Modeling and Analysis on Access Control for Device-to-Device Communications in Cellular Network: A Network-Calculus-Based Approach

Jun Huang, *Member, IEEE*, Yi Sun, Zi Xiong, Qiang Duan, *Member, IEEE*, Yanxiao Zhao, *Member, IEEE*, Xianghui Cao, *Member, IEEE*, and Wei Wang, *Member, IEEE*

***Abstract*—Device-to-device (D2D) communication plays a crucial role in improving the performance of cellular systems, and it is expected to be an innovative technology for next-generation wireless systems. Although significant progress has been made toward cellular and D2D coexistence, the issue of access control for D2D communications in the cellular network has received limited attention. In this paper, we address this issue by employing the network calculus (NC) theory for the first time. We propose a multipriority model, which assigns the strictly highest priority to cellular users and multiple levels of priority to D2D users to characterize the communication requests' access. The proposed model facilitates interference avoidance between cellular and D2D communications and, thus, enhances the quality of service (QoS) of the cellular system. We also apply the NC theory to analyze the worst-case performance of service rate, delay, and backlog for processing communication requests of cellular and D2D users. Both theoretical and experimental results demonstrate that the proposed model is effective and applicable to characterize the access control for D2D communications underlaying cellular networks.**

***Index Terms*—Access control, device-to-device (D2D), multipriority, network calculus (NC), performance analysis.**

Manuscript received September 1, 2013; revised July 14, 2014 and November 28, 2014; accepted February 27, 2015. Date of publication March 12, 2015; date of current version March 10, 2016. This work was supported by the National Science Foundation of China under Grant 61309031 and Grant 61272400, by the National "973" Program under Grant 2012CB315803, by the Postdoctoral Science Foundation of China under Grant 2014M551740, by the Program for Innovation Team Building at Institutions of Higher Education in Chongqing under Grant KJTD201310, by the National Science Foundation of Chongqing under Grant cstc2013jcyjA40026, by the S&T Research Program of Chongqing Municipal Education Commission under Grant KJ130523, and by the Chongqing University of Posts and Telecommunications (CQUPT) Research Fund for Young Scholars under Grant A2012-79.

J. Huang, Y. Sun, and Z. Xiong are with the Department of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xiaoniuadmin@gmail.com).

Q. Duan is with the Department of Information Science and Technology, The Pennsylvania State University, Abington, PA 19001 USA (e-mail: qduan@psu.edu).

Y. Zhao is with the Department of Electrical and Computer Engineering, South Dakota School of Mines and Technology, Rapid City, SD 57701 USA (e-mail: yanxiao.zhao@sdsmt.edu).

X. Cao is with the State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China, and also with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: xhcao@iipc.zju.edu.cn).

W. Wang is with the Department of Computer Science, San Diego State University, San Diego, CA 92115 USA (e-mail: wwang@mail.sdsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2015.2412154

I. Introduction

EVER-growing mobile multimedia services pose a great challenge on the higher data rate of wireless communications, which calls for a new communication scheme underlaying the current cellular systems [1], [2]. *Device-to-Device* (D2D) communication is an emerging technology that has been introduced as a beneficial complement to cellular infrastructures [3]–[5]. The D2D communication paradigm allows mobile devices to directly communicate with each other without the assistance of infrastructures and reuse the radio resource of cellular networks, thus mitigating the system overhead, increasing the spectrum utilization, and improving the cellular coverage [6]. Recent advancements from both academia and industry have indicated that D2D technology has been attracting considerable interest. For example, the *Third-Generation Partnership Project* (3GPP) [7] and *Long-Term Evolution* (LTE) and *LTE-Advanced* (LTE-A) projects consider employing D2D as a potential solution for supporting growing communication demands [8]. Therefore, D2D plays a vital role in improving the performance of cellular systems, and it is expected to be an innovative technology for the next-generation wireless systems [9].

When D2D users share the spectrum resources with cellular users, a major technical concern lies in the mutual interference between two groups. While there have been various approaches proposed to address intra- and intercell interferences [10]–[13], access control as an initial step toward resources coordination for avoiding interference, however, has yet to receive attention. Access control in cellular networks with D2D underlay allows a central controller, typically implemented at the base station (BS), to process requests from different types of users for establishing connections and allocating resources to meet various requirements. The performance of access control in processing connection requests has a significant impact on the quality of service (QoS) provided to D2D users. For example, the delay introduced by access control for processing a connection request influences the total delay performance of real-time communications, particularly for interactive communication sessions.

Modeling the access control for D2D underlaying cellular networks faces a special challenge: the diversity of D2D communications requires a general modeling approach that is applicable to diversified D2D communication schemes, such as D2D communications using cellular, dedicated, and reuse modes

0018-9545 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

[14]. This challenge requires a sophisticated performance analysis technique that can obtain the thorough understanding of D2D access control. *Network calculus* (NC) has been recently developed as a powerful means to address communication system modeling and performance analysis. This employs the envelope method to describe both arrival traffic and service capacity and adopts the systematic method to analyze network performance. Compared with traditional queueing theory techniques, NC specifies traffic load and characterizes system service capabilities via envelopes that are agnostic to system implementations, thereby providing an effective method to develop general modeling and analysis for various networking scenarios. A prominent advantage of NC is the worst-case end-to-end QoS bound calculation, such as delay and backlog. NC not only can be used to study QoS in a more accurate manner but can be used to predict and analyze the network performance locally or globally as well.

Unfortunately, there are no adequate studies in the literature on the performance of the system access control perceived by the cellular and D2D users when they request connection setups. Such a performance and the ensuing insight indeed play important roles in the subsequent radio resource scheduling and allocation procedure. Therefore, it is crucial to develop analytical modeling and analysis on access control for D2D underlaying cellular networks to obtain thorough understanding about the performance, which are the main research objectives of this paper. Note that our research goal is to evaluate the performance of the access control procedure, which processes user requests for setting up connections of D2D and cellular communications, rather than the performance for data transmission scheduling. That is, our study focuses on the control plane performance instead of the data plane performance.

In this paper, we model and analyze the access control for D2D communications in the cellular network from an NC theoretic perspective. To the best of our knowledge, this work is the first attempt that applies NC to address the modeling and performance analysis issue on D2D access control. The contributions of this paper are summarized as follows.

- We propose a multipriority model to characterize the access control of cellular and D2D communications in a single cell. In this model, cellular communication requests are endowed with the strictly highest priority, while D2D communication requests are differentiated with distinct priorities. Unlike existing works dedicated to the nonpreemptive queues, the proposed model is configured with both preemptive and nonpreemptive queues. Hence, this model allows evaluating the access performance, which facilitates the interference avoidance between cellular and D2D users and the enhancement of the QoS for the entire system.
- We apply NC to explore the worst-case access performance of cellular and D2D communications based on the proposed model. The performance bounds of service rate, delay, and backlog are obtained. These bounds reflect the achievable performance for system access experienced by the cellular and D2D users when they request connection setup.
- We conduct thorough numerical simulations to verify theoretical analysis. From simulation results, we find that the performance of low-priority D2D requests is impacted not only by the cellular requests but by the higher priority D2D requests as well. The results also demonstrate that the proposed model is valid and applicable to characterize the access control for D2D underlaying the cellular networks.

The rest of this paper is organized as follows: Section II briefly reviews the related works. Section III presents the system framework and models the access control for D2D underlaying cellular networks using NC. Section IV analyzes the performance of both cellular and D2D communications and obtains the related performance bounds. The numeric experiments are presented in Section V to validate the theoretical analysis. Section VI concludes this paper.

II. Related Work

The proliferation of D2D technology brings new challenges to the interference coordination with cellular communications. Although various solutions have been proposed to address the coexistence of cellular and D2D users [6], [10]–[15], with [16]–[18] particularly providing multiple resource sharing/access approaches for D2D links, the management and control for communication requests of D2D and cellular users have yet to receive much attention. In addition, prior works lack a direct insight about the achievable performance for system access perceived by the cellular and D2D users when they request connection setup.

The multipriority model based on the NC theory has attracted considerable interests in recent years. In general, these models can be classified into two categories: deterministic and stochastic. For the former, Xie *et al.* [19] developed an NC approach based on max-plus algebra and used it to derive the worst-case bound of a network system with nonpreemptive priority scheduling. Qian *et al.* [20] built analytical models for traffic flows under strict priority queue and weighted round robin scheduling in on-chip wormhole networks. Using these models, they derived the per-flow end-to-end delay bound. Sofack and Boyer [21] addressed the QoS for a residual flow in a context of aggregation with nonpreemptive fixed-priority scheduling, where they mainly considered the strict residual service curve. Li *et al.* [22] studied the generic cell rate algorithm and used a hierarchical scheduling policy of fixed priority combined with round robin to derive end-to-end delay bounds. Hua and Liu [23] provided end-to-end delay bounds of transmitting heterogeneous flows, using the deficit round robin (DRR) scheduling policy on the switch output, and compared the DRR, the first-in–first-out (FIFO) method, and the static priority in terms of transmission delays and fairness. Zhang *et al.* [24] investigated how to guarantee a deterministic transfer delay for cyclic control data in switched industrial Ethernet, in which real-time data are assigned the priority to access network resources. Schmitt and Zdarsky [25] exploited a case for a simple alternative in providing QoS in packet-switched networks based on strict priority queuing. Bouillard and Junier [26] presented several

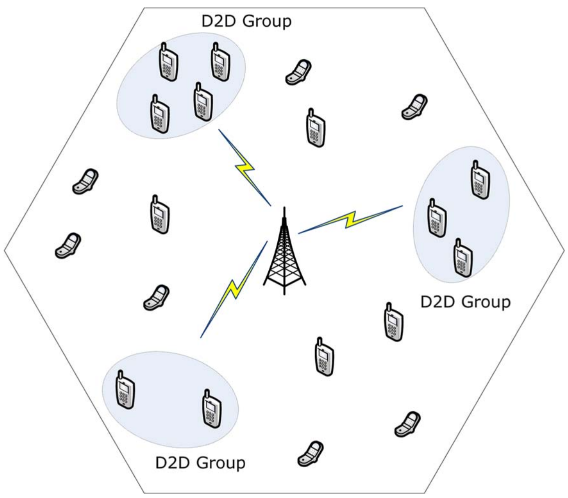


Fig. 1. D2D communications as an underlay in a cell.

methods to compute worst-case delay upper bounds in tandem networks, where flows are ordered according to fixed priorities.

In addition to the aforementioned deterministic multipriority models, stochastic multipriority models have also been explored. Fang *et al.* [27] analyzed the stochastic per-flow delay and backlog in a DiffServ-based wireless sensor network (WSN) and proved that the DiffServ-based WSN can provide better delay and backlog guarantees to data flows with higher priority. A generalized stochastically bounded burstiness model for prioritizing the network traffic with self-similar property is presented in [28]. Gao and Jiang [29] applied stochastic NC to characterize a cognitive radio network, where a queueing model with two priorities for primary and cognitive users is considered.

In our recent work [30], we proposed a multipriority model based on the deterministic NC theory. We aimed to gain tighter performance bounds by assuming that flows arrive in different intervals and in different order. Nevertheless, like aforementioned researches, the work in [30] was dedicated to the nonpreemptive model, while leaving the preemptive model untouched. In this paper, we propose a hybrid model to handle both nonpreemptive and preemptive cases and, as an initial research effort in this area, apply such hybrid model to the access control of D2D communications underlaying the cellular network.

III. Modeling Access Control for Device-to-Device Communications Underlaying the Cellular Network

A. System Overview

We consider the scenario shown in Fig. 1, in which D2D communications form an underlay that coexists with cellular communications in a single cell. In this scenario, we assume that the cellular and D2D users use uplink resources for data transmission and all communications follow a centralized control protocol, i.e., the BS serves as the mediator that establishes connections for both cellular and D2D users. Upon receiving a connection request from a cellular user, the BS sets up a direct uplink from the user to the BS. When a D2D user requests a connection, the BS assists it to discover the peers, to pair the candidate peers, and to setup a connection.

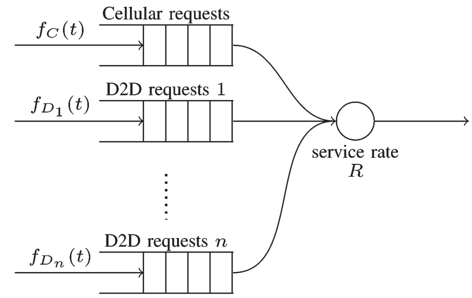


Fig. 2. Multipriority access control model for cellular and D2D communications.

Since the BS is capable of obtaining location information about user equipment in an easy way, the peer discovery and pairing procedures can be accomplished through a certain signaling approach [31], which is not detailed here due to the space limitation. For the link establishment procedure, the BS classifies the requests and endows each of them a priority according to the communication pattern. The rules for request classification and prioritization are given in the following subsection.

B. Access Control Model

To characterize the behavior of communication requests to access the central controller at a BS, we propose a priority-based access control model, as shown in Fig. 2. The model consists of $n+1$ queues, where the top queue is designated for cellular requests, and the rest are designed for D2D requests. The priorities of queues are assigned from high to low in a top-down manner. The top queue in the model is enabled to strictly have the highest priority over the rest for guaranteeing the privilege of cellular communications, i.e., when a cellular request arrives, it is served by the BS immediately if the BS is idle. Since it takes a while for the D2D request to withdraw from the BS, the cellular request waits for a period up to the tolerable maximum, if the BS is serving a D2D request. Moreover, we assume that the idle interval between two consecutive cellular requests is short enough, so that the cellular request will continue to be served by the BS until all requests have been served as long as the channels are available. Note that the priority queues in Fig. 2 are for connection requests that are submitted by users for both cellular and D2D communications. These queues are located at and managed by the BS, which processes the requests and sets up communication connections.

The model described earlier characterizes the most common scenario for D2D underlaying cellular networks since D2D technology is proposed to considerably improve the performance and spectrum utilization of cellular systems [15]. Note that the model is fully capable of being extended to handle other D2D scenarios, such as emergency, by granting D2D users with higher priority than cellular users when a cellular user is in outage due to low signal quality. Although employing priority queues to the BS may incur overhead and latency, they are ignorable, and the proposed multipriority model can facilitate the interference avoidance between cellular and D2D users and significantly enhance the QoS for D2D communications.

When some channels are idle in the system, they will be scheduled for cellular communication requests even if the D2D request queues are not empty. In this case, the interference between cellular and D2D users is avoided. The model is independent to the D2D communication patterns, and thus, it is general and applicable to characterize the access control of both cellular and D2D requests in practical settings. Suppose, for example, that the model has three queues. It is known that D2D communications can be typically classified into cellular, reuse, and dedicated communication modes based on their locations in a cell. The last mode is excluded because D2D communications in the dedicated mode utilize reserved channels and result in no competition with other communications. Therefore, the communication requests from cellular and D2D can be classified into three queues, whose priority assignments are described as follows. The cellular communication has to be guaranteed first; hence, the cellular requests are assigned the highest priority. D2D communications in the cellular mode are interference free to the cellular users, whereas D2D communications may cause interference to them. Therefore, D2D requests with the cellular mode are entitled to receive a higher priority than D2D requests with the reuse mode. In short, three queues filled with cellular communication, D2D with the cellular mode (D2D communication in the cellular mode), and D2D with the reuse mode are allocated priorities from high to low in order.

In the next section, we analyze the performance for each type of request flow based on the model. We will obtain the in-depth insights into access control performance. For the sake of convenience, we use the term "requests" and "request flow" interchangeably throughout the rest of this paper.

IV. Performance Analysis for Access Control of Cellular and Device-to-Device Communications

Here, we first introduce the definitions and notations that will be used in the analysis, and then, we carry out the performance analyses by applying the NC theory to the presented model. In particular, we explore the performance of service rate, delay, and the backlog bounds for both cellular and D2D communications.

A. Definitions and Notations

The theory of NC was first developed by Chang [32] and Cruz [33], [34] and then extended by others (e.g., [35]–[38]). The theory formulates the multipriority model by virtue of the following definitions.

Definition 1—(Wide-Sense Increasing Sequences): If $F = \{f(t) | f(t) = 0,\ \forall t < 0; f(0) \geq 0; f(s) \leq f(t),\ \forall s \leq t, s, t \in [0, +\infty]\}$, we say that F is wide-sense increasing sequences.

Definition 2—(Min-Plus Convolution): Let f and g be two functions or sequences of F. The min-plus convolution of f and g is the function $(f \otimes g)(t) = \inf_{0 \leq u \leq t}[f(u) + g(t-u)]$, if $t < 0$, $f \otimes g = 0$.

Definition 3—(Arrival Curve): Given a wide-sense increasing function α defined for $t \geq 0$, we say that a flow R is constrained by α if and only if for all $s \leq t$ such that $R(t) - R(s) \leq \alpha(t-s)$. We say that R has α as an arrival curve or that R is α-smooth.

TABLE I
Frequently Used Notations

f_C	cellular request flow
f_{D_i}	the i-th D2D request flow, i.e., D2D request flow with priority i
$\alpha_C(t)$	arrival curve of a cellular request flow
$\alpha_{D_i}(t)$	arrival curve of the i-th D2D request flow
$\beta(t)$	service curve
$\beta_C(t)$	service curve for a cellular request flow
$\beta_{D_i}(t)$	service curve for the i-th D2D request flow
$A_i(t)$	sum of arrive curves of D2D request flow(s) whose priority are higher than i
$L_{\max}^{D_i}$	maximum request length of the i-th D2D flow
$L'_{\max}$	maximum request length of D2D flows whose priorities are lower than i
$R_C(t)$, $R_C^*(t)$	input and output cumulative function of a cellular request flow
$R_{D_i}(t)$, $R_{D_i}^*(t)$	input and output cumulative function of the i-th D2D request flow
s_C	backlog start time of a cellular request flow
s_{D_i}	backlog start time of the i-th D2D request flow
$P_C(t)$, $Q_C(t)$	delay and backlog bound of a cellular request flow
$P_{D_i}(t)$, $Q_{D_i}(t)$	delay and backlog bound of the i-th D2D request flow

Since the arrival curve is often regulated by the leaky bucket function, we will use affine arrival curves $\gamma_{r,b}$, defined by $\gamma_{r,b}(t) = rt + b$ for $t > 0$ and 0 otherwise. Parameters b and r are called the burst tolerance and rate. Note that the leaky bucket function herein is used for illustration; other arrival curves can be employed for the following performance analysis in a similar way.

Definition 4—(Service Curve): Consider a system S and a flow through S with input and output function R and R^*. We say that S offers to the flow a service curve β if and only if β is wide-sense increasing, i.e., $\beta(0) = 0$ and $R^* \geq R \otimes \beta$.

The service curve that a system offers to a flow gives the lower bound of the service capacity that the system guarantees to the flow. Since a service curve is given as a function of time that is defined by the relationship between arrival and departure traffic of a flow at the system, it is agnostic to system implementations, thus providing a general approach to characterizing the service capability of any system.

Assuming that a flow constrained by the arrival curve α traverses a system that offers a service curve β, then we have the following definition.

Definition 5—(Delay Bound): The delay bound $P(t)$ is expressed as

$$P(t) \leq h(\alpha, \beta) \tag{1}$$

$$h(\alpha, \beta) = \sup_{t \geq 0} \{\inf \{d \geq 0 : \alpha(t) \leq \beta(t+d)\}\}. \tag{2}$$

Definition 6—(Backlog Bound): The backlog bound $Q(t)$ is expressed as

$$Q(t) = R(t) - R^*(t) \leq \sup\nolimits_{s \geq 0} \{\alpha(s) - \beta(s)\}. \tag{3}$$

Table I lists the frequently used notations throughout the paper.

B. Analysis on Service Curve

Suppose that f_C is the request flow of cellular communications and that f_{D_i}, $1 \leq i \leq n$ is the request flow of a

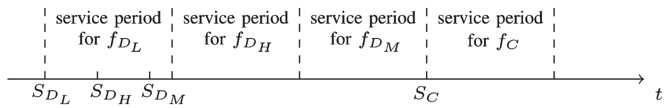


Fig. 3. Illustrative example with typical service order and backlog start time.

D2D user, we say that f_{D_j} has a higher priority than f_{D_i} if $1 \le j \le i \le n$. We assume that f_{D_i} is constrained by the arrival curve $\alpha_{D_i}(t) = r_{D_i}t + b_{D_i}$, $1 \le i \le n$; f_C is constrained by the arrival curve $\alpha_C(t) = r_C t + b_C$; the service curve is $\beta(t) = R \cdot [t-0]^+$, where R is assumed to be the minimum service rate for the worst-case performance calculation [33]. Let $A_i(t) = \sum_{i>j} \alpha_{D_j}(t)$ be the sum of arrive curves whose flow priorities are higher than i, $L'_{\max} = \max_{i<j}\{L^{D_j}_{\max}\}$ denotes the maximum request length of D2D flows whose priorities are lower than i. The following theorem gives the service curve of the model.

Theorem 1: The service curve of f_{D_i} is

$$\begin{aligned}\beta_{D_i}(t) &= \beta(t) - A_i(t) - L'_{\max} - \alpha_C(t)\\ &= \beta(t) - \sum_{i>j} \alpha_{D_j}(t) - L'_{\max} - \alpha_C(t)\\ &= R \cdot t - \sum_{i>j} (r_{D_j}t + b_{D_j}) - L'_{\max} - (r_C t + b_C)\\ &= \left(R - \sum_{i>j} r_{D_j} - r_C\right)\left[t - \frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum r_{D_j} - r_C}\right]^+ \end{aligned} \quad (4)$$

and the service curve of $f_C(t)$ is

$$\beta_C(t) = \beta(t) = R \cdot [t-0]^+. \quad (5)$$

Proof: Without loss of generality, we would like to use a representative example to show how these service curves are obtained. Assume that the D2D communications consists of three priority requests f_{D_L}, f_{D_M}, f_{D_H} denoting the low-priority, medium-priority, and high-priority flows in D2D communications, respectively. Denote f_C as a cellular request flow. s_{D_L}, s_{D_M}, s_{D_H}, and s_C are the time that requests f_{D_L}, f_{D_H}, f_{D_M}, and f_C start to backlog; the service order and the backlog start time are shown in Fig. 3.

Let us first analyze the performance of f_{D_L}. Supposing that all queues are empty in the initial time and f_{D_L} arrives first, then f_{D_L} would be served immediately. If f_{D_H} and f_{D_M} arrive at system when the BS is serving f_{D_L}, the BS would continue to serve the current request of f_{D_L} and then serve f_{D_H} and f_{D_M}. While a cellular request flow f_C arrives when the system is serving f_{D_M}, the model will turn to serve f_C and then f_{D_M}. Flow f_{D_L} will be served after all the other flows being served. That is, within the time interval $(s_{D_L}, t]$, the system output is $R \cdot (t - s_{D_L})$; therefore, we have

$$\begin{aligned} &R^*_{D_L}(t) - R^*_{D_L}(s_{D_L})\\ &\quad = R \cdot (t - s_{D_L}) - \left[R^*_{D_H}(t) - R^*_{D_H}(s_{D_L})\right]\\ &\qquad - \left[R^*_{D_M}(t) - R^*_{D_M}(s_{D_L})\right] - \left[R^*_C(t) - R^*_C(s_{D_L})\right] \quad (6)\\ &0 \le R^*_{D_H}(t) - R^*_{D_H}(s_{D_L}) = R^*_{D_H}(t) - R_{D_H}(s_{D_L})\\ &\quad \le R_{D_H}(t) - R_{D_H}(s_{D_L}) \le \alpha_{D_H}(t - s_{D_L}). \quad (7)\end{aligned}$$

Similarly

$$\begin{aligned} &0 \le R^*_{D_M}(t) - R^*_{D_M} \le \alpha_{D_M}(t - s_{D_L})\\ &0 \le R^*_C(t) - R^*_C \le \alpha_C(t - s_{D_L}). \end{aligned} \quad (8)$$

Through (6)–(8), we can derive

$$\begin{aligned} &R^*_{D_L}(t) - R_{D_L}(s_{D_L})\\ &\quad = R^*_{D_L}(t) - R^*_{D_L}(s_{D_L})\\ &\quad \ge K_{D_L}(t - s_{D_L}), \quad (9)\\ &R^*_{D_L}(t)\\ &\quad \ge R_{D_L}(s_{D_L}) + K_{D_L}(t - s_{D_L})\\ &\quad \ge \inf_{0 \le s_{D_L} \le t} \{R_{D_L}(s_{D_L}) + K_{D_L}(t - s_{D_L})\}\\ &\quad = (R_{D_L} \otimes K_{D_L})(t) \quad (10)\end{aligned}$$

in which $K_{D_L}(t - s_{D_L}) = [R \cdot (t - s_{D_L}) - \alpha_{D_H}(t - s_{D_L}) - \alpha_{D_M}(t - s_{D_L}) - \alpha_C(t - s_{D_L})]^+$.

Therefore, the service curve of f_{D_L} is $K_{D_L}(t - s_{D_L})$ if $K_{D_L}(t - s_{D_L})$ is a wide-sense increasing function. Denoting the arrival curves of f_{D_M}, f_{D_H}, and f_C by $\alpha_{D_M} = \gamma_{r_{D_M}, b_{D_M}}$, $\alpha_{D_H} = \gamma_{r_{D_H}, b_{D_H}}$, and $\alpha_C = \gamma_{r_C, b_C}$, respectively, in this example, we can write the service curve as a rate–latency function $\beta_{R_{D_L}, T_{D_L}}(t)$, with rate

$$R_{D_L} = R - r_{D_M} - r_{D_H} - r_C \quad (11)$$

and latency

$$T_{D_L} = \frac{b_{D_M} + b_{D_H} + b_C}{R - r_{D_M} - r_{D_H} - r_C}. \quad (12)$$

Next, let us explore the performance of f_{D_M}. Since s_{D_M} falls into the service period of f_{D_L}, when f_{D_H} arrives in this period, f_{D_M} will start to be served after the current requests of f_{D_L} and f_{D_H} are processed. In Fig. 3, it is shown that, if f_C arrives in the service period of f_{D_M}, f_C will be served within a very short period, regardless of whether the service for f_{D_M} requests is completed or not. f_{D_M} will obtain the service until the service of f_C is finished. Within the period of $(s_{D_M}, t]$, the system output is $R \cdot (t - s_{D_M})$; thus, we have

$$\begin{aligned} R^*_{D_M}(t) - R^*_{D_M}(s_{D_M}) &= R \cdot (t - s_{D_M}) - L^{D_L}_{\max}\\ &\quad - \left[R^*_{D_H}(t) - R^*_{D_H}(s_{D_M})\right] - \left[R^*_C(t) - R^*_C(s_{D_M})\right]. \end{aligned} \quad (13)$$

Likewise, we can obtain

$$\begin{aligned} R^*_{D_M}(t) &\ge R_{D_M}(S_{D_M}) + K_{D_M}(t - S_{D_M})\\ &= R_{D_M}(s_{D_M}) + [R \cdot (t - s_{D_M})\\ &\qquad - L^{D_L}_{\max} - \alpha_{D_H}(t) - \alpha_C(t)]^+\\ &\ge R_{D_M}(s_{D_M}) \otimes K_{D_M}(t - s_{D_M}). \end{aligned} \quad (14)$$

If $K_{D_M}(t - s_{D_M})$ is a wide-sense increasing function, the service curve of f_{D_M} could be described as a rate–latency function $\beta_{R_{D_M}, T_{D_M}}(t)$, with

$$R_{D_M} = R - r_{D_H} - r_C \tag{15}$$

$$T_{D_M} = \frac{L_{\max}^{D_L} + b_{D_H} + b_C}{R - r_{D_H} - r_C}. \tag{16}$$

Now, we investigate the performance of f_{D_H}. Because s_{D_H} is in the service period of f_{D_L}, the system would serve f_{D_H} after serving the current request of f_{D_L}. In this case, the system output during the time interval $(s_{D_H}, t]$ is $R \cdot (t - s_{D_H})$; thus, we have

$$R_{D_H}^*(t) - R_{D_H}^*(s_{D_H}) = R \cdot (t - s_{D_H}) - L_{\max}^{D_L}. \tag{17}$$

Then, we have

$$\begin{aligned} R_{D_H}^*(t) &\geq R_{D_H}(s_{D_H}) + K_{D_H}(s_{D_H}) \\ &= R_{D_H}(s_{D_H}) + R \cdot (t - s_{D_H}) - L_{\max}^{D_L} \\ &\geq R_{D_H}(s_{D_H}) \otimes K_{D_H}(s_{D_H}). \end{aligned} \tag{18}$$

Therefore, the service curve of f_{D_H} is $\beta_{R_{D_H}, T_{D_H}}(t)$ with

$$R_{D_H} = R \tag{19}$$

$$T_{D_H} = \frac{L_{\max}^{D_L}}{R}. \tag{20}$$

Since the cellular request flow f_C is preemptive to D2D communications, the system will stop serving the current D2D request and start to serve f_C. Within the time interval $(s_C, t]$, the system output is $R \cdot (t - s_C)$, i.e.,

$$R_C^*(t) - R_C^*(s_C) = R \cdot (t - s_C). \tag{21}$$

This suggests that

$$\begin{aligned} R_C^*(t) &\geq R_C(s_C) + K_C(s_C) \\ &= R_C(s_C) + R \cdot (t - s_C) \\ &\geq R_C(s_C) \otimes K_C(s_C). \end{aligned} \tag{22}$$

Thus, the service curve of f_C is $\beta_{R_C, T_C}(t)$, where $R_C = R$ and $T_C = 0$.

The aforementioned analysis indicates that the minimum service curve of f_{D_i} depends on f_C and D2D flows, whose priority is higher than i, as well as the maximum length request of D2D flows, whose priority is lower than i. Thus, we can calculate the accumulative output of f_{D_i} during the period $(s_{D_i}, t]$ by the following inequality:

$$\begin{aligned} R_{D_i}^*(t) - R_{D_i}^*(s_{D_i}) \geq R \cdot (t - s_{D_i}) - \sum_{i>j} \Big[R_{D_j}^*(t) \\ - R_{D_j}^*(s_{D_i}) \Big] - \max_{i<j} \left\{ L_{\max}^{D_j} \right\} - \left[R_C^*(t) - R_C^*(s_{D_i}) \right]. \end{aligned} \tag{23}$$

That is

$$\begin{aligned} R_{D_i}^*(t) &\geq R_{D_i}(s_{D_i}) + K_{D_i}(t - s_{D_i}) \\ &= R_{D_i}(s_{D_i}) \\ &\quad + \Bigg[R \cdot (t - s_{D_i}) - \sum_{i>j} \alpha_{D_j}(t - s_{D_i}) \\ &\qquad - \max_{i<j} \left\{ L_{\max}^{D_j} \right\} - \alpha_c(t - s_{D_i}) \Bigg]^+ \\ &\geq R_{D_i}(s_{D_i}) \otimes K_{D_i}(t - s_{D_i}). \end{aligned} \tag{24}$$

If $K_{D_i}(t - s_{D_i})$ is a generalized increasing function, the service curve should be

$$\begin{aligned} \beta_{D_i}(t) &= K_{D_i}(t - s_{D_i}) \\ &= \left(R - \sum_{i>j} r_{D_j} - r_C \right) \\ &\quad \times \left[t - \frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C} \right]^+. \end{aligned} \tag{25}$$

As the cellular communication f_C exhibits a preemptive priority when compared with D2D communications, the output accumulation function of f_C during time period $(s_C, t]$ turns out to be (21), and the service curve is $\beta_{R_C, T_C}(t) = R \cdot [t - 0]^+$ accordingly.

Note that the example discussed earlier is only used for the purpose of illustration. Other arrival patterns of cellular and D2D flows will lead to a similar proof based on the same conditions.

This completes the proof. ■

C. Performance Bound Calculation

Through the service curves derived in the previous subsection, we can obtain the worst-case performance bounds of service rate, delay, and backlog, for access control. According to Theorem 1, we immediately have the following corollary.

Corollary 1: The service rate bound, i.e., the minimum service rates for cellular and D2D requests are R and $R - \sum_{i>j} r_{D_j} - r_C$, respectively.

Corollary 1 indicates that the minimum service rate of a D2D request flow is affected by the arrival rate of high-priority flows. This is natural because the higher the arrival rate of high-priority flows, the more requests of high-priority flows, which results in the lower service rate for low-priority flows.

To obtain the delay and backlog bound, we assume that the arrival rate and service rate satisfy the following constraint: $\sum_{i=1}^{n} r_{D_i} + r_C \leq R$. Fig. 4(a) shows that the arrive rate satisfies constraints, whereas Fig. 4(b) shows the opposite. In addition, Fig. 4 implies that we can obtain the performance bounds if and only if the constraints are respected by the arrive rate; otherwise, the bounds tend to be infinity large.

Theorem 2: The delay bound of the ith D2D request flow is

$$P_{D_i}(t) = \frac{b_{D_i}}{R - \sum_{i>j} r_{D_j} - r_C} + \frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C}.$$

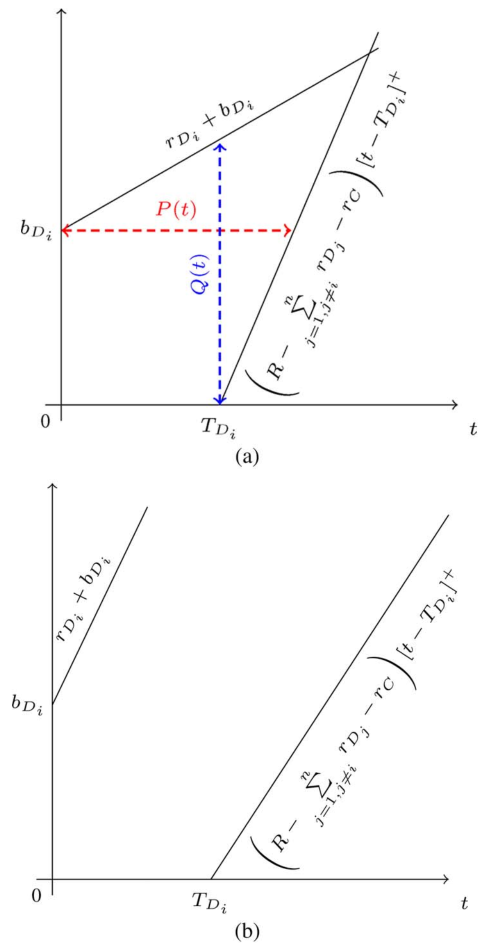


Fig. 4. Arrival rate constraints.

Proof: According to (2), the worst-case delay bound of f_{D_i} is the horizontal deviation of the arrival curve $\alpha_{D_i} = \gamma_{r_{D_i},b_{D_i}}$ and the service curve $\beta_{R_{D_i},T_{D_i}}$. Thus

$$\begin{aligned}P_{D_i}(t) &= h(\alpha_{D_i}, \beta_{R_{D_i},T_{D_i}})\\ &= \sup_{t\geq 0}\{\inf\{d \geq 0 : \alpha_{D_i}(t) \leq \beta_{R_{D_i},T_{D_i}}(t+d)\}\}\\ &= \inf\{d \geq 0 : \alpha_{D_i}(0) \leq \beta_{R_{D_i},T_{D_i}}(d)\}\\ &= \inf\left\{d \geq 0 : b_{D_i} \leq \left(R - \sum_{i>j} r_{D_j} - r_C\right)\cdot d\right.\\ &\qquad \left. - \sum_{i>j} b_{D_j} - L'_{\max} - b_C\right\}\\ &= \inf\left\{d \geq 0 : d \geq \frac{b_{D_i}}{R - \sum_{i>j} r_{D_j} - r_C}\right.\\ &\qquad \left. + \frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C}\right\}\\ &= \frac{b_{D_i}}{R - \sum_{i>j} r_{D_j} - r_C} + \frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C}.\end{aligned} \tag{26}$$

This completes the proof. ∎

Theorem 3: The delay bound of the cellular request flow is

$$P_C(t) = \frac{b_C}{R}.$$

Proof: The worst-case delay bound of f_C is the horizontal deviation of the arrival curve $\alpha_C = \gamma_{r_C,b_C}$ and the service curve β_{R_C,T_C}. Following the same approach to the proof of Theorem 2, the desired equation can be proved by the worst-case delay bounds of D2D communications. This completes the proof. ∎

Theorems 2 and 3 reveal that the delay bound of D2D requests is controlled by several parameters, most of which belong to high-priority flows. The delay bound of cellular requests is only related to R and b_C. We now proceed to analyze the worst-case backlog bounds.

Theorem 4: The backlog bound of the ith D2D request flow is

$$Q_{D_i}(t) = b_{D_i} + r_{D_i}\cdot\left(\frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C}\right).$$

Proof: As (3) indicates, the maximum backlog bound of f_{D_i} is the vertical deviation of the arrival curve $\alpha_{D_i} = \gamma_{r_{D_i},b_{D_i}}$ and the service curve $\beta_{R_{D_i},T_{D_i}}$. Thus

$$\begin{aligned}Q_{D_i}(t) &= R_{D_i}(t) - R^*_{D_i}(t)\\ &\leq \sup_{t\geq 0}\{\alpha_{D_i}(t) - \beta_{R_{D_i},T_{D_i}}(t)\}\\ &= \alpha_{D_i}(T_{D_i}) - \beta_{R_{D_i},T_{D_i}}(T_{D_i})\\ &= \alpha_{D_i}(T_{D_i})\\ &= b_{D_i} + r_{D_i}\cdot\left(\frac{\sum_{i>j} b_{D_j} + L'_{\max} + b_C}{R - \sum_{i>j} r_{D_j} - r_C}\right).\end{aligned} \tag{27}$$

This completes the proof. ∎

Theorem 5: The backlog bound of the cellular request flow is

$$Q_C(t) = b_C.$$

Proof: The maximum backlog bound of f_C is the horizontal deviation of the arrival curve $\alpha_C = \gamma_{r_C,b_C}$ and the service curve β_{R_C,T_C}. Following the same approach to the proof of Theorem 4, the desired equation can be proved by the maximum backlog bounds of D2D communications. This completes the proof. ∎

Theorems 4 and 5 suggest that the backlog bound of a D2D request flow is influenced not only by the arrival curve of itself but also by that of high-priority flows. In contrast, the backlog bound of the cellular flow is only impacted by b_C.

V. Simulation Results

In the previous section, we derived the service curves for both D2D and cellular requests, as well as the worst-case delay and backlog bounds for requests processing. Here, we evaluate those performance bounds and discuss the impacting factors on them through extensive numerical experiments. We also conduct simulations for the proposed model, and the results can provide in-depth insights into the communication connection setup performance.

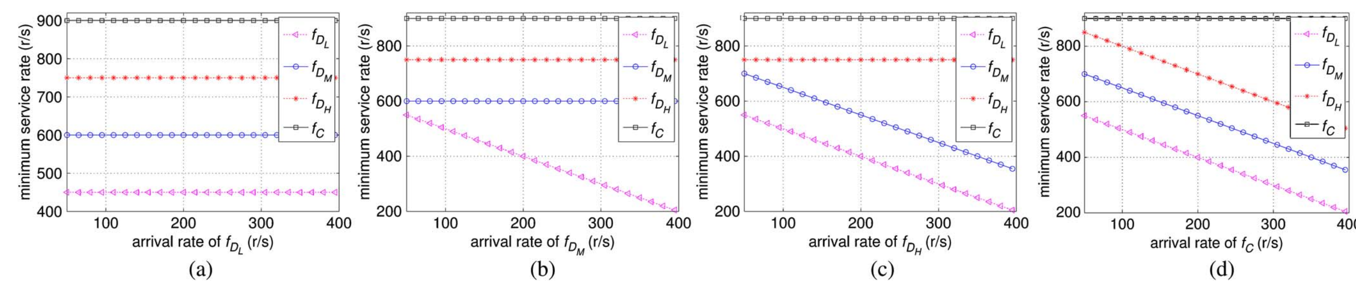


Fig. 5. Minimum service rate versus different arrival rates.

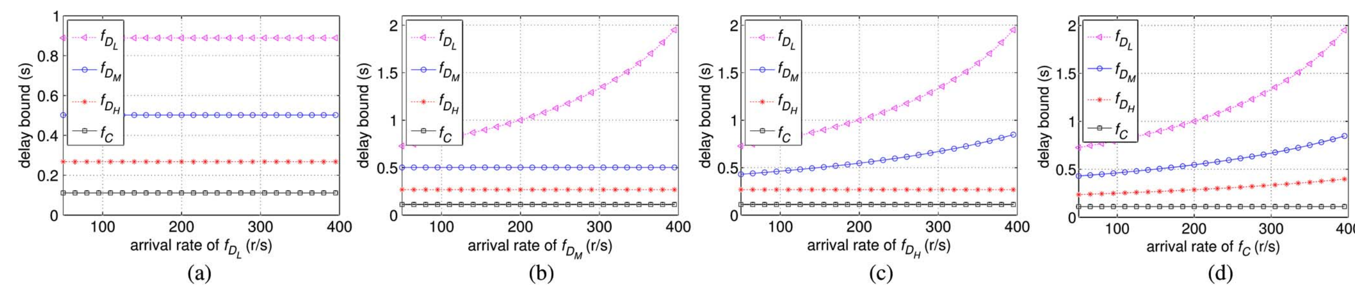


Fig. 6. Delay bound versus different arrival rates.

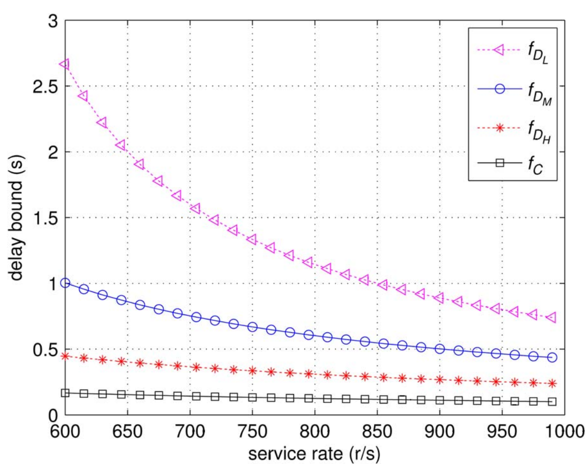


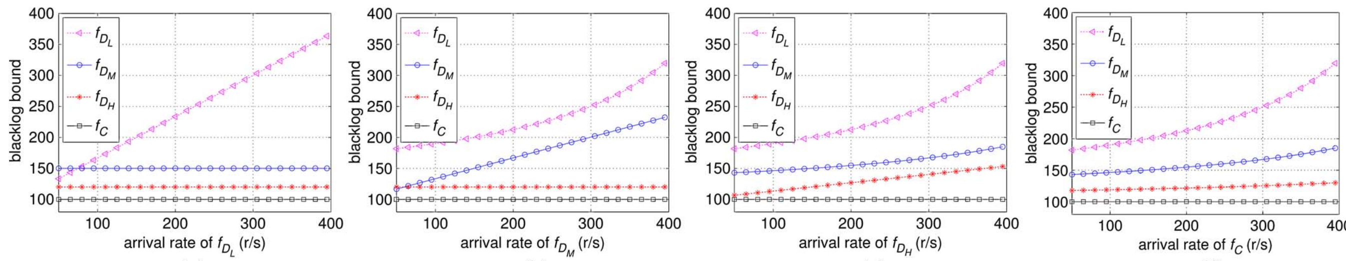
Fig. 7. Delay bound versus service rate.

Following prior conventions, we assume that D2D communications have three request flows with different priorities f_{D_L}, f_{D_M}, and f_{D_H}, which represent the low, medium, and high priority, respectively. The request flow of cellular communications is denoted by f_C. The burst parameter of leaky bucket is set to be 100 requests/s (r/s) with $L_{\max} = 1$ request. The evaluation results are shown in Figs. 5–9.

Fig. 5 shows the variance of the service rate of request flows, with respect to different arrival rates. Fig. 5(a)–(d) shows the service rate in the condition of (a) $R = 900$ r/s, $r_{D_M} = r_{D_H} = r_C = 150$ r/s, $r_{D_L} \in [50, 400]$ r/s; (b) $R = 900$ r/s, $r_{D_L} = r_{D_H} = r_C = 150$ r/s, $r_{D_M} \in [50, 400]$ r/s; (c) $R = 900$ r/s, $r_{D_L} = r_{D_M} = r_C = 150$ r/s, $r_{D_H} \in [50, 400]$ r/s, and (d) $R = 900$ r/s, $r_{D_L} = r_{D_M} = r_{D_H} = 150$ r/s, $r_C \in [50, 400]$ r/s, respectively. In this figure, we can observe that the minimum service rate of low-priority flows decreases as the arrival rate of high-priority flows increases. For example, in Fig. 5(c), the minimum service rate of D2D request flows f_{D_L} and f_{D_M} decline when the arrival rate of f_{D_H} achieves higher. However, the minimum service rate of f_{D_H} and f_C stays the same. This suggests that the minimum service rate of low-priority request flow is affected by the arrival rate of high-priority request flows, which is consistent with the analytical results on service rate stated in Corollary 1.

Fig. 6 shows the changes of the delay bounds of request flows, with respect to different arrival rates, under the same condition as that in Fig. 5. Fig. 6(a)–(c) shows the impact of arrival rate on delay bounds of D2D request flows f_{D_L}, f_{D_M}, f_{D_H}, respectively, and Fig. 6(d) shows those of cellular request flow f_C. As shown in these figures, the delay bound of cellular request flow is tighter than that of D2D request flows, and within D2D communications, the delay bound of high-priority flows has smaller value than that of low-priority flows. This is to say that the cellular requests have better delay guarantee than D2D requests. While for D2D request flows, one with high priority also has better delay guarantee than a low-priority flow. Moreover, with higher arrival rate, their delay bound gap widens. Fig. 6(a)–(c) shows that, with larger arrival rates of f_{D_i}, the delay bound of lower priority flow would significantly increase, i.e., the influence of increasing arrival rates imposed on the lower priority flows is more notable. In Fig. 6(d), we find that the increasing arrival rate of f_C leads to a larger delay bound of D2D flows, as well as a greater impact on the lower priority flows.

Fig. 7 depicts the delay bounds of request flows, with respect to different service rates, when $r_{D_L} = r_{D_M} = r_{D_H} = r_C = 150$ r/s, $R \in [600, 1000]$ r/s. When the arrival rates are unchanged, the delay bounds of request flows drop in varying degrees with regard to the growth of the service rate; particularly for the D2D request flows, they show a shaper decline. Therefore, the results shown in Figs. 6 and 7 conform to our previous analyses in Theorems 2 and 3.



(a) (b) (c) (d)

Fig. 8. Backlog bound versus different arrival rates.

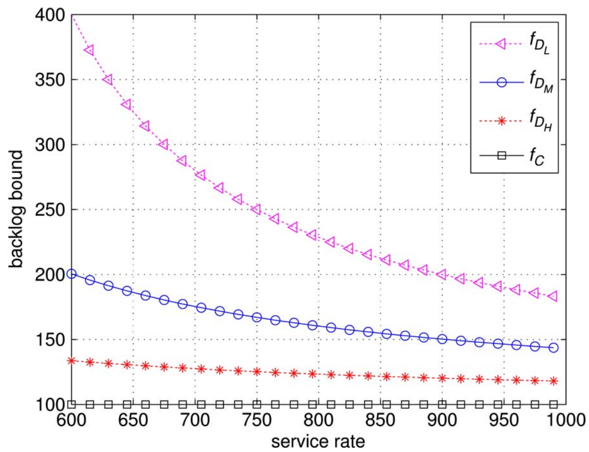


Fig. 9. Backlog bound versus service rate.

Fig. 8 illustrates the impact of arrival rate of request flows on the maximum backlog bound under the same condition as that in Fig. 5. Fig. 8(a)–(c) shows the impact of arrival rate on the backlog bound of D2D request flows f_{D_i}, whereas Fig. 8(d) shows that of cellular request flow f_C. In Fig. 8(a)–(c), we notice that the increase of the arrival rate of f_{D_i} results in a great impact on its own backlog bound and an even greater value than the impact of lower priority. According to Fig. 8(d), the arrival rate of f_C shows larger influence on D2D communications' backlog bound than that on itself. Moreover, the cellular request flow will represent the minimum backlog bound on the condition that f_{D_i} reaches the same arrival rate, and the medium priority, as well as high priority, of D2D requests turn to obtain better backlog bounds. Otherwise, f_{D_i} with low arrival rate may lead to a low backlog bound.

Fig. 9 shows the backlog bound of each request flow, with respect to different service rates, when $r_{D_L} = r_{D_M} = r_{D_H} = r_C = 150$ r/s, $R \in [600, 1000]$ r/s. The same insight as that for delay bounds can be gained. With the increase of service rate, the backlog of D2D request flows will deteriorate to some extent and cause a great effect on D2D data flows but will not affect the cellular request flow. With the consistent growth of service rate, the impact on D2D tends to be reduced accordingly. Hence, the data presented in Figs. 8 and 9 confirm the theoretical results stated in Theorems 4 and 5.

In summary, both the arrival rate and the service rate have substantial effects on performance bounds. The performance bounds of D2D request flows are affected not only by service rate but by the arrival rate of both high-priority D2D and cellular request flows as well. Meanwhile, the cellular request flow is only impacted by its own arrival and service rate. In addition, the performance bound of cellular request flow is significantly better than that of D2D communications, and the performance bound of high-priority D2D requests is also significantly higher than that of low-priority D2D requests.

In the second set of experiments, OMNet++ [39] is employed to conduct simulations for validating the proposed model further. Similar to the aforementioned numerical experiments, we set up a simulation scenario that has four request flows with different priorities representing cellular request flow and D2D requests with high, medium, and low priority. The request flows arrive at four priority queues, with an identical rate of 150 r/s. The server processes the requests complying with the policy defined in Section III-B. We set the service rate to be within the interval (600, 1000) following the uniform distribution with interval (600, 1000) r/s; the simulation time is set to be 300 s. During the simulations, the quantity metrics, including queueing delay and queue length, are traced.

The simulation results, in terms of queueing delay and queue length, are shown in Figs. 10 and 11, respectively. Within each figure, subfigure (a) compares the performance of four flows, and subfigure (b) is a zoom-in of subfigure (a) focusing on the performance of f_{D_M}, f_{D_H}, and f_C in a more distinguishable manner. Since the request flow of f_C is not completely processed by the exact 300 s, the simulation is actually terminated at 324.5 s. Thus, we present the results in a slightly longer time span. The curves in this two figures indicate that f_C has the lowest queueing delay and queue length, f_{D_H} has higher performance than f_{D_M}, and f_{D_L} has the highest delay and length. This is aligned with the insight obtained in the aforementioned theoretical results. In addition, we observe that the maximum queueing time and queue length of four request flows satisfy the delay and backlog bounds obtained in Figs. 7 and 9. This again confirms that the simulation results are consistent with the theoretical results, thus demonstrating the validity of the proposed multipriority model.

VI. Conclusion

In this paper, we have investigated the access control for D2D communications underlaying cellular networks. A multipriority model for characterizing the access control of cellular and D2D communications in a single cell is proposed. In the model, cellular communications are endowed with the strictly highest

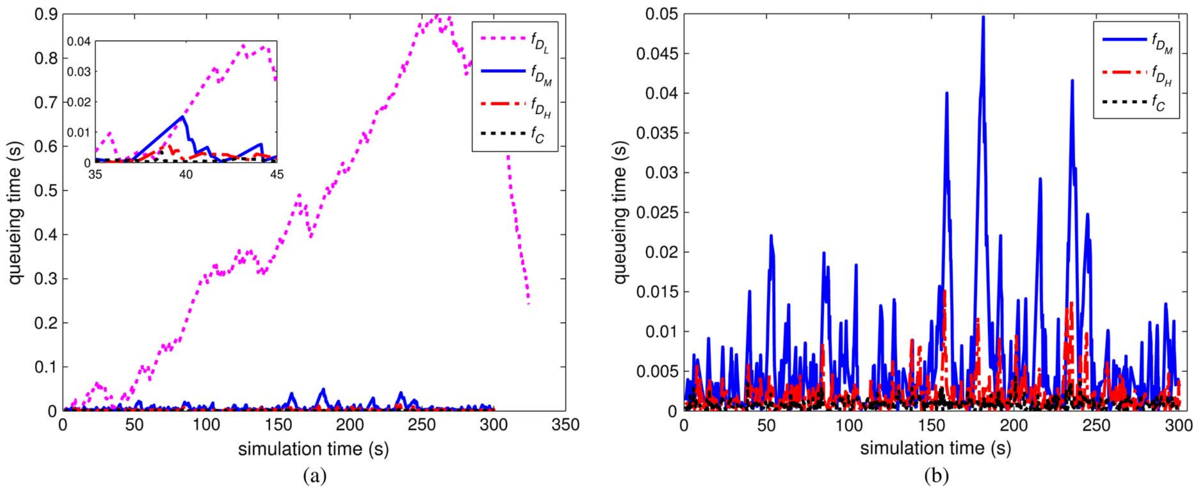


Fig. 10. Queue delay comparisons.

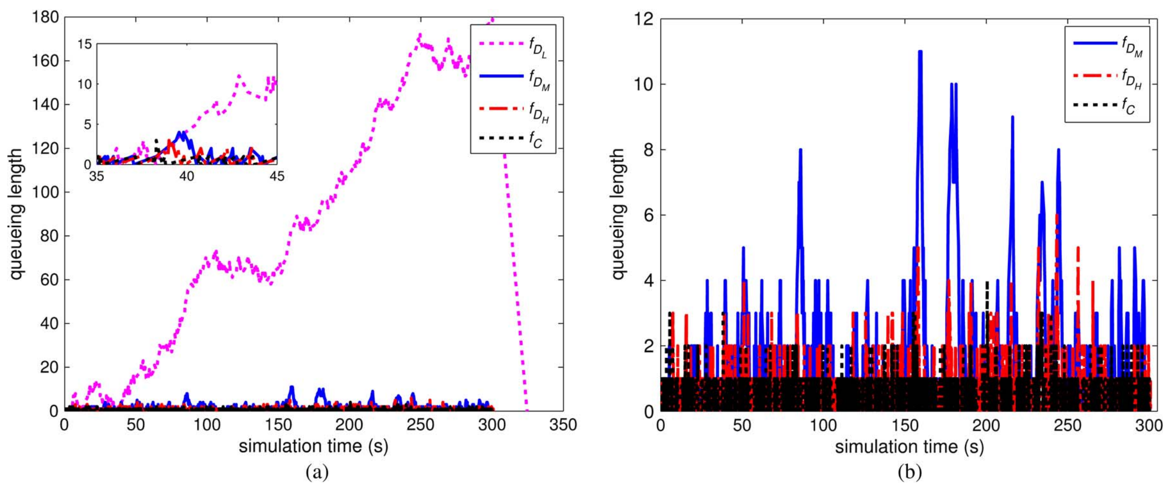


Fig. 11. Queue length comparisons.

priority, whereas D2D communications are differentiated with distinct priorities. As a result, the proposed model is able to avoid the interference between cellular and D2D users and enhance the QoS for D2D communications. We have evaluated the worst-case access performance of cellular and D2D communications based on the proposed model using the theory of NC. The performance bounds including service rate, delay, and backlog are derived. We have also conducted thorough numerical simulations to verify our proposed model. The results have shown that the D2D performance bounds are affected not only by its service rate but by the arrival rate of the high-priority flows and cellular request flow as well. Meanwhile, the cellular requests are only impacted by their own arrival and service rate. The performance bound of cellular requests is significantly better than that of D2D communications, and the performance bound of high-priority D2D requests is also significantly higher than that of low-priority requests. The results have further confirmed the theoretical analysis, thus demonstrating the validity of the proposed model and the associated analyses.

REFERENCES

[1] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[2] S. Abeta, "Toward LTE commercial launch and future plan for LTE enhancements (LTE-Advanced)," in *Proc. IEEE ICCS*, Nov. 2010, pp. 146–150.

[3] 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.

[4] L. Wei, R. Hu, Y. Qian, and G. Wu, "Enable device-to-device communications underlaying cellular networks: Challenges and research aspects," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 90–96, Jun. 2014.

[5] D. Feng *et al.*, "Device-to-device communications in cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 49–55, Apr. 2014.

[6] G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.

[7] 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.

[8] M. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: Challenges, solutions, and future directions," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86–92, May 2014.

[9] F. Boccardi, R. Heath, 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.
[10] G. Fodor, D. Della Penda, M. Belleschi, M. Johansson, and A. Abrardo, "A comparative study of power control approaches for device-to-device communications," in *Proc. IEEE ICC*, Jun. 2013, pp. 6008–6013.
[11] P. Phunchongharn, E. Hossain, and D. Kim, "Resource allocation for device-to-device communications underlaying LTE-advanced networks," *IEEE Wireless Commun.*, vol. 20, no. 4, pp. 91–100, Aug. 2013.
[12] C. Xu *et al.*, "efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 348–358, Sep. 2013.
[13] Y. Pei and Y.-C. Liang, "Resource allocation for device-to-device communications overlaying two-way cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3611–3621, Jul. 2013.
[14] K. Doppler, C.-H. Yu, C. Ribeiro, and P. Janis, "Mode selection for device-to-device communication underlaying an LTE-advanced network," in *Proc. IEEE WCNC*, Apr. 2010, pp. 1–6.
[15] K. Doppler, M. Rinne, C. Wijting, C. 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, Dec. 2009.
[16] C.-H. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.
[17] Y. Li, D. Jin, J. Yuan, and Z. Han, "Coalitional games for resource allocation in the device-to-device uplink underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3965–3977, Jul. 2014.
[18] J. Huang, Y. Zhao, and K. Sohraby, "Resource allocation for intercell device-to-device communication underlaying cellular network: A game-theoretic approach," in *Proc. 23rd ICCCN*, Aug. 2014, pp. 1–8.
[19] J. Xie and M. Xie, "Delay bound analysis in real-time networks with priority scheduling using network calculus," in *Proc. IEEE ICC*, Jun. 2013, pp. 2469–2474.
[20] Y. Qian, Z. Lu, and Q. Dou, "QoS scheduling for NoCs: Strict priority queueing versus weighted round robin," in *Proc. ICCD*, Oct. 2010, pp. 52–59.
[21] W. Sofack and M. Boyer, "Non preemptive static priority with network calculus," in *Proc. Int. Conf. ETFA*, Sep. 2011, pp. 1–8.
[22] J. Li, H. Guan, J. Yao, G. Zhu, and X. Liu, "Performance enhancement and optimized analysis of the worst case end-to-end delay for AFDX networks," in *Proc. Int. Conf. GreenCom*, Nov. 2012, pp. 301–310.
[23] Y. Hua and X. Liu, "Scheduling design and analysis for end-to-end heterogeneous flows in an avionics network," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 2417–2425.
[24] Q. Zhang, Y. Cai, D. Gu, and W. Zhang, "Determine the maximum closed-loop control delay in switched industrial Ethernet using network calculus," in *Proc. Amer. Control Conf.*, Jun. 2006, pp. 4870–4875.
[25] J. Schmitt and F. Zdarsky, "A case for simplicity in providing network quality of service: Class-based strict priority queueing," in *Proc. IEEE ICON*, Nov. 2004, pp. 809–813.
[26] A. Bouillard and A. Junier, "Worst-case delay bounds with fixed priorities using network calculus," in *Proc. Int. ICST Conf. Perform. Eval. Methodologies Tools*, May 2011, pp. 381–390.
[27] S. Fang, L. Yu, J. Zheng, and Q. Zhu, "Stochastic QoS performance analysis of DiffServ-based wireless sensor network," in *Proc. IEEE/CIC ICCC*, Aug. 2013, pp. 270–274.
[28] C. Luo and L. Yu, "Performance bounds analysis on different scheduling algorithms for networks with self-similar traffic," in *Proc. IEEE ICCC*, Aug. 2012, pp. 103–108.
[29] Y. Gao and Y. Jiang, "Performance analysis of a cognitive radio network with imperfect spectrum sensing," in *Proc. IEEE Conf. INFOCOM Workshops*, Mar. 2010, pp. 1–6.
[30] J. Huang *et al.*, "Multi-priority scheduling using network calculus: Model and analysis," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 1486–1491.
[31] 3GPP, "Study on LTE device to device proximity services," Third-Generation Partnership Project, Sophia Antipolis Cedex, France, 2014. [Online]. Available: http://www.3gpp.org/ftp/specs/archive/36_series/36.843/
[32] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.
[33] R. Cruz, "A calculus for network delay, Part I: Network elements in isolation," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114–131, Jan. 1991.
[34] R. Cruz, "A calculus for network delay, Part II: Network analysis," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 132–144, Jan. 1991.
[35] R. Agrawal, R. L. Cruz, C. Okino, and R. Rajan, "Performance bounds for flow control protocols," *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, pp. 310–323, Jun. 1999.
[36] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. London, U.K.: Springer-Verlag, 2008.
[37] J.-Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. New York, NY, USA: Springer-Verlag, 2001, vol. 2050.
[38] J.-Y. Le Boudec, "Application of network calculus to guaranteed service networks," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1087–1096, May 1998.
[39] OMNeT++, 2014. [Online]. Available: http://www.omnetpp.org/

Jun Huang (M'12) received the B.S. degree in computer science from Hubei University of Automotive Technology, Shiyan, China, in 2005; the M.S. degree (with honor) in computer science from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2009; and the Ph.D. degree (with honor) from the Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China, in 2012.

He is currently an Associate Professor with the Department of Communication and Information Engineering, Chongqing University of Posts and Telecommunications. From September 2010 to September 2011, he was a Visiting Researcher with the Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan. From December 2013 to December 2014, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, South Dakota School of Mines and Technology, Rapid City, SD, USA. He has published more than 50 refereed journal/conference papers. His current research interests include network optimization and control, quality of service, etc.

Dr. Huang is a member of the Association for Computing Machinery (ACM). He received the Best Paper Runner-Up Award from the ACM Symposium on Applied Computing 2014 and the Best Paper Award from the Asia Future Internet Forum (AsiaFI) 2011.

Yi Sun is currently working toward the M.S. degree with the Department of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China.

His current research interests include device-to-device optimization.

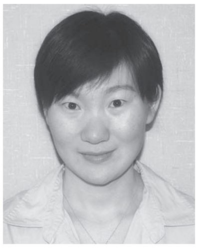
Zi Xiong received the B.E. and M.S. degrees in communication and information engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2011 and 2014, respectively.

Her current research interest includes device-to-device and network calculus.

Qiang Duan (M'03) received the B.S. degree in electrical and computer engineering and the M.S. degree in telecommunications and electronic systems and the Ph.D. degree in electrical engineering from The University of Mississippi, University, MS, USA.

He is an Associate Professor with The Pennsylvania State University, Abington, PA, USA. His general research interest involves data communications and computer networking. Currently, his active research areas include future Internet architecture, network virtualization, network as a service, software-defined network, and cloud computing. He has published over 50 journal articles and conference papers and authored five book chapters.

Dr. Duan serves as an Associate Editor for several international research journals and has served on the technical program committees for numerous international conferences, including the IEEE Global Telecommunications Conference, the IEEE International Conference on Communications, the IEEE International Conference on Computer Communications and Networks, the IEEE International Conference on Advanced Information Networking and Applications (AINA), the IEEE Wireless Communications and Networking Conference, etc.



Yanxiao Zhao (M'10) received the Ph.D. degree from Old Dominion University, Norfolk, VA, USA, in May 2012.

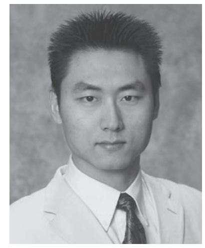
She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, South Dakota School of Mines and Technology, Rapid City, SD, USA, which she joined in August 2012. Her research interests include protocol design, performance evaluation, and experimental implementation for various communication and networking systems. She has published over 30 refereed journal and conference papers.

Dr. Zhao is a member of the Association for Computing Machinery. She received the Best Paper Award at the 2009 International Conference on Wireless Algorithms, Systems, and Applications (WASA). She serves as an Editor for several international journals and as a Technical Program Committee Member for several international conferences.

Xianghui Cao (S'08–M'11) received the B.S. and Ph.D. degrees in control science and engineering from Zhejiang University, Hangzhou, China, in 2006 and 2011, respectively.

During 2007–2009, he was a Visiting Scholar with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL, USA. Since 2012, he has been a Senior Research Associate with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, USA. He is also with the State Key Laboratory of Industrial Control Technology, Zhejiang University. His research interests include cyberphysical systems, wireless network performance analysis, networked estimation and control, and network security.

Dr. Cao received the Best Paper Runner-Up Award from the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) in 2014. He serves as an Associated Editor for several journals, including the Korean Society for Internet Information (KSII) *Transactions on Internet and Information Systems*, the *Security and Communication Networks* (Wiley), and the *International Journal of Ad Hoc and Ubiquitous Computing*; as a Publicity Cochair for ACM MobiHoc 2015; and as a Technical Program Committee Member for a number of conferences, including the IEEE Global Telecommunications Conference, the IEEE International Conference on Communications, the IEEE Vehicular Technology Conference, etc.



Wei Wang (M'10) received the Ph.D. degree in computer engineering from the University of Nebraska–Lincoln, Lincoln, NE, USA, in 2009.

He is an Assistant Professor with the Department of Computer Science, San Diego State University, San Diego, CA, USA. His major research interests include wireless sensor networks, multimedia computing, information security, and educational robotics.

Dr. Wang received two Best Paper Awards from the IEEE Wireless Communications and Networking Conference in 2008 and the Annual Simulation Symposium (ANSS) in 2011. He serves as an Associate Editor for the Wiley *Security in Communication Networks* Journal; as the Guest Editor for three Special Issues of the Hindawi *International Journal of Distributed Sensor Networks* (IJDSN) on Energy-Efficient Sensor Networks, Underwater Wireless Sensor Networks, and Data Dissemination in Vehicular Environments; as the Program Chair for the Association for Computing Machinery Research in Adaptive and Convergent Systems during 2014–2015; as the Workshop Cochair for the Institute for Computer Sciences, Social Informatics, and Telecommunications Engineering (ICST) International Conference on Body Area Networks (BodyNets) for 2013; as the Chair for the IEEE International Conference on Computer and Information Technology Mobile Multimedia Communications (CIT-MMC) Track for 2012; as the Vice Chair for the IEEE International Conference on Communication Technology Next-Generation Networking and Future Internet (ICCT-NGN) Track for 2011; as the Program Chair for the ICST International Workshop on Mobile Multimedia Networking in conjunction with Mobilware (IWMMN) for 2010; and as a Technical Program Committee Member for many international conferences, such as the IEEE Global Telecommunications Conference, the IEEE International Conference on Communications, and the IEEE Wireless Communications and Networking Conference.