

Improving User Perceived QoS in D2D Networks via Binary Quantile Opportunistic Scheduling

Yicong Wang and Gustavo de Veciana

Department of Electrical and Computer Engineering, The University of Texas at Austin

Email: yicong.wang@utexas.edu, gustavo@ece.utexas.edu

Abstract—State-of-the-art D2D schedulers’ performance can be evaluated from at least two different perspectives: the sum throughput and the fraction of satisfied users/applications. In this paper we revisit the performance of such schedulers, e.g., FlashLinQ/ITLinQ, showing they strike particular trade-offs between these two metrics. Our analysis and simulations show that the sum-rate benefits of such schedulers come at the expense of fairness, which under high densities can lead to a substantial fraction of unsatisfied users. This motivates a proposed opportunistic scheduler design, Binary Quantile (BQ) scheduling, which further exploits temporal channel variations via a low overhead distributed mechanism and realizes substantial improvements in user/application performance. We further show that an adapting version of BQ scheduling where link quantile thresholds are adjusted based on achieved throughput can further achieve substantial improvement in user/application level satisfaction which is robust to heterogeneity in the network topology.

I. INTRODUCTION

Device-to-device (D2D) communication has been proposed as a key technology to meet the fast growing demand for mobile traffic for future communication systems [1][2]. By enabling two users in close proximity to connect directly, D2D links can short-cut the two links (uplink/downlink) required to connect through infrastructures. This in turn improves resource utilization and link density. D2D networks are expected to work in a distributed manner since the link density, which can be much higher than that in cellular systems, makes it hard to perform centralized scheduling and interference management without a high signaling overhead. Meanwhile, previous contention-based distributed scheduling methods, e.g., CSMA/CA, fall short when the user density is high.

These issues have motivated recent innovations on distributed D2D scheduling mechanisms, e.g., FlashLinQ [3] and IT-LinQ [4]. As compared with previous distributed D2D schedulers, FlashLinQ and ITLinQ consider the Signal-to-Interference Ratio (SIR) at each link versus simply controlling the Signal-to-Noise Ratio (SNR) or Interference-to-Noise Ratio (INR) in scheduling links. Links with a high SNR have a higher likelihood to be scheduled and the sum throughput of the network can be improved. One drawback with such schedulers, particularly when all links have equal priority in contention, is that unfairness among users is exacerbated since links with good channels are given more transmission opportunities. When there is a large variation in channel quality, links will experience large variations in Quality of Service (QoS) and links with relatively poor channels will barely meet their QoS requirements.

The relative performance of schedulers is heavily dependent on the application scenario and the metrics used. Most performance evaluations of D2D schedulers use metrics for

system performance, such as the network sum rate [3][4] or system spectral efficiency [5][6]. Such metrics indicate the total traffic that the D2D network might be able to offload from the cellular network but fail to reveal the QoS experienced by individual users. Furthermore, improving system metrics, such as the sum rate or the average number of links scheduled in each slot, may lead to shifting resources from links with poor channels to links with good channels, aggravating unfairness among links. This poses a challenge to making a more comprehensive evaluation of schedulers, and especially obviates the need for a good selection of metrics.

One way to further increase network capacity and improve users’ QoS is to use opportunistic scheduling in the time domain, i.e., to exploit temporal variations in the channel gain. Notice that SIR-based schedulers such as FlashLinQ and ITLinQ are already opportunistic in that link scheduling is channel-aware, i.e., links with good channels have more transmission opportunities, but there is still room for improvement. To further exploit temporal channel variations, in this paper we suggest time-domain opportunistic scheduling for FlashLinQ and ITLinQ, e.g., quantile-based opportunistic scheduling. The quantile of the channel gain of a link i at time t , q_i^t , is defined as,

$$q_i^t = G_i(|h_{ii}|^2), \quad (1)$$

where h_{ii} is the channel gain of link i at time t , $G_i(\cdot)$ is the cumulative density function (CDF) of channel gain. The CDF may change over time and we assume this change in CDF is slow and can be tracked. Channel quantile measures the relative quality of the current channel, compared to the link’s own channel instead of channels of other links. By setting links’ contention priorities based on their current channel gain’s quantile, links with higher quantiles have a higher likelihood to be scheduled, which translates to increases in the SNR of scheduled links without changing their INR, if we assume the interference channels are independent from communication channels. The opportunistic schedulers proposed in this paper extend the ideas presented in [7], in which opportunistic scheduling was applied to CSMA type protocols.

Contributions We explore the performance of state-of-the-art D2D schedulers from two different perspectives: sum throughput, and user/link level satisfaction. We highlight that a scheduler will instantiate a particular choice for the trade-off between the two perspectives. We then show that through a combination of power control, opportunistic scheduling and individual link adaptation, one can substantially extend the user-level satisfaction while achieving a fairly high sum throughput. In the process we develop an approach to efficiently realizing opportunistic user scheduling via a threshold based priority contention mechanism. We further show the threshold can be adapted so as to enable links to efficiently meet target QoS.

To the best of our knowledge, this is the first work to explore the MAC design space in terms of trade-offs between sum rate and individual user satisfaction for a range of possible protocols, using different scheduling criterion, power control, opportunism and QoS-centric adaptation.

Related Work Opportunistic scheduling has been proposed to improve various existing protocols, starting with opportunistic Aloha [8][9], where a link contends only when its channel gain exceeds a threshold. A similar qualification-based scheduler is studied for ad hoc networks employing CSMA in [7]. However, there are some drawbacks with such schedulers. The qualification threshold needs to be optimized as the network topology (density) changes. When the links experience fast fading, it is possible that no links in a neighborhood qualify, resulting in low resource reuse. Also, for scenarios with heterogeneous links, optimization of such thresholds is difficult. Quantile-based scheduling [7] avoids such problems by using the quantile of channel quality rather than the absolute value. Such schedulers choose the links with “best” channel in their neighborhood to transmit and improve the performance of the network while guaranteeing that users have the same level of fairness in contention as their non-opportunistic counterparts. All links participate in the scheduling every slot, thus no slot is wasted even all the links are experiencing relatively bad channels. [10][11] studied quantile-based scheduling for cellular networks (see [12] for WLAN) where a centralized controller schedules users. [7] applied quantile-based scheduling to CSMA in ad hoc networks and analyzed the scheduler’s performance using stochastic geometry tools.

An important technique to help achieve performance trade-offs among users and satisfy users’ QoS requirements we consider in this paper is power control. The power control method we will use in our simulations is the square root channel inverse power control mechanism studied in [13]. An ON-OFF distributed power control for D2D underlaid cellular network is studied in [14].

Organization The rest of the paper is organized as follows. In Section II, we give a brief introduction to FlashLinQ and ITLinQ and characterize the unfairness that arises among heterogeneous links. We discuss how to evaluate and improve user QoS and fairness in Section III. In Section IV, we show how quantile-based scheduling can be introduced in FlashLinQ and ITLinQ. We further present a user-level QoS driven adaptive opportunistic scheduler. Section V provides an evaluation of different scheduling methods and Section VI concludes our paper.

II. D2D SCHEDULING

In this section, we first give a brief introduction to FlashLinQ and ITLinQ. We then analyze and evaluate their performance and exhibit the unfairness achieved among heterogeneous links, which motivates our work towards improving the QoS seen by D2D links/users.

A. FlashLinQ

The key mechanism underlying FlashLinQ occurs at the beginning of each slot, when each link performs an OFDM based measurement and scheduling of links in a distributed

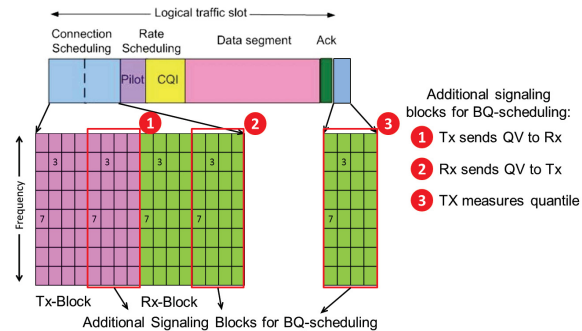


Fig. 1. Frame structure of FlashLinQ and ITLinQ, including the additional signaling blocks required by BQ and AQT (Block 1, Block 2 and Block 3).

manner. The frame structure of FlashLinQ is shown in Fig. 1, excluding the additional signaling blocks (circled in the red boxes) for opportunistic scheduling. During each slot, each link is assigned a unique randomly generated priority index. Denote by B_i a random variable corresponding to the priority index of link i , and b_i as a realization of B_i . For two links i, j with priority index b_i, b_j , link i has a higher priority than link j if $b_i < b_j$. The scheduling consists of two phases, *Phase 1* and *Phase 2*. In each phase, a link is scheduled with a tone to send its pilot according to its priority index.

In Phase 1, a link checks whether it is strongly interfered by links with higher priorities. Let TX_i and RX_i denote the transmitter and receiver of link i . TX_i sends a pilot using its transmit power, $P_t(i)$, on its own tone and receivers measure SNR and INR from all transmitters. Link i survives Phase 1 if the sum interference from links with higher priorities is not too high, i.e., for some threshold γ_{RX} ,

$$\frac{P_t(i) \cdot |h_{ii}|^2}{\sum_{j \text{ s.t. } b_j < b_i} P_t(j) \cdot |h_{ji}|^2} \geq \gamma_{RX}, \quad (2)$$

where h_{ji} is the channel gain of the channel from TX_j to RX_i .

In Phase 2, a link avoids interfering links with higher priorities. Receivers of links that survive Phase 1 transmit using an *inverse echo power* [3] and transmitters measure the channels to estimate the interference they have on other links. Link i is scheduled in Phase 2 if it survives Phase 1 and the SIR TX_i causes to links of higher priority is below some threshold γ_{TX} , i.e., for all j s.t. $b_j < b_i$ and j survives Phase 1,

$$\frac{\text{SNR}_j}{\text{INR}_{ij}} \geq \gamma_{TX}. \quad (3)$$

B. ITLinQ

ITLinQ can achieve within a constant gap of the whole information theoretic capacity region, if for any link i , the product of maximum INR it receives and maximum INR it generates to other links, is no larger than its own SNR, i.e., for all i ,

$$\text{SNR}_i \geq \max_{j \neq i} \text{INR}_{ji} \cdot \max_{j \neq i} \text{INR}_{ij}. \quad (4)$$

Distributed ITLinQ scheduling uses a similar two-phase scheduling algorithm as FlashLinQ to schedule a subset of links following a qualification criterion motivated by (4). In

Phase 1, link i is not strongly interfered by high-priority links if the following holds for all j s.t. $b_j < b_i$,

$$\text{INR}_{ji} \leq M \cdot \text{SNR}_i^\eta, \quad (5)$$

where $M \geq 1$ and $\eta \in [0.5, 1]$ are constants. In Phase 2, link i is scheduled if it does not interfere with high-priority links, i.e., for all j s.t. $b_j < b_i$ and j survives Phase 1,

$$\text{INR}_{ij} \leq M \cdot \text{SNR}_i^\eta. \quad (6)$$

C. Performance of Heterogeneous Links

A D2D scheduler may result in unfairness among links when links are heterogeneous and the design and optimization objectives are to maximize the sum rate. To explore this problem, we study the QoS of heterogeneous links under FlashLinQ and ITLinQ, which is absent in the existing works [3][4][6]. [5] and [6] provide an analytical model to evaluate the sum rate of FlashLinQ and ITLinQ, but transmit power and channel fading is not considered when modeling the probability that a link is scheduled. Our model is to qualitatively understand how unfairness arises in D2D schedulers like FlashLinQ and ITLinQ. We note that unfairness arises in these schemes even though they randomize access priority amongst contending links.

Let us first compute the probability that a typical link, Link 0, is scheduled to transmit. The channel gain H_{ji} between TX_j and RX_i is modeled as follows,

$$H_{ji} = K_{ji} \cdot \sqrt{D_{ji}^{-\alpha}}, \quad (7)$$

where K_{ji} is a random variable denoting the fast fading of the channel from TX_j to RX_i , α is the path loss exponent, D_{ji} is the distance between TX_j and RX_i . We assume that K_{ji} is independent and identically distributed (i.i.d.). The transmit power $P_t(i)$ is only related to channel state between TX_i and RX_i , i.e., D_{ii} and K_{ii} . The locations of the transmitters follow a homogeneous Poisson Point Process (HPPP) with density λ and the locations of the receivers follow an HPPP with density λ but independent of the locations of transmitters. For analysis purpose, we assume the priority index, B_i , is i.i.d. and uniformly distributed on $[0, 1]$. If Link 0 has a priority index b_0 , then on average there is a fraction b_0 of all links with lower priority index and thus higher priority in scheduling. The locations of transmitters (same for receivers) of links with higher priority follow a thinned HPPP with density $b_0\lambda$. We further assume that whether a link passes Phase 1 is independent of other links, i.e., after Phase 1, the locations of receivers of remaining links follow an independently thinned HPPP. Based on our assumptions, the state of links are uniquely decided by $U_i = (P_t(i), D_{ii}, K_{ii}, B_i)$, $i = 1, 2, \dots$, which are i.i.d., and let a set of random variables, $U = (P_t, D, K_C, B)$, have the same distribution as U_i . We further let K_I and K_O denote random variables having the same distribution of K . K_I corresponds to the fading of a typical channel from a transmitter to RX_0 , and K_O corresponds to that of a channel from TX_0 to a receiver other than RX_0 .

Let us first consider FlashLinQ. For Phase 1, we simplify condition in Eq. (2) to

$$\frac{\text{SNR}_i}{\text{INR}_{ji}} \geq \gamma_{\text{RX}}, \quad \forall j \text{ s.t. } b_j < b_i. \quad (8)$$

Link 0 would survive Phase 1 if there are no transmitters with lower priority index that are close enough to RX_0 , i.e., for all i s.t. $b_i < b_0$,

$$D_{i0} > R_{i0},$$

where R_{i0} is the minimum distance such that TX_i does not interfere with RX_0 , i.e., based on Eq. (8) and our channel model we have,

$$R_{i0} = \left(\frac{P_t(i)|K_{i0}|^2\gamma_{\text{RX}}}{P_t(0)|K_{00}|^2} \right)^{\frac{1}{\alpha}} D_{00}. \quad (9)$$

In each slot, each TX_i has an interfering disc for Link 0, Ξ_{i0} , a disc centered at TX_i with radius R_{i0} . Link 0 would be interfered by TX_i if $b_i < b_0$ and RX_0 lies within Ξ_{i0} . The random discs Ξ_{i0} 's are independent of each other and the locations of transmitters with priorities higher than Link 0 follow HPPP, thus the interfering discs can be modeled by a Boolean Model [15].

Let us consider the probability that Link 0 survives Phase 1, given the state of Link 0 is $u_0 = (p_{t,0}, d_0, k_0, b_0)$. Denote by $N_{\text{FLQ}}^{\text{RX}}$ a random variable corresponding to the number of transmitters that have higher priorities than Link 0 and interfere with RX_0 . Let $N_{\text{FLQ}}^{\text{RX}, u_0}$ be a random variable whose distribution is that of $N_{\text{FLQ}}^{\text{RX}}$ given $U_0 = u_0$. Denote by Ξ_{u_0} a random set having the same distribution as the interfering discs of links with priority indexes lower than b_0 , and R_{u_0} the radius of Ξ_{u_0} . According to the Boolean Model, $N_{\text{FLQ}}^{\text{RX}, u_0}$ follows a Poisson distribution with mean $b_0\lambda E[|\Xi_{u_0}|]$, where $|\Xi_{u_0}| = \pi R_{u_0}^2$ is the area of Ξ_{u_0} . Using Eq. (9) and conditioning on $U_0 = u_0$, one obtains

$$E[N_{\text{FLQ}}^{\text{RX}, u_0}] = \frac{b_0\lambda\pi\gamma_{\text{RX}}^{\frac{2}{\alpha}}d_0^2}{p_{t,0}^{\frac{2}{\alpha}}|k_0|^{\frac{4}{\alpha}}} E[|K_I|^{\frac{4}{\alpha}}] E[P_t^{\frac{2}{\alpha}} | B < b_0], \quad (10)$$

The probability that Link 0 survives Phase 1 is $P(N_{\text{FLQ}}^{\text{RX}, u_0} = 0) = e^{-E[N_{\text{FLQ}}^{\text{RX}, u_0}]}$ and we define a function $f_{\text{FLQ}}(V) = e^{-E[N_{\text{FLQ}}^{\text{RX}} | U_0 = V]}$, which gives the probability that a typical link passes Phase 1 of FlashLinQ given the user state is V .

In Phase 2, each receiver that passes Phase 1 and has a priority index lower than b_0 can be associated a protection disc for Link 0, such that Link 0 interferes with that receiver if TX_0 falls into that protection disc. We assume that whether a link i survives Phase 1 is independent of the status of other links and the probability only depends on U_i , which is given by $f_{\text{FLQ}}(U_i)$. The locations of receivers of the links contending in Phase 2 follow an independently thinned HPPP thus once again a Boolean Model can be used. Similar to Phase 1, we let $N_{\text{FLQ}}^{\text{TX}}$ be a random variable corresponding to the number of receivers which have lower priority index than Link 0 and are interfered by TX_0 , $N_{\text{FLQ}}^{\text{TX}, u_0}$ be a random variable with the same distribution as $N_{\text{FLQ}}^{\text{TX}}$ given $U_0 = u_0$. $N_{\text{FLQ}}^{\text{TX}, u_0}$ follows a Poisson distribution with mean,

$$E[N_{\text{FLQ}}^{\text{TX}, u_0}] = b_0\lambda\pi\gamma_{\text{TX}}^{\frac{2}{\alpha}}p_{t,0}^{\frac{2}{\alpha}} E[|K_O|^{\frac{4}{\alpha}}] \cdot E \left[\frac{D^2 f_{\text{FLQ}}(U)}{P_t^{\frac{2}{\alpha}} |K_C|^{\frac{4}{\alpha}}} \Big| B < b_0 \right]. \quad (11)$$

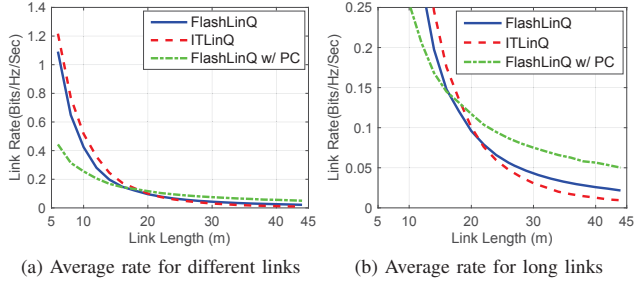


Fig. 2. (a) shows the average rate of links with different lengths under FlashLinQ, ITLinQ and FlashLinQ with power control (PC) shown in (15). (b) scales the y-axis of (a) to better compare the rates of long links.

The probability that Link 0 survives Phase 2 is $e^{-E[N_{\text{FLQ}}^{\text{TX},u_0}]}$. The probability that Link 0 is scheduled given $U_0 = u_0$ is then given by, $p_{\text{scheduled}}^{\text{FLQ},u_0} = e^{-E[N_{\text{FLQ}}^{\text{RX},u_0}] - E[N_{\text{FLQ}}^{\text{TX},u_0}]}$.

Similarly for ITLinQ, given $U_0 = u_0$, we can define and compute $E[N_{\text{ITQ}}^{\text{RX},u_0}]$, $E[N_{\text{ITQ}}^{\text{TX},u_0}]$, f_{ITQ} to get,

$$E[N_{\text{ITQ}}^{\text{RX},u_0}] = \frac{b_0 \lambda \pi d_0^{2\eta}}{\gamma_{\text{ITQ}}^{\frac{2}{\alpha}} |k_0|^{\frac{4\eta}{\alpha}}} E[|K_I|^{\frac{4}{\alpha}}] E[P_t^{\frac{2}{\alpha}} | B < b_0], \quad (12)$$

$$E[N_{\text{ITQ}}^{\text{TX},u_0}] = \frac{b_0 \lambda \pi p_{t,0}^{\frac{2}{\alpha}} d_0^{2\eta}}{\gamma_{\text{ITQ}}^{\frac{2}{\alpha}} |k_0|^{\frac{4\eta}{\alpha}}} E[|K_O|^{\frac{4}{\alpha}}] E[f_{\text{ITQ}}(U) | B < b_0], \quad (13)$$

where $\gamma_{\text{ITQ}} = M p_{t,0}^{\eta} (N_0 W)^{1-\eta}$, N_0 is the noise spectral density, W is the bandwidth. The probability that Link 0 is scheduled is $p_{\text{scheduled}}^{\text{ITQ},u_0} = e^{-E[N_{\text{ITQ}}^{\text{RX},u_0}] - E[N_{\text{ITQ}}^{\text{TX},u_0}]}$.

Unfairness in transmission opportunity. In FlashLinQ, $E[N_{\text{FLQ}}^{\text{RX},u_0}] \propto d_0^2$, while in ITLinQ, $E[N_{\text{ITQ}}^{\text{RX},u_0}] + E[N_{\text{ITQ}}^{\text{TX},u_0}] \propto d_0^{2\eta}$. Such results indicate that in both schemes, links with good channels (short links) enjoy a higher transmission opportunity than links with poor channels (longer links). In principle, ITLinQ provides better fairness than FlashLinQ in terms of transmission opportunities since $2\eta \leq 2$. Shadowing can change link quality and this may relieve the unfairness problem in the long term. However, shadowing changes at slow time scales and users would still suffer unfairness on shorter time scales.

Now we consider another important metric impacting user QoS, the quality of the channel when a link is scheduled. We observe that ITLinQ works differently from FlashLinQ and the channel quality for a successfully scheduled link can be poor. In ITLinQ, the “target SIR”,

$$\text{SIR} = \frac{\text{SNR}}{\text{INR}} \approx \frac{\text{SNR}}{M \cdot \text{SNR}^{\eta}}, \quad (14)$$

varies with the SNR thus a link with a low SNR may have a very low “target SIR”. Additionally, in Phase 2 of ITLinQ, when links back off to avoid interfering with high priority links, a link compares the interference it causes to others with the SNR of its own channel, see Eq. (6). A link with a good channel may be allowed to transmit even if it strongly interferes with a link with poorer channel thus a link with poorer channel can barely meet its target SIR in ITLinQ.

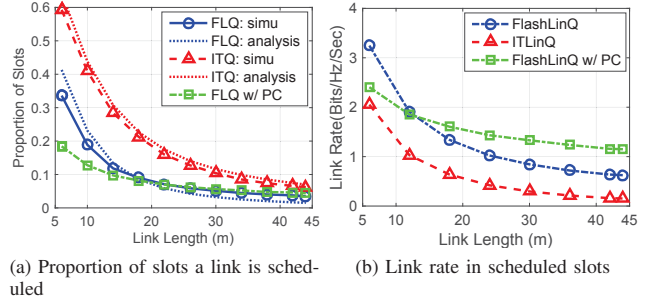


Fig. 3. Comparison of QoS for links with different lengths: (a) illustrates the proportion of slots allocated to each link; (b) shows the average rate in the scheduled slots.

Let us compare the QoS of different links via simulation. 1000 links are randomly placed in an area of $1000m \times 1000m$. The density is $\lambda = 1 \times 10^{-3}$, which is a moderate density. Link lengths are uniformly distributed over $[5, 45]m$. The target SIR is $\gamma_{\text{TX}} = \gamma_{\text{RX}} = 5$ dB for FlashLinQ and $M = 10$, $\eta = 0.7$ for ITLinQ, a “target SIR” of 5.3 dB for SNR = 55 dB. Other settings of the simulation are described in Section V.

The average link rate for the two schedulers are as follows: 0.1784 for FlashLinQ, 0.1992 for ITLinQ. Fig. 2 shows the rate achieved by links with different lengths and link rate varies substantially with link length. ITLinQ provides a higher network sum rate than FlashLinQ by giving shorter links higher rates, but long links suffer. In Fig. 3 we further compare the links QoS using the proportion of slots allocated to each link and the link rate in scheduled slots. In Fig. 3(a) we show the probability of being scheduled for links of different lengths and compare the simulation results with our analysis. The proportion of scheduled slots decrease roughly exponentially with link length and our analysis predicts the same trend. In Fig. 3(b), we notice that although both schemes attempt to guarantee a minimum SIR in their scheduling criterion, the link rate is not strictly guaranteed and in particular long links can hardly meet the rate associated with the target SIR. The performance of these schedulers is sensitive to parameters and experiment settings, but our simulation results suggest that the SIR may not well controlled, a fact which can be ignored if only system metrics are used in the performance evaluation but is likely unacceptable from the point of view of individual links’ QoS.

Our analytical and simulation results unfortunately suggest that FlashLinQ and ITLinQ achieve higher sum rate by shifting resources to shorter links.

III. MEETING USER QoS REQUIREMENTS IN D2D NETWORKS

The unfairness among links raises a question: *If trade-offs are being made among heterogeneous links, what is a good way to evaluate the overall performance of D2D schedulers?*

We suggest that both overall system and user-level QoS metrics need to be used in evaluating D2D schedulers. System metrics such as network sum rate and average number of scheduled links show the total amount of traffic the D2D network can carry, while user QoS metrics show whether the trade-off achieved by the scheduler better serves most of the

users. For metrics evaluating user QoS, user-level satisfaction is a good candidate, which we define as follows,

Definition 1. *User-level satisfaction, S , is the proportion of users for which the minimum QoS requirements are met after scheduling, i.e.,*

$$S = \frac{N_{\text{satisfied}}}{N_{\text{total}}},$$

where $N_{\text{satisfied}}$ is the number of users whose minimum QoS requirements is satisfied, N_{total} is the total number of users.

Possible QoS requirements include: average Shannon rate, proportion of slots of meeting target SIR and variation of rate across slots, etc. A combination of system metrics and metrics measuring the QoS of individual users can better reveal the trade-offs among users and thus provide a good way to evaluate the performance of different D2D schedulers.

Methods to improve user QoS. Serving the QoS requirements of heterogeneous links is a challenging problem for distributed D2D schedulers. One method to compensate for the co-existence of heterogeneous links is power control. By letting shorter links work at lower power levels, shorter links can save energy and reduce their interference on longer links. In fact, the work in [13] suggests that choosing a transmit power which is inversely proportional to the square root of channel gain, i.e.,

$$P \propto (|h|^2)^{-\beta} \text{ for } \beta = 1/2, \quad (15)$$

maximizes the number of links whose SIR exceeds a fixed threshold given that each transmitter only knows its own link parameters. Such power control is used in FlashLinQ and the results in Fig. 2 and Fig. 3 show how resources allocated across heterogeneous links are rebalanced when it is used.

A natural idea to achieve better trade-offs among users is to tune parameters, either of the system or of individual links. The effect of using different γ_{RX} and γ_{TX} is discussed in [3]. The fair ITLinQ in [4] uses different M and η based on the SNR of users. A detailed analysis of parameters in opportunistic scheduling for CSMA/CA can also be found in [7]. Different parameters achieve different trade-offs between channel SIR and the probability of transmitting. However, tuning parameters may not resolve the unfairness intrinsic to the mechanisms of FlashLinQ and ITLinQ, and the resulting link SIR may still be poor.

Another method to improve the capacity of the network, especially the rate of links with relatively poor channels, is opportunistic scheduling. FlashLinQ and ITLinQ are already opportunistic in favoring links with good channel, while we can further take advantage of the temporal channel variations by giving higher priorities to links with *relatively good* channels. We give a full description of such opportunistic scheduling in the next section and evaluate its performance in Section V.

IV. OPPORTUNISTIC SCHEDULING IN D2D NETWORKS

In this section, we show how quantile-based scheduling can be introduced into FlashLinQ and ITLinQ and present a distributed binary quantile-based scheduling that requires limited signaling overhead. We further discuss how such schedulers

can adapt their behavior to better meet users' heterogeneous QoS requirements.

A. Quantile-based Scheduling

Instead of giving links randomly selected priorities, we consider giving priority to links based on the links' current quantiles, which is defined in Eq. (1). Estimating the CDF of channel gain only requires each link to record the channel gain of previous slots. See [10] for a study of estimation of quantiles and penalties associated with noisy estimation. We only change how scheduling priorities are set in FlashLinQ and ITLinQ, while the two-phase scheduling remains unchanged. In the ideal case there is a centralized scheduler that collects channel quantile information and assigns priorities to links in every slot, where the priority order of links is decreasing in the order of links' quantiles, i.e.,

$$b_i < b_j, \text{ for all } i, j \text{ s.t. } q_j^t < q_i^t. \quad (16)$$

We refer to such a scheduler as the *ideal quantile-based opportunistic scheduler* (iQT for short). In theory, the quantile of each link is uniformly distributed on $[0, 1]$, i.e., $Q_i^t \sim \text{unif}(0, 1), \forall i, t$, and can be assumed to be mutually independent of each other, thus iQT has the same level of fairness in terms of how priorities are allocated to links as a scheme that assign those at random. However, iQT guarantees that links with higher channel quantiles have higher priorities than links with lower quantiles, thus the SNR of links with high priority in iQT is statistically higher than the links with same priority in the original algorithm while the distribution of INR remains the same (assuming that fading of different channels are mutually independent). As a result, the sum rate and average number of scheduled links will be higher under iQT.

A centralized scheduler that collects the exact quantiles of links and schedules links in each slot is not efficiently implementable in D2D networks, thus we propose *binary quantile-based scheduling* (BQ) to perform quantile-based scheduling in a distributed way at reasonable signaling cost.

How BQ works. In BQ, the priority index, B_i , is still randomly allocated. Each link is allocated an additional pair of tones, as illustrated in Fig. 1, to broadcast a one-bit binary quantile value, which is defined as follows,

$$qv_i^t = \mathbb{1}(q_i^t \geq \epsilon_i), \quad (17)$$

where $\epsilon_i \in [0, 1]$ is the threshold to quantize channel quantile q_i^t . ϵ_i can be assigned by the system or selected by users to better meet their QoS requirements. To send this one-bit value, we only need an ON-OFF signaling: for link i , TX_i and RX_i transmit with power $P_t(i)$ on the additional tones if $qv_i^t = 1$; or do not transmit if $qv_i^t = 0$. If $qv_i^t > qv_j^t$, link i has higher priority than link j . If $qv_i^t = qv_j^t$, link i and link j compare their priority index and link i has a higher priority than link j if $b_i < b_j$. A problem we need to solve is that each transmitter needs to know its quantile for the current slot at the beginning of each slot thus we propose to add a signaling block at the end of each slot for receivers to send pilots to transmitters to measure the channel gain.

Signaling cost of BQ. Fig. 1 illustrates the frame structure of BQ. Two additional signaling blocks are added to the link scheduling period: Block 1 gives each transmitter one

tone to broadcast its binary quantile value, while Block 2 is reserved for receivers to broadcast their binary quantile values to transmitters. At the end of each frame, we add a similar OFDM based-signaling block, Block 3, for receivers to send pilots to transmitters for estimation of channel quantile in the next slot. Notice that extra signaling we describe here is the worst-case scenario. If the original signaling block has multiple tones for each link, then we only need Block 3 for channel estimation at the end of each slot. Furthermore, Block 3 can be omitted if the coherence time of the channel is large and we can estimate channel quantile based on the measurement from the previous slot. To account for overheads of BQ-based scheduling, in our simulations we assume each signal block takes 4% time of a slot, and the average rate of links in BQ are penalized for the 12% extra signaling overhead.

Another advantage of BQ over iQT is that BQ is more flexible than iQT: each link may adjust the threshold used to compute qv_i^t , ϵ_i , based on its own QoS requirements and channel quality. This advantage enables the *adaptive binary quantile-based scheduling* presented next.

B. Adaptive Binary Quantile-based Scheduling

Adaptive binary quantile-based scheduling (AQT for short) is a variation on BQ. AQT aims to satisfy users' QoS requirements by adjusting the resources allocated to each user while improving network throughput with opportunistic scheduling. We shall assume that a link i has a target average rate $r_{\text{target},i}$ as its QoS requirements. By monitoring the current average link rate, \bar{r}_i , we can verify if the target is being met.

In AQT, if the current achieved average rate of link i is below its target rate, $\bar{r}_i < r_{\text{target},i}$, link i decreases its threshold ϵ_i to get more slots with high binary quantile value; if $\bar{r}_i > r_{\text{target},i}$, ϵ_i is increased to possibly spare more resources for other links. We further constrain ϵ_i to stay within an interval $[\epsilon_{i,\text{low}}, \epsilon_{i,\text{high}}]$: $\epsilon_{i,\text{high}}$ guarantees that all links benefit from opportunistic scheduling while $\epsilon_{i,\text{low}}$ is used to prevent a link with poor channel from taking too much resource. Algorithm 1 in the panel exhibits the mechanism used for updating the threshold, ϵ_i , in AQT. The parameter s is the step size for updating the average rate, c_i is the step size for updating ϵ_i .

Algorithm 1 Threshold Update Algorithm for AQT

```

In each slot, each link  $i$  do the following:
 $\bar{r}_i \leftarrow (1 - s)\bar{r}_i + sr_i$ 
if  $\bar{r}_i(t) < r_{\text{target},i}$  then
     $\epsilon_i \leftarrow \epsilon_i - c_i \cdot (r_{\text{target},i} - \bar{r}_i) \cdot (\epsilon_i - \epsilon_{i,\text{low}})$ 
     $\epsilon_i \leftarrow \max(\epsilon_i, \epsilon_{i,\text{low}})$ 
else
     $\epsilon_i \leftarrow \epsilon_i + c_i \cdot (\bar{r}_i - r_{\text{target},i}) \cdot (\epsilon_{i,\text{high}} - \epsilon_i)$ 
     $\epsilon_i \leftarrow \min(\epsilon_i, \epsilon_{i,\text{high}})$ 
end if
if  $\epsilon_i = \epsilon_{i,\text{low}}$  and  $\bar{r}_i < \kappa_i \cdot r_{\text{target},i}$  for  $N_i$  slots then
    link  $i$  lowers  $r_{\text{target},i}$  or stop transmitting (optional)
end if

```

FlashLinQ and ITLinQ also provide mechanisms to support user QoS requirements. In [3] the authors propose to assign multiple pairs of tones to each D2D link and D2D links can then choose different tones based on their QoS requirements,

e.g., queue-length or packet delay. As mentioned earlier, the authors in [4] also propose a fair ITLinQ to achieve better fairness among links. The signaling overhead of AQT is similar to that proposed in FlashLinQ but AQT can achieve the gains from opportunistic scheduling. AQT only alters the chance that each link is scheduled in each slot and does not change the parameters used in scheduling thus AQT can be further combined with parameter tuning to provide a more robust support for different users' QoS requirements.

The adaptive scheduler described here does not guarantee an immediate response on the time scale of a slot, i.e., lowering ϵ_i does not guarantee that link i gets into the high priority group in the following slots due to the unpredictable nature of fast fading. In order to meet more strict QoS requirements, one may use multiple levels of quantile value in scheduling and/or use quantiles based on other factors in addition to channel quantile, e.g., packet delay.

V. EVALUATION AND ANALYSIS

In this section, we study the performance of different channel-aware opportunistic scheduling methods and compare them with non-opportunistic versions. We will further show how different trade-offs are made among links by different scheduling methods.

A. Qualitative Analysis

Let us first consider the power control in Eq. (15). We can derive from Eq. (10) and Eq. (11) that $E[N_{\text{FLQ}}^{\text{RX},u_0}] \propto d_0^{2-2\beta}$ and $E[N_{\text{FLQ}}^{\text{TX},u_0}] \propto d_0^{2\beta}$ when power control is used (ignoring fast fading), compared to $E[N_{\text{FLQ}}^{\text{RX},u_0}] \propto d_0^2$ when transmit power is fixed. $E[N_{\text{FLQ}}^{\text{RX},u_0}]$ is more uniform among different links but $E[N_{\text{FLQ}}^{\text{TX},u_0}]$ is now larger for long links. Actually, $E[N_{\text{FLQ}}^{\text{RX},u_0}]$ is generally larger than $E[N_{\text{FLQ}}^{\text{TX},u_0}]$ as only a fraction of the links survive the first phase of scheduling, thus unfairness among links is relieved, as can be seen in Fig. 3(a).

Transmission opportunity under BQ. BQ scheduling only changes how priorities are assigned to links, i.e., for a link i , the joint distribution of priority index B_i and fading K_{ii} . Link i will have a small priority index if its channel is relatively good, thus B_i and $|K_{ii}|$ are negatively correlated. Notice that when link density is high, a link is most likely scheduled when it has a high priority, thus we study the performance of the typical Link 0 when b_0 is small.

In FlashLinQ, $E[N_{\text{FLQ}}^{\text{RX},u_0}] \propto |k_0|^{-\frac{4}{\alpha}}$ and $E[N_{\text{FLQ}}^{\text{TX},u_0}] \propto E[|K_C|^{-\frac{4}{\alpha}} | B < b_0]$. In BQ-FlashLinQ, both $|k_0|^{-\frac{4}{\alpha}}$ and $E[|K_C|^{-\frac{4}{\alpha}} | B < b_0]$ are positively correlated to b_0 , thus Link 0 sees fewer high priority links interfering with it and the probability of being scheduled is higher than that in FlashLinQ if b_0 is small. If the link length is large, or the link density is high, Link 0 is mostly scheduled when it has a small priority index, and it can enjoy more transmission opportunities in BQ-FlashLinQ. In ITLinQ, $E[N_{\text{ITQ}}^{\text{RX},u_0}] + E[N_{\text{ITQ}}^{\text{TX},u_0}] \propto |k_0|^{-\frac{4\gamma}{\alpha}}$ and the same analysis applies.

Rate for scheduled links under BQ. BQ-FlashLinQ may reduce the rate as γ is fixed but a scheduled link may see more interference as BQ-FlashLinQ schedules more links on a slot.

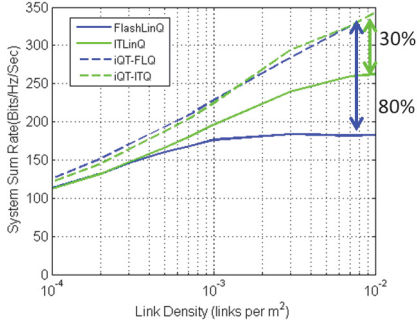


Fig. 4. Sum Shannon rate of network for different link densities.

BQ-ITLinQ, on the other hand, increases the rate in scheduled slots as the “target SIR” increases with $|K_{ii}|$.

The actual scheduling of FlashLinQ and ITLinQ is more complex than our model, thus the analysis here only roughly explains how power control and BQ scheduling work.

B. Simulation Results and Discussion

We consider a network of N D2D links which are randomly dropped in a square area of $1000m \times 1000m$. The path loss between two nodes is modeled by ITU-1411 LOS model with an antenna height of 1.5m as in [3][4]. To eliminate boundary effects, we only consider the performance of links whose midpoints fall into the central $600m \times 600m$ square area. The link lengths have a uniform distribution on $[5, 45]$ m. The carrier frequency is selected to be 2.4GHz and the bandwidth is 5 MHz. The noise power spectral density is -174 dBm/Hz and the max transmit power is 20 dBm with an antenna gain of -2.5 dBi for transmitters and receivers and the noise figure is 7 dB. Fast fading is modeled differently for links of different length: Rayleigh fading for links longer than 72 m; Rician distribution with $K = 4$ dB for links shorter than 72 m. Fast fading of channels is independent in both space and time. Shadowing was not considered in our simulations. All transmitters send at full power if no power control is applied. For power control, we assume the actual channel gain $|h_{ii}|^2$ is not available at the beginning of each slot due to fast fading and the transmit power is decided by the path loss of the channel, i.e.,

$$P_t(i) \propto l_{ii}^{-\beta}, \quad \beta \in [0, 1], \quad (18)$$

where l_{ii} is the path loss of channel between TX_i and RX_i . All links are directional, i.e., from transmitters to receivers, and the rate of a link of certain length is the average rate over different network topologies. FlashLinQ will use a threshold $\gamma_{RX} = \gamma_{RX} = 5$ dB while for ITLinQ, $M = 10$, $\eta = 0.7$.

On 2.4GHz band, the D2D network may interfere with other networks, e.g., WiFi. Possible solutions to managing interference include working on reserved band and sharing the band with other networks in a Time Division Multiple Access way. In this work, we focus on the performance of D2D networks when no other interferers present.

Fig. 4 illustrates how the sum Shannon rate changes as link density increases. We use iQT as a benchmark for possible gains reaped from opportunistic scheduling since every link has perfect knowledge of link quantiles in iQT. As shown in the figure, opportunistic scheduling shows substantial gain over

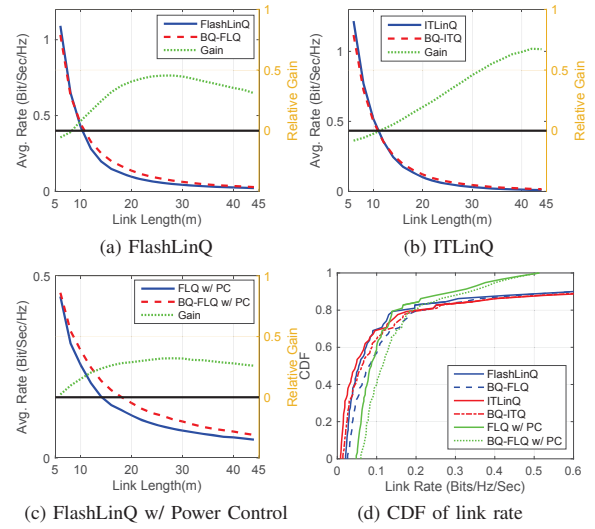


Fig. 5. Average rate of links with different lengths for FlashLinQ (a), ITLinQ (b) and FlashLinQ with power control (c). The relative gain in rate, i.e., $r_{BQ}/r_{base} - 1$. A relative gain of 0 means the rate in BQ and the original scheduler are the same. (d) shows the CDF of average rate of links.

baseline algorithms. Furthermore, the gain from opportunistic scheduling increases with link density if the distribution of link lengths remains the same as link density increases. The reason behind this trend is that as link density increases, the system has more links to choose from, thus the average quantile of the scheduled links increases. These characteristics are in accordance with the performance of opportunistic scheduling on CSMA [7], indicating that opportunistic scheduling is effective in improving the system capacity of dense D2D networks. For the rest of performance evaluation, we set $N = 1000$, which is equivalent to a density of 1×10^{-3} links per m^2 .

Let us consider the average rate achieved by different links, the relative gain of BQ and the CDF of average rate of links, see Fig. 5. For FlashLinQ, BQ increases the rate of 75% of the links by at least 30%, while for ITLinQ, opportunistic scheduling does not increase the system sum rate much, but can greatly improve the average rate of long links. FlashLinQ with power control achieves much better fairness among links and BQ improves the average rate of all links in this case. As we discussed in Section IV-A, our assumption of extra overhead is the worst case. If the non-opportunistic schedulers already provide each link with multiple pairs of tones in scheduling, then we only need to penalize BQ for signaling block at the end which takes 4% of a slot. In this case, 25% gain in our simulation results becomes 35%, and 50% gain becomes 62%.

We further study the QoS of different links in Fig. 6. In Fig. 6(a), we observe that the BQ-FlashLinQ increase the proportion of slots allocated to all links, with long links enjoying a higher relative gain. BQ-ITLinQ only increase scheduled slots slightly (about 2%) and mainly for long links. These results are in consistent with our analysis. In Fig. 6(b), we observe that BQ scheduling increases the rate for ITLinQ, especially for long links. For FlashLinQ, the rate remains almost the same while for FlashLinQ with power control, the rate in scheduled slots decreases.

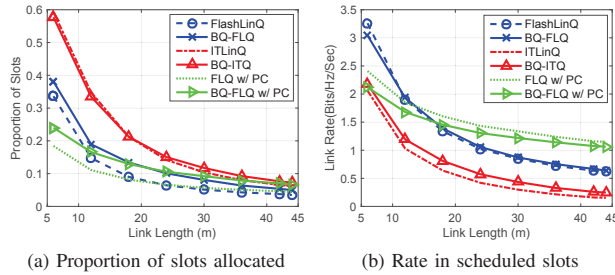


Fig. 6. Links QoS of BQ and non-opportunistic schedulers.

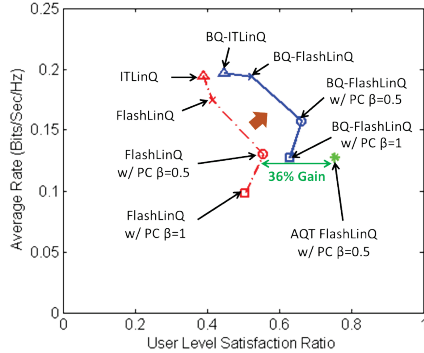


Fig. 7. Working points of different distributed D2D schedulers that treat interference as noise. β is the parameter power control (PC), see Eq. (18).

Fig. 7 exhibits the trade-off between average link rate (system performance) and user-level satisfaction. The link density and link length distribution stay the same but half of the links have a minimum rate requirements of 0.05 Bits/Sec/Hz, the other half have a minimum average rate of 1.5 Bits/Sec/Hz. Connecting the operational performance points of different scheduling methods, we can have an idea of the operational region for schemes with and without opportunistic priority selection (BQ). The dashed line shows the original operational points for FlashLinQ and ITLinQ while the solid line contains the operational points of their opportunistic counterparts. Opportunistic scheduling improves the rate of most users thus both sum rate and satisfaction ratio are improved. This shows that opportunistic scheduling will expand the feasible region of D2D network instead of simply striking different trade-offs among different links. To further improve user-level satisfaction, our adaptive protocol (AQT) can be used to make smart trade-off among heterogeneous users. Compared with non-opportunistic FlashLinQ, adaptive opportunistic scheduling improves user satisfaction ratio by 36% without sacrificing sum rate. AQT can also be adjusted to support other QoS requirements like packet delay and file transfer completion. Notice that AQT will not work well if the user's QoS requirements is beyond the operational region as all users would contend for resource and may decrease the satisfaction ratio. Some other methods are required to guarantee user QoS, i.e., reject users from joining the network.

VI. CONCLUSION

In this paper, we studied the performance of distributed D2D scheduling algorithms, FlashLinQ and ITLinQ, and revealed the problem with user QoS requirements. A distributed

opportunistic scheduling method was proposed and its adaptive version was applied to improve user-level satisfaction. An evaluation of performance tradeoffs and comparison of performance was given. Using the metric we propose for evaluating user QoS, we show that opportunistic versions of D2D schedulers expand the operational region of D2D networks and that adaptive opportunistic scheduling combined with simple channel inverse power control can substantially increase the proportion of users meeting their QoS requirements robustly. We believe a focus on link performance should be emphasized in future work on the design of D2D schedulers.

ACKNOWLEDGEMENT

The research in this paper was supported in part by NSF Grant CNS-1343383 and an Intel grant.

REFERENCES

- [1] 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, Apr. 2014.
- [2] F. Boccardi, R. W. Heath Jr., A. Lozano, T. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74-80, Feb. 2014.
- [3] X. Wu, S. Tavildar, S. Shakkottai, T. Richardson, J. Li, R. Laroia, and A. Jovicic, "FlashLinQ: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1215-1228, Aug. 2013.
- [4] N. Naderializadeh, and A. S. Avestimehr, "ITLinQ: A new approach for spectrum sharing in device-to-device communication systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1139-1151, Jun. 2014
- [5] G. George, R. K. Mungara and A. Lozano, "Optimum exclusion regions for interference protection in device-to-device wireless networks," in *Proc. WiOpt*, May 2015, pp. 94-101.
- [6] R. K. Mungara, X. Zhang, A. Lozano and R. W. Heath, "Performance Evaluation of ITLinQ and FlashLinQ for Overlaid Device-to-Device Communication," in *Proc. IEEE ICC Workshop on Device-to-Device Communication for Cellular and Wireless Networks (ICC15)*, Jun. 2015.
- [7] Y. Kim, F. Baccelli, and G. de Veciana, "Spatial reuse and fairness of mobile ad hoc networks with channel-aware CSMA protocols," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4139-4157, Jul. 2014.
- [8] S. Weber, J. G. Andrews, and N. Jindal, "The effect of fading, channel inversion and threshold scheduling on ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4127-4149, Nov. 2007.
- [9] F. Baccelli, B. Blaszczyszyn, and P. Muhlethaler, "Stochastic Analysis of Spatial and Opportunistic Aloha," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1105-1119, Sep. 2009.
- [10] S. Patil and G. de Veciana, "Measurement-based opportunistic scheduling for heterogeneous wireless systems," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2745-2753, Sep. 2009.
- [11] D. Park, H. Kwon, and B. Lee, "Wireless packet scheduling based on the cumulative distribution function of user transmission rates," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1919-1929, Nov. 2005.
- [12] C. Hwang and J. M. Cioffi, "Using opportunistic CSMA/CA to achieve multi-user diversity in wireless LAN," in *Proc. IEEE Global Telecommun. Conf.*, 2007, pp. 4952-4956.
- [13] F. Baccelli, J. Li, T. Richardson, and S. Shakkottai, "On optimizing CSMA for wide area ad-hoc networks," in *Proc. WiOpt*, 2011, pp. 354-359.
- [14] N. Lee, X. Lin, J. G. Andrews, and R. W. Heath, "Power Control for D2D Underlaid Cellular Networks: Modeling, Algorithm and Analysis," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp 1-13, Jan. 2015.
- [15] F. Baccelli and B. Blaszczyszyn, "Stochastic geometry and wireless networks, volume I - theory," *Foundations and Trends in Networking*, vol. 3, no. 3-4, pp. 249-449, 2009.