A DYNAMIC GRAPH OPTIMIZATION FRAMEWORK FOR MULTIHOP DEVICE-TO-DEVICE COMMUNICATION UNDERLAYING CELLULAR NETWORKS

YONG LI, CHAOMING SONG, DEPENG JIN, AND SHENG CHEN

Yong Li and Depeng Jin are with Tsinghua University.

Chaoming Song is with the University of Miami.

Sheng Chen is with the University of Southampton and King Abdulaziz University.

This work is supported by the National Basic Research Program of China (973 Program) (No. 2013CB329001), National Nature Science Foundation of China (No. 61301080, No. 91338203, and No. 91338102), National High Technology Research and Development Program (No. 2013AA013501 and No. 2013AA013505), Chinese National Major Scientific and Technological Specialized Project (No. 2013ZX03002001), and China's Next Generation Internet (No. CNGI-12-03-007).

ABSTRACT

With emerging demands for local area and popular content sharing services, multihop device-to-device communication is conceived as a vital component of next-generation cellular networks to improve spectral reuse, bring hop gains, and enhance system capacity. Ripening these benefits depends on fundamentally understanding its potential performance impacts and efficiently solving several main technical problems. Aiming to establish a new paradigm for the analysis and design of multihop D2D communications, in this article, we propose a dynamic graph optimization framework that enables the modeling of large-scale systems with multiple D2D pairs and node mobility patterns. By inherently modeling the main technological problems for multihop D2D communications, this framework benefits investigation of theoretical performance limits and studying the optimal system design. Furthermore, these achievable benefits are demonstrated by examples of simulations under a realistic multihop D2D communication underlaying cellular network.

INTRODUCTION

As one of the next-generation wireless communication systems, Long Term Evolution-Advanced (LTE-A) supports mobile content downloading [1]. To meet the increasing demands for local area services of popular content downloading, device-to-device (D2D) communication is proposed as a key component of LTE-A, which enables devices to communicate directly, and is an underlay to the cellular network for improving spectral efficiency [2–4]. Under the control of a base station (BS), user equipment (UE) can transmit data to each other over direct links using the cellular resources instead of through BSs. In these operator controlled LTE networks, with the unified control of access authentication, connection control, resource allocation, and lawful interception of information, D2D communications take place within the cellular coverage area and even outside the coverage of BSs [2]. Most context-aware applications that involve discovering and communicating with nearby devices can benefit from D2D communication by reducing communication cost, since it enables physical proximity communication, which saves power while improving spectral efficiency [3]. Furthermore, in the vast application of popular Internet content downloading and sharing, multihop D2D communications will play a vital role in decreasing redundant cellular direct transmissions by enabling multihop sharing. Thus, it is expected that multihop D2D communication will be a key feature supported by next-generation cellular networks [4].

Although D2D communication may enhance spectral efficiency and increase system capacity, it also causes interference to the cellular network as the result of spectrum sharing, and multihop transmissions depend on resource sharing and allocation schemes. Furthermore, multihop D2D communications occur only when mobile users are physically in close proximity. Thus, it depends on how often the devices are in physical proximity communication with each other. In other words, underlaying multihop D2D communication opportunities and hence system performance depend on node mobility patterns. Unsurprisingly, it is challenging to analyze and design such a complicated system, which involves dynamic human behaviors and complex communication relations.

Current works often consider D2D communication technical problems under a very restricted cellular system setting, for example, consisting of only four nodes — a pair of D2D UEs, a cellular UE, and a BS [4–6]. Within such a small and simplified network scenario, these existing studies can only deal with some individual aspects of the underlaying D2D cellular network. However, none of these existing contributions are able to

 1536-1284/14/$25.00 © 2014 IEEE

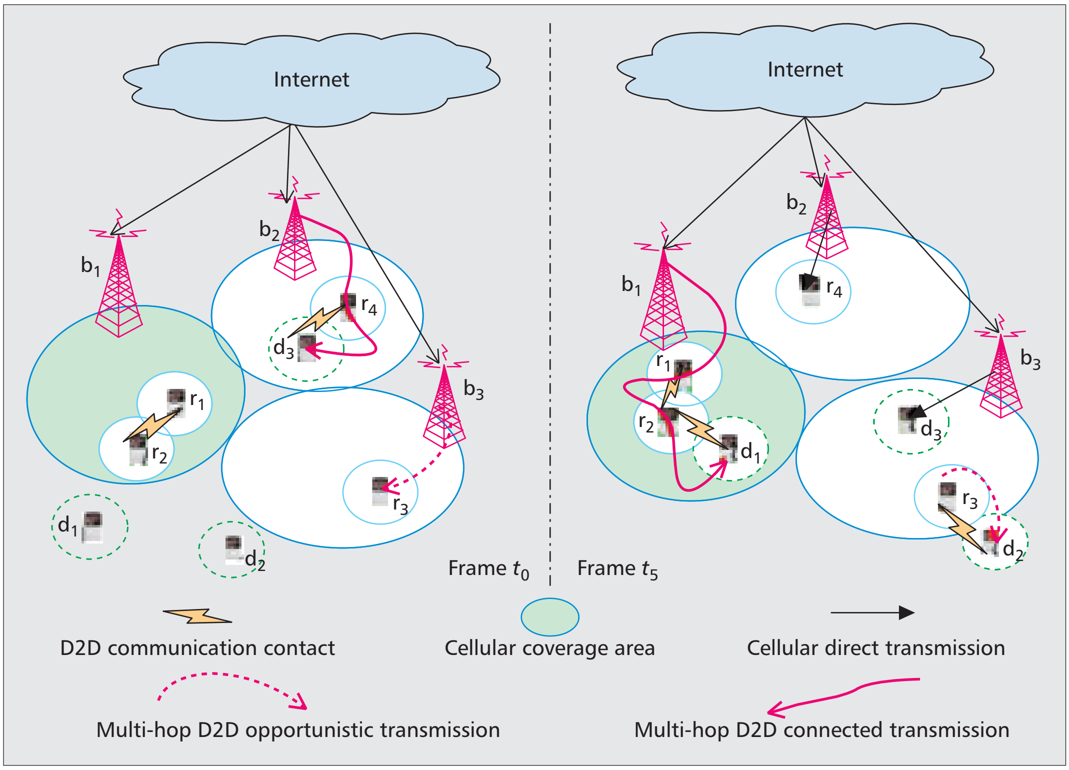

Figure 1. Illustration of a multihop D2D communication underlaying cellular network at different timeframes of t_0 and t_5, where there are three cellular BSs, b_1, b_2, and b_3, and seven mobile UEs. Among these mobile UEs, there are four relaying nodes, r_1 to r_4, the communication ranges of which are denoted by solid thin circles, and three receiver nodes, d_1, d_2, and d_3, the communication ranges of which are denoted by dotted thin circles. In this multihop D2D underlay communication system, a receiver can receive the data from the cellular network directly (cellular transmission mode) or from the cellular network via some relaying nodes (multihop D2D connected or opportunistic transmission mode).

The requested data may be delivered from the corresponding data servers to the related users either via direct cellular transmissions, if the users are under the cellular coverage, or via D2D enabled multi-hop transmissions, when the users requesting the data are in physical proximity with some other users.

tackle the underlaying problem as a whole, and therefore they are unable to quantify the real potential of multihop D2D enabling a cellular network under a realistic network scenario with tens of hundreds of nodes and D2D pairs. Against this background, in this article, we aim to establish a new paradigm for the design and analysis of multihop D2D communications. Specifically, we propose a dynamic graph optimization framework for general multihop D2D communication underlaying cellular. This framework enables the modeling of a large-scale system with multiple D2D pairs and node mobility patterns, which sets up a realistic multihop D2D communication framework for investigating theoretical performance limits and studying the optimal system design. The model of a dynamic graph and its corresponding optimization [10] have been utilized in the performance analysis and optimization in wireless vehicular networks [11], and more. In this article, we exploit it in the analysis and design of multihop D2D communications; fundamentally, it enables us to answer the following challenging questions: how multihop D2D can improve the cellular network system performance and what the potential effects of multihop D2D are.

We structure the article as follows. After providing a system overview, we introduce the dynamic graph optimization framework for modeling the multihop underlaying D2D cellular network system. After analyzing the capability of the proposed dynamic graph optimization framework in investigating the challenging technical problems, we introduce its two main applications in theoretical performance bound analysis and optimal system design. Finally, we quantitatively assess the capability of the proposed dynamic graph optimization framework by simulation that targets a realistic multihop D2D communication underlaying cellular network.

System Overview

A typical example of a multihop D2D communication underlaying cellular network is illustrated in Fig. 1, where the cellular network provides coverage over a certain region through the BS deployment. The UEs are mobile nodes, and their positions as well as access states change over time. Therefore, at different timeframes, their access and physical (location) relationships are different. Here, a *timeframe* is loosely used to mark a system time period during which no access or physical state change occur. For example, Fig. 1 displays the access and physical relationships at different system timeframes of t_0 and t_5, respectively.

A typical application for the multihop D2D communication underlaying cellular network is content downloading and sharing. In such a system, the BSs are connected to the data servers in the Internet. The UEs requesting data send their requests to the relevant data servers via the

It should be emphasized that the D2D connected and opportunistic transmission modes are inherently multihop. Clearly, either mode must have at least two hops, with the first hop from a BS to a relay and the second hop from this relay to a receiver.

cellular network. The requested data may be delivered from the corresponding data servers to the related users either via direct cellular transmissions, if the users are under the cellular coverage, or via D2D-enabled multihop transmissions when the users requesting the data are in physical proximity of some other users.

In such a content sharing system, some UEs are requesting data, and they are referred to as data *receivers*. Other UEs that currently are not retrieving data for themselves may participate in data transmission by receiving the data from the BSs and then transmitting them to the relevant receivers via D2D communication. These UEs are referred to as *relays*. Incentives for stimulating UEs to act as relays can be given by using some micro-payment scheme, or the operator can offer the relays a reduced cost for their services or better quality of service [7]. Thus, the UEs with energy and storage resources will participate in the data transmission as relays under the assumption of rationality. Note that our model does not make any assumptions about these incentives, and only the UEs that are willingly participating in the content transmission will be treated as relays. In the example depicted in Fig. 1, there are three cellular BSs marked as b_1 to b_3, three receivers known as d_1, d_2, and d_3, and four relays called r_1, r_2, r_3, and r_4. Apart from the original and intuitive method of *cellular direct transmission*, there are two other important transmission modes to efficiently overcome the weak coverage and signal attenuation problems of a cellular network, which are defined below with the aid of Fig. 1.

Multihop D2D connected transmission: In this mode, a connected path from a cellular BS via some relays to a receiver is established by taking advantage of the physical proximity of communicating devices, and the mobile data are transmitted via the D2D enabled connected path with multiple hops to the targeted receiver. The example illustrated in Fig. 1 shows that during timeframe t_0, BS b_2 is transmitting data to receiver d_3 via the two-hop connected path with the aid of relay r_4, while during timeframe t_5, BS b_1 is transmitting data to d_1 using the three-hop connected path $r_1 \rightarrow r_2 \rightarrow d_1$.

Multihop D2D opportunistic transmission: Due to the mobility of receivers and relays, a D2D connected path is prone to be broken. However, a relay can store the received data in its buffer, and transmit the data to the relevant receiver or other relays when communication contact arises. Since UEs are inherently mobile, this opportunistic communication mode is capable of enhancing the system performance. In Fig. 1, over timeframes t_0–t_5, receiver d_2 is outside the coverage of any BS. During timeframe t_0, relay r_3 receives data for d_2 from BS b_3 and stores the data in its buffer. When the communication opportunity occurs between r_3 and d_2 during timeframe t_5, r_3 transmits the data to d_2.

It should be emphasized that the D2D connected and opportunistic transmission modes are inherently multihop. Clearly, either mode must have at least two hops with the first hop from a BS to a relay and the second hop from this relay to a receiver. In order to model the D2D communication system with these two types of multihop transmissions, and fundamentally understand the complex relations between the D2D communication and the underlaying system factors such as resource allocation and human mobility, we present a dynamic graph model. Based on this graph model, we propose a generalized dynamic-graph-based optimization framework for investigating multihop D2D communication underlaying cellular networks.

Dynamic Graph Optimization Framework

Graph Construction

The graph model for the above underlaying D2D cellular network should include all the possible transmission opportunities of cellular direct transmission, and multihop D2D connected and opportunistic transmissions. Assume that there are B cellular BS nodes, labeled by the set of $\mathcal{B} = \{b_1, b_2, \cdots, b_{\mathcal{B}}\}$, D data receivers, labeled $\mathcal{D} = \{d_1, d_2, \cdots, d_{\mathcal{D}}\}$, and R data relaying nodes, labeled $\mathcal{R} = \{r_1, r_2, \cdots, r_{\mathcal{R}}\}$, which are involved in multihop D2D communication to help the data transmission to receivers.

Let us first form the graph of all the nodes, including BSs, relays, and receivers, in the underlaying D2D cellular network within a certain timeframe by representing each network participating node as a vertex in the graph and adding directional edges between nodes according to the BSs' coverage and UEs' D2D contacts. As within a timeframe no access and state change occurs, such a graph has a static topology for the duration of the timeframe. For example, considering the underlaying D2D cellular network illustrated in Fig. 1, the constructed static graphs for timeframe t_0 and t_5 are shown in Figs. 2a and 2b, respectively. Furthermore, we can associate each edge with a real value to represent the temporal link transmission rate. Given all the vertices and edges, we thus build the transmission graph for the underlaying D2D cellular communication in some timeframe, which is a weighted directional graph representing the spatial distribution of the network topology involving all the BSs and UEs.

To describe the UEs' dynamic accessing relationships with the BSs as well as the time-varying communication opportunities and multihop paths among the relays and receivers, we identify the communication contact events between a pair of nodes, which may be BS and UE, relay and receiver, or relay and relay. These contact events include five types: cellular accessing starts, cellular accessing ends, D2D contact starts, D2D contact ends, and link quality level changes. These communication contact events, which are assumed to sequentially occur at different time points of $p_0, p_1, \cdots, p_N$, divide the continuous time into time periods. The time period between two successive events defines a timeframe. Specifically, the time period between p_{l-1} and p_1 is labeled as timeframe t_l, with the initial timeframe t_0 defining the time before p_0. All the timeframes are therefore labeled t_0, t_1, …, t_N. Within each timeframe, no contact event

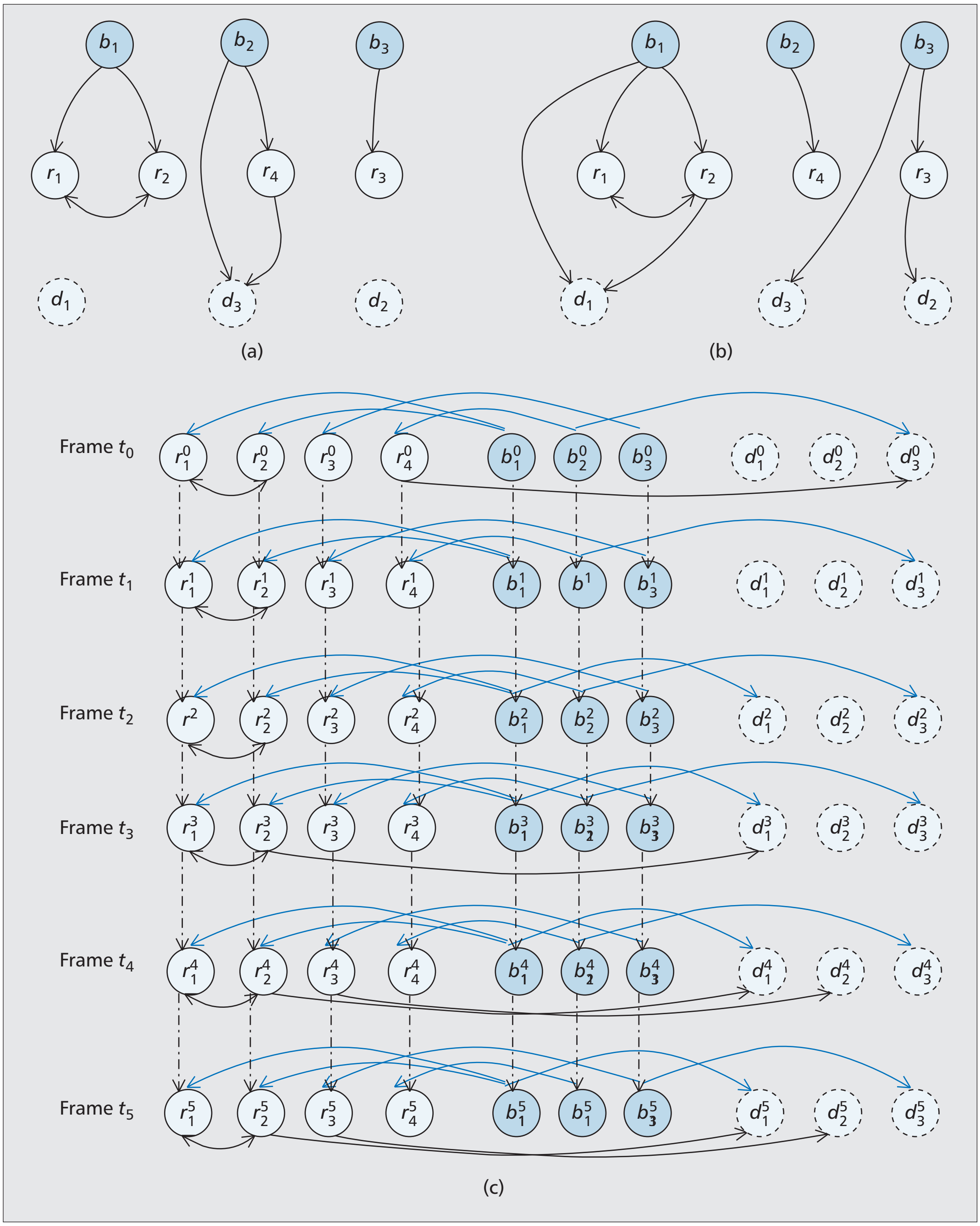

Figure 2. The dynamic underlaying D2D cellular communication graph generated by the events of opportunistic contact and access relationship that are given in Fig. 1 and Table 1: a) static graph in timeframe t_0; b) static graph in timeframe t_5; c) dynamic graph for the system varying from frame t_0 to frame t_5.

Since the nodes may store the contents in their local buffers at a timeframe and then transmit them in the coming frames, we need to model the time evolution of data buffering for relays and similarly for BSs in order to represent the sequential time evolution of the static graphs.

occurs and no link quality changes. In other words, within each timeframe, the states of the nodes and the commutation contacts do not change. Thus, for each timeframe, we can obtain a static graph to represent the system using the method stated before. Since the nodes may store the contents in their local buffers in a timeframe and then transmit them in the coming frames, we need to model the time evolution of data buffering for relays and similarly for BSs in order to represent the sequential time evolution of the static graphs. For this aim, a directional edge with certain weight, which corresponds to the buffer size of a node, is drawn for the same BS or relay between two successive frames.

Given all the vertices and edges, we thus build the transmission graph for the multihop D2D communication underlaying cellular communication, which is a weighted directional dynamic graph representing the access and contact distribution of the network topology. Consider the underlaying D2D cellular network illustrated in Fig. 1. For this example, assume that there are no link quality changing events, and we have counted all the network events of cellular access start and end as well as D2D contact start and end at time points p_0 to p_4, which

Time point	Event	Meanings
p_0	$r_4 \nleftrightarrow d_3$	Contact between r_4 and d_3 ends
p_1	$d_1 \rightarrow b_1$	d_1 moves into coverage of b_1
p_2	$r_2 \leftrightarrow d_1$	Contact between r_2 and d_1 starts
p_3	$r_3 \leftrightarrow d_2$	Contact between r_3 and d_2 starts
p_4	$d_3 \nrightarrow b_2$, $d_3 \rightarrow b_3$	d_3 moves out coverage of b_2 and moves into b_3

Table 1. Network events for the underlaying D2D cellular network involving three BSs (b_1, b_2, b_3), three receivers (d_1, d_2, d_3), and four relays (r_1, r_2, r_3, r_4).

are summarized in Table 1. Figure 2c depicts the dynamic graph for this underlaying D2D cellular network. Note that this graph includes all the BSs and UEs, and it also allows us to capture all the possible data transmission modes.

Optimization Objective

Based on the graph model for multihop D2D communication underlaying cellular networks, we form an optimization framework by expressing the maximization objective and analyzing the constraints. Since usually the optimization goal is to maximize the data transmission in the system, we can optimize any combined flows on the vertices in the dynamic graph according to the specific requirements. Take the content downloading system as an example; its goal is to maximize the total data amount received by the receivers, which is a max-flow goal. In order to achieve this objective, we introduce two virtual vertices, S_{V} and D_{V}, to represent the source and destination of the total flow over the graph. To model the flow from the source to destination, we add the edges from the source to all the BS vertices before timeframe t_0, and introduce the edges from all the receiver vertices to the destination. In this way, we obtain the directional connected graph as shown in Fig. 3, on which the data transmission is modeled as the flow from S_{V} to D_{V} over the total communication timeframes of $\{t_0, t_1, \cdots, t_N\}$, which represents the total amount of transmitted data by the D2D communication underlaying cellular network. For any edge (a, b) in the graph, we further denote the flow over the edge connecting the two vertices of a and b as $f(a, b)$. Then the objective of maximizing the total transmitted data can be expressed as

$$\max \sum_{l=0}^{N} \sum_{i=1}^{D} f\left(d_i^l, D_{\mathrm{V}}\right),$$

where d_i^l denotes the vertex of receiver d_i at timeframe t_l.

System Constraints

We also need to consider the constraints of the above flow maximization problem, which include the constraints of flow conservation and system resources.

Flow Conservation — For any vertex in the graph, flow conservation means that the amount of incoming flow must equal the amount of outgoing flow. In the system, there are three different types of vertices, BSs, relays, and receivers, which have different content transmission behaviors that influence the flow. Thus, we need to analyze the constraints of flow conservation for these different types of nodes according to their actual behaviors.

Transmission Rate and Channel Access — The spectrum bandwidth is allocated between the D2D and cellular communications, and the allocated resource directly influences the data transmission rates. Therefore, the weight of each edge, the "flow" rate of each directional edge, is directly associated with the allocated resource. In the case in which D2D communications of all pairs use the same licensed band of cellular communication, we denote the spectrum resources that are allocated directly to cellular communication and D2D communication as x_i^l and y_i^l, respectively, for each timeframe t_l and each cellular BS b_i. Since the total resource for the coverage area of b_i is limited to a value denoted by b_i, we have the following constraint on the resource allocation $x_i + y_i \leq B_i$, $1 \leq i \leq C$, $0 \leq 1 \leq N$. Under the allocated spectrum x_i^l and y_i^l with a channel model like Rayleigh fading, we can express the maximum achievable average data rate for the cellular direct transmission link to user u and D2D link g under the coverage of under b_i, respectively, as

$$R_u = x_i^l \log_2\left(1 + P_{b_i} \zeta_u^{-\ell} \left|h_{b_i,u}\right|^2 \Big/ N_0\right) \text{and}$$

$$R_g = y_i^l \log_2\left(\frac{1 + P_g \zeta_g^{-\ell} \left|h_g\right|^2}{\left(\sum_{g \in \mathcal{G} \backslash g} P_{g'} \zeta_{g',g}^{-\ell} \left|h_{g',g}\right|^2 + N_0\right)}\right),$$

where P_{b_i}, P_g, and $P_{g'}$ are the transmitted signal power of BS b_i, two D2D pairs g and g', respectively, ζ_u and ζ_g are the distance of the cellular transmission and D2D links, ℓ is the path loss exponent, N_0 is the power of the receiver noise, which is assumed to be the additive white Gaussian noise (AWGN), $|h_{bi,u}|^2$ and $|h_{g',g}|^2$ denote the average power or second-order statistic of the Rayleigh fading channel linking b_i and u and D2D pairs, respectively, and $\mathcal{G}$ is the set of all the D2D communication pairs in the coverage area of b_i. In the above derivations, interference between different D2D pairs is considered, which influences the achievable communication rate. Thus, the interference is inherently modeled in the proposed framework. At the same time, since the transmission rates of both the cellular and D2D links are proportional to the allocated resource x_i^l or y_i^l, transmission resources are added as optimization variables without explicit increase for the problem complexity. For direct cellular communication with UEs, the total transmitted flow to all the UEs during a timeframe should satisfy the system resource constraints on the transmission rates and the given timeframe duration. As for the

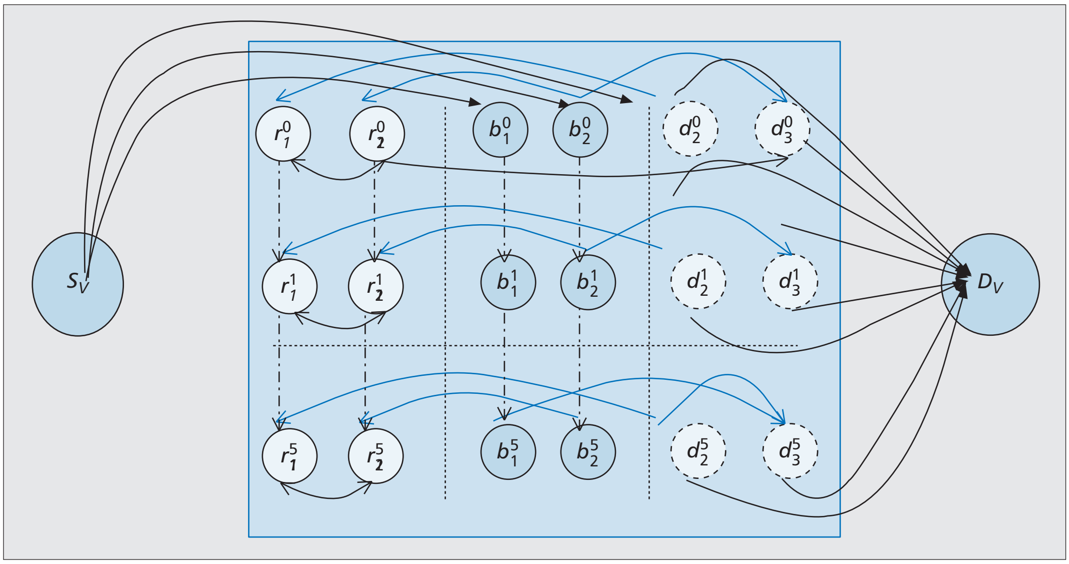

Figure 3. Illustration of the flow maximization formulation for the optimal throughput content downloading problem in the multihop D2D cellular network.

Since the D2D communication uses the same air interface as cellular communication, a receiver can only operate in either the D2D mode or the direct cellular mode. Furthermore, in the D2D mode, a decision must be made whether to use multihop connected or opportunistic transmission.

D2D communication, we also need to limit the transmitted content flows among the "connected" UEs in each timeframe to satisfy the system resource constraints on the transmission rates and the given timeframe duration as well as to meet the interference requirements for channel access.

System Resources — As stated before, we also need to consider the fact that the limited system transmission resources are divided between cellular direct and D2D communications. Sometimes, the number of operating BSs in the system is also limited to save energy, and we also need to take this into consideration in the problem formulation.

Algorithm and Solution

Combining the objective and the constraints introduced above, we can form maximization problems for the multihop underlaying D2D cellular system with the decision variables consisting of link flows and allocated resources. It is worth pointing out that the objective flow maximization,

$$\max \sum_{l=0}^{N} \sum_{i=1}^{D} f\left(d_i^l, D_V\right),$$

is a linear composition of individual flows of $f(d_i^l, D_V)$, while the associated constraints of flow conservation and resource allocation introduced above are linear constraints. For the remaining constraints on channel access that cannot be expressed in linear forms, by using the reformulation linearization technique (RLT) [8] we can transform them into linear expressions of the decision variables. Thus, the formulated maximization problem falls into the category of linear programming problems, which can be solved by linear programming techniques (i.e., using the existing optimization toolkits, such as CPLEX and YALMIP [9]). In the operator controlled underlaying D2D cellular networks, the BS has full control over the allocated resources and needs to inform the D2D pairs and cellular users of the scheduled resources for data transmission via L1/L2 control signaling, such as a physical downlink control channel (PDCCH) [4]. Thus, by the centralized resource allocation approach, system-wide optimal throughput can be achieved.

Capability of the Optimization Framework

In the above dynamic graph optimization framework, several technical problems, such as mode selection and resource allocation, are inherently modeled in the optimization formulation. With simple extensions, this dynamic graph optimization framework is capable of modeling other important technical problems for multihop D2D communication underlaying cellular networks. In this section, we briefly analyze how we can model and analyze these important problems in the proposed framework by considering communication mode selection, BS deployment and scheduling, and storage management and allocation.

Communication Mode Selection

Since D2D communication uses the same air interface as cellular communication, a receiver can only operate in either D2D mode or direct cellular mode. Furthermore, in D2D mode, a decision must be made whether to use multihop connected or opportunistic transmission. Given all the possible transmission modes involving all the UEs, the task of communication mode selection is how to utilize them in order to maximize the data transmission throughput.

In the dynamic graph optimization formulation, a cellular direct transmission corresponds to the flow from a BS vertex to a UE vertex, which is a one-hop transmission. A D2D transmission represents the data flow originating from a BS vertex, via one or more relay vertices at the same timeframe or subsequent timeframes, to a receiver vertex, which is a multihop flow occurring on some "connected" or "oppor-

The solution of the optimization problem yields the optimal resource sharing and mode selection for the cellular network with the given set of users. Thus, it represents the theoretical performance upper bound to the D2D-enabled cellular system.

tunistic" multi-edges. All these transmission modes are naturally modeled by the flows $f(a, b)$ occurring on all the possible edges (a, b) existing in the graph. Thus, we can utilize this dynamic graph optimization framework to investigate the transmission mode selection problem by optimizing the flow allocation on the edges of the graph.

Base Station Deployment and Scheduling

Since D2D communication can transmit data to nodes even without infrastructure coverage, BSs can adjust their coverage or even shut down completely in order to save energy while providing acceptable services by taking viable D2D alternatives [12]. For service providers and carriers, how to place the BSs given a set of possible candidate places as well as how to dynamically schedule the BSs given a network with groups of users and a set of operational BS candidates is an important problem. In particular, given available BSs and data transmission opportunities between BSs and UEs as well as between UEs, the BS Scheduling task is to dynamically schedule the BS operation, that is, to decide which BSs should be in operation and which should be shut down, in order to reduce energy consumption but still attain the maximum data transmission throughput from the operating BSs to all the UEs.

To model the above challenging problem in our proposed optimization framework, we first construct a graph with all the BSs as the vertices as usual, and then define a variable to indicate the operational status of BS i, where $1 \le i \le B$ at timeframe t_l, which is denoted by z_i^l. Specifically, candidate BS i is scheduled to be operational during timeframe t_l if $z_i^l = 1$; otherwise, $z_i^l = 0$. By adding all the operating BSs as a constraint and setting the corresponding weights of the edges connecting with the BSs that are shut down to a zero value, the BS deployment and scheduling can be incorporated into the dynamic graph optimization framework, and this allows us to study the optimal BS placement or scheduling policies by solving the corresponding flow maximization problems.

Storage Management and Allocation

Multihop D2D communication offers the benefits of saving communication power and improving the spectral efficiency, but requires the UEs to contribute local storage to buffer the data. Moreover, a decision process is needed to efficiently utilize the distributed storage in order to enhance the opportunities for user collaboration. Under this context, how to manage the distributed storage contributed by the UEs and how to allocate them efficiently for content access are among the important problems that need to be modeled and solved by some effective optimization tools.

The dynamic graph optimization framework enables us to investigate these challenging problems with some simple extensions. In the constructed graph, the edges connecting the UE vertices at consecutive timeframes represent the same UEs at different times. Thus, they can be utilized to represent the capability or possibility that the nodes physically carry data with their local storage during movement in time. By simply associating these edges with appropriate weight values, we can model how much storage each UE contributes. Furthermore, in a multi-content sharing system, by dividing the flow $f(a, b)$ over edge (a, b) into the sub-flows $f_m(a, b)$, each corresponding to the buffer size or transmission rate for an individual content, we can investigate the optimal storage policies for UEs in a D2D-enabled multi-content sharing system.

Application Scenarios and Challenges

Theoretical Performance Bound Analysis

One of the most important applications of our proposed dynamic graph optimization framework is to analyze the theoretical performance upper bound for the multihop D2D communication underlaying cellular network. It is clear that in order to obtain the fundamental performance bound achievable by multihop D2D communication, it is essential to first find the optimal solutions for the system (optimal resource allocation, optimal mode selection, etc.) under the practical network setting of a large number of users with realistic mobility patterns. From a system engineering viewpoint, in analyzing the theoretical performance bound, we can make some realistic assumptions: the availability of the preemptive knowledge of BS positions and user mobility trajectories. With the aid of this available system information, we can cast the multihop D2D communication underlaying cellular system as a dynamic graph optimization problem that maximizes the system utilities (e.g., content downloading throughput, wireless network capacity). The solution of the optimization problem yields the optimal resource sharing and mode selection for the cellular network with the given set of users. Thus, it represents the theoretical performance upper bound to the D2D-enabled cellular system.

In particular, targeting the mobile content downloading system as an example, by formulating a max-flow optimization problem that maximizes the content downloading flows from all the BSs to the content receivers through all the possible transmission modes, and multihop D2D connected and opportunistic transmissions under the dynamic graph optimization framework, we obtain the theoretical upper bound to the achievable system content downloading performance. This enables us to further investigate the fundamental problems of how D2D communication improves system performance and what the potential effect of the multihop D2D communication may be. Specifically, using a realistic mobility model and trace-driven simulations, we can evaluate the effects of the different system settings on the performance of the mobile content downloading system and reveal the fundamental influence of D2D communication.

System Design and Optimization

From the above discussion on the capability of the proposed optimization framework in solving technical problems of multihop D2D communications, we perceive another intuitive application, which is to solve difficult optimization problems in protocol and algorithm design.

Indeed, for the spatial-temporal dynamic multi-hop D2D communication system, it is hard to model and formulate the system under a realistic environment with tens of hundreds of cellular users and multiple D2D pairs. The proposed dynamic graph optimization framework may provide this well organized structure to model and investigate such difficult problems. However, some challenges need to be tackled in order to exploit this optimization framework in designing and analyzing these large-scale practical systems.

First, in the formulation of a dynamic graph optimization problem, constructing the transmission graph needs the information of the UEs' trajectories and the channel state information of cellular and D2D links. However, in practice, we do not have that information at the stage of calculating optimal solutions. In a practical system, the UEs are usually carried by human beings who exhibit regularity in daily mobility. Existing study [13] reveals that there is strong regularity in daily human and vehicular mobility (driven by humans) in terms of both temporal and spatial dimensions, and it finds 93 percent potential predictability in user mobility across the whole user base [13]. This regularity may potentially be exploited in prediction algorithm design [14]. Consequently, using mobility prediction and including the prediction information may help construct transmission graphs with the introduction of probabilistic graph-based representation. How to use daily regularity and prediction to design large-scale practical systems with the aid of real-life human mobility traces remains an open problem.

Second, under the large-scale practical network setting with tens of hundreds of nodes, the algorithms to solve the formulated optimization problems are also challenging to design. In such cases, usually the optimal solutions are obtained with very large overhead of computing time and cost. Thus, heuristic algorithms with low computational complexity and competitive performance are in demand. This is another open problem in fully utilizing the proposed dynamic graph optimization framework, and further investigation is needed.

Case Study and Performance Evaluation

In this section, we use a cellular network deployment to quantitatively analyze and assess the capability of the proposed optimization framework in revealing system performance bound and designing an optimization solution.

Targeted System

We target a cellular network by considering BS deployment with real-world human mobility traces of *Orlando* [15], which were obtained from volunteers who spent their Thanksgiving or Christmas holidays in the city of Orlando, Florida. In the trace collection, the GPS devices take readings of the users' current positions every 10 s and record them into a daily track log with a position accuracy of better than 3 m 95 percent of the time [15]. All the users are divided into relays and receivers with a ratio of 5:3, and their mobility is governed by the collected traces. In the network, multiple BSs are deployed to yield about 30 candidate locations to deploy BSs. For the spectrum sharing model, all the D2D communications use the same allocated frequency that is part of the cellular network's spectrum resource, while the D2D communication and cellular communication occur on different frequency channels. We limit the maximum node transmission range for D2D communication to 50 m. Other network parameters are based on the wireless propagation settings given in [16].

We concentrate on investigating the application of utilizing the dynamic graph optimization framework in the optimal BS scheduling design; that is, given the number of operational BSs, optimally choosing the operational BSs from the 30 candidate BSs (the remaining BSs are shut down) in order to save energy while providing acceptable services [12]. In this scenario, the service providers look for the opportunity to shut down entire underutilized BSs, and the corresponding load or other delay-tolerant service would be transferred to their neighboring cells or/and to the D2D communications outside the cellular coverage area. Under the above simulation settings, we generate the transmission graph and formulate the received content maximization problem with the decision variables of BS scheduling under the dynamic graph optimization framework, which as explained previously is naturally built with optimal decisions of mode selection and spectrum allocation. Thus, the output of this flow maximization problem implies *optimal* BS scheduling. In order to show the advantages of our optimization framework, we compare our optimal BS scheduling with three other strategies:

- **Random** scheduling, which randomly chooses the operational BSs according to a uniform distribution among the candidate BSs
- **Max node** scheduling, which chooses the BSs that cover large numbers of users to be operational
- **Max time** scheduling, which selects the BSs that cover the users over long average staying time as the operational BSs

Note that max node and max time are two efficient schemes for dynamic BS operation [12]. Thus, the compared schemes reflect the state of the art for BS scheduling.

Results Analysis

We use the mobility data to generate the dynamic transmission graph and evaluate the system performance in terms of content downloading throughput. The results generated are shown in Fig. 4, in comparison with the content downloading throughput obtained by the other three BS scheduling schemes. As can be clearly seen in Fig. 4, among the different BS scheduling strategies, our optimal scheme attains the best performance, and it significantly outperforms the max node and max time when there are fewer than 15 operational or deployed BSs. When the number of operational BSs increases to 15, the performance of max node and max time are closely matched to our optimal scheme. This is because 15 deployed BSs are sufficient for this simulated network. The results of Fig. 4 show that how the

In the network, multiple BSs are deployed to yield about 30 candidate locations to deploy BSs. For the spectrum sharing model, all the D2D communications use the same allocated frequency that is part of the cellular network's spectrum resource, while the D2D communication and the cellular communication occur on different frequency channels.

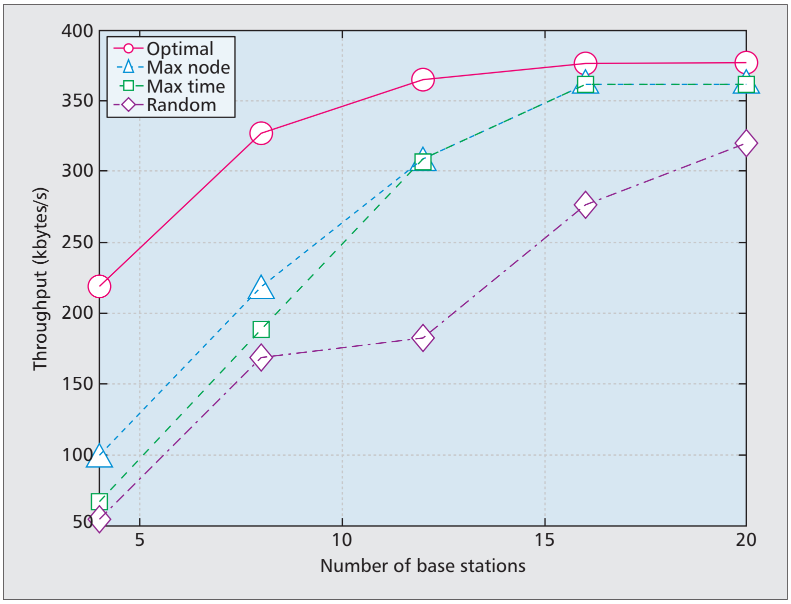

Figure 4. Comparison of content downloading throughput with the number of operational BSs increasing from 3 to 20.

BSs are deployed has more significant influence on the achievable system performance when the number of deployed BSs is limited, and our optimal BS scheduling solution is particularly effective in such limited BS coverage network scenarios.

Next, we further analyze the optimal BS scheduling solution of the dynamic graph optimization framework to investigate the obtained performance bound of the targeted system. Specifically, we examine the amounts of data transmission achieved by the three different modes of direct cellular transmission, multihop D2D connected, and opportunistic transmissions, respectively. The throughputs achieved by the different transmission modes in the optimal BS scheduling solution obtained by solving the flow maximization problem are shown in Fig. 5a. We observe that the importance of opportunistic transmission increases with the decrease of the number of operational BSs, and it accounts for about 40 percent of the total network throughput when the number of operational BSs is 4. This clearly demonstrates the performance enhancing capability of D2D communication. It is also clear that most of the data transmitted by multihop D2D communication are via the D2D opportunistic mode, and the D2D connected mode is only responsible for a small amount of the data transmitted. This indicates that the connected path from a BS with the assistance of some relays to a receiver may not be the most efficient method of data transmission.

To further observe the important role played by the multihop D2D communication in data transmission, we plot the ratio of the data transmitted by the two D2D transmission modes over the total transmitted data, and the results obtained by the four different BS scheduling solutions are depicted in Fig. 5b. We observe that the D2D transmission fraction attains the largest value under our optimal solution. Specifically, the D2D mobile data transmission in the system accounts for about 44 percent of the total data flow when the number of BSs is 4, and this decreases to about 28 percent when the number of BSs increases to 20. These results demonstrate the important role of multihop D2D transmission in the targeted simulation system, which further reveals the capability of our framework to analyze the performance bound of the multihop D2D communication underlaying cellular network.

Conclusions

We have proposed a dynamic graph optimization framework that exploits the dynamic transmission graph to model the spatio-temporal dynamics in terms of mobile users in the multihop D2D communication underlaying system. The proposed framework enables optimal algorithm design and performance bound analysis for the multihop underlaying D2D communication system. This study thus provides a useful model and optimization methodology for multihop D2D communication, and opens up a new research direction for the analysis and design of next-generation multihop transmission enabled cellular networks.

References

[1] S. Sesia, I. Toufik, and M. Baker, Eds., *LTE — the UMTS Long Term Evolution: From Theory to Practice*, Wiley, 2009.

[2] K. Doppler *et al.*, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," *IEEE Commun. Mag.*, vol. 47, no. 12, Dec. 2009, pp. 42–49.

[3] G. Fodor *et al.*, "Design Aspects of Network Assisted Device-to-Device Communications," *IEEE Commun. Mag.*, vol. 50, no. 3, Mar. 2012, pp. 170–77.

[4] L. Lei *et al.*, "Operator Controlled Device-to-Device Communications in LTE-Advanced Networks," *IEEE Wireless Commun.*, vol. 19, no. 3, June 2012, pp. 96–104.

[5] C.-H. Yu *et al.*, "Resource Sharing Optimization for Device-to-Device Communication Underlaying Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, Aug. 2011, pp. 2752–63.

[6] H. Min *et al.*, "Capacity Enhancement Using an Interference Limited Area for Device-to-Device Uplink Underlaying Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, Dec. 2011, pp. 3995–4000.

[7] B. Han *et al.*, "Mobile Data Offloading through Opportunistic Communications and Social Participation," *IEEE Trans. Mobile Computing*, vol. 11, no. 5, May 2012, pp. 821–34.

[8] H. D. Sherali and C. H. Tuncbilek, "A Global Optimization Algorithm for Polynomial Programming Problems Using a Reformulation-Linearization Technique," *J. Global Optimization*, vol. 2, no. 1, 1992, pp. 101–12.

[9] J. Löfberg, "YALMIP: A Toolbox for Modeling and Optimization in MATLAB," *Proc. 2004 IEEE Int'l. Symp. Computer Aided Control Systems Design*, Taipei, Taiwan, Sept. 2–4, 2004, pp. 284–89.

[10] F. Harary and G. Gupta, "Dynamic Graph Models," *Mathematical and Computer Modelling*, 1997.

[11] F. Malandrino, C. Casetti, and C. Chiasserini, "Content Downloading in Vehicular Networks: What Really Matters," *Proc. IEEE INFOCOM '11*, Shanghai, Beijing, Apr. 10–15, 2011.

[12] E. Oh *et al.*, "Toward Dynamic Energy-Efficient Operation of Cellular Network Infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, June 2011, pp. 56–61.

[13] C. Song *et al.*, "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, 2010, pp. 1018–21.

[14] W. Su, S.-J. Lee, and M. Gerla, "Mobility Prediction and Routing in Ad Hoc Wireless Networks," *Int'l. J. Network Management*, vol. 11, no. 1, 2001, pp. 3–30.

[15] K. R. Lee *et al.*, "Slaw: A New Mobility Model for Human Walks," *Proc. IEEE INFOCOM '09*, Rio de Janeiro, Brazil, April 19–25, 2009, pp. 855–63.

[16] C. Xu *et al.*, "Efficiency Resource Allocation for Device-to-Device Underlay Communication Systems: A Reverse Iterative Combinatorial Auction Based Approach," *IEEE JSAC*, vol. 31, no. 6, Sept. 2013, pp. 348–58.

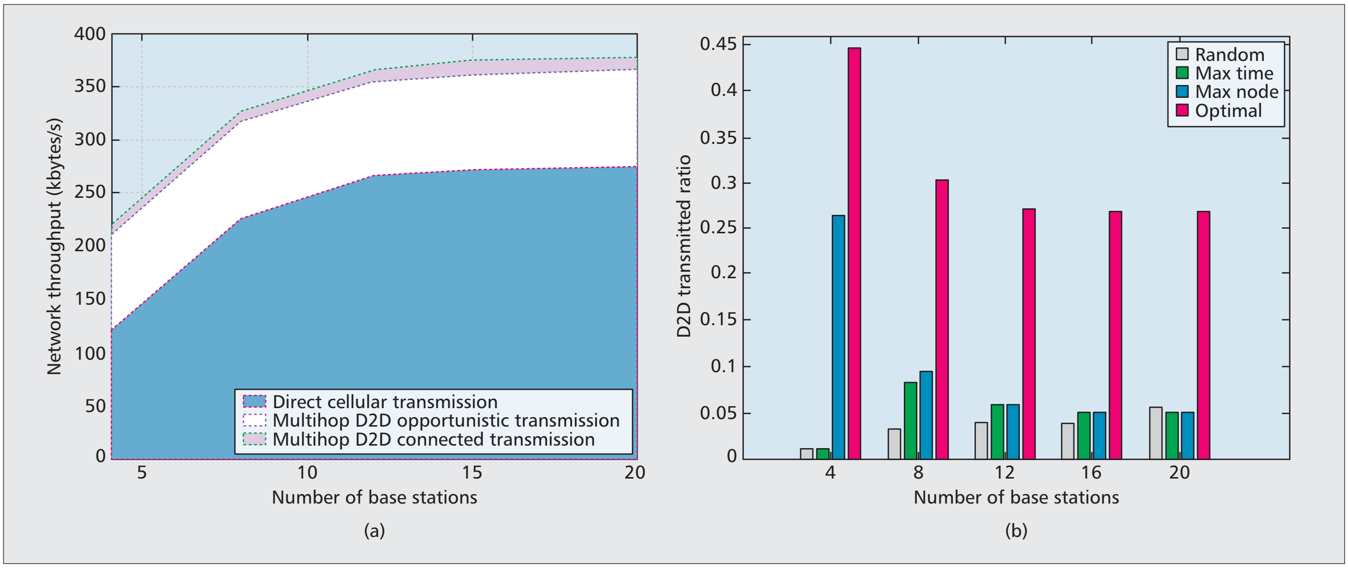

Figure 5. Multihop D2D performance bound analysis for the simulated scenario: a) throughputs achieved by the different transmission modes in the optimal BS scheduling solution of the dynamic graph optimization framework; b) comparison of the multihop D2D transmission fractions achieved by different BS scheduling solutions.

Biographies

YONG LI [M'09] (liyong07@tsinghua.edu.cn) received his B.S. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and his Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. In July to August of 2012 and 2013, he was a visiting research associate with Telekom Innovation Laboratories and the Hong Kong University of Science and Technology, respectively. From December 2013 to March 2014, he was a visiting scientist with the University of Miami, Coral Gables, Florida. He is currently a faculty member of the Department of Electronic Engineering, Tsinghua University. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet. He has served as a Technical Program Committee (TPC) Chair for the WWW workshop of Simplex 2013 and on the TPCs of several international workshops and conferences. He is a Guest Editor of ACM/Springer Mobile Networks and Applications, Special Issue on Software-Defined and Virtualized Future Wireless Networks. He is currently an Associate Editor of *EURASIP Journal on Wireless Communications and Networking*. He was a recipient of the Outstanding Postdoctoral Researcher, Outstanding Ph.D. Graduates, and Outstanding Doctoral Thesis awards from Tsinghua University. His research is granted by the Young Scientist Fund of the Natural Science Foundation of China, the Postdoctoral Special Fund of China, and industry companies Hitachi, ZET, and others.

CHAOMING SONG (csong@physics.miami.edu) is an assistant professor in the Physics Department at the University of Miami. He received his Ph.D. degree in physics from the City University of New York in 2008. His research interests lie in the intersection of statistical physics, network science, biological science, and computational social science, broadly exploring patterns behind petabytes of data. One of his recent activities in this area is aiming to understand the fundamental properties of human mobility and interactions on various scales. He is also actively contributing to the network science area, an interdisciplinary field studying complex interactions that aims to connect phenomena by merging different fields into a universal description. Currently he is serving as an editorial board member of *Nature: Scientific Reports* in the complex system section.

DEPENG JIN (jindp@tsinghua.edu.cn) received his B.S. and Ph.D. degrees from Tsinghua University in 1995 and 1999, respectively both in electronics engineering. He is an associate professor at Tsinghua University and vice chair of Department of Electronic Engineering. He was awarded the National Scientific and Technological Innovation Prize (Second Class) in 2002. His research fields include telecommunications, high-speed networks, ASIC design, and future Internet architecture.

SHENG CHEN [M'90, SM'97, F'08] (sqc@ecs.soton.ac.uk) obtained his B.Eng. degree from the East China Petroleum Institute, Dongying, in January 1982, and his Ph.D. degree from City University London, United Kingdom, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, D.Sc., from the University of Southampton, United Kingdom. From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh. and Portsmouth, all in the United Kingdom. Since 1999, he has been with the Department of Electronics and Computer Science, University of Southampton, where he currently holds the post of professor in Intelligent Systems and Signal Processing. He is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia. He is a Chartered Engineer and a Fellow of IET. His recent research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization. He has published over 470 research papers. He is an ISI highly cited researcher in the engineering category (March 2004).