can find the desired file from its own cache goes to zero as q and M go to infinity. This avoids the trivial gain from self-caching in terms of asymptotic performance. Moreover, similar to [8], we assume that different users making the requests on the same file would request different segments of the file, which avoids the gain from naive multicasting. Also, we assume file segments delivered to users have the same size.

We define an outage as an occurrence where a user cannot obtain its desired file within one hop. Suppose we are given a realization of the placement of the user locations P according to the binomial point process. In addition, we are given a realization of file requests F and a realization of file placement G of users according to the popularity distribution $P_r(\cdot)$ and caching policy $P_c(\cdot)$, respectively. Denote $\mathcal{U}$ as the index set of users. We can then define T_u as the throughput of user $u \in \mathcal{U}$ under a feasible single-hop file delivery scheme:

$$T_u = \frac{1}{T}\sum_{t=1}^{T} c_u A_u(t), \quad (2)$$

where T is the number of time-slots for the transmission, c_u is the link rate of user u, and $A_u(t)$ is the link activation indicator of user u at time-slot t, where $A_u(t) = 1$ if the link of user u is scheduled at time-slot t; otherwise $A_u(t) = 0$. We then define the average throughput of user u as $\overline{T}_u = \mathbb{E}_{\mathsf{P},\mathsf{F},\mathsf{G}}[T_u]$, where the expectation is taken over the placement of user locations P, file requests F of users, the file placement of users G, and the file delivery scheme. Finally, we define the average throughput of a user in the network as

$$T = \min_{u \in \mathcal{U}} \overline{T}_u. \quad (3)$$

When the number of users in the network is N, we define

$$N_o = \sum_{u \in \mathcal{U}} \mathbf{1}\{\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}] = 0\} \quad (4)$$

as the number of users in outage, where $\mathbf{1}\{\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}] = 0\}$ is the indicator function such that the value is 1 if $\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}] = 0$; otherwise the value is 0. Intuitively, $\mathbf{1}\{\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}] = 0\}$ is equal to zero when the file delivery scheme cannot deliver the desired file to user u. We note that the expectation of $\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}]$ is taken over the file delivery scheme. The outage probability is then defined as

$$p_o = \frac{1}{N}\mathbb{E}_{\mathsf{P},\mathsf{F},\mathsf{G}}[N_o] = \frac{1}{N}\sum_{u \in \mathcal{U}} \mathbb{P}\left(\mathbb{E}[T_u \mid \mathsf{P},\mathsf{F},\mathsf{G}] = 0\right). \quad (5)$$

In this paper, we focus on the regime that the outage probability is either negligibly small or converging to zero when N and M go to infinity. This corresponds to the consideration that the outage probability should be small in practice.

A. Channel Models

We consider two types of physical models. Suppose there is a transmitter (TX)-receiver (RX) pair u, where user u is serving as the TX and user $u^{(\mathrm{r})}$ is serving as the RX of user u. We denote x_u and $x_{u^{(\mathrm{r})}}$ as the locations of user u and $u^{(\mathrm{r})}$, respectively. We denote Γ_{Co} as the set of users transmitting at the same time-frequency resource of user u. Then, the first physical model, referred to as "bounded physical model", defines the link rate for the TX-RX pair u as [10], [15]:

$$R(u,u^{(\mathrm{r})}) = \begin{cases} B_u \log_2(1+\vartheta), & C(u,u^{\mathrm{r}}) \geq \vartheta \\ 0, & C(u,u^{\mathrm{r}}) < \vartheta \end{cases}, \quad (6)$$

where

$$C(u,u^{\mathrm{r}}) = \log_2\left(1 + \frac{P_u l_{uu^{(\mathrm{r})}}}{B_u N_0 + \sum_{k\neq u, k\in\Gamma_{\mathrm{Co}}} P_k l_{ku^{(\mathrm{r})}}}\right); \quad (7)$$

ϑ is some constant according to the delivery mechanism; N_0 is the noise power spectral density; B_u is the bandwidth used for communication between users u and $u^{(\mathrm{r})}$; P_u is the power of user u; and $l_{uu^{(\mathrm{r})}} = \frac{\chi}{\left(d_{uu^{(\mathrm{r})}}\right)^{\alpha}}$ is the path (power) gain between users u and $u^{(\mathrm{r})}$, where $d_{uu^{(\mathrm{r})}} = |x_u - x_{u^{(\mathrm{r})}}|$ is the distance between users u and $u^{(\mathrm{r})}$, $\chi > 0$ is some calibration factor, and $\alpha > 2$ is the pathloss coefficient.

The second physical model, referred to as "generalized physical model", defines the link rate for the TX-RX pair u as [10], [15]:

$$R(u,u^{(\mathrm{r})}) = B_u \log_2\left(1 + \frac{P_u l_{uu^{(\mathrm{r})}}}{B_u N_0 + \sum_{k\neq u, k\in\Gamma_{\mathrm{Co}}} P_k l_{ku^{(\mathrm{r})}}}\right). \quad (8)$$

III. Throughput–Outage Scaling Law Outer Bound Analysis with Bounded Physical Model

In this section, we provide the throughput-outage outer bound under the bounded physical model. To derive the outer bound, we first provide the following two lemmas:

Lemma 1 (Lemma 4 in [5]): When $N = \omega(M)$ users are uniformly distributed within a network with unit size, the probability to have $N_{\rm D}$ users within an area of size $A = o\left(\frac{N_{\rm D}}{N}\right)$ is upper bounded by $o(1)$.

Lemma 2 (Lemma 5 in [5]): Suppose $\gamma < 1$. We then have the following results: when a user in the network searches through $n_{\rm s} = o\left(\frac{M}{S}\right)$ different users, we obtain $p_{\rm miss}(n_{\rm s}) \geq 1 - o(1)$. Furthermore, when a user in the network searches through $n_{\rm s} = \rho' M$ different users for some ρ', we have the following results: (i) $p_{\rm miss}(n_{\rm s}) \geq \epsilon_1'(\rho')$ if $\rho' = \Theta(1)$, where $\epsilon_1'(\rho')$ can be arbitrarily small as ρ' is large enough; and (ii) $p_{\rm miss}(n_{\rm s}) \geq (1-\gamma)e^{-(S\rho'-\gamma)}$ if $\rho' = \omega(1)$.

From Lemmas 1 and 2, we conclude that to have a non-vanishing probability for a user to obtain the desired file (i.e., $p_{\rm miss}(n)$ does not go to 1), the distance between the source and destination is at least $\Theta\left(\sqrt{\frac{\rho' M}{SN}}\right)$ with high probability (w.h.p). As a result, we have the following theorem:

Theorem 1: Let $M \to \infty$, $N \to \infty$, and $q \to \infty$. Suppose $\gamma < 1$ and $q = \mathcal{O}(M)$. Consider the bounded physical model. Then, when $\rho' = \Omega(1)$ is large enough, the throughput-outage performance of the network is dominated by:

$$T(P_o) = \Theta\left(\frac{S}{\rho' M}\right), P_o = \Theta(e^{-\rho'}). \tag{9}$$

Proof. See Appendix A. □

Remark 1: Theorem 1 shows that when $\gamma < 1$, M is the dominant factor while q does not impact the asymptotic scaling law. In addition, it shows that the throughput-outage performance bound considering the bounded physical model has the same scaling law as the throughput-outage performance considering the protocol model [8], [9]. Note that in contrast to the physical model here which enables the link-level power allocation and scheduling, the protocol model and the approaches in [8], [9] only consider the simple clustering network and the system-level changing of the cluster size. As a result, this indicates that the link-level power allocation and scheduling cannot improve the scaling law, i.e., the growth rate, of the throughput-outage performance asymptotically when requests of users are served with equal quality. However, in practice, the constant factor of the throughput-outage performance might be improved by a good power control and link scheduling approach.

IV. Throughput–Outage Scaling Law Outer Bound Analysis with Generalized Physical Model

Here, we provide the throughput-outage outer bound analysis under the generalized physical model. We start the analysis by providing Theorem 2 which describes a transport capacity upper bound.

Theorem 2: We denote the set of TX-RX pairs as Γ and define r_u as the communication distance for the TX-RX pair u. Let $M \to \infty$, $N \to \infty$, and $q \to \infty$. Suppose $\gamma < 1$ and $q = \mathcal{O}(M)$. We let $R_0 = \epsilon_0\sqrt{\frac{\rho' M}{SN}}$, where ϵ_0 is some small constant. We denote the transport capacity of the network consisting of Γ, defined in terms of meter-bits/s, as $C_\Gamma = \sum_{u\in\Gamma} r_u C_u$ in T seconds, where C_u is the average throughput (bits/s) of user u. We denote the set of TX-RX pairs which have the largest powers among the TX-RX pairs in their corresponding time-frequency resources as $\mathcal{W}$. Under the generalized physical model, C_Γ is upper bounded as:

$$C_\Gamma \leq B\overline{C}_{\mathcal{W}} + B\overline{C}_{\Gamma_{R_0}} + B\frac{\log_2(e)}{\epsilon_0}\sqrt{\frac{SN}{\rho' M}}\left(\alpha\left(3\sqrt{2}+1\right) + 2(2(\sqrt{2}+1))^\alpha\right), \tag{10}$$

where B is the total bandwidth of the network; $\overline{C}_{\mathcal{W}}$ is the average transport capacity efficiency, defined in terms of meter-bits per second per Hz, of the TX-RX pairs in $\mathcal{W}$; $\overline{C}_{\Gamma_{R_0}}$ is the average transport capacity efficiency of TX-RX pairs that are not in $\mathcal{W}$ and have communication distances smaller than R_0; $P_{\rm max}$ is the maximum power that a user can use for transmission; $B_{\rm s}$ and τ are the minimum amount of bandwidth and time duration for a TX-RX pair to be scheduled to communicate. The definitions of $\overline{C}_{\mathcal{W}}$ and $\overline{C}_{\Gamma_{R_0}}$ are formally given below (28) at the end of Appendix B.

Proof. See Appendix B. □

To have a negligibly small outage probability, according to Lemmas 1 and 2, we need the TX-RX pairs to have communication distances $r_u = \Theta\left(\sqrt{\frac{\rho' M}{SN}}\right)$ with high probability. This indicates that the average distance of a bit that is transported in the network is $\overline{L} = \Theta\left(\sqrt{\frac{\rho' M}{SN}}\right)$. From (10), we first observe that the upper bound can be optimized by maximizing $\overline{C}_{\mathcal{W}}$ and $\overline{C}_{\Gamma_{R_0}}$. When optimizing $\overline{C}_{\mathcal{W}}$, we aim to have larger $r_w, \forall w \in \mathcal{W}$. In addition, we know that, w.h.p., the communication distances of users are $r_u = \Theta\left(\sqrt{\frac{\rho' M}{SN}}\right), \forall u \in \Gamma$. It follows that we can assume without loss of optimality $r_w = \Theta\left(\sqrt{\frac{\rho' M}{SN}}\right), \forall w \in \mathcal{W}$. Therefore, we obtain:

$$C_\Gamma \leq B\Theta\left(\sqrt{\frac{\rho' M}{SN}}\log_2\left(1 + \frac{P_{\rm max}}{N_0 B_{\rm s}}\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}\right)\right) + B\overline{C}_{\Gamma_{R_0}} + B\frac{\log_2(e)}{\epsilon_0}\sqrt{\frac{SN}{\rho' M}}\left(\alpha\left(3\sqrt{2}+1\right) + 2(2(\sqrt{2}+1))^\alpha\right), \tag{11}$$

We then denote λ as the average throughput per user (bits/s) in T sec; $\Gamma_{\rm tot}^{R_0} = \{u : |x_u - x_{u^{\rm r}}| < R_0, u \in \Gamma \setminus \mathcal{W}\}$; and let

$$C_{P_{\rm max}} = B\Theta\left(\sqrt{\frac{\rho' M}{SN}}\log_2\left(1 + \frac{P_{\rm max}}{N_0 B_{\rm s}}\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}\right)\right). \tag{12}$$

By definition, we thus obtain

$$\begin{aligned} C_\Gamma &= \lambda N\overline{L} = \lambda N \Theta\left(\sqrt{\frac{\rho' M}{SN}}\right) = \lambda \Theta\left(\sqrt{\frac{\rho' M N}{S}}\right) \\ &\overset{(a)}{\leq} C_{P_{\max}} + \sum_{u\in\Gamma_{\text{tot}}^{R_0}} \epsilon_0 \sqrt{\frac{\rho' M}{SN}}\lambda \\ &+ B\frac{\log_2(e)}{\epsilon_0}\sqrt{\frac{SN}{\rho' M}}\left(\alpha\left(3\sqrt{2}+1\right)+2(2(\sqrt{2}+1))^\alpha\right) \end{aligned} \tag{13}$$

where (a) is because $r_u < R_0, \forall u \in \Gamma_{\text{tot}}^{R_0}$ and segments delivered for users have the same size. It follows that

$$\begin{aligned} \lambda \leq & C_{P_{\max}}\sqrt{\frac{S}{\rho' M N}} + \lambda\epsilon_0\frac{1}{N}\sum_{u\in\Gamma_{\text{tot}}^{R_0}} 1 \\ & + B\frac{\log_2(e)}{\epsilon_0}\frac{S}{\rho' M}\left(\alpha\left(3\sqrt{2}+1\right)+2(2(\sqrt{2}+1))^\alpha\right). \end{aligned} \tag{14}$$

Suppose the number of users having communication distances smaller than R_0 is ϵN, where $\epsilon = \mathcal{O}(1) < 1$. This leads to

$$\begin{aligned} \lambda \leq & \frac{B}{(1-\epsilon_0\epsilon)}\frac{1}{N}\log_2\left(1+\frac{P_{\max}}{N_0 B_{\text{s}}}\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}\right) \\ & + \frac{B\log_2(e)\left(\alpha\left(3\sqrt{2}+1\right)+2(2(\sqrt{2}+1))^\alpha\right)}{\epsilon_0(1-\epsilon_0\epsilon)}\frac{S}{\rho' M}. \end{aligned} \tag{15}$$

Recall that, w.h.p., the communication distances of TX-RX pairs is $\Theta\left(\sqrt{\frac{\rho' M}{SN}}\right)$ and ϵ is small when ϵ_0 is small. Thus, $(1-\epsilon_0\epsilon)$ should be some constant. Then, from (15) and Lemma 2, we obtain the following theorem:

Theorem 3: Let $M \to \infty$, $N \to \infty$, and $q \to \infty$. Suppose $\gamma < 1$. We consider the generalized physical model. Then, when $\rho' = \Omega(1)$ is large enough, the throughput-outage performance of the network is dominated by (16) on the bottom of this page.

Remark 2: Note that when $N \to \infty$, $T(P_o)$ can be unbounded because the term $\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}$ in $T(P_o)$ in Theorem 3 can go to infinity. This unreasonable result is brought by having the cases that the signal-to-interference-plus noise ratio (SINR) becomes unbounded due to the unbounded path gain. Thus, to correctly interpret the result in Theorem 3, we should consider that the term $\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}$ in $T(P_o)$ in Theorem 3 is upper bounded by 1, which corresponds to the physical reality of the pathloss laws.

Remark 3: Identical to Theorem 1, Theorem 3 shows that if the maximum transmit power is some constant, the throughput-outage performance bound considering the generalized physical model has the same scaling law as the throughput-outage performance considering the protocol model [8], [9]. This again indicates that the link-level power allocation and scheduling cannot improve the the scaling law of the throughput-outage performance asymptotically. However, when we allow the maximum instantaneous transmit power to be infinity while the average power is still some constant, Theorem 3 suggests that the asymptotic performance might be improved if $\frac{S}{\rho' M} = o(\frac{\log_2(N)}{N})$. Such improvement could be possible if we let a user to exclusively transmit with the power being $\Theta(N)$ once every $\Theta(N)$ time-slots.

V. Conclusions and Discussion

This paper investigates the asymptotic throughput-outage performance for cache-aided D2D networks under two common physical models. The results show that the asymptotic performance of the network under physical models is identical to that under protocol model. This indicates that the throughput-outage performance cannot be changed asymptotically by using link-level power control and scheduling.

Our results are derived under the assumptions that different TX-RX instances are forced to transmit at the same rate and that the users are uniformly distributed within the network. Therefore, our analysis can be considered as the *statistically worst-case* analysis, and different scaling laws might be derived if different assumptions are considered.

Appendix A
Proof of Theorem 1

Proof. We first note that since ϑ is a constant, the link rate is finite and upper bounded. Therefore, there is no benefit of increasing the power to infinity. Besides, we need the transmit power to be larger than some constant to have link rate larger than zero. As a result, without loss of optimality, we can assume that the powers of users in the network is upper bounded by some constant $\nu_{\text{upp}} = \Theta(1)$ and lower bounded by some constant $\nu_{\text{low}} = \Theta(1)$, i.e., $\nu_{\text{low}} \leq P_u \leq \nu_{\text{upp}}, \forall u$. Then, we consider a TX-RX pair u. From (6), we can obtain:

$$\frac{P_u \frac{\chi}{d_{uu^{(\text{r})}}^\alpha}}{N_0 + \sum_{k\neq u, k\in\Gamma_{\text{Co}}} P_k \frac{\chi}{d_{ku^{(\text{r})}}^\alpha}} \geq 2^\vartheta - 1 = \beta, \tag{17}$$

where Γ_{Co} is the set of TX-RX pairs that communicate at the same time-frequency resource as the TX-RX pair u and $\beta = \Theta(1)$ is some constant. Since $N_0 > 0$ and $\nu_{\text{low}} \leq P_u \leq \nu_{\text{upp}}, \forall u$, we have:

$$\frac{\nu_{\text{upp}}|x_u - x_{u^{(r)}}|^{-\alpha}}{\sum_{k\neq u, k\in\Gamma_{\text{Co}}} \nu_{\text{low}}|x_k - x_{u^{(r)}}|^{-\alpha}} \geq \beta. \tag{18}$$

Then, suppose we consider only interfering TXs k that satisfy $|x_k - x_{u^{(r)}}| = \mathcal{O}\left(\sqrt{\frac{\rho' M}{SN}}\right)$ and denote the set of these TXs as Γ_{Co}^u. Besides, by Lemmas 1 and 2, we know that to have

$$T(P_o) = \Theta\left(\frac{B}{N}\log_2\left(1+\frac{P_{\max}}{N_0 B_{\text{s}}}\frac{\chi}{\left(\frac{\rho' M}{SN}\right)^{\frac{\alpha}{2}}}\right) + B\log_2(e)\left(\alpha\left(3\sqrt{2}+1\right)+2(2(\sqrt{2}+1))^\alpha\right)\frac{S}{\rho' M}\right), P_o = \Theta\left(e^{-\rho'}\right). \tag{16}$$

$$C_{\Gamma_{st}} = \sum_{u\in\Gamma_{st}} r_u c_u = \sum_{u\in\Gamma_{st}} r_u \tau B_{\mathrm{s}} \log_2\left(1+\frac{P_u l_{uu^{(\mathrm{r})}}}{N_0 B_{\mathrm{s}} + \sum_{k\in\Gamma_{st},k\neq u} P_k l_{ku^{(\mathrm{r})}}}\right)$$
$$\leq r_{w_{st}}\tau B_{\mathrm{s}}\log_2\left(1+\frac{P_{\max} l_{w_{st}w_{st}^{(\mathrm{r})}}}{N_0 B_{\mathrm{s}}}\right) + \tau B_{\mathrm{s}}\left[\sum_{u\in\Gamma_{st}^{R_0}} r_u c_u + \log_2(e)\sum_{j>0}\sum_{u\in\Gamma_{st}^{R_j}} r_u \log_e\left(1+\frac{P_u l_{uu^{(\mathrm{r})}}}{\sum_{k\in\Gamma_{st},k\neq u} P_k l_{ku^{(\mathrm{r})}}}\right)\right]. \quad (20)$$

the outage probability $p_o = \Theta(e^{-\rho'})$, the distance between the TX and RX is $\Theta\left(\sqrt{\frac{\rho' M}{SN}}\right)$ with high probability. It follows that for a pair u, w.h.p., we have:

$$\frac{\nu_{\mathrm{upp}}\Theta\left(|\frac{\rho' M}{SN}|^{-\alpha/2}\right)}{\sum_{k\in\Gamma_{\mathrm{Co}}^u}\nu_{\mathrm{low}}\Theta\left(|\frac{\rho' M}{SN}|^{-\alpha/2}\right)} \geq \beta. \quad (19)$$

Therefore, we need $|\Gamma_{\mathrm{Co}}^u| \leq \beta_{\mathrm{number}} = \Theta(1)$ to let the TX-RX pair u to have a non-zero link rate. Finally, if we split the network into equally-sized square clusters in which the side length of each cluster is $d = \Theta\left(\sqrt{\frac{\rho' M}{Sn}}\right)$, we then have in total a number $\Theta\left(\frac{Sn}{\rho' M}\right)$ of such clusters. Then, by the above arguments, we know that in each cluster, we can at most have $\beta_{\mathrm{number}}+1$ different TX-RX pairs to communicate. Therefore, the sum throughout of the network is at most $(\beta_{\mathrm{number}}+1)\Theta\left(\frac{SN}{\rho' M}\right)$. It follows that the throughput per user is $\Theta\left(\frac{S}{\rho' M}\right)$, where the corresponding outage probability is characterized as in Lemma 2. □

Appendix B
Proof of Theorem 2

Proof. Suppose that the T-second duration of the transmission is split into different time-slots, in which the duration of a time-slot is τ sec. We also assume that the bandwidth is equally split into H different sub-channels. We denote the bandwidth of a sub-channel as B_{s} and denote the set of TX-RX pairs that transmit in sub-channel s and time-slot t as Γ_{st}. Note that a TX-RX pair might be scheduled in more than one time-frequency resource. The total transport capacity for TX-RX pairs in sub-channel s in time-slot t, defined in bit-meters, is provided in (20) on the top of this page, where w_{st} is the user who has the largest transmission power in Γ_{st}, i.e., $P_{w_{st}} \geq P_u, \forall u \in \Gamma_{st}$; $\Gamma'_{st} = \Gamma_{st}\setminus\{w_{st}\}$; and

$$\Gamma_{st}^{R_j} = \{u : R_{j-1} \leq r_u < R_j, u \in \Gamma'_{st}\}, \quad (21)$$

where $R_j = \epsilon_0\sqrt{\frac{\rho' M}{SN}}\cdot 2^j, \forall j \in \mathbb{N}$. We now want to compute

$$C_{R_j} = \sum_{u\in\Gamma_{st}^{R_j}} r_u \log_e\left(1+\frac{P_u l_{uu^{(\mathrm{r})}}}{\sum_{k\in\Gamma_{st},k\neq u} P_k l_{ku^{(\mathrm{r})}}}\right), \forall j > 0. \quad (22)$$

To do this, we split the network into equally-sized square clusters whose size length is R_j. Then, we collect all clusters that contain at least one RX of the TX-RX pairs in $\Gamma_{st}^{R_j}$ and denote the set of such clusters as Φ_j. Consequently, $|\Phi_j| \leq \frac{1}{R_j^2}$. We denote the gth cluster in Φ_j as ϕ_g^j, where $1 \leq g \leq |\Phi_j|$ and denote the number of RXs located in ϕ_g^j as n_g^j. We denote the communication distance corresponding to RX h_g^j in ϕ_g^j as r_{gh}^j; denote the transmit power of the TX corresponding to RX h_g^j in ϕ_g^j as P_{gh}^j, where $h = 1, 2, ..., n_g^j$; denote I_{gh}^j as the total interference at RX h_g^j in ϕ_g^j; and let the indices of the RXs in ϕ_g^j follow the descending order of the transmit powers of their corresponding TX, i.e, $P_{gh}^j \geq P_{gh'}^j, \forall h > h'$. Then, we know:

$$C_{R_j} \leq \underbrace{\sum_{g=1}^{|\Phi_j|} r_{g1}^j \log_e\left(1+\frac{P_{g1}^j\frac{\chi}{(r_{g1}^j)^\alpha}}{I_{g1}^j}\right)}_{(a)} + \underbrace{\sum_{g=1}^{|\Phi_j|}\sum_{h=2}^{n_g^j} r_{gh}^j \log_e\left(1+\frac{P_{gh}^j\frac{\chi}{(r_{gh}^j)^\alpha}}{I_{gh}^j}\right)}_{(b)} \quad (23)$$

We start computing (23) by first computing (a) in it. To do this, without loss of generality, we assume $P_{g1}^j \leq P_{g'1}^j$ if $g < g'$. We define $d_g^j = \min\left(\{\sqrt{2}\}\cup\{|x_{1_{g'}^j} - x_{1_g^j}| : 1 \leq g' \leq |\Phi_j|, g' \neq g, g < g'\}\right)$. Then, it should be noted that w_{st}, which is the user who has the largest transmit power in Γ_{st} must be located within a distance of $\sqrt{2}$ from any RX $1_g^j, \forall g$. Therefore, according to the definition of d_g, for RX 1_g^j, there must be a TX to locate within the distance of d_g from it. Moreover, such a TX must have a transmit power that is at least as large as P_{g1}^j. This thus leads to

$$I_{g1}^j \geq P_{g1}^j\frac{\chi}{(d_g)^\alpha} \geq P_{g1}^j\frac{\chi}{(d_g+R_j)^\alpha}. \quad (24)$$

To proceed, we provide Lemmas 3 and 4 as following:

Lemma 3: Suppose that the network is split into $\frac{1}{R_j}\times\frac{1}{R_j}$ number of equally-sized square clusters, in which each cluster has size R_j^2, where $\frac{1}{R_j}$ is a power of 2. Let $\Delta = \{\delta_1, ..., \delta_G\}$ be a set of G points, where $G \leq \frac{1}{R_j^2}$, such that each cluster contains at most one point from Δ. For $g = 1, 2, ..., G$, we define $d_g = \min\left(\{\sqrt{2}\}\cup\{|x_{\delta_{g'}} - x_{\delta_g}| : 1 \leq g' \leq G, g' \neq g, g < g'\}\right)$. Then, $\sum_{g=1}^G d_g \leq 3\sqrt{2}\frac{1}{R_j} - 2\sqrt{2}$.

Proof. This is obtained by applying Lemma 4.2 in [15]. □

Lemma 4: When $x > 0$ and $\alpha \geq 1$, $\log_e(1+x^\alpha) \leq \alpha x$.

Proof. $\log_e(1+x^\alpha) \leq \alpha\log_e(1+x) \leq \alpha x$. □

From (24) and Lemmas 3 and 4, we can upper bound (a) in (23) as in (25) on the top of next page, where (a) is due to Lemma 3 and (b) is because $|\Phi_j| \leq \frac{1}{R_j^2}$. Now, we upper bound

$$\sum_{g=1}^{|\Phi_j|} r_{g1}^j \log_e\left(1+\frac{P_{g1}^j\frac{\chi}{(r_{g1}^j)^\alpha}}{I_{g1}^j}\right) \le \sum_{g=1}^{|\Phi_j|} r_{g1}^j \log_e\left(1+\frac{P_{g1}^j\frac{\chi}{(r_{g1}^j)^\alpha}}{P_{g1}^j\frac{\chi}{(d_g+R_j)^\alpha}}\right) = \sum_{g=1}^{|\Phi_j|} r_{g1}^j \log_e\left(1+\frac{(d_g+R_j)^\alpha}{(r_{g1}^j)^\alpha}\right) \le \sum_{g=1}^{|\Phi_j|} r_{g1}^j \alpha \frac{d_g+R_j}{r_{g1}^j}$$
$$= \alpha\left(|\Phi_j|R_j + \sum_{g=1}^{|\Phi_j|} d_g\right) \overset{(a)}{\le} \alpha\left(|\Phi_j|R_j + 3\sqrt{2}\frac{1}{R_j} - 2\sqrt{2}\right) \overset{(b)}{\le} \alpha\left(\frac{1}{R_j} + 3\sqrt{2}\frac{1}{R_j} - 2\sqrt{2}\right) \le \frac{\alpha}{R_j}\left(1+3\sqrt{2}\right). \tag{25}$$

$$\sum_{g=1}^{|\Phi_j|}\sum_{h=2}^{n_g^j} r_{gh}^j \log_e\left(1+\frac{P_{gh}^j\frac{\chi}{(r_{gh}^j)^\alpha}}{I_{gh}^j}\right) \le \sum_{g=1}^{|\Phi_j|}\sum_{h=2}^{n_g^j} r_{gh}^j \log_e\left(1+\frac{2P_{gh}^j}{P_g^j}\frac{((\sqrt{2}+1)R_j)^\alpha}{(r_{gh}^j)^\alpha}\right)$$
$$\le \sum_{g=1}^{|\Phi_j|}\sum_{h=2}^{n_g^j} R^j \log_e\left(1+(\sqrt{2}+1)^\alpha\frac{2P_{gh}^j}{P_g^j}\frac{2^\alpha R_{j-1}^\alpha}{R_{j-1}^\alpha}\right) \overset{(a)}{\le} \sum_{g=1}^{|\Phi_j|}\sum_{h=2}^{n_g^j} R^j(2(\sqrt{2}+1))^\alpha\frac{2P_{gh}^j}{P_g^j} = 2R^j(2(\sqrt{2}+1))^\alpha \sum_{g=1}^{|\Phi_j|}\frac{\sum_{h=2}^{n_g^j}P_{gh}^j}{P_g^j}$$
$$\le 2R^j(2(\sqrt{2}+1))^\alpha \sum_{g=1}^{|\Phi_j|}\frac{\sum_{h=1}^{n_g^j}P_{gh}^j}{P_g^j} \overset{(b)}{=} 2R^j(2(\sqrt{2}+1))^\alpha \sum_{g=1}^{|\Phi_j|} 1 = 2(2(\sqrt{2}+1))^\alpha R^j|\Phi_j| \overset{(c)}{\le} 2(2(\sqrt{2}+1))^\alpha\frac{1}{R_j}. \tag{27}$$

(b) of (23). We define $P_g^j = \sum_{h=1}^{n_g^j} P_{gh}^j$. Then, we observe that when we consider only TXs whose RXs are within the same cluster of RX $u^{(\mathrm{r})}$, the distances between those TXs and RX $u^{(\mathrm{r})}$ must be upper bounded by $\sqrt{2}R_j + R_j$, where the first term is because the largest distance between any receive in the same cluster is $\sqrt{2}R_j$ and the second term is because the the communication distance of the TX-RX pair in $\Gamma_{st}^{R_j}$ is upper bounded by R_j. In addition, since we have $P_{g1}^j \ge P_{gh}^j, \forall h \ge 2$ we know $P_g^j \ge 2P_{gh}^j, \forall h \ge 2$ leading to $P_g^j - P_{gh}^j \ge \frac{P_g^j}{2}, \forall h \ge 2$. By above results, we obtain

$$I_{gh}^j \ge (P_g^j - P_{gh}^j)\frac{\chi}{(\sqrt{2}R_j+R_j)^\alpha} \ge \frac{P_g^j}{2}\frac{\chi}{(\sqrt{2}R_j+R_j)^\alpha}. \tag{26}$$

Recall that $R_{j-1} \le r_{gh} < R_j$ according to the definition of $\Gamma_{st}^{R_j}$. Also, notice that $R_j = 2R_{j-1}$ by definition. It follows by using (26) that (b) of (23) can be upper bounded as in (27) on the top of this page, where (a) is because $\log_e(1+x) \le x$ when $x > 0$; (b) is because $\sum_{h=1}^{n_g^j} P_{gh}^j = P_g^j$ by definition; and (c) is because $|\Phi_j| \le \frac{1}{R_j^2}$.

We note that results in (25) and (27) can be applied when $j > 0, \forall j$. Then, by using (23), (25), (27), and $R_j = 2R_{j-1}$, we can obtain:

$$\sum_{j>0}\sum_{u\in\Gamma_{st}^{R_j}} r_u \log_e\left(1+\frac{P_u l_{uu^{(\mathrm{r})}}}{\sum_{k\in\Gamma_{st},k\ne u} P_k l_{ku^{(\mathrm{r})}}}\right)$$
$$\le \left(\alpha\left(3\sqrt{2}+1\right) + 2(2(\sqrt{2}+1))^\alpha\right)\frac{1}{\epsilon_0}\sqrt{\frac{SN}{\rho' M}}. \tag{28}$$

By combining (20) and (28) and summing contributions in all time-slots and sub-channels, we can obtain Theorem 2, where $\mathcal{W} = \{w_{st}, \forall s, t\}$,

$$\overline{C}_{\mathcal{W}} = \frac{\tau B_{\mathrm{s}} \sum_t \sum_s r_{w_{st}} \log_2\left(1+\frac{P_{\max} l_{w_{st}w_{st}^{(\mathrm{r})}}}{N_0 B_{\mathrm{s}}}\right)}{BT}, \tag{29}$$

and

$$\overline{C}_{\Gamma_{R_0}} = \frac{\tau B_{\mathrm{s}} \sum_t \sum_s \sum_{u\in\Gamma_{st}^{R_0}} r_u c_u}{BT}. \tag{30}$$

□

REFERENCES

[1] 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, April 2013.

[2] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1710–1732, 2018.

[3] S. Gitzenis, G. S. Paschos, and L. Tassiulas, "Asymptotic laws for joint content replication and delivery in wireless networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2760–2776, 2012.

[4] L. Qiu and G. Cao, "Popularity-aware caching increases the capacity of wireless networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 1, pp. 173–187, 2019.

[5] M.-C. Lee, M. Ji, and A. F. Molisch, "Optimal throughput–outage analysis of cache-aided wireless multi-hop D2D networks," *IEEE Trans. Commun.*, Dec. 2020, (In press).

[6] A. Liu, V. K. N. Lau, and G. Caire, "Cache-induced hierarchical cooperation in wireless device-to-device caching networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 4629–4652, June 2018.

[7] N. Golrezaei, A. D. Dimakis, and A. F. Molisch, "Scaling behavior for device-to-device communications with distributed caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4286–4298, July 2014.

[8] M. Ji, G. Gaire, and A. F. Molisch, "The throughput-outage tradeoff of wireless one-hop caching networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6833–6859, December 2015.

[9] M.-C. Lee, M. Ji, A. F. Molisch, and N. Sastry, "Throughput-outage analysis and evaluation of cache-aided D2D networks with measured popularity distributions," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 11, pp. 5316–5332, November 2019.

[10] F. Xue and P. R. Kumar, "Scaling laws for ad hoc wireless networks: an information theoretic approach," *Foundations and Trends® in Networking*, vol. 1, no. 2, pp. 145–270, 2006.

[11] L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for d2d-assisted wireless caching networks," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2438–2452, June 2016.

[12] M. Choi, A. F. Molisch, and J. Kim, "Joint distributed link scheduling and power allocation for content delivery in wireless caching networks," *IEEE Trans. Wireless Commun.*, 2020.

[13] M. Hefeeda and O. Saleh, "Traffic modeling and proportional partial caching for peer-to-peer systems," *IEEE/ACM Trans. Netw.*, vol. 16, no. 6, pp. 1447–1460, December 2008.

[14] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," June 2015.

[15] A. Agarwal and P. R. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 3, pp. 71–81, 2004.