This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

Delay-Optimal Dynamic Mode Selection and Resource Allocation in Device-to-Device Communications - Part I: Optimal Policy

Lei Lei *Member, IEEE*, Yiru Kuang, Nan Cheng *Student Member, IEEE*, Xuemin (Sherman) Shen *Fellow, IEEE*, Zhangdui Zhong and Chuang Lin *Senior Member, IEEE*

***Abstract*—In this paper (Part I and Part II), we investigate the optimal dynamic mode selection and resource allocation to minimize the average end-to-end delay under dropping probability constraint for an Orthogonal Frequency Division Multiple Access (OFDMA) cellular network with device-to-device (D2D) communications. Different from the previous studies which mostly focus on infinite backlog traffic model, we consider dynamic data arrival with non-saturated buffers and formulate the resource control problem in D2D communications into an infinite horizon average reward constraint Markov decision process (CMDP) in Part I. The CMDP characterizes the dynamic interference between D2D links and cellular links based on their varying backlogged states, the dynamic route selection, and the coupled interactions between uplink and downlink resource allocations. We propose the general form of the optimal policy. In particular, it is proved that the optimal delay respective to all feasible randomized policies is attained by either a deterministic policy or a simple mixed policy which randomizes between two deterministic policies. Therefore, the determination of optimal randomized policy essentially becomes the determination of one or two deterministic policies, which can be obtained by an equivalent Bellman's equation with reduced state space. Simulation results show that the optimal policy based on the CMDP model outperforms the conventional CSI-only scheme and throughput-optimal scheme in stability sense.**

***Index Terms*—Device-to-Device Communication; Mode Selection; Resource Allocation; Markov Decision Process**

I. Introduction

Device-to-device (D2D) communications commonly refer to a type of technologies that enable devices to communicate directly with each other without the communication infrastructure, e.g., access points (APs) or base stations (BSs). Bluetooth and WiFi-Direct are the two most popular D2D techniques, both working in the unlicensed industrial, scientific and medical (ISM) bands. Cellular networks, on the other hand, do not support direct over-the-air communications between user devices. However, with the emergence of context-aware applications and the accelerating growth of Machine-to-Machine (M2M) applications, D2D communications play a more important role since it facilitates the discovery of geographically close devices and enables direct communications between these proximate devices so as to increase communication capability and reduce communication delay and power consumption [1]–[4]. To seize the emerging market that requires D2D communications, the mobile operators and vendors are accepting D2D as a part of fourth generation (4G) Long Term Evolution (LTE)-Advanced standard in 3rd Generation Partnership Project (3GPP) Release 12 [5]. Moreover, as the future fifth generation (5G) cellular networks are envisioned to support 100 times higher number of connecting devices and user data rate, D2D is also considered as one of the pieces of the 5G jigsaw puzzle in order to offload proximity services from the cellular networks [6], [7]. Compared with traditional D2D techniques, e.g., Bluetooth and WiFi-Direct, network assisted D2D communications can work in the licensed band of cellular networks with more controllable interference. Moreover, the network infrastructure can assist the user equipments (UEs) in various key functions of D2D communications, such as new peer discovery, physical layer procedures, and radio resource control, which make it different from traditional D2D technologies, such as WiFi direct.

Mode selection and resource allocation are two important resource control functions in network assisted D2D communications. Compared with the resource control problem in traditional cellular networks, there are a number of unique issues to address to obtain resource optimization in D2D communications.

1) *Route Selection and intra-cell resource reuse*: A pair of D2D user equipments (UEs) can either communicate directly over-the-air using the D2D Mode, or communicate via the base station (BS) using the Cellular Mode. Specifically, the data between the D2D UEs will be routed along a one-hop route of D2D link (direct over-the-air link) in D2D Mode and a two-hop route of cellular links in Cellular Mode. Moreover, if a pair of D2D UEs work in the D2D Mode, the D2D link may reuse radio resources with other cellular or D2D links in order to improve resource utilization, if the interference between these links is acceptable. Two resource sharing modes

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Manuscript received Jan. 12, 2015; revised March 22, 2015; accepted June 8, 2015. This work was supported by the National Natural Science Foundation of China (No. 61272168, No. U1334202, No. 61472199), the State Key Laboratory of Rail Traffic Control and Safety (No. RCS2014ZT10), Beijing Jiaotong University, and the Key Grant Project of Chinese Ministry of Education (No. 313006).

L. Lei, Y. Kuang and Z. Zhong are with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China.

N. Cheng and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada.

C. Lin is with the Department of Computer Science and Technology, Tsinghua University, Beijing, China.

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

are defined as D2D overlay Mode and D2D underlay Mode, depending on whether D2D links and cellular links use orthogonal or non-orthogonal resources. Therefore, mode selection is needed to select the optimal mode (i.e., D2D overlay/D2D underlay/Cellular Mode) for a pair of D2D UEs either semi-statically at the time-scale of connection establishement/release, or dynamically per time slot [8]. Dynamic mode selection can capture and utilize the fast fading effects of wireless channels opportunistically to improve performance, while it involves more computation complexity and communication overhead. Dynamic mode selection is performed jointly with resource allocation, which is responsible for selecting the set of links for data transmissions at the beginning of each time slot. This poses new challenges for resource allocation function to deal with the route selection and intra-cell interference problems, which do not exist in current fourth generation (4G) Long Term Evolution (LTE) cellular networks.

2) *Joint uplink and downlink resource optimization*: If a pair of D2D UEs work in the Cellular Mode, the end-to-end performance of the two-hop route including one cellular uplink and one cellular downlink should be optimized, instead of separately optimizing the performance of each hop. This joint resource optimization of uplink and downlink transmissions are not considered in traditional cellular networks.

When dealing with the above issues, most related research works assume that the D2D and cellular users are saturated with infinite backlogs and focus only on optimizing the PHY layer performance metrics such as sum throughput and power consumption, where the resource control functions only consider the Channel State Information (CSI) information. In practice, data arrival process at the users is dynamic, and performance metrics such as delay and dropping probability are also very important, especially for real-time and delay-sensitive services, such as voice conversation, video streaming, and interactive gaming [9]. Existing research on resource allocation and scheduling in wireless networks [10] show that algorithms under the infinite backlog traffic model considering only the channel state information are not sufficient to ensure queue stability or guarantee packet delay/Quality-of-Service (QoS) requirement under the dynamic packet arrival setting. Therefore, both the CSI and queue state information (QSI) should be taken into account in the resource control policies. By making use of the QSI information, the resources can be allocated more judiciously as the users' instantaneous transmission requirements are also considered in addition to their transmission capabilities. For example, we can avoid allocating resources to a user with a good channel state but few data in the queue. Therefore, system performance in terms of throughput, delay and dropping probability etc. can all be improved.

Delay-aware resource control with bursty traffic has received little attention in network assisted D2D communications, since it is a non-trivial problem involving both queueing theory (to model the queue dynamics) and information theory (to model the physical layer dynamics). In this paper, we consider an Orthogonal Frequency Division Multiple Access (OFDMA) cellular network with one BS, multiple D2D UE pairs, and cellular UEs with uplink or downlink transmission. Our objective is to design an optimal dynamic mode selection and resource allocation algorithm to minimize the average end-to-end delay under the constraint of packet dropping probability for network assisted D2D communications with bursty traffic. Specifically, the contributions of this paper mainly lie in the following aspects:

1) *Queuing Model Formulation*: We develop a queuing model whose underlying system state dynamics evolves as a controlled Markov chain, where the system state includes the joint queue state of the queues at the UEs for uplink transmission and the queues at the BS for downlink transmission as well as the joint channel state of all the D2D links, cellular uplinks and cellular downlinks. *The main contribution of the queuing model lies in the introduction of two important concepts to characterize the unique features of D2D communications. The first concept is radio resource group (RRG), which defines a group of links that may reuse radio resources to characterize the intra-cell resource reuse. Therefore, the channel state of a link can be represented by a tuple including its Adaptive Modulation and Coding (AMC) states in all the RRGs that this link belongs to. The second concept is link constraint set of a queue, which defines the set of servers for the queue in different routes to characterize route selection.*
2) *CMDP Framework*: Based on the queuing model, a general constrained Markov Decision Process (CMDP) framework for the dynamic optimization of mode selection and resource allocation in D2D communications over frequency-selective fading channel with AMC scheme in the physical layer under bursty traffic model is provided. *The main contributions of the CMDP framework include: (1) The transition kernel of the controlled Markov chain takes into account the coupling relationship between the uplink and downlink resource allocation, which is a unique feature in D2D communications. (2) The cost function is given based on the closed-form expressions for end-to-end performance metrics such as average delay and dropping probability as functions of steady-state probabilities of the controlled Markov chain. (3) Although we focus on the delay-optimal resource control with dropping probability constraint in this paper, the framework can also be applied to study other CMDP problems with different optimization objectives and constraints, e.g., maximize the sum throughput subject to the delay constraint.*
3) *General Form of Optimal Policy*: We utilize the Lagrangian approach to turn the CMDP problem into an unconstraint Markov Decision Process (MDP) problem, and establish the strong duality result over the space of randomized policy. *The contribution related to the determination of the optimal policy mainly lies*

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

in providing its general form, which allows us to solve the MDP problem over the space of deterministic policy instead of the space of randomized policy. Specifically, we prove the existence of an optimal policy, which is either a deterministic policy or a mix of two deterministic policies, equivalent to choosing independently one of two deterministic policies at each epoch by the toss of a (biased) coin. To solve the MDP model over the space of deterministic policy, we derive an equivalent Bellman's equation with reduced state space. We show by simulations that the optimal policy based on the equivalent Bellman's equation achieves significant gain compared to various baselines such as the conventional CSI-only control and the throughput optimal control (MaxWeight algorithm).

The remainder of the paper is organized as follows. The related work is summarized in Section II. We develop a general network model for network assisted D2D communications with nodes, links, connections and queues in Section III. In Section IV, we formulate a queuing model and identify the system state space and action space for the underlying controlled Markov process, whose transition kernel is derived. In Section V, we elaborate the MDP problem formulation for dynamic mode selection and resource allocation and derive the optimal control policy by offline value iteration algorithm. In Section VI, we discuss the simulated performance. Finally, we highlight the main results in Section VII.

II. Related Work

A. Resource Control for D2D Communications

Resource control for D2D communications has been widely studied in recent years [11]–[19]. The authors in [11] consider the case where one cellular UE and a pair of D2D UEs share the the radio resources, and propose to make mode selection decision based on the estimated throughput performance of D2D overlay, D2D underlay and Cellular Modes, assuming the optimum power control and resource allocation algorithms are adopted under every mode. A more general network model with multiple cellular UEs and D2D UE pairs is considered in [12], where the UE positions are modeled by random spatial Poisson point process. The authors in [12] derive an optimal distance threshold for choosing between D2D Mode and Cellular Mode that minimizes the transmit power, and an optimal fraction of spectrum that should be dedicated to/shared with D2D links for D2D overlay/D2D underlay Mode under a weighted proportional fair utility function. The authors in [13] address the problem of dynamic mode selection and resource allocation, and mainly focus on the interference control and management between D2D links and cellular links such that they can efficiently reuse the radio resources whenever the interference is small. In [14], the joint mode selection and resource allocation problem is studied to maximize the overall system throughput while guaranteeing the Signal to Interference and Noise Ratio (SINR) of both D2D and cellular links. However, all the above research works assume the infinite backlog traffic model without considering important QoS metrics such as delay and dropping probability.

B. Delay-Aware Resource Control in Wireless Networks

In general, there are various approaches to deal with delay-aware control problem. The first approach converts average delay constraints into equivalent average rate constraints using the large deviation theory and solves the optimization problem using a purely information theoretical formulation based on the rate constraints [20]. While this approach allows potentially simple solutions, the resulting control policies are only functions of the CSI and such policies are good only for the large delay regime where the probability of empty queues is small. The second approach utilizes the notion of Lyapunov stability [10] and the derived MaxWeight algorithm can be directly adapted to the D2D scenario. Compared with the first approach, the derived policies in the second approach are adaptive to both the CSI and QSI and are throughput-optimal (in stability sense). However, stability is only a weak form of delay performance and derived policies may not have good delay performance especially in the small delay regime. A more systematic approach in dealing with delay-optimal resource control in general delay regime is the MDP approach [21]. However, there is little work in existing literature that applies the MDP model to address the resource control problem in network assisted D2D communications [9], [22].

III. Network Model

In this section, we develop a general network model for dynamic mode selection and resource allocation in network assisted D2D communications. We consider a Frequency Division Duplex (FDD) OFDMA system. The whole uplink or downlink spectrum is divided into N_{F} equal size subchannels. A subchannel in the uplink (resp. downlink) spectrum shall be referred to as uplink (resp. downlink) subchannel in the rest of the paper. Moreover, we assume that D2D links share uplink resources with cellular uplinks [1]. We focus on the intra-cell interference between D2D links and cellular links, and adopt the static inter-cell interference model [23], which assumes that inter-cell interference is randomized and/or fractional frequency reuse is used to mitigate inter-cell interference, so that the sum of total inter-cell interference power can be simply treated as another Gaussian-like noise. The important symbols used in this paper is summarized in Table I.

A. Nodes, Links, and Connections

Consider an OFDMA cellular network with D2D communications capability, where there are D D2D UE pairs, C_{u} cellular UEs (CUEs) with uplink communications and C_{d} CUEs with downlink communications in a single cell. A D2D UE pair consists of a source D2D UE (src. DUE) and a destination D2D UE (dest. DUE) within direct over-the-air communications range with each other, which is formed through the various neighbor/peer/service discovery mechanisms proposed in literature. Fig.1 illustrates a simple example with $D = C_{\mathrm{u}} = C_{\mathrm{d}} = 1$, i.e., there are one pair of DUEs (src. DUE 1 and dest. DUE 2), one uplink CUE (CUE 3), and one downlink CUE (CUE 4). Time is slotted and each time slot has an equal length.

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

TABLE I
SUMMARY OF IMPORTANT SYMBOLS USED

Category	Symbol	Definition
Constant	N_{F}	The number of subchannels in the uplink or downlink spectrum
	D	The number of D2D UE pairs
	C_{u}	The number of cellular UEs with uplink communications
	C_{d}	The number of cellular UEs with downlink communications
	C	The number of connections
	N_{Q}	The queue capacity in number of bits or packets
Set	$\mathcal{N}$	The set of nodes i, $\mathcal{N} := \{0, 1, ..., N\} = \{0\} \bigcup \mathcal{N}_{\mathrm{D}} \bigcup \mathcal{N}_{\mathrm{Cu}} \bigcup \mathcal{N}_{\mathrm{Cd}}$
	$\mathcal{N}_{\mathrm{D}}$	The set of D pairs of DUEs, $\mathcal{N}_{\mathrm{D}} := \{1, \ldots, 2D\}$
	$\mathcal{N}_{\mathrm{Ds}}$	The set of D src. DUEs, $\mathcal{N}_{\mathrm{Ds}} := \{1, 3, \ldots, 2D-1\}$
	$\mathcal{N}_{\mathrm{Dd}}$	The set of D dest. DUEs, $\mathcal{N}_{\mathrm{Dd}} := \{2, 4, \ldots, 2D\}$
	$\mathcal{N}_{\mathrm{Cu}}$	The set of C_{u} uplink CUEs, $\mathcal{N}_{\mathrm{Cu}} := \{2D+1, \ldots, 2D+C_{\mathrm{u}}\}$
	$\mathcal{N}_{\mathrm{Cd}}$	The set of C_{d} downlink CUEs, $\mathcal{N}_{\mathrm{Cd}} := \{2D+C_{\mathrm{u}}+1, \ldots, 2D+C_{\mathrm{u}}+C_{\mathrm{d}}\}$
	$\mathcal{L}$	The set of eligible transmission links (i,j), $\mathcal{L} := \mathcal{L}_{\mathrm{D}} \bigcup \mathcal{L}_{\mathrm{Cu}} \bigcup \mathcal{L}_{\mathrm{Cd}}$
	$\mathcal{L}_{\mathrm{D}}$	The set of D eligible D2D links, $\mathcal{L}_{\mathrm{D}} := \{(i, i+1) \vert i \in \mathcal{N}_{\mathrm{Ds}}\}$
	$\mathcal{L}_{\mathrm{Cu}}$	The set of eligible cellular uplinks, $\mathcal{L}_{\mathrm{Cu}} := \{(i,0) \vert i \in \mathcal{N}_{\mathrm{Cu}} \bigcup \mathcal{N}_{\mathrm{Ds}}\}$
	$\mathcal{L}_{\mathrm{Cd}}$	The set of eligible cellular downlinks, $\mathcal{L}_{\mathrm{Cd}} := \{(0,i) \vert i \in \mathcal{N}_{\mathrm{Cd}} \bigcup \mathcal{N}_{\mathrm{Dd}}\}$
	$\mathcal{C}$	The set of connections $\mathcal{C} := \{1, \ldots, C\} = \mathcal{C}_{\mathrm{D}} \bigcup \mathcal{C}_{\mathrm{Cu}} \bigcup \mathcal{C}_{\mathrm{Cd}}$
	$\mathcal{C}_{\mathrm{D}}$	The set of D D2D connections, $\mathcal{C}_{\mathrm{D}} = \{1, \ldots, D\}$
	$\mathcal{C}_{\mathrm{Cu}}$	The set of C_{u} cellular uplink connections, $\mathcal{C}_{\mathrm{Cu}} = \{D+1, \ldots, D+C_{\mathrm{u}}\}$
	$\mathcal{C}_{\mathrm{Cd}}$	The set of C_{u} cellular downlink connections, $\mathcal{C}_{\mathrm{Cd}} = \{D+C_{\mathrm{u}}+1, \ldots, D+C_{\mathrm{u}}+C_{\mathrm{d}}\}$
	$\mathcal{L}_c$	The set of all links (i,j) that connection c data is allowed to use
	Θ	The set of queues in the system
	$\mathcal{L}_i^{(c)}$	The set of links (i,j) that serve queue $q_i^{(c)}$
	$\mathcal{B}_u$	A RRG, which is a set of links that can be scheduled for transmission simultaneously
	$\mathcal{U}$	The index set of RRGs
	$\mathcal{U}_{ij}$	The index set of RRGs that contain link (i,j), $\mathcal{U}_{ij} := \{u \vert (i,j) \in \mathcal{B}_u, u \in \mathcal{U}\}$
	$\mathcal{S}$	The system state space
	$\mathcal{H}$	The channel state space
	$\mathcal{Q}$	The queue state space
	$\mathcal{A}_{\mathrm{x}}$	The action space
Variable	(i,j)	A transmission link from node i to node j
	$A_{c,t}$	The amount of new connection c data arrived to its source node during time slot t
	λ_{c}	The mean arrival rate of connection c
	$q_i^{(c)}$	The queue maintained at node i for connection c
	$Q_{i,t}^{(c)}$	The length of $q_i^{(c)}$ at the beginning of time slot t
	$x_{u,t}^{(m)}$	The subchannel allocation for RRG $\mathcal{B}_u$ at time slot t, $x_{u,t}^{(m)} \in \{0,1\}$
	$r_{i,t}^{(c)}$	The instantaneous data rate of queue $q_i^{(c)}$ during time slot t
	$r_{ij,t}^{(m,u)}$	The instantaneous data rate of link (i,j) on subchannel m when RRG $\mathcal{B}_u$ is scheduled
	$SINR_{ij,t}^{(m,u)}$	The SINR of link (i,j) on subchannel m when RRG $\mathcal{B}_u$ is scheduled
State	$\mathbf{S}_t$	The global system state at time slot t, $\mathbf{S}_t = (\mathbf{H}_t, \mathbf{Q}_t)$
	$\mathbf{Q}_t$	The QSI at time slot t, $\mathbf{Q}_t := \{Q_{i,t}^{(c)} \vert q_i^{(c)} \in \Theta\}$
	$\mathbf{H}_t$	The CSI at time slot t, $\mathbf{H}_t := \{\mathbf{H}_{ij,t} \vert (i,j) \in \mathcal{L}\}$
	$\mathbf{H}_{ij,t}$	The CSI of link (i,j) at time slot t
Action	$\mathbf{x}$	The subchannel allocation action, $\mathbf{x} := \{x_u^{(m)} \in \{0,1\} \vert m \in \{1, 2, \ldots, N_{\mathrm{F}}\}, u \in \mathcal{U}\}$

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

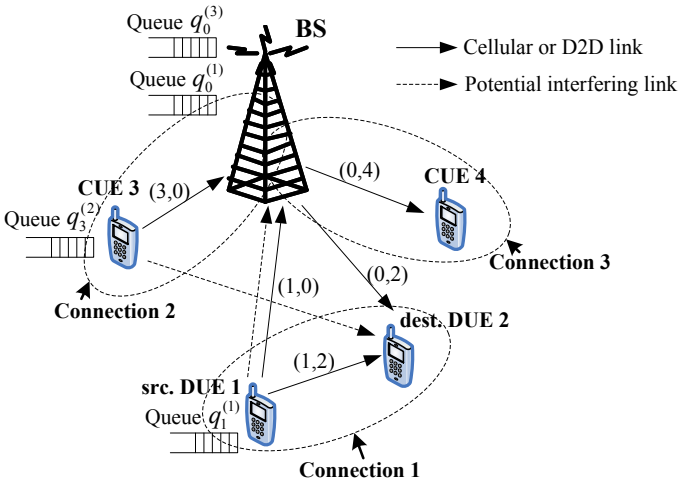

Fig. 1. A simple OFDMA cellular network with D2D communications.

The above OFDMA cellular network with D2D communications can be formulated as a general network model with a set $\mathcal{N}$ of nodes and a set $\mathcal{L}$ of transmission links. Define $\mathcal{N} := \{0, 1, \ldots, N\}$, where node 0 represents the base station (BS) and nodes $1, \ldots, N$ represent the UEs. Let $\mathcal{N}_{\rm D} := \{1, \ldots, 2D\}$ be the set of DUEs of the D D2D pairs. Let $\mathcal{N}_{\rm Ds} := \{1, 3, \ldots, 2D-1\}$ and $\mathcal{N}_{\rm Dd} := \{2, 4, \ldots, 2D\}$ be the sets of src. DUEs and dest. DUEs, respectively, where node $i \in \mathcal{N}_{\rm Ds}$ (resp. $j \in \mathcal{N}_{\rm Dd}$) is the src. DUE (resp. dest. DUE) of D2D pair $\lceil i/2 \rceil$ (resp. $j/2$). Let $\mathcal{N}_{\rm Cu} := \{2D+1, \ldots, 2D+C_{\rm u}\}$ and $\mathcal{N}_{\rm Cd} := \{2D+C_{\rm u}+1, \ldots, 2D+C_{\rm u}+C_{\rm d}\}$ be the set of $C_{\rm u}$ uplink CUEs and $C_{\rm d}$ downlink CUEs, respectively (with $N = 2D+C_{\rm u}+C_{\rm d}$). We use i or j to denote the index of a node within $\mathcal{N}$ (i.e., $i, j \in \mathcal{N}$) in the rest of the paper.

Each transmission link represents a communication channel for direct transmission from a given node i to another node j, and is labeled by (i, j) (where $i, j \in \mathcal{N}$). Note that link (i, j) is distinct from link (j, i). The link set $\mathcal{L}$ is composed of three non-overlapping subsets, where $\mathcal{L}_{\rm D} := \{(i, i+1) | i \in \mathcal{N}_{\rm Ds}\}$ is the set of D2D links, $\mathcal{L}_{\rm Cu} := \{(i, 0) | i \in \mathcal{N}_{\rm Cu} \bigcup \mathcal{N}_{\rm Ds}\}$ is the set of cellular uplinks, and $\mathcal{L}_{\rm Cd} := \{(0, i) | i \in \mathcal{N}_{\rm Cd} \bigcup \mathcal{N}_{\rm Dd}\}$ is the set of cellular downlinks.

All data that enter the network are associated with a particular connection which defines the source and destination of the data. Let $\mathcal{C}_{\rm D} = \{1, \ldots, D\}$ represent the set of D D2D connections, $\mathcal{C}_{\rm Cu} = \{D+1, \ldots, D+C_{\rm u}\}$ represent the set of $C_{\rm u}$ cellular uplink connections, and $\mathcal{C}_{\rm Cd} = \{D+C_{\rm u}+1, \ldots, D+C_{\rm u}+C_{\rm d}\}$ represent the set of $C_{\rm d}$ cellular downlink connections. Define $\mathcal{C} := \{1, \ldots, C\} = \mathcal{C}_{\rm D} \bigcup \mathcal{C}_{\rm Cu} \bigcup \mathcal{C}_{\rm Cd}$ (with $C = D + C_{\rm u} + C_{\rm d}$) as the set of all connections in the network. We use c to denote the index of a connection within $\mathcal{C}$ (i.e., $c \in \mathcal{C}$) in the rest of the paper.

Define the link constraint set for a connection c as the set of all links that the connection data is allowed to use. Obviously, $\mathcal{L}_c = \{(D+c, 0)\}$ for any cellular uplink connections $c \in \mathcal{C}_{\rm Cu}$, and $\mathcal{L}_c = \{(0, D+c)\}$ for any cellular downlink connections $c \in \mathcal{C}_{\rm Cd}$, since there is only a single-hop route between the CUE and the BS for these connections. For the D2D connection, since the data can be transmitted either via the single hop route of D2D link or the two-hop route of cellular links, and the decision is made dynamically at each time slot, we have $\mathcal{L}_c = \{(2c-1, 2c), (2c-1, 0), (0, 2c)\}$. The source node and destination node of a connection along with its link constraint set are given in Table II.

B. Queues

The data from a connection c is transmitted hop by hop along the route(s) of the connection to its destination node. Each node i along the route(s) of connection c maintains a queue $q_i^{(c)}$ for storing its data except for the destination node, since the data is considered to exit the network once it reaches the destination. Define Θ as the set of queues in the system. We assume each queue has a finite capacity of $N_{\rm Q} < \infty$ (in number of bits or packets).

The set of queues can be divided into two non-overlapping disjoint sets according to whether a queue is maintained by an UE or the BS. Every src. DUE ($i = 2c-1$) and uplink CUE ($i = c+D$) maintains a queue $q_i^{(c)}$ for the corresponding D2D connection or uplink cellular connection $c \in \mathcal{C}_{\rm D} \bigcup \mathcal{C}_{\rm Cu}$, which is referred to as an *uplink queue*. On the other hand, the BS maintains a set of queues $q_0^{(c)}$ for all the downlink cellular connections $c \in \mathcal{C}_{\rm Cd}$ and D2D connections $c \in \mathcal{C}_{\rm D}$, which are referred to as *downlink queues*. Let $\Theta_{\rm u}$ and $\Theta_{\rm d}$ denote the set of uplink queues and downlink queues, respectively.

1) Mapping between connections and queues: Every cellular connection has only one queue. Define $\Theta_{\rm Cu} = \{q_{(c+D)}^{(c)} | c \in \mathcal{C}_{\rm Cu}\}$ and $\Theta_{\rm Cd} = \{q_0^{(c)} | c \in \mathcal{C}_{\rm Cd}\}$ as the set of queues for cellular uplink connections and cellular downlink connections, respectively. On the other hand, every D2D connection has two queues including one uplink queue and one downlink queue. Define $\Theta_{\rm D-u} = \{q_{(2c-1)}^{(c)} | c \in \mathcal{C}_{\rm D}\}$ and $\Theta_{\rm D-d} = \{q_0^{(c)} | c \in \mathcal{C}_{\rm D}\}$ as the set of uplink queues and downlink queues for D2D connections, respectively. Obviously, we have $\Theta_{\rm u} = \Theta_{\rm Cu} \bigcup \Theta_{\rm D-u}$ and $\Theta_{\rm d} = \Theta_{\rm Cd} \bigcup \Theta_{\rm D-d}$.

2) Mapping between queues and links: Define the per queue link constraint set of a queue $q_i^{(c)}$ as $\mathcal{L}_i^{(c)} = \{(i, j) \in \mathcal{L}_c\}$, i.e., all the links from node i within link constraint set $\mathcal{L}_c$. The data from a queue $q_i^{(c)}$ can only be transmitted via links in $\mathcal{L}_i^{(c)}$, which is given for different queues in Table III. Note that there is only one link in $\mathcal{L}_i^{(c)}$ for all the queues except those $q_i^{(c)} \in \Theta_{\rm D-u}$, i.e., the uplink queues for D2D connections, which can be served by either D2D link $(2c-1, 2c)$ or cellular uplink $(2c-1, 0)$.

C. Resource Reuse Group

We define a Resource Reuse Group (RRG) $\mathcal{B}_u$ as the subset of links $(i, j) \in \mathcal{L}$ that can be scheduled for transmission simultaneously on any subchannel in a time slot. Therefore, a RRG for an uplink subchannel may contain at most one cellular uplink and one or more D2D links. On the other hand, an RRG for a downlink subchannel can contain one and only one cellular downlink. Let $\mathcal{U}$ represent the set of RRG indexes, $\mathcal{U}_{\rm u}$ and $\mathcal{U}_{\rm d}$ represent the subsets of RRG indexes for uplink and downlink subchannels, respectively. Therefore, we have

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

TABLE II
CONNECTIONS, NODES AND LINK CONSTRAINT SETS

Connection c	Node i		Link Constraint Set $\mathcal{L}_c$
	Source Node	Destination Node	
D2D connection $c \in \mathcal{C}_{\mathrm{D}}$	$2c-1$	$2c$	$\{(2c-1,2c),(2c-1,0),(0,2c)\}$
Cellular uplink connection $c \in \mathcal{C}_{\mathrm{Cu}}$	$c+D$	0	$\{(c+D,0)\}$
Cellular downlink connection $c \in \mathcal{C}_{\mathrm{Cd}}$	0	$c+D$	$\{(0,c+D)\}$

TABLE III
CONNECTIONS, QUEUES AND PER QUEUE LINK CONSTRAINT SETS

Connection c	Uplink Queue $q_i^{(c)} \in \Theta_{\mathrm{u}}$	Per Queue Link Constraint Set $\mathcal{L}_i^{(c)}$	Downlink Queue $q_0^{(c)} \in \Theta_{\mathrm{d}}$	Per Queue Link Constraint Set $\mathcal{L}_0^{(c)}$
Cellular uplink connection $c \in \mathcal{C}_{\mathrm{Cu}}$	$q_{c+D}^{(c)} \in \Theta_{\mathrm{Cu}}$	$\{(c+D,0)\}$		
Cellular downlink connection $c \in \mathcal{C}_{\mathrm{Cd}}$			$q_0^{(c)} \in \Theta_{\mathrm{Cd}}$	$\{(0,c+D)\}$
D2D connection in Hybrid RM $c \in \mathcal{C}_{\mathrm{D}}$	$q_{2c-1}^{(c)} \in \Theta_{\mathrm{D-u}}$	$\{(2c-1,2c),(2c-1,0)\}$	$q_0^{(c)} \in \Theta_{\mathrm{D-d}}$	$\{(0,2c)\}$

$\mathcal{U}=\{\mathcal{U}_{\mathrm{u}},\mathcal{U}_{\mathrm{d}}\}$. We use u to denote the index of a RRG within $\mathcal{U}$ (i.e., $u \in \mathcal{U}$) in the rest of the paper. For any link $(i,j) \in \mathcal{L}$, define $\mathcal{U}_{ij} := \{u|(i,j) \in \mathcal{B}_u, u \in \mathcal{U}\}$ as the index set of RRGs that contain link (i,j).

Assumption 1 (assumption on resource reuse group)**.** *The set of RRGs in the network are determined and updated at the time scale of connection setup/release, when there are links added/deleted from the network. Therefore, it is static at the time scale of subchannel allocation as considered in this paper, where the nodes, links, and connections are fixed in the network model.*

Remark 1 (resource reuse group partitioning)**.** *Since no resource reuse is allowed between cellular downlinks, there is a one-to-one mapping between downlink RRGs and cellular downlinks, i.e., there are* $|\mathcal{U}_{\mathrm{d}}| = |\mathcal{L}_{\mathrm{Cd}}|$ *downlink RRGs, and* $|\mathcal{U}_{ij}| = 1$, $\forall (i,j) \in \mathcal{L}_{\mathrm{Cd}}$. *On the other hand, since resource reuse is allowed between a cellular uplink and multiple D2D links, the total number of uplink RRGs is* $|\mathcal{U}_{\mathrm{u}}| = (|\mathcal{L}_{\mathrm{Cu}}|+1) \times 2^{|\mathcal{L}_D|} - 1$, *which is much larger than the total number of cellular uplinks and D2D links. Therefore, it is desirable to reduce* $|\mathcal{U}_{\mathrm{u}}|$ *in order to reduce the complexity of subchannel allocation algorithm. This can reasonably be achieved by deleting those uplink RRGs with too much interference between the links from* $\mathcal{U}_{\mathrm{u}}$. *For example, one simple method is to delete any RRG from* $\mathcal{U}_{\mathrm{u}}$ *if the average Signal to Interference and Noise Ratio (SINR) of any link within the RRG is below a certain threshold after resource reuse.*

D. Dynamic Mode Selection and Subchannel Allocation

In each time slot, an uplink (resp. downlink) subchannel can be allocated to at most one uplink (resp. downlink) RRG for uplink (resp. downlink) transmission. Let $m \in \{1,\ldots,N_{\mathrm{F}}\}$ denote the index of a subchannel. Note that subchannel m can be either the m-th uplink subchannel or the m-th downlink subchannel. In the rest of the paper, we will not explicitly indicate whether m denotes an uplink or downlink subchannel when no ambiguity shall be caused. Let $x_{u,t}^{(m)} \in \{0,1\}$ denote the subchannel allocation for RRG $\mathcal{B}_u$, $u \in \mathcal{U}$ at time slot t, where $x_{u,t}^{(m)} = 1$ if subchannel m is allocated to RRG $\mathcal{B}_u$, and $x_{u,t}^{(m)} = 0$ otherwise. Note that when RRG $\mathcal{B}_u$ is an uplink (resp. downlink) RRG, i.e., $u \in \mathcal{U}_{\mathrm{u}}$ (resp. $u \in \mathcal{U}_{\mathrm{d}}$), m denotes the m-th uplink (resp. downlink) subchannel in $x_{u,t}^{(m)}$. Therefore, we have the constraint that $\sum_{u\in\mathcal{U}_{\mathrm{u}}} x_{u,t}^{(m)} \le 1$ and $\sum_{u\in\mathcal{U}_{\mathrm{d}}} x_{u,t}^{(m)} \le 1$ for any $m \in \{1,\ldots,N_{\mathrm{F}}\}$. We assume that a RRG is scheduled for transmission only when all its links have non-empty queues.

A queue $q_i^{(c)}$ is scheduled in time slot t when at least one RRG $\mathcal{B}_u$ containing a link (i,j) in its link constraint set $\mathcal{L}_i^{(c)}$ is scheduled on any subchannel. From Table III we know that except for the uplink queues of D2D connections $\{q_{(2c-1)}^{(c)}|c \in \mathcal{C}_{\mathrm{D}}\}$, the per-queue link constraint set $\mathcal{L}_i^{(c)}$ of every other queue contains only one link. When mode selection of a D2D connection c is performed dynamically at each time slot, the problem becomes deciding whether to schedule the D2D link $(2c-1,2c)$ or the cellular uplink $(2c-1,0)$ to serve the queue $q_{(2c-1)}^{(c)}$, which can be solved by designing a subchannel allocation function. Therefore, the delay-optimal dynamic mode selection and subchannel allocation problem can be solved by only considering the design of delay-optimal subchannel allocation algorithm.

Remark 2 (simultaneous selection of D2D Mode and Cellular Mode)**.** *Both D2D link* $(2c-1,2c)$ *and cellular uplink* $(2c-1,0)$ *may be scheduled simultaneously to serve the queue* $q_{(2c-1)}^{(c)}$ *at any time slot, since orthogonal subchannels may be allocated to both links. From the mode selection perspective, this means that both D2D Mode and Cellular Mode can be selected simultaneously at a time slot, although only one of the modes can be selected for a single subchannel.*

E. Instantaneous Data Rate

1) Potential Interference Link: For any link $(i,j) \in \mathcal{L}_{\mathrm{D}} \bigcup \mathcal{L}_{\mathrm{Cu}}$, we define its potential interfering link as the communication channel from the transmitter of any link that belongs to the same RRG with link (i,j) to the receiver of node j. Define $\mathcal{I}_{ij} := \{I_{i'j}|(i',j') \in \mathcal{B}_u \backslash \{i,j\}, u \in \mathcal{U}_{ij}\}$ as the set of potential interfering links of link (i,j), where

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

$I_{i'j}$ denotes the potential interfering link from the transmitter of node i' to the receiver of node j. An interfering link is 'potential' since it only exists when the corresponding RRG is scheduled for transmission. Since there are two categories of links, i.e., transmission links and potential interfering links, all links mentioned are referred to the transmission links by default in the rest of the paper.

2) Instantaneous SINR of Link (i,j): Assume that the instantaneous channel gain comprising the path loss, shadowing and fast fading effects of the wireless channel from the transmitter of node $i \in \mathcal{N}$ to the receiver of node $j \in \mathcal{N}$ on any subchannel m remains constant within a time slot and i.i.d. between time slots, the value of which at time slot t is denoted by $G_{ij,t}^{(m)}$. Let $p_{ij,t}^{(m)}$ be the transmission power of link $(i,j) \in \mathcal{L}$ on subchannel m at time slot t. Assume that every scheduled link on a downlink subchannel always transmits at constant power $p_{ij,t}^{(m)} = P_{\max}^{\mathrm{BS}}/N_{\mathrm{F}}$, where $P_{\max}^{\mathrm{BS}}$ is the maximum transmit power of the base station. Moreover, assume that every scheduled link on an uplink subchannel always transmits at constant power $p_{ij,t}^{(m)} = P_{\max}^{\mathrm{UE}}/N_{\mathrm{F}}$, where $P_{\max}^{\mathrm{UE}}$ is the maximum transmit power of a UE. Let $\gamma_{ij,t}^{(m)} := \frac{p_{ij,t}^{(m)} G_{ij,t}^{(m)}}{N_{ij}^{(m)}}$ represent the Signal to Noise Ratio (SNR) of link $(i,j) \in \mathcal{L}$ on subchannel m, where $N_{ij}^{(m)}$ is the noise power. The SINR value of a link (i,j) on a subchannel m may or may not equal its SNR value, depending on whether there are any other links that are scheduled simultaneously with link (i,j) on that subchannel and thus causing interference to each other. Specifically, considering that RRG $\mathcal{B}_u$ is scheduled at time slot t on subchannel m, i.e., $x_{u,t}^{(m)} = 1$, we have

$$\begin{aligned} SINR_{ij,t}^{(m,u)} &= \frac{p_{ij,t}^{(m)} G_{ij,t}^{(m)}}{N_{ij,t}^{(m)} + \sum_{(i',j') \in \mathcal{B}_u \setminus \{(i,j)\}} p_{i'j',t}^{(m)} G_{i'j,t}^{(m)}} \\ &= \frac{\gamma_{ij,t}^{(m)}}{1 + \sum_{(i',j') \in \mathcal{B}_u \setminus \{(i,j)\}} \gamma_{i'j,t}^{(m)}}, \ \forall \ (i,j) \in \mathcal{B}_u. \end{aligned} \tag{1}$$

3) CSI of Link (i,j): Define $r_{ij,t}^{(m,u)}$ to be the instantaneous data rate of link (i,j) on subchannel m when RRG $\mathcal{B}_u$, $\forall u \in \mathcal{U}_{ij}$ is scheduled. We assume that AMC is used, where the SINR values are divided into K non-overlapping consecutive regions [24]. For any $k \in \{1,\ldots,K\}$, if the SINR value $SINR_{ij,t}^{(m,u)}$ of link (i,j) falls within the k-th region $[\Gamma_{k-1}, \Gamma_k)$, the corresponding data rate $r_{ij,t}^{(m,u)}$ of link (i,j) is a fixed value R_k according to the selected modulation and coding scheme in this state. Obviously, $\Gamma_0 = 0$ and $\Gamma_K = \infty$. Also, we have $R_1 = 0$, i.e., no packet is transmitted in channel state 1 to avoid the high transmission error probability.

Definition 1 (definition of CSI)**.** *Define the CSI of link (i,j) to be $\mathbf{H}_{ij,t} := \{H_{ij,t}^{(m,u)} | m \in \{1,\ldots,N_{\mathrm{F}}\}, u \in \mathcal{U}_{ij}\}$, where $H_{ij,t}^{(m,u)}$ denotes the channel state of link (i,j) on subchannel m when RRG $\mathcal{B}_u$ is scheduled. Specifically, $H_{ij,t}^{(m,u)} = k$ if $SINR_{ij,t}^{(m,u)}$ is between $[\Gamma_{k-1}, \Gamma_k)$. Therefore, we have*

$$r_{ij,t}^{(m,u)} = R_{H_{ij,t}^{(m,u)}}. \tag{2}$$

Remark 3 (CSI per cellular downlink)**.** *Since each downlink RRG contains only one cellular downlink and there is no interference between the cellular downlinks, the SINR value of a cellular downlink equals to its SNR value, and $H_{ij,t}^{(m,u)} = H_{ij,t}^{(m)} = k$ if $\gamma_{ij,t}^{(m)} \in [\Gamma_{k-1}, \Gamma_k)$.*

4) Instantaneous Data Rate of Queues and Links: Let $r_{i,t}^{(c)}$ be the instantaneous data rate of queue $q_i^{(c)}$ during time slot t[1], which is equal to the sum of the instantaneous data rate $r_{ij,t}$ of the scheduled link $(i,j) \in \mathcal{L}_i^{(c)}$ on all the N_{F} uplink or downlink subchannels at time slot t, i.e.,

$$r_{i,t}^{(c)} = \sum_{(i,j) \in \mathcal{L}_i^{(c)}} r_{ij,t}, \tag{3}$$

$$r_{ij,t} = \sum_{m=1}^{N_{\mathrm{F}}} r_{ij,t}^{(m)}, \tag{4}$$

where $r_{ij,t}^{(m)}$ is the instantaneous data rate of link (i,j) on subchannel m at time slot t

$$r_{ij,t}^{(m)} = \sum_{u \in \mathcal{U}_{ij}} x_{u,t}^{(m)} r_{ij,t}^{(m,u)}, \tag{5}$$

and $r_{ij,t}^{(m,u)}$ can be determined by the CSI of link (i,j) according to (2). Therefore, $r_{ij,t}^{(m)} = r_{ij,t}^{(m,u)}$ if RRG $\mathcal{B}_u$ containing link (i,j) is scheduled on subchannel m, and $r_{ij,t}^{(m)} = 0$ if none of the RRGs containing link (i,j) is scheduled on subchannel m.

F. Queuing Dynamics

Let $A_{c,t}$ denote the amount of new connection c data[2] that exogenously arrives to its source node during time slot t. We assume that the data arrival process is i.i.d. over time slots following general distribution $f_A(n)$ with average arrival rate $\mathbf{E}[A_{c,t}] = \lambda_c$. Let $A_{i,t}^{(c)}$ denote the amount of data arrived to node i for connection c during time slot t. When $q_i^{(c)} \in \Theta_{\mathrm{u}} \bigcup \Theta_{\mathrm{Cd}}$, node i is the source node of connection c, and $A_{i,t}^{(c)} = A_{c,t}$. Otherwise, when $q_0^{(c)} \in \Theta_{\mathrm{D-d}}$, it is the second-hop queue of connection c, and $A_{0,t}^{(c)}$ depends on the data departure process of the corresponding uplink transmission on cellular uplink $((2c-1),0)$.

Only when a queue is scheduled shall it move the data out of the queue for transmission. Let $Q_{i,t}^{(c)}$ denote the length of $q_i^{(c)}$ at the beginning of time slot t. If $Q_{i,t}^{(c)}$ is less than $r_{i,t}^{(c)}$ during time slot t, padding bits shall be transmitted along with the data. However, the amount of useful data transmitted from $q_i^{(c)}$ during time slot or the throughput of $q_i^{(c)}$ is defined as

$$T_{i,t}^{(c)} = \min[Q_i^{(c)}, r_i^{(c)}]. \tag{6}$$

Moreover, the amount of useful data transmitted via link (i,j) during time slot or the throughput of link (i,j) is defined

[1]The instantaneous data rate can take units of bits/slot or packets/slot. The latter is appropriate when all the packets have fixed length and the achievable data rates are constrained to integral multiples of the packet size.

[2]The data can take units of bits or packets. The latter is appropriate when all the packets have fixed length.

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

for any link within the link constraint set of queue $q_i^{(c)} \in \Theta_{\rm d} \bigcup \Theta_{\rm Cu}$ as

$$T_{ij,t} = \min[Q_i^{(c)}, r_{ij,t}],\ \forall (i,j) \in \mathcal{L}_i^{(c)}. \tag{7}$$

For any queue $q_i^{(c)} \in \Theta_{\rm D-u}$, there are two links $(2c-1, 0)$ and $(2c-1, 2c)$ within its link constraint set $\mathcal{L}_i^{(c)}$ as given in Table III. Both links may be scheduled simultaneously since different sets of subchannels may be allocated to them. We assume that the data in the queue is first assigned to link $(2c-1, 0)$ and then the remaining data left in the queue (if any) shall be assigned to link $(2c-1, 2c)$. According to the above data assignment rule, we have that $T_{(2c-1)0,t}$ obeys (7), while $\forall q_{(2c-1)}^{(c)} \in \Theta_{\rm D-u}$

$$T_{(2c-1)(2c),t} = \min[Q_i^{(c)} - T_{(2c-1)0,t}, r_{(2c-1)(2c),t}]. \tag{8}$$

Arriving data are placed in the queue throughout the time slot t and can only be transmitted during the next time slot $t+1$. If the queue length reached the buffer capacity $N_{\rm Q}$, the subsequent arriving data will be dropped. According to the above assumption, the queuing process evolves as follows:

$$Q_{i,t+1}^{(c)} = \min\left[N_{\rm Q}, \max[0, Q_{i,t}^{(c)} - r_{i,t}^{(c)}] + A_{i,t}^{(c)}\right]. \tag{9}$$

IV. Queuing Model

A. Model Description

Based on the above network model, a queuing model is developed as illustrated in Fig.2. With a slight abuse of notation, we use (i,j) to denote the server in the queuing model corresponding to link (i,j). We also use a black circle and a white circle to illustrate a server corresponding to a cellular link and a D2D link, respectively.

As the set of connections can be divided into three non-overlapping subsets, i.e., $\mathcal{C}_{\rm D}$, $\mathcal{C}_{\rm Cu}$, and $\mathcal{C}_{\rm Cd}$, the queues and servers in the general queuing model can also be divided accordingly. For any cellular uplink or downlink connection, i.e., $c \in \mathcal{C}_{\rm Cu} \bigcup \mathcal{C}_{\rm Cd}$, since there is only one single-hop route, its queuing model has a single queue with a data arrival process of mean λ_c, and a single server. For any D2D connection $c \in \mathcal{C}_{\rm D}$, since the data can be either transmitted via the one-hop route or two-hop route, the system can be formulated as a two-stage tandem queuing model. Specifically, there is a queue $q_{2c-1}^{(c)}$ having two stage-1 servers corresponding to link $(2c-1, 2c)$ and link $(2c-1, 0)$, respectively, and a queue $q_0^{(c)}$ having a stage-2 server corresponding to link $(0, 2c)$. The packets arrive with mean λ_c at the queue $q_{2c-1}^{(c)}$, and those served by server $(2c-1, 0)$ will join $q_0^{(c)}$ immediately after they receive service from the stage-1 server, and upon completion of service at the stage-2 server left the system. On the other hand, the packets in $q_{2c-1}^{(c)}$ served by server $(2c-1, 2c)$ will leave the system directly upon completion.

B. System State

The *global system state* of the above queuing model at time slot t can be characterized by the aggregation of the system CSI and system QSI, i.e., $\mathbf{S}_t = (\mathbf{H}_t, \mathbf{Q}_t)$. The system QSI is

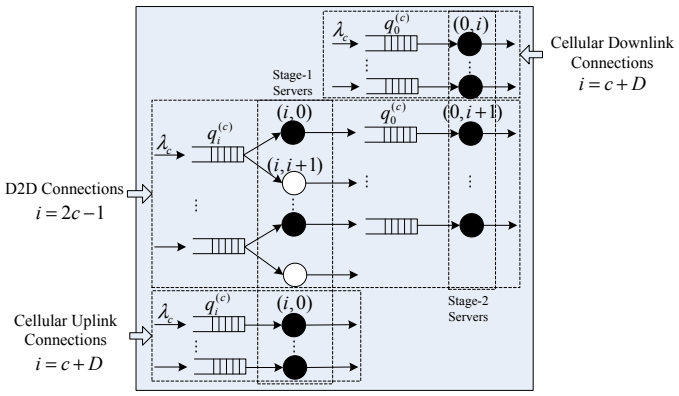

Fig. 2. Queuing model for the general network model.

defined as $\mathbf{Q}_t := \left\{Q_{i,t}^{(c)} | q_i^{(c)} \in \Theta\right\}$, which is a vector consisting of the lengths of all the queues at the beginning of time slot t. The system CSI is defined as $\mathbf{H}_t := \{\mathbf{H}_{ij,t} | (i,j) \in \mathcal{L}\}$, where $\mathbf{H}_{ij,t}$ denotes the channel state of link (i,j) in time slot t as given in Definition 1.

The global system state can be represented as the union of *uplink system state* $\mathbf{S}_{{\rm u},t}$ and *downlink system state* $\mathbf{S}_{{\rm d},t}$, i.e., $\mathbf{S}_t = \mathbf{S}_{{\rm u},t} \bigcup \mathbf{S}_{{\rm d},t}$. The uplink system state (resp. downlink system state) can be characterized by the aggregation of the uplink CSI (resp. downlink CSI) and the uplink QSI (resp. downlink CSI), i.e., $\mathbf{S}_{{\rm u},t} = (\mathbf{H}_{{\rm u},t}, \mathbf{Q}_{{\rm u},t})$ (resp. $\mathbf{S}_{{\rm d},t} = (\mathbf{H}_{{\rm d},t}, \mathbf{Q}_{{\rm d},t})$), where uplink CSI (resp. downlink CSI) is a vector consisting of the channel states of all the cellular uplinks and D2D links (resp. cellular downlinks) denoted as $\mathbf{H}_{{\rm u},t} := \{\mathbf{H}_{ij,t} | (i,j) \in \mathcal{L}_{\rm Cu} \bigcup \mathcal{L}_{\rm D}\}$ (resp. $\mathbf{H}_{{\rm d},t} := \{\mathbf{H}_{ij,t} | (i,j) \in \mathcal{L}_{\rm Cd}\}$); while uplink QSI (resp. downlink QSI) is a vector consisting of the lengths of all the uplink queues (resp. downlink queues), denoted as $\mathbf{Q}_{{\rm u},t} := \left\{\mathbf{Q}_{i,t}^{(c)} | q_i^{(c)} \in \Theta_{\rm u}\right\}$ (resp. $\mathbf{Q}_{{\rm d},t} := \left\{\mathbf{Q}_{i,t}^{(c)} | q_i^{(c)} \in \Theta_{\rm d}\right\}$). For every queue $q_i^{(c)} \in \Theta$, we define its local system state as $\mathbf{S}_{i,t}^{(c)} := \left(\mathbf{H}_{i,t}^{(c)}, Q_{i,t}^{(c)}\right)$, where $\mathbf{H}_{i,t}^{(c)} := \left\{\mathbf{H}_{ij,t} | (i,j) \in \mathcal{L}_i^{(c)}\right\}$. Note that $\bigcup_{q_i^{(c)} \in \Theta_{\rm u}} \mathbf{S}_{i,t}^{(c)} = \mathbf{S}_{{\rm u},t}$ and $\bigcup_{q_i^{(c)} \in \Theta_{\rm d}} \mathbf{S}_{i,t}^{(c)} = \mathbf{S}_{{\rm d},t}$.

Let $\mathcal{S} = \mathcal{H} \times \mathcal{Q}$ be the full system state space, $\mathcal{S}_{\rm u}$ be the uplink system state, $\mathcal{S}_{\rm d}$ be the downlink system state, and $\mathcal{S}_i^{(c)} = \mathcal{H}_i^{(c)} \times \mathcal{Q}_i^{(c)}$ be the local system state space of queue $q_i^{(c)}$. For any downlink queue $q_i^{(c)} \in \Theta_{\rm d}$, the cardinality of its local system state space is $|\mathcal{S}_i^{(c)}| = K^{N_{\rm F}} \times (N_{\rm Q}+1)$. For any uplink queue $q_i^{(c)} \in \Theta_{\rm u}$, $|\mathcal{S}_i^{(c)}| = \prod_{(i,j) \in \mathcal{L}_i^{(c)}} K^{N_{\rm F}|\mathcal{U}_{ij}|} \times (N_{\rm Q}+1)$, since the channel state space size for any link $(i,j) \in \mathcal{L}_{\rm Cu} \bigcup \mathcal{L}_{\rm D}$ depends on the number of RRGs that it belongs to. Therefore, the cardinalities of the downlink and uplink system state spaces, and full system state space can be derived as $|\mathcal{S}_{\rm d}| = (K^{N_{\rm F}} \times N_{\rm Q})^{|\Theta_{\rm d}|}$, $|\mathcal{S}_{\rm u}| = \prod_{q_i^{(c)} \in \Theta_{\rm u}} (|\mathcal{S}_i^{(c)}|)$ and $|\mathcal{S}| = |\mathcal{S}_{\rm d}| \times |\mathcal{S}_{\rm u}|$, respectively.

Assumption 2 (observation of system state in BS)**.** *We assume that the BS maintains the global system state. Specifically, the BS measures the CSI of cellular uplinks. The downlink CUEs and dest. DUEs measure the CSI of cellular downlinks and*

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

D2D links and report these information to the BS. The BS has the knowledge of downlink QSI, while the src. DUEs and uplink CUEs report the uplink QSI information to the BS. Compared with the existing 4G LTE system, the only additional signaling overhead is for reporting the CSI of D2D links from dest. DUEs to the BS.

C. Control Policy

As discussed in Section II.D, dynamic mode selection is implicitly determined by the subchannel allocation function, so that we only focus on the latter. In each time slot, the resource controller observes the system state $\mathbf{S}_t$ and chooses a subchannel allocation action from the set of allowable actions in the action space $\mathcal{A}_{\rm x}$. A subchannel allocation action $\mathbf{x}$ is composed of an uplink subchannel allocation action $\mathbf{x}_{\rm u} \in \mathcal{A}_{\rm xu}$ and a downlink subchannel allocation action $\mathbf{x}_{\rm d} \in \mathcal{A}_{\rm xd}$, i.e., $\mathbf{x} := (\mathbf{x}_{\rm u}, \mathbf{x}_{\rm d})$, where $\mathbf{x}_{\rm u} := \left\{x_u^{(m)} \in \{0,1\} | u \in \mathcal{U}_{\rm u}, m \in \{1,\ldots,N_{\rm F}\}\right\} \in \mathcal{A}_{\rm xu}$ and $\mathbf{x}_{\rm d} := \left\{x_u^{(m)} \in \{0,1\} | u \in \mathcal{U}_{\rm d}, m \in \{1,\ldots,N_{\rm F}\}\right\} \in \mathcal{A}_{\rm xd}$. Since at most one RRG $\mathcal{B}_u$ can be allocated on any uplink or downlink subchannel, there are $(|\mathcal{U}_{\rm u}|+1)^{N_{\rm F}}$ (resp. $(|\mathcal{U}_{\rm d}|+1)^{N_{\rm F}}$) actions in the set $\mathcal{A}_{\rm xu}$ (resp. $\mathcal{A}_{\rm xd}$).

A control policy prescribes a procedure for action selection in each state at all decision epoches t. We consider stationary Markovian control policies. A control policy can be either *deterministic* or *randomized*. Let $\mathcal{H}_{\rm D}$ and $\mathcal{H}_{\rm R}$ be the space of deterministic policy and randomized policy, respectively. A deterministic control policy $\Omega \in \mathcal{H}_{\rm D}$ is a mapping $\mathcal{S} \to \mathcal{A}$ from the state space to the action space, which is given by $\Omega(\mathbf{S}) = \mathbf{a} \in \mathcal{A},\ \forall \mathbf{S} \in \mathcal{S}$. A randomized control policy $\Omega \in \mathcal{H}_{\rm R}$ is a mapping $\mathcal{S} \to \mathcal{P}(\mathcal{A})$ from the state space to the set of probability distributions on the action space, which is given by $\Omega_{\rm R}(\mathbf{S}) = \{\phi_{\mathbf{S}}(\mathbf{a}) | \mathbf{a} \in \mathcal{A}\},\ \forall \mathbf{S} \in \mathcal{S}$. A deterministic control policy may be regarded as a special case of a randomized control policy in which the probability distribution on the set of actions is degenerate.

D. State Transition Probability and Steady-State Probability

The system behavior of the above queuing model can be represented by the discrete-time Markov chain (DTMC) $\{\mathbf{S}_t\}_{t=0,1,\ldots} := \{(\mathbf{H}_t, \mathbf{Q}_t)\}_{t=0,1,\ldots}$. Given a system state $\mathbf{S}_t$ and an action $\mathbf{x}$ at time slot t, the state transition probability of the DTMC is given by

$$\begin{aligned}\Pr.\{\mathbf{S}_{t+1}|\mathbf{S}_t,\mathbf{x}\} &= \Pr.\{\mathbf{H}_{t+1}|\mathbf{H}_t\}\Pr.\{\mathbf{Q}_{t+1}|\mathbf{S}_t,\mathbf{x}\} \\ &= \Pr.\{\mathbf{H}_{t+1}\}\Pr.\{\mathbf{Q}_{t+1}|\mathbf{S}_t,\mathbf{x}\}. \end{aligned} \quad (10)$$

First, we derive the queue state transition probability $\Pr.\{\mathbf{Q}_{t+1}|\mathbf{S}_t,\mathbf{x}\}$. According to (9), the conditional probability of $Q_{i,t+1}^{(c)}$ given the system state $\mathbf{S}_t$ and an action $\mathbf{x}$ can be derived as

$$\begin{aligned}&\Pr.\{Q_{i,t+1}^{(c)}|\mathbf{S}_t,\mathbf{x}\} = \Pr.(A_{i,t}^{(c)} = n), \\ &\quad \text{if } Q_{i,t+1}^{(c)} = \min\left[N_{\rm Q}, \max[0, Q_{i,t}^{(c)} - r_{i,t}^{(c)}] + n\right]. \end{aligned} \quad (11)$$

The value of $\Pr.(A_{i,t}^{(c)} = n)$ in (11) depends on the queue $q_i^{(c)} \in \Theta$. As discussed in Section II.F, for any queue $q_i^{(c)} \in \Theta_{\rm u} \bigcup \Theta_{\rm Cd}$ where node i is the source node of connection c, its data arrival process equals $A_{c,t}$, which has a general distribution $f_A(n)$ with mean λ_c. Otherwise, for any $q_0^{(c)} \in \Theta_{\rm D-d}$, its data arrival process depends on the data departure process of link $(2c-1,0)$ in the uplink transmission. Therefore, we examine the probability that n units of data are transmitted via link $(2c-1,0)$. Given the local system state $\mathbf{S}_{(2c-1)}^{(c)}$ of $q_{(2c-1)}^{(c)}$ and the subchannel allocation action $\mathbf{x}$, the throughput of link $(2c-1,0)$ during a time slot $T_{(2c-1)0,t}$ is known according to (7).

Therefore, we have

$$\begin{aligned}&\Pr.(A_{i,t}^{(c)} = n) \\ &= \begin{cases} f_A(n), & \text{if } q_i^{(c)} \in \Theta_{\rm u} \bigcup \Theta_{\rm Cd}, \\ 1, & \text{if } q_i^{(c)} \in \Theta_{\rm D-d} \text{ and } n = T_{(2c-1)0,t}, \\ 0, & \text{if } q_i^{(c)} \in \Theta_{\rm D-d} \text{ and } n \neq T_{(2c-1)0,t}. \end{cases} \end{aligned} \quad (12)$$

The queue state transition probability $\Pr.\{\mathbf{Q}_{t+1}|\mathbf{S}_t,\mathbf{x}\}$ can be derived as the product of $\Pr.\{Q_{i,t+1}^{(c)}|\mathbf{S}_t,\mathbf{x}\}$ over all queues $q_i^{(c)} \in \Theta$ as

$$\Pr.\{\mathbf{Q}_{t+1}|\mathbf{S}_t,\mathbf{x}\} = \prod_{q_i^{(c)} \in \Theta} \Pr.\{Q_{i,t+1}^{(c)}|\mathbf{S}_t,\mathbf{x}\}. \quad (13)$$

Next, we derive the channel state transition probability $\Pr.\{\mathbf{H}_{t+1}\}$. When $\mathbf{H}_{ij,t} = \{\mathbf{H}_{ij,t}^{(m)}\}_{m\in\{1,\ldots,N_{\rm F}\}}$ where $\mathbf{H}_{ij,t}^{(m)} = \mathbf{k}_{ij}^{(m)} = \{k_{ij}^{(m,u)}\}_{u\in\mathcal{U}_{ij}}$ with $k_{ij}^{(m,u)} \in \{1,\ldots,K\}$, we have $SINR_{ij,t}^{(m,u)} \in \left[\chi_{(k_{ij}^{(m,u)}-1)}, \chi_{k_{ij}^{(m,u)}}\right)$. From (1), it can be seen that for any given $m \in \{1,\ldots,N_{\rm F}\}$ and $u \in \mathcal{U}_{ij}$, the SINR value $SINR_{ij,t}^{(m,u)}$ and thus the channel state $H_{ij,t}^{(m,u)} = k_{ij}^{(m,u)}$ of link (i,j) depends on the SNR value $\gamma_{ij,t}^{(m)}$ of link (i,j) and the 'virtual SNR' values $\gamma_{i'j,t}^{(m)}$ of its interfering links $I_{i'j},\ (i',j') \in \mathcal{B}_u \backslash \{(i,j)\}$. Define $\vec{\gamma}_{ij,t}^{(m,u)} := \{\gamma_{i'j}^{(m)} | (i',j') \in \mathcal{B}_u\}$ as the (virtual) SNR vector of link (i,j) on subchannel m considering only its potential interfering links when RRG $\mathcal{B}_u$ is scheduled. Therefore, given $SINR_{ij,t}^{(m,u)} \in \left[\chi_{(k_{ij}^{(m,u)}-1)}, \chi_{k_{ij}^{(m,u)}}\right)$, $\vec{\gamma}_{ij,t}^{(m,u)}$ at time slot t belongs to the convex polyhedron $\Upsilon_{k_{ij}^{(m,u)}} := \{\vec{\gamma}_{ij,t}^{(m,u)} | \gamma_{ij}^{(m)} - \chi_{(k_{ij}^{(m,u)}-1)} \sum_{(i',j')\in\mathcal{B}_u\backslash\{(i,j)\}} \gamma_{i'j}^{(m)} \geq \chi_{(k_{ij}^{(m,u)}-1)}, \gamma_{ij}^{(m)} - \chi_{k_{ij}^{(m,u)}} \sum_{(i',j')\in\mathcal{B}_u\backslash\{(i,j)\}} \gamma_{i'j}^{(m)} < \chi_{k_{ij}^{(m,u)}}, \vec{\gamma}_{ij,t}^{(m,u)} \geq 0\}$. The (virtual) SNR regions corresponding to the channel state $k_{ij}^{(m,u)}$ and $k_{ij}^{(m,u)}+1$ are separated by the hyperplane $\gamma_{ij} - \chi_{k_{ij}^{(m,u)}} \sum_{(i',j')\in\mathcal{B}_u\backslash\{(i,j)\}} \gamma_{i'j}^{(m)} = \chi_{k_{ij}^{(m,u)}}$. Next, define $\vec{\gamma}_{ij,t}^{(m)} := \{\gamma_{i'j}^{(m)} | (i',j') \in \bigcup_{u\in\mathcal{U}_{ij}} \mathcal{B}_u\}$ as the (virtual) SNR vector of link (i,j) on subchannel m considering all its potential interfering links in the set $\mathcal{I}_{ij}$. Since $\mathbf{H}_{ij,t}^{(m)} = \mathbf{k}_{ij}^{(m)}$, $\vec{\gamma}_{ij,t}^{(m)}$ belongs to the convex polyhedron $\Upsilon_{\mathbf{k}_{ij}^{(m)}} = \bigcap_{u\in\mathcal{U}_{ij}} \Upsilon_{k_{ij}^{(m,u)}}$. Therefore, the steady-state probability that $\mathbf{H}_{ij,t}^{(m)} = \mathbf{k}_{ij}^{(m)}$ can

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

be derived as

$$\Pr.(\mathbf{H}_{ij}^{(m)}) = \int_{\Upsilon_{\mathbf{k}_{ij}^{(m)}}} f(\vec{\gamma}_{ij}^{(m)}) \mathrm{d}\vec{\gamma}_{ij}^{(m)} = \int_{\Upsilon_{\mathbf{k}_{ij}^{(m)}}} \prod_{(i',j') \in \bigcup_{u \in \mathcal{U}_{ij}} \mathcal{B}_u} (f(\gamma_{i'j}) \mathrm{d}\gamma_{i'j}). \quad (14)$$

where $f(\vec{\gamma}_{ij}^{(m)})$ is the joint probability distribution function (pdf) of $\{\gamma_{i'j,t}\}_{(i',j') \in \bigcup_{u \in \mathcal{U}_{ij}} \mathcal{B}_u}$, and $f(\gamma_{i'j})$ is the pdf of $\gamma_{i'j,t}$. The second equality is due to the independence between the r.v. elements in the set $\vec{\gamma}_{ij}^{(m)}$. Given $\Pr.(\mathbf{H}_{ij}^{(m)})$, the global channel state transition probability can be derived as

$$\Pr.\{\mathbf{H}_{t+1}\} = \prod_{(i,j) \in \mathcal{L}} \prod_{m \in \{1,\ldots,N_{\mathrm{F}}\}} \Pr.(\mathbf{H}_{ij}^{(m)}). \quad (15)$$

Given a deterministic control policy $\Omega \in \mathcal{H}_{\mathrm{D}}$, since the action $\mathbf{x}_t$ under every system state $\mathbf{S}_t$ is determined, we can directly derive $\Pr.\{\mathbf{S}_{t+1}|\mathbf{S}_t, \Omega(\mathbf{S}_t)\}$. On the other hand, given a randomized control policy $\Omega \in \mathcal{H}_{\mathrm{R}}$, we can derive the system state transition probability as $\Pr.\{\mathbf{S}_{t+1}|\mathbf{S}_t, \Omega(\mathbf{S}_t)\} = \sum_{\mathbf{a} \in \mathcal{A}} \Pr.\{\mathbf{S}_{t+1}|\mathbf{S}_t, \mathbf{a}\} \phi_{\mathbf{S}_t}(\mathbf{a})$. Let $\mathbf{S}^{(y)}$ denote the y-th system state within the state space. Define the transition probability matrix $\mathbf{P}^{\Omega} = [\Pr.\{\mathbf{S}_{t+1} = \mathbf{S}^{(y)}|\mathbf{S}_t = \mathbf{S}^{(z)}, \Omega(\mathbf{S}^{(z)})\}], y, z \in \{1, \ldots, |\mathcal{S}|\}$ and the steady-state probability matrix $\boldsymbol{\pi}^{\Omega} = [\pi^{\Omega}_{\mathbf{S}^{(z)}}], z \in \{1, \ldots, |\mathcal{S}|\}$, where $\pi^{\Omega}_{\mathbf{S}^{(z)}} = \lim_{t \to \infty} \Pr.\{\mathbf{S}_t = \mathbf{S}^{(z)}\}$. Each element of the transition probability matrix $\mathbf{P}^{\Omega}$ can be derived. Then, the stationary distribution of the ergodic process $\{\mathbf{S}_t\}_{t=0,1,\ldots}$ can be uniquely determined from the balance equations.

E. Performance Metrics

Given $\boldsymbol{\pi}^{\Omega}$, the end-to-end performance measures such as the mean throughput, the average delay and the dropping probability for all the connections can be derived.

1) Average queue length: The average queue length of queue $q_i^{(c)}$ equals

$$\overline{Q}_i^{(c)} = \mathbf{E}^{\pi(\Omega)}[Q_i^{(c)}]. \quad (16)$$

2) Mean throughput: Denote $\overline{T}_c$ as the end-to-end mean throughput of connection $c \in \mathcal{C}$ and $\overline{T}_{ij}$ as the mean throughput of link (i,j). The relationships between the throughput of a connection c and those of the links in its link constraint set $\mathcal{L}_c$ are given below

$$\overline{T}_c = \begin{cases} \overline{T}_{(c+D)0}, & \text{if } c \in \mathcal{C}_{\mathrm{Cu}}, \\ \overline{T}_{0(c+D)}, & \text{if } c \in \mathcal{C}_{\mathrm{Cd}}, \\ \overline{T}_{(2c-1)(2c)} + \overline{T}_{0(2c)}, & \text{if } c \in \mathcal{C}_{\mathrm{D}}. \end{cases} \quad (17)$$

The first two equations are due to fact that every cellular connection consists of one hop. The third equation is because the data of a D2D connection can be transmitted either via the single-hop route or the two-hop route, and thus its throughput should be the sum throughput of the two routes.

According to the discussion above, we need to derive the mean throughput of related links in order to obtain the end-to-end mean throughput of a connection c . The mean throughput of a link $(i,j) \in \mathcal{L}$ can be derived as

$$\overline{T}_{ij} = \mathbf{E}^{\pi(\Omega)}\left[T_{ij}\left(\mathbf{S}_i^{(c)}, \Omega(\mathbf{S})\right)\right], \quad (18)$$

where $T_{ij}(\mathbf{S}_i^{(c)}, \Omega(\mathbf{S}))$ is the throughput of link (i,j) given in (7) and (8).

3) Average delay: Denote $\overline{D}_c$ as the end-to-end average delay of connection $c \in \mathcal{C}$ and $\overline{D}_i^{(c)}$ as the average delay of queue $q_i^{(c)}$. The relationships between the average delay of a connection c and those of the queues along its route(s) are given below

$$\overline{D}_c = \begin{cases} \overline{D}_{c+D}^{(c)}, & \text{if } c \in \mathcal{C}_{\mathrm{Cu}}, \\ \overline{D}_0^{(c)}, & \text{if } c \in \mathcal{C}_{\mathrm{Cd}}, \\ \overline{D}_{2c-1}^{(c)} + \overline{D}_0^{(c)} \frac{\overline{T}_{0(2c)}}{\overline{T}_{(2c-1)(2c)} + \overline{T}_{0(2c)}}, & \text{if} c \in \mathcal{C}_{\mathrm{D}}. \end{cases} \quad (19)$$

Similar to derivation of mean throughput, it is straightforward to see that the first two equations are due to the fact that every cellular connection consists of one hop, while the third equation is because among all the served data for a D2D connection c, $\frac{\overline{T}_{(2c-1)(2c)}}{\overline{T}_{(2c-1)(2c)} + \overline{T}_{0(2c)}}$ fraction of data are transmitted via the single hop route which has an average delay of $\overline{D}_{2c-1}^{(c)}$, while $\frac{\overline{T}_{0(2c)}}{\overline{T}_{(2c-1)(2c)} + \overline{T}_{0(2c)}}$ fraction of data are transmitted via the two-hop route which has an average delay of $\overline{D}_{2c-1}^{(c)} + \overline{D}_0^{(c)}$.

The average delay $\overline{D}_i^{(c)}$ for any queue $q_i^{(c)}$ can be calculated according to Little's Law as

$$\overline{D}_i^{(c)} = \overline{Q}_i^{(c)} / \overline{T}_i^{(c)}, \quad (20)$$

which is the average amount of time between the arrival and departure of a data unit in queue $q_i^{(c)}$, where

$$\overline{T}_i^{(c)} = \mathbf{E}^{\pi(\Omega)}\left[T_i^{(c)}\left(\mathbf{S}_i^{(c)}, \Omega(\mathbf{S})\right)\right], \quad (21)$$

is the mean throughput of queue $q_i^{(c)}$, which equals the effective arrival rate of queue $q_i^{(c)}$, i.e., the average rate at which the packets enter queue $q_i^{(c)}$. Note that $T_i^{(c)}(\mathbf{S}_i^{(c)}, \Omega(\mathbf{S}))]$ can be derived from (6).

4) Dropping probability: Denote d_c as the dropping probability of connection $c \in \mathcal{C}$, which can be estimated as

$$d_c = \frac{\text{Average \# of data units dropped in a time slot}}{\text{Average \# of data units arrived in a time slot}} = 1 - \frac{\text{Average \# of data units transmitted in a time slot}}{\text{Average \# of data units arrived in a time slot}} = 1 - \frac{\overline{T}_c}{\lambda_c}, \quad (22)$$

where the mean throughput $\overline{T}_c$ of connection c can be derived from (17).

V. Problem Formulation and Solution

A. Problem Formulation

Our objective is to optimize the subchannel allocation policy so as to minimize the average weighted sum delay of all the connections subject to dropping probability constraints. Note that the dropping probability is an important QoS metric indicating the reliability of the system, and has also been used as QoS constraint in other works [21], [25]. On the other hand, the problem formulation and solution method in this section

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

and also that in Part II of this work can also be applied to study other CMDP problems with different optimization objectives and constraints, e.g., maximize the sum throughput subject to the delay constraint.

Problem 1. *The delay-optimal subchannel allocation design can be formulated as the constrained optimization problem in* (23)

$$\min_{\Omega\in\mathcal{H}_{\rm R}} \sum_{c=1}^{C} \omega_c \bar{D}_c \tag{23}$$

$$\text{s.t. } d_c \le d_{\max},\ \forall c \in \mathcal{C},$$

where ω_c $(c = 1,\ldots,C)$ *are the weights of the delays of connections* c. $\bar{D}_c$ *and* d_c *can be derived from* (19) *and* (22)*, respectively.*

In Problem 1, the subchannel allocation policy Ω influences the behavior of a probabilistic system as it evolves through time. The goal is to choose a sequence of actions which causes the system to perform optimally with respect to the time-average or expected system performance. Therefore, it is a dynamic optimization problem and we would like to formulate and solve it as a CMDP model. To apply the CMDP model, we need to define four elements, i.e., state space $\mathcal{S}$, action space $\mathcal{A}_{\rm x}$, state transition probability $\Pr.\{\mathbf{S}^{(y)}|\mathbf{S}^{(z)},\mathbf{x}\}$, and a set of cost functions, which include one cost function $g_0(\mathbf{S},\mathbf{x})$ related to the optimization objective and an additional set of cost functions $g_c(\mathbf{S},\mathbf{x})$, $c \in \mathcal{C}$ related to the C constraints in Problem 1. Suppose the above elements are defined, the infinite horizon CMDP model for problem 1 can be formulated as

$$\min_{\Omega\in\mathcal{H}_{\rm R}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}^{\Omega}[g_0(\mathbf{S}_t,\Omega(\mathbf{S}_t))] \tag{24}$$

$$\text{s.t. } \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}^{\Omega}[g_c(\mathbf{S}_t,\Omega(\mathbf{S}_t))] \le d_{\max},\ \forall c \in \mathcal{C},$$

where under any unichain policy we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}^{\Omega}[g_c(\mathbf{S}_t,\Omega(\mathbf{S}_t))] = \mathbf{E}^{\pi(\Omega)}[g_c(\mathbf{S},\Omega(\mathbf{S}))],\ \forall c \in \mathcal{C},$$

and $\mathbf{E}^{\pi(\Omega)}[x]$ denotes the expectation operation taken w.r.t. the unique steady-state distribution induced by the given policy Ω.

The former three elements of the CMDP model have been defined in Section III.B-D, respectively. Comparing (23) with (24), we cannot directly derive the cost function $g_0(\mathbf{S},\mathbf{x})$, since the form $\frac{\mathbf{E}[x]}{\mathbf{E}[y]}$ exists when combining (19), (20), (21) and (16) in order to derive the average delay $\bar{D}_c$ of every connection in (23). To solve this problem, we replace the denominator in the R.H.S. of (19) which is in the form of $\mathbf{E}[y]$ with $\lambda_c(1-d_{\max})$. This approximation is reasonable because the term $\mathbf{E}[y]$ represents the average throughput of either a connection or its stage-1 server, and we thus have $\mathbf{E}[y] \ge \lambda_c(1-d_{\max})$ due to the constraint on dropping probability. In this way, we can define

$$g_0(\mathbf{S},\mathbf{x}) = \frac{Q_i^{(c)}}{\lambda_c(1-d_{\max})}, \tag{25}$$

$$g_c(\mathbf{S},\mathbf{x}) = \begin{cases} 1-\frac{T_{(c+D)0}}{\lambda_c}, & \text{if } c\in\mathcal{C}_{\rm Cu},\\ 1-\frac{T_0(c+D)}{\lambda_c}, & \text{if } c\in\mathcal{C}_{\rm Cd},\\ 1-\frac{T_{(2c-1)(2c)}+T_{0(2c)}}{\lambda_c}, & \text{if } c\in\mathcal{C}_{\rm D}. \end{cases} \tag{26}$$

In the rest of paper, we will refer to the CMDP problem defined in (24) whenever we refer to Problem 1.

In order to solve the CMDP Problem 1, we cast this dynamic optimization problem as an abstract 'static' optimization problem over a close convex set of measures. It is shown in [27] that the MDP problem belongs to the convex programming (in fact, infinite dimensional linear programming) problems, so that Lagrangian duality method can be used to take the constraints of CMDP problem into account by augmenting the objective function with a weighted sum of the constraint functions.

For any given nonnegative Lagrangian Multipliers (LMs) $\boldsymbol{\eta} = \{\eta_c | c \in \mathcal{C}\}$, we define the Lagrangian function of Problem 1 as

$$L(\Omega,\eta) = \mathbf{E}^{\pi(\Omega)}[g(\mathbf{S},\Omega(\mathbf{S}))] + \sum_{c\in\mathcal{C}} \lambda_c\eta_c(1-d_{\max})^2, \tag{27}$$

where

$$\begin{aligned} g(\mathbf{S},\Omega(\mathbf{S})) = &\sum_{c\in\mathcal{C}_{\rm Cu}} \left(\omega_c Q_{(c+D)}^{(c)} - \eta_c(1-d_{\max})T_{(c+D)0}\right)\\ &+ \sum_{c\in\mathcal{C}_{\rm Cd}} \left(\omega_c Q_0^{(c)} - \eta_c(1-d_{\max})T_{0(c+D)}\right)\\ &+ \sum_{c\in\mathcal{C}_{\rm D}} \big(\omega_c(Q_{(2c-1)}^{(c)} + Q_0^{(c)})\\ &- \eta_c(1-d_{\max})(T_{(2c-1)(2c)} + T_{0(2c)})\big). \end{aligned} \tag{28}$$

Therefore, Problem 1 can be decomposed into the following two subproblems:

$$\textbf{Subproblem 1-1: } G(\boldsymbol{\eta}) = \min_{\Omega\in\mathcal{H}_{\rm R}} L(\Omega,\boldsymbol{\eta}),$$

$$\textbf{Subproblem 1-2: } G(\boldsymbol{\eta}^*) = \max_{\boldsymbol{\eta}} G(\boldsymbol{\eta}).$$

where $G(\boldsymbol{\eta})$ is the corresponding Lagrange dual function and Subproblem 1-2 is the dual problem.

The following theorem establishes the strong duality result of the CMDP Problem 1 over the space of randomized policy $\mathcal{H}_{\rm R}$.

Theorem 1 (Strong Duality Over $\mathcal{H}_{\rm R}$)**.** *If there exists a strictly feasible policy* $\Omega \in \mathcal{H}_{\rm R}$ *in Problem 1 such that*

$$d_c = \mathbf{E}^{\pi(\Omega)}[g_c(\mathbf{S},\Omega(\mathbf{S}))] < d_{\max},\ \forall c = \{1,\ldots,C\}, \tag{29}$$

then there exists an optimal stationary randomized policy Ω^* *and a vector of finite nonnegative LMs* $\boldsymbol{\eta}^*$ *such that* Ω^* *minimizes the ergodic cost* $L(\Omega^*,\boldsymbol{\eta}^*)$ *and the following 'saddle point condition' holds:*

$$L(\Omega,\boldsymbol{\eta}^*) \ge L(\Omega^*,\boldsymbol{\eta}^*) \ge L(\Omega^*,\boldsymbol{\eta}). \tag{30}$$

Therefore, Ω^* *is the primal optimal (i.e., solving Problem 1),* $\boldsymbol{\eta}^*$ *is the dual optimal (i.e., solving the dual problem) and the duality gap is zero. By solving the dual problem (i.e., Subproblem 1-2), we can obtain the primal optimal* Ω^*.

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

Proof: The CMDP problem 1 is an infinite dimensional linear programming problem [27], which is a special type of convex problem. Moreover, (29) guarantees that the slater's condition is satisfied. Thus, Theorem 1 is proved using standard optimization theory. ■

In the above discussion, we focus on randomized policies. However, deterministic policies are much simpler to implement and evaluate. The following theorem shows that if the constraint can be satisfied by any randomized policy, the optimization over all randomized policies is attained by either a deterministic policy or a simple mixed policy. Before we introduce Theorem 2, two optimal deterministic policies Ω_1^* and Ω_2^* with respect to two unconstraint MDP problems are introduced, where $\Omega_1^* = \min_{\Omega\in\mathcal{H}_D} \mathbf{E}^{\pi(\Omega)}[g_0(\mathbf{S},\Omega(\mathbf{S}))]$ and $\Omega_2^* = \min_{\Omega\in\mathcal{H}_D} \mathbf{E}^{\pi(\Omega)}[\max_{c\in\mathcal{C}} g_c(\mathbf{S},\Omega(\mathbf{S}))]$. Note that the first unconstraint MDP problem is the unconstraint version of Problem 1 to minimize the average delay, while the second unconstraint MDP problem is to minimize the maximum dropping probability of all the connections. We will refer to Ω_1^* as the unconstraint delay optimal policy and Ω_2^* as the unconstraint drop optimal policy in the rest of the paper.

Lemma 1. *There exists two open sets Λ_1 and Λ_2 with $\mathbf{0} \in \Lambda_1 \subseteq \Lambda_2 \subset \mathbf{R}^{C+}$, such that the boundaries of Λ_1 and Λ_2 are defined by all the mean arrival rate vector $\boldsymbol{\lambda} = \{\lambda_c\}_{c\in\mathcal{C}}$ that result in $\max_{c\in\mathcal{C}} d_c = d_{\max}$ using policy Ω_1^* and Ω_2^*, respectively.*

Proof: The proof is straightforward since for any given policy Ω, the dropping probability of every connection increases with the increasing mean arrival rate vector[3] $\boldsymbol{\lambda}$, and the value of $\max_{c\in\mathcal{C}} d_c$ achieved by Ω_2^* is the minimal value over all randomized policies by definition and no larger than that achieved by Ω_1^*. ■

Now, we introduce a third deterministic policy Ω_3^* which is the optimal policy of Problem 1 over the space of deterministic polices $\mathcal{H}_{\rm D}$, i.e., $\Omega_3^* = \arg\min_{\Omega\in\mathcal{H}_{\rm D}} L(\Omega,\boldsymbol{\eta}^*)$.

Theorem 2 (General Form of Optimal Policy for Problem 1)**.** *For CMDP Problem 1, there exists two open sets $\Lambda_1 \subseteq \Lambda_2 \subset \mathbf{R}^{C+}$ as defined in Lemma 1, such that*

1) *When $\boldsymbol{\lambda} \in \Lambda_1$, the optimization of Problem 1 over all randomized policies is attained by the deterministic policy Ω_3^* or equivalently Ω_1^*, since the two policies are essentially the same in this case.*
2) *When $\boldsymbol{\lambda} \in \Lambda_2 \bigcap \overline{\Lambda_1}$, the optimization of Problem 1 over all randomized policies is attained by*
   a) *a deterministic policy Ω_3^*, if $\max_{c\in\mathcal{C}} d_c(\Omega_3^*) = d_{\max}$;*
   b) *a mixed policy, if $\max_{c\in\mathcal{C}} d_c(\Omega_3^*) < d_{\max}$, which is equivalent to choosing independently at each time slot between the two deterministic policies Ω_1^* and Ω_2^* by the throw of a biased coin with probability p such that $p\max_{c\in\mathcal{C}} d_c(\Omega_1^*) + (1-p)\max_{c\in\mathcal{C}} d_c(\Omega_2^*) = d_{max}$.*
3) *When $\boldsymbol{\lambda} \in \overline{\Lambda_2}$, there exists no strictly feasible policy for Problem 1.*

[3]vector inequalities in this paper are understood as pointwise

Proof: The proof is given in Appendix A. ■

B. Problem Solution

According to Theorem 2, the optimal policy for Problem 1 Ω^* is either a deterministic policy Ω_3^* or a simple mix of two deterministic policies Ω_1^* and Ω_2^*. Therefore, we will first focus on solving Problem 1 over the space of deterministic policy $\mathcal{H}_{\rm D}$ to derive Ω_3^*. The solution method can also be applied to obtain the optimal policies Ω_1^* and Ω_2^* for the unconstraint MDP problems. Then, an algorithm (Algorithm 1) will be proposed to derive the optimal policy Ω^* for Problem 1 based on the deterministic policies Ω_1^*, Ω_2^* and Ω_3^*.

1) Solving Problem 1 over the space of deterministic policy to derive Ω_3^*: Given the LMs $\boldsymbol{\eta}^*$, Subproblem 1-1 is a classical infinite horizon average reward MDP problem, which can be solved by the Bellman's equation [21].

$$\theta + V(\mathbf{S}^{(z)}) = \min_{\Omega(\mathbf{S}^{(z)})\in\mathcal{H}_{\rm D}} \Big\{ g(\mathbf{S}^{(z)},\Omega(\mathbf{S}^{(z)})) + \sum_{\mathbf{S}^{(y)}\in\mathcal{S}} \Pr.[\mathbf{S}^{(y)}|\mathbf{S}^{(z)},\Omega(\mathbf{S}^{(z)})] V(\mathbf{S}^{(y)}) \Big\}, \forall \mathbf{S}^{(z)} \in \mathcal{S}, \tag{31}$$

where $V(\mathbf{S}^{(z)})$ is the value function representing the average reward obtained following policy Ω from each state $\mathbf{S}^{(z)}$, while θ represents the optimal average reward per period for a system in steady-state.

As a remark, note that the Bellman's equation (31) represents a series of fixed-point equations, where the number of equations are determined by the number of value functions $V(\mathbf{S}^{(z)})$, which is $|\mathcal{S}|$. Theoretically, the BS can use the brute force value iteration method to offline solve (31) and derive the optimal control policy, in which $|\mathcal{S}|$ value functions need to be stored and the computation complexity is $O(|\mathcal{S}|^2|\mathcal{A}_{\rm x}|)$ in one iteration. Therefore, the offline value iteration algorithm is too complicated to compute due to curse of dimensionality, i.e., the exponential growth of the cardinality of the system state space and the large dimension of the control action space involved.

In order to reduce the state space of the above MDP, we construct an equivalent Bellman's equation. Let $\mathbf{Q}^{(\dot{y})}$ and $\mathbf{Q}^{(\dot{z})}$ denote the $\dot{y}$-th and $\dot{z}$-th queue states within the queue state space, respectively, where $\dot{y},\dot{z} \in \{1,\ldots,|\mathcal{Q}|\}$. We first define the partitioned actions of a policy Ω as follows.

Definition 2 (definition of partitioned actions)**.** *Given a control policy Ω, we define*

$$\Omega(\mathbf{Q}^{(\dot{z})}) = \{\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})|\forall \mathbf{H}\} \subseteq \mathcal{A}_{\rm x}$$

as the collection of $|\mathcal{H}|$ actions, where every action is mapped by policy Ω from a system state with given QSI $\mathbf{Q}^{(\dot{z})}$, and a different realization of CSI $\mathbf{H} \in \mathcal{H}$.

Lemma 2. *The control policy obtained by solving the original Bellman's equation* (31) *is equivalent to the control policy*

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

obtained by solving the reduced-state Bellman's equation (32)

$$\theta + V(\mathbf{Q}^{(\dot{z})}) = \min_{\Omega(\mathbf{Q}^{(\dot{z})})\in\mathcal{H}_{\mathrm{D}}} \Big\{ g(\mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{Q}^{(\dot{z})})) + \sum_{\mathbf{Q}^{(\dot{y})}\in\mathcal{Q}} \Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{Q}^{(\dot{z})})] V(\mathbf{Q}^{(\dot{y})}) \Big\}, \forall \mathbf{Q}^{(\dot{z})} \in \mathcal{Q}, \tag{32}$$

where $V(\mathbf{Q}^{(\dot{y})}) = \mathbf{E}_{\mathbf{H}}\Big[V(\mathbf{H}, \mathbf{Q}^{(\dot{y})})|\mathbf{Q}^{(\dot{y})}\Big] = \sum_{\mathbf{H}\in\mathcal{H}} \Pr.[\mathbf{H}] V(\mathbf{H}, \mathbf{Q}^{(\dot{y})})$ *is the conditional expectation of value function* $V(\mathbf{S})$ *taken over the channel state space* $\mathcal{H}$ *given the queue state* $\mathbf{Q}^{(\dot{y})}$*, while* $g(\mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{Q}^{(\dot{z})})) = \mathbf{E}_{\mathbf{H}}\Big[g(\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})}))|\mathbf{Q}^{(\dot{z})}\Big]$ *and* $\Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{Q}^{(\dot{z})})] = \mathbf{E}_{\mathbf{H}}\Big[\Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})]|\mathbf{Q}^{(\dot{z})}\Big]$ *are conditional expectations of cost function* $g(\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})}))$ *and transition probability* $\Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})]$ *taken over the channel state space* $\mathcal{H}$ *given the queue state* $\mathbf{Q}^{(\dot{z})}$*, respectively.*

Proof: The proof is given in Appendix B. ■

Remark 4 (complexity of solving equivalent Bellman's equation with offline value iteration)**.** *If we use the offline value iteration algorithm to solve* (32)*, the BS needs to store* $|\mathcal{Q}|$ *value functions and the computation complexity is* $O(|\mathcal{Q}||\mathcal{S}||\mathcal{A}_{\mathrm{x}}|)$ *in one iteration. The amount of information stored on BS and the computation complexity is greatly reduced compared to the original Bellman's equation in* (31)*, so that it is possible to obtain the optimal policy in a simple network such as the one in Fig.1 and study its properties. However, since the state space still grows exponentially with the number of queues, we will propose a practical solution in Part II of this work using linear value approximation and online stochastic learning to deal the curse of dimensionality problem in solving Problem 1.*

In order to learn the correct $\boldsymbol{\eta}^*$, we use a gradient ascent in the dual (i.e., Lagrange multiplier) space in view of subproblem 1-2. Therefore, the LM vector $\boldsymbol{\eta}$ is initiated as $\boldsymbol{\eta}_0 = \mathbf{0}$ and updated iteratively. In the l-th iteration, the LM vector $\boldsymbol{\eta}_l$ is updated according to

$$\eta_{c,l+1} = \eta_{c,l} + (\mathbf{E}^{\pi(\Omega^*(\boldsymbol{\eta}_l))}[g_c(\mathbf{S}, \Omega(\mathbf{S}))] - d_{\max}),\ \forall c = 1, \ldots, C, \tag{33}$$

where $\Omega^*(\boldsymbol{\eta}_l)$ is the optimal policy obtained by solving the unconstraint Subproblem 1-1 using the value iteration algorithm based on the equivalent Bellman's equation with LM vector $\boldsymbol{\eta}_l$.

2) Optimal policy Ω^* *for Problem 1:* The following Algorithm 1 can be used to derive Ω^* according to Theorem 2, given the deterministic policies Ω_1^*,

Algorithm 1 Derive the optimal policy for Problem 1 (Ω^*)

```
Derive Ω_3^* based on the reduced-state Bellman's equation
if no feasible policy exists then
    End
else if max_{c∈C} d_c(Ω_3^*) = d_max then
    Ω^* ⇐ Ω_3^*
else if max_{c∈C} d_c(Ω_3^*) < d_max then
    Derive Ω_1^* based on the reduced-state Bellman's equation
    if max_{c∈C} d_c(Ω_1^*) = max_{c∈C} d_c(Ω_3^*) then
        Ω^* ⇐ Ω_3^* = Ω_1^*
    else if max_{c∈C} d_c(Ω_1^*) > d_max then
        Derive Ω_2^* based on the reduced-state Bellman's equation
        Ω^* ⇐ a mixed policy which randomizes between Ω_1^* and Ω_2^* as given in Theorem 2.
    end if
end if
```

VI. Simulation Results

In this section, we compare the performance of optimal policy for Problem 1 derived by Algorithm 1 with two other reference subchannel allocation algorithms. One is the CSI-only algorithm, in which the RRG selection is only adaptive to CSI and a subchannel is allocated to the RRG with the maximum sum over all its link transmission rates at every time slot. The other is the MaxWeight algorithm [10], which is adaptive to both CSI and QSI and a subchannel is allocated to the RRG with maximum sum over all its links of the product of the corresponding differential backlog and transmission rate. We develop discrete event system-level simulator for D2D communications system with dynamic packet arrivals using Matlab, and all experiments are run on 3.4GHz PC with 8GHz RAM. In the simulations, we consider Poisson packet arrival with mean arrive rate λ and fixed packet size of 1080 bits at the source node. Moreover, we consider a wireless network employing adaptive M-ary quadrature amplitude modulation (M-QAM) with convolutional coding which has six channel states for all transmission links. The SINR thresholds for the channel states are given in Table II of [26]. We assume the Rayleigh fading channel and the number of packets transmitted in a time slot under different channel states, i.e., R_k with $k = 1, 2, 3, 4, 5, 6$ are set to $0, 1, 2, 3, 6, 9$, respectively. The carrier frequency and the time slot duration ΔT are set to 2GHz and 1ms, respectively. Due to the complexity and the large required memory of solving even the reduced-state equivalent Bellman's equation, we consider the simple network in Fig.1 with one D2D connection, one cellular uplink connection and one cellular downlink connection. The buffer size is set to be $N_{\mathrm{Q}} = 3$ packets and only one subchannel is considered. In this case, the cardinality of the state space $|\mathcal{S}|$ and action space $|\mathcal{A}_{\mathrm{x}}|$ are 1536 and 9, respectively. Therefore, the computation complexity of the original Bellman's equation is $O(2\mathrm{e}7)$ in one iteration, and the computation complexity of the equivalent Bellman's equation is $O(3\mathrm{e}6)$ in one iteration. We assign equal weights of the delays on each connection.

We simulate a circular cell with a BS (node 0) in the center and the cell radius is 500m. All the CUEs and src. DUEs are uniformly distributed in the cell area at random, whereas the dest. DUEs are distributed uniformly upon a disk centered by their corresponding src. DUEs with a radius of R (the maximum distance of D2D links is R). The statistics are collected over multiple realizations of the position of the UEs. The distance between the transmitter of node i and

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

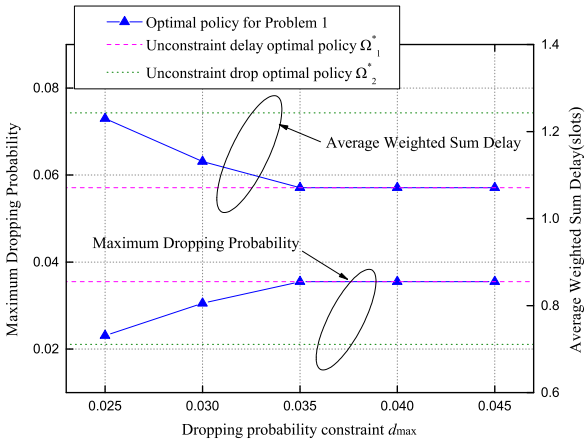

Fig. 3. The average weighted sum delay and the maximum dropping probability over all connections versus the dropping probability constraint $d_{\max}$ with $\lambda = 1$ packets/slot and $R = 100$m.

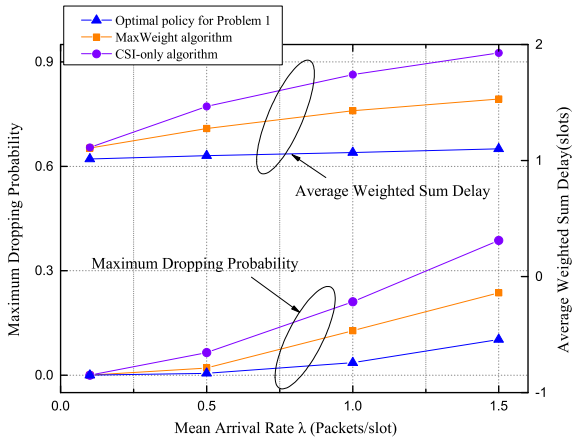

Fig. 4. The average weighted sum delay and the maximum dropping probability over all connections versus the mean arrive rate λ with $d_{\max} = 0.1$ and $R = 100$m.

receiver of node j is denoted by l_{ij}. We consider the path loss channel model $28 + 40\log_{10} l_{ij}[m]$ for any wireless channel between a pair of UEs, while the path loss channel model $15.3 + 37.6\log_{10} l_{ij}[m]$ for any wireless channel between the BS and a UE. The transmission power of the BS and the UE are set to be 46dBm and 23dBm, respectively. We normalize the uplink noise power and downlink noise power, respectively, so that the SNR is 0dB for any cellular uplink or downlink at the cell border.

Fig.3 shows the average weighted sum delay and the maximum dropping probability over all connections versus the dropping probability constraint $d_{\max}$ of the optimal policy for Problem 1 with $\lambda = 1$ packets/slot and $R = 100$m. Moreover, the performance of the unconstraint delay optimal policy Ω_1^* and unconstraint drop optimal policy Ω_2^* are also illustrated to provide bounds for the optimal policy for Problem 1. It can be observed that the maximum dropping probability increases and the average weighted sum delay decreases slowly as the dropping probability constraint $d_{\max}$ becomes looser. The unconstraint delay optimal policy Ω_1^* provides the lower bound of average weighted sum delay and upper bound of maximum dropping probability for the optimal policy for Problem 1, since the objective functions of the unconstraint delay optimal problem and Problem 1 are the same, while the former has a larger domain than the latter. As illustrated in Fig.3, when $d_{\max}$ exceeds the maximum dropping probability of Ω_1^*, i.e., 0.035, the performance of the optimal policy for Problem 1 remain the same as those achieved by Ω_1^*, since the unconstraint delay optimal policy is included in the domain of Problem 1 when $d_{\max} \geq 0.035$ and will always be chosen irrespective of $d_{\max}$. On the other hand, the unconstraint drop optimal policy Ω_2^* provides the lower bound of maximum dropping probability and upper bound of average weighted sum delay for the optimal policy for Problem 1 due to the objective function of Ω_2^*. When $d_{\max}$ is lower than the maximum dropping probability of Ω_2^*, i.e., 0.021, no feasible policy exists for Problem 1 and its domain is empty

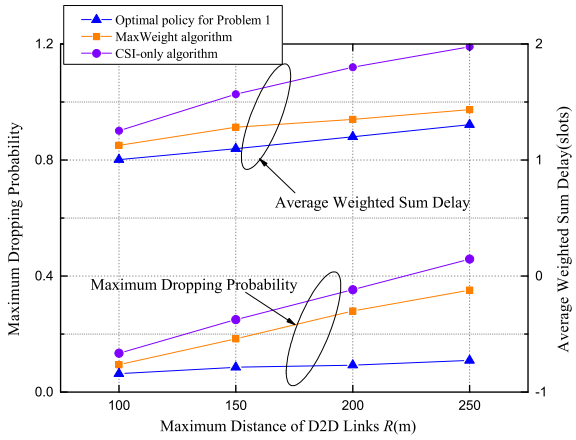

Fig. 5. The average weighted sum delay and the maximum dropping probability over all connections versus the maximum distance of D2D links R with $\lambda = 1.5$ packets/slot and $d_{\max} = 0.1$.

since the maximum dropping probabilities of all the other policies must be larger than that of Ω_2^* and thus also larger than $d_{\max}$. From Fig.3, it can be observed that the optimal policy for Problem 1 ensures the dropping probability at the cost of delay, and can achieve the optimal delay within the dropping probability constraint whenever possible.

Fig.4 shows the average weighted sum delay and the maximum dropping probability over all connections versus the mean arrive rate λ of the optimal policy for Problem 1, the CSI-only algorithm and the Maxweight algorithm with $d_{\max} = 0.1$ and $R = 100$m. It can be observed that the performance of the optimal policy for Problem 1 achieves lower average weighted sum delay compared with the two reference algorithms and always keeps the maximum dropping probability within $d_{\max}$. Although the MaxWeight algorithm achieves nearly the same maximum dropping probability in light traffic load regime, when λ gradually increases, the optimal policy for Problem 1 achieves a better performance.

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

Fig.5 shows the average weighted sum delay and the maximum dropping probability over all connections versus the maximum distance of D2D links R of the optimal policy for Problem 1, the CSI-only algorithm and the Maxweight algorithm with $\lambda = 1.5$ packets/slot and $d_{\max} = 0.1$. The performance of all the three algorithms deteriorates with increasing R, since the channel quality of the D2D link becomes worse. Moreover, the performance gaps between the two baseline algorithms and the optimal policy for Problem 1 increase with increasing R. This is because when the DUEs are close to each other, the D2D link will have a very good channel quality and the D2D mode will always be chosen. In this case, the packets of D2D connection can be transmitted effectively leaving its queue empty most of the time, which means that more resources can be allocated to the cellular uplink connection. Thus, a low average delay and dropping probability can be achieved by all the three algorithms. However, as the distance between the D2D pair increases, the link quality of all the links may not have a distinct difference. In this case, an effective mode selection and resource allocation algorithm can make a significant improvement of the achieved performance.

VII. Conclusion

In this paper, we studied a delay-optimal dynamic mode selection and resource allocation problem under dropping probability constraint for network assisted D2D communications with bursty traffic arrival, where the data are transmitted over frequency-selective fading channel with AMC scheme in the physical layer. To formulate the above problem into an infinite horizon average reward CMDP, we first developed a queuing model whose underlying system state dynamics evolves as a controlled Markov chain, where the system state includes the joint queue state of the queues at the UEs for uplink transmission and the queues at the BS for downlink transmission as well as the joint channel state of all the D2D links, cellular uplinks and cellular downlinks. The transition kernel of the controlled Markov chain was derived, taking into account the coupling relationship between the uplink and downlink resource allocation. Moreover, closed-form expressions for end-to-end performance metrics such as average delay and dropping probability were given as functions of steady-state probabilities of the controlled Markov chain, based on which the cost function of MDP model was given. We provided two regions of mean arrival rate vector, in which the strong duality result and the existence of optimal deterministic policy can be guaranteed, respectively. Moreover, we proved that the optimal policy is no more complex than the randomization between two deterministic policies. In order to reduce the state space of the CMDP model, we proposed an equivalent Bellman's equation. Simulation results showed that the optimal control policy based on our CMDP model outperforms the conventional CSI-only scheme and throughput-optimal scheme (MaxWeight algorithm). In part II of this paper, we will propose a practical solution which uses linear value approximation and online stochastic learning to deal the curse of dimensionality problem in solving the CMDP and achieves near-optimal performance.

Appendix

A. Proof of Theorem 2

The proof is based on the following Lemma from [29].

Lemma 3. *For an unconstraint MDP problem, suppose the state space and action space are finite, the cost function are bounded, and the model is unichain, then there exists an optimal deterministic policy that achieves the objective over all randomized policies.*

By Lemma 3, both deterministic policies Ω_1^* and Ω_2^* achieve the optimal objective over all randomized policies for their respective unconstraint MDP problem.

We first prove Theorem 2-1), where for any $\boldsymbol{\lambda} \in \Lambda_1$, the deterministic policy $\Omega_1^* \in \mathcal{H}_D$ that achieves the unconstraint minimum of Problem 1 is also a strictly feasible policy under the constraint of Problem 1. In this case, the optimal LM $\boldsymbol{\eta}^* = \mathbf{0}$ and $\Omega_1^* \in \mathcal{H}_D$ also achieves the minimum of Problem 1 over all the randomized policies according to Lemma 3, which completes the proof of Theorem 2-1).

We next prove Theorem 2-2), where for any $\boldsymbol{\lambda} \in \Lambda_2 \bigcap \overline{\Lambda_1}$, we have (i) the deterministic policy $\Omega_1^* \in \mathcal{H}_D$ that achieves the unconstraint minimum of Problem 1 is not a feasible policy for Problem 1; (ii) there exists a deterministic feasible policy $\Omega_2^* \in \mathcal{H}_D$ for Problem 1. According to [28][Theorem 4.4], the optimal constraint policy for a CMDP problem satisfying condition (i) and (ii) is either a deterministic policy Ω_3^* if the equality of the constraint is attained by Ω_3^* or a convex combination of the above two deterministic policies Ω_1^* and Ω_2^*, which completes the proof of Theorem 2-2).

Theorem 2-3) follows directly from (29) of Theorem 1.

B. Proof of Lemma 2

$$
\begin{aligned}
&\theta + V(\mathbf{H}, \mathbf{Q}^{(\dot{z})}) \ \forall \mathbf{H} \in \mathcal{H},\ \mathbf{Q}^{(\dot{z})} \in \mathcal{Q},\\
&= \min_{\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})} \Big\{ g(\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})}))\\
&+ \sum_{\mathbf{H}', \mathbf{Q}^{(\dot{y})}} \Pr.[\mathbf{H}', \mathbf{Q}^{(\dot{y})} | \mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})] V(\mathbf{H}', \mathbf{Q}^{(\dot{y})}) \Big\}\\
&\overset{(a)}{=} \min_{\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})} \Big\{ g(\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})) + \sum_{\mathbf{Q}^{(\dot{y})}}\\
&\Pr.[\mathbf{Q}^{(\dot{y})} | \mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})] \big(\sum_{\mathbf{H}'} \Pr.(\mathbf{H}') V(\mathbf{H}', \mathbf{Q}^{(\dot{y})}) \big) \Big\}\\
&\overset{(b)}{=} \min_{\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})} \Big\{ g(\mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})}))\\
&+ \sum_{\mathbf{Q}^{(\dot{y})}} \Pr.[\mathbf{Q}^{(\dot{y})} | \mathbf{H}, \mathbf{Q}^{(\dot{z})}, \Omega(\mathbf{H}, \mathbf{Q}^{(\dot{z})})] V(\mathbf{Q}^{(\dot{y})}) \Big\},
\end{aligned}
$$

where (a) is due to (10) by the i.i.d. assumption of CSI over time slots, (b) is due to the definition $V(\mathbf{Q}^{(\dot{z})})$ given in Section V.B.

Taking the conditional expectation (conditioned on $\mathbf{Q}^{(\dot{z})}$)

0018-9545 (c) 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2444795, IEEE Transactions on Vehicular Technology

on both sides of the equation above, we have

$$\begin{aligned}
&\theta + V(\mathbf{Q}^{(\dot{z})}) \ \forall \mathbf{Q}^{(\dot{z})} \in \mathcal{Q}, \\
=&\mathbf{E}_{\mathbf{H}}\Big[\min_{\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})} \big\{g(\mathbf{H},\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})) \\
&+\sum_{\mathbf{Q}^{(\dot{y})}} \Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{H},\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{H},\mathbf{Q}^{(\dot{z})})]V(\mathbf{Q}^{(\dot{y})})\big\}\Big] \\
\overset{(c)}{=}&\min_{\Omega(\mathbf{Q}^{(\dot{z})})} \Big\{g(\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{Q}^{(\dot{z})})) \\
&+\sum_{\mathbf{Q}^{(\dot{y})}} \Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{Q}^{(\dot{z})})]V(\mathbf{Q}^{(\dot{y})})\Big\},
\end{aligned}$$

where (c) is due to the definition of "conditional reward" $g(\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{Q}^{(\dot{z})}))$ and "conditional transition probability" $\Pr.[\mathbf{Q}^{(\dot{y})}|\mathbf{Q}^{(\dot{z})},\Omega(\mathbf{Q}^{(\dot{z})})]$ given in Section IV.B.

REFERENCES

[1] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 96-104, June 2012.

[2] J. Liu *et al*., "Device-to-Device communications achieve efficient load balancing in LTE-Advanced networks," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 57-65, Apr. 2014.

[3] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-Smartphone: realizing multi-hop Device-to-Device communications," *IEEE Wireless Commun.*, vol. 52, no. 4, pp. 56-65, Apr. 2014.

[4] K. Zheng, S. Ou, J. Alonso-Zarate, M. Dohler, and F. Liu, "Challenges of massive access in highly dense LTE-advanced networks with machine-to-machine communications," *IEEE Wireless Communications*, vol. 21, no. 3, pp.12-18, 2014.

[5] 3GPP, "3rd generation partnership project; technical specification group RAN; Study on LTE device to device proximity services; Radio aspects (Release 12)," TR 36.843 V12.0.1, Sept. 2014.

[6] M. N. 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.

[7] S. Mumtaz, K. M. S. Huq, and J. Rodriguezdirect, "Mobile-to-mobile communication: paradigm for 5G," *IEEE Wireless Commun.*, vol. 21, no. 5, pp. 14-23, Oct. 2014.

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

[9] W. Wang and V. K. N. Lau, "Delay-aware cross-layer design for device-to-device communications in future cellular systems," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 133-139, June 2014.

[10] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks", *Foundations and Trends in Networking*, vol. 1, no. 1, pp. 1-144, 2006.

[11] C.-H. Yu *et al*., "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752-2763, Aug. 2011.

[12] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*,vol. pp, no. 99, pp. 1-14, Sept. 2014

[13] C.-P. Chien, Y.-C. Chen, and H.-Y. Hsieh, "Exploiting spatial reuse gain through joint mode selection and resource allocation for underlay device-to-device communications," *2012 15th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 80-84, Sept. 2012.

[14] G. Yu *et al*., "Joint mode selection and resource allocation for device-to-device communications," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3814-3824, Nov. 2014.

[15] W. Dan et al., "Energy-efficient resource sharing for mobile device-to-device multimedia communications, *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2093-2103, June 2014.

[16] 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 in Commun.*, vol. 31, no. 9, pp. 348-358, Sept. 2013.

[17] L. Song *et al*., "Game-theoretic resource allocation methods for Device-to-Device (D2D) communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136-144, Jun. 2014.

[18] C. Xu, L. Song, and Z. Han, Resource Management for Device-to-Device Underlay Communications. Springer Briefs in Computer Science, 2014.

[19] N. Cheng, N. Lu, N. Zhang, T. Yang, X. Shen, and J.W. Mark, "Vehicle-assisted device-to-device data delivery for smart grid," IEEE Trans. on Veh. Technol., to appear.

[20] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630-643, July 2003.

[21] Y. Cui, V. K. N. Lau, and R. Wang, "A survey on delay-aware resource control for wireless systems-large deviation theory, stochastic lyapunov drift, and distributed stochastic learning," *IEEE Trans. Info. Theory*, vol. 58, no. 3, pp. 1677-1701, March 2012.

[22] L. Lei, Y. Kuang, X. Shen, C. Lin, and Z. Zhong, "Resource control in network assisted device-to-device communications: solutions and challenges," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 108-117, June 2014.

[23] H. Kim, G. d. Veciana, X. Yang, and M. Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Networking*, vol. 20, no. 1, pp. 177-190, Feb. 2012.

[24] K. Zheng, F. Liu, L. Lei, C. Lin, and Y. Jiang, "Stochastic Performance Analysis of a Wireless Finite-State Markov Channel," *IEEE Trans. Wireless Communications*, vol. 12, no. 2, pp. 782-793, 2013.

[25] R. Wang and V. K. Lau, "Delay-Aware Two-Hop Cooperative Relay Communications via Approximate MDP and Stochastic Learning," *IEEE Trans. Info. Theory*, vol. 59, no. 11, pp. 7645-7670, Nov. 2013.

[26] Q. Liu, S. Zhou, and G. B. Giannakis, "Queueing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 50, no. 3, pp. 1142-1153, May 2005.

[27] V. S. Borkar, "Convex analytic methods in Markov decision processes", ser. Handbook of Markov Decision Processes, F.A. Schwartz, Ed. Norwell, MA, USA: Kluwer, 2001, pp.347-375.

[28] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint," *Journal of Mathematical Analysis and Applications*, vo. 112, pp. 236-252, 1985.

[29] M. L. Puterman, Markov decision processes: discrete stochastic dynamic programming. New York, NY., USA: Wiley, 2005.

0018-9545 (c) 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.