eDirect: Energy-Efficient D2D-Assisted Relaying Framework for Cellular Signaling Reduction

Xiaomeng Yi, Li Pan, Yanqi Jin, Fangming Liu*, *Senior Member, IEEE,*
and Minghua Chen, *Senior Member, IEEE*

***Abstract*—Mobile Instant Messaging (IM) apps, such as WhatsApp and WeChat, frequently send heartbeat messages to remote servers to maintain their always-online status. Periodic heartbeat messages are small in size, but their transmissions incur heavy signaling traffic due to frequently establishing and releasing communication channels between Base Stations (BSs) and smartphones, known as the signaling storm. Meanwhile, smartphones also need to activate the cellular data communication module frequently for transmitting short heartbeat messages, resulting in substantial energy consumption. To address these issues, we present eDirect, an energy-efficient D2D-assIsted Relaying framEwork for Cellular signaling reducTion. eDirect selects active smartphones as relays to opportunistically collect heartbeat messages from nearby smartphones using energy-efficient D2D communication. The collected heartbeat messages are transmitted to the BS in an aggregated manner to reduce cellular signaling traffic. Based on the beating frequencies and deadlines of the collected heartbeat messages, eDirect schedules transmissions of the collected heartbeat messages to minimize signaling overhead and energy consumption while meeting the deadline constraints. We implement and evaluate our solution on Android smartphones. The results from real-world experiments show that our solution reduces signaling traffic by at least 50% and energy consumption by up to 36%.**

***Index Terms*—Energy-efficient Optimization, Signaling Storm, D2D, Heartbeat Message.**

I. Introduction

MOBILE Instant Messaging (IM) apps (e.g., WeChat, Line, and WhatsApp) are dominating in smartphones. These IM apps periodically send short signaling messages, called heartbeat messages, to remote servers to maintain the always-online status. Heartbeat messages are small in size but large in quantity. In Table I, we illustrate the fraction of heartbeat messages in the total number of transmitted messages in several popular apps, by summarizing the analysis in [1] and [2]. Nearly 50% of the messages are heartbeat messages.

TABLE I
Heartbeat Message Percentages in Popular Apps.

App	WeChat	WhatsApp	QQ	Facebook
Heartbeats	50%	61.9%	52.6%	48.4%

The frequent heartbeat transmissions have caused serious problems for both mobile network operators and smartphone users. For mobile network operators, frequent heartbeat transmissions caused by heavy smartphone usage in crowded areas often lead to serious overload in control channel [3], which is also known as the problem of *signaling storm*. All the transmissions over cellular networks require establishing Radio Resource Control (RRC) connections before data transmissions. The establishment and release of RRC connections cause substantial cellular signaling traffic. Frequent change of RRC state caused by heartbeat transmissions consumes a large amount of resource in control channels. For example, WeChat, one of the most prevalent IM apps in China, sends heartbeat messages every 270 seconds in Android. Its periodic heartbeat transmission, though accounts for only 10% of the cellular data traffic, occupies 60% of cellular signaling traffic, according to China Mobile (the biggest mobile operator in China). On the other hand, the heartbeat transmission results in substantial energy consumption for smartphones [4]. The network interface lingers in the high power state after a heartbeat transmission, wasting substantial energy. A smartphone spends at least 6% of its battery capacity in sending heartbeat messages even with only one IM app running [1]. In conclusion, the frequent heartbeat transmissions lead to two major issues: rapid drainage of the mobile device's battery and a signaling traffic surge in cellular control channels.

To reduce cellular signaling traffic, one approach is to send heartbeat messages in an aggregated fashion, without missing their deadlines. Some strategies, such as extending the beating frequencies of heartbeat messages, or delaying heartbeat messages and piggybacking them with other messages, are proposed in [2]. In addition, there are a lot of existing works that change the RRC mechanism to reduce RRC state transitions. These solutions not only cause extra energy consumption but also are hard to deploy for mobile network operators [5]. It is also conceivable to use Device-to-Device (D2D) communication[1] to collect heartbeat messages from nearby smartphones and send them in an aggregated

This work was supported in part by the NSFC under Grant 61722206 and 61761136014 and 392046569 (NSFC-DFG) and 61520106005, in part by National Key Research & Development (R&D) Plan under grant 2017YFB1001703, in part by the Fundamental Research Funds for the Central Universities under Grant 2017KFKJXX009 and 3004210116, and in part by the National Program for Support of Top-notch Young Professionals in National Program for Special Support of Eminent Professionals. (Corresponding author: Fangming Liu)

X. Yi, L. Pan, Y. Jin, and F. Liu are with the National Engineering Research Center for Big Data Technology and System, the Services Computing Technology and System Lab, Cluster and Grid Computing Lab in the School of Computer Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, China. E-mail: {xiaomengyi, panl, jinyanqi, fmliu}@hust.edu.cn.

M. Chen is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin N.T., Hong Kong, China. E-mail: minghua@ie.cuhk.edu.hk

[1]D2D communication, a promising paradigm in next generation cellular technologies, refers to direct communication between two mobile clients.

fashion to the BS so as to reduce cellular signaling traffic.

Based on the above discussions, in this paper, we propose eDirect, an energy-efficient D2D-assIsted Relaying framEwork for Cellular signaling reducTion. We set two roles, relay and User Equipment (UE), for participating smartphones in the framework. Relays, selected by the mobile network operator among the participating smartphone users, collect heartbeat messages from nearby UE(s) via D2D communication, schedule these messages, and transmit them to BS in an aggregated fashion. It is a win-win-win scheme for mobile network operators, relay users, and UE users. For mobile network operators, it can reduce the signaling traffic in cellular control channels at a low cost. For relay users, they can get payback from operators as they help infrastructure of mobile network reduce signaling traffic using its own connectivity and energy resource. For UE users, they can reduce the energy consumption in heartbeat transmissions since the D2D approaches, e.g., Bluetooth or Wi-Fi Direct, are more energy-efficient in short heartbeat transmission than the cellular approach.

Although the idea of relaying heartbeat messages via D2D connection reads simple, there are two key challenges we need to address. First, because of the inherent mobility of smartphones, the relay has to deal with the unpredictable heartbeat messages from the UE(s). Without message scheduling, the relay would suffer from excessive energy consumption for more data transmissions. Second, improper D2D pairs might consume more energy than the traditional cellular approach. In order to solve the first issue, we design a scheduling mechanism for transmitting the collected heartbeat messages, which aims to reduce the energy consumption while satisfying heartbeat deadline constraints. As for the second issue, we design a matching mechanism for UEs to find "proper" relays and determine whether to use D2D relaying to send heartbeat messages.

Furthermore, we implement a heartbeat relaying system with the proposed D2D-assisted framework on the Android platform, and evaluate its performance in reducing the cellular signaling traffic for mobile operators and saving energy for smartphone users. The prototype system consists of three main modules: D2D Detector, Message Monitor, and Message Scheduler. We utilize Wi-Fi Direct as the D2D connection technique in heartbeat transmissions for its compatibility among Android smartphones and considerable communication distance. Message Scheduler of UEs and relays varies in the implementation. The one of relays makes decisions on sending time of these collected heartbeat messages to achieve minimum energy consumption and maximum cellular signaling traffic reduction.

To summarize, our contributions are:

- We design a D2D-based heartbeat relaying framework, eDirect, to assist heartbeat transmission. The framework collects and schedules heartbeat messages among nearby smartphones through D2D communication.
- We design a D2D peer matching mechanism for the proposed framework. The matching mechanism helps UEs with discovery of proper relays, aiming to reduce energy waste in D2D peer discoveries and D2D transmission failures.
- We design a heartbeat message scheduling mechanism for the proposed framework. The scheduling mechanism determines the waiting time for heartbeat messages in an online fashion. The effective scheduling improves benefits of D2D relaying and reduces the total energy cost in heartbeat transmissions.
- We implement eDirect on the Android smartphone. In real-world measurement experiments, our system can achieve more than 50% signaling traffic reduction and up to 36% energy saving.

The remainder of this paper is organized as follows. In Section II, we introduce the background of the signaling storm problem and D2D techniques, as well as explaining the motivation of using a D2D approach to solve the problem. In Section III, we introduce the design of the D2D relaying framework and highlight the smartphone matching processing in Section IV and heartbeat message scheduling processing in Section V. In Section VI, we present the implementation of the prototype. Section VII evaluates the performance of the D2D relaying framework. Section VIII reviews related work. Section IX concludes the paper and discusses about the future work.

II. Background and Motivation

A. Heartbeat Message

Heartbeat messages are used to support real-time communications or push notification services for IM apps. IM servers set expiration timers to determine whether a client is online or not [3]. In order to maintain the online status, IM apps send heartbeat messages frequently to reset the expiration timers. Heartbeat messages are small in size, large in quantity, and do not require replies. For example, the heartbeat messages of QQ, WeChat, and WhatsApp are sent every 300 seconds, 270 seconds, and 240 seconds, respectively. Their sizes are 378 Bytes, 74 Bytes and 66 Bytes, respectively [1]. Some IM apps utilize the existing protocols or push service platforms, similar to apple push notification service (APN) in IOS, to implement heartbeat mechanism. For example, Google Talk is based on XMPP protocol [6], and Facebook Messenger uses MQTT protocol [7]. There are also many IM apps implementing the heartbeat mechanism using proprietary solutions.

B. Signaling Storm

The problem of a large volume of signaling traffic, leading to severe service outage or degraded network performance, is known as signaling storm. Signaling is used for information exchange about the establishment and control of a telecommunication channel and management of the network [8]. In mobile networks, RRC state machine, which is used to allocate the limited radio resources, is implemented in GPRS, EVDO, UMTS, and LTE Networks [9]. For LTE Networks, there are two main states: CONNECTED, a higher-power state, and IDLE, a lower-power state [5]. Before mobile clients send messages, terminals need to establish RRC connections. When data transmission is completed, the RRC connection is released. The establishment and release of RRC connection cause state transitions [10]. Periodic heartbeat transmissions, resulting from numerous apps attempting to stay online, incur frequent state transitions, and significant signaling traffic creating the signaling storm.

Mobile network operators have incentives to reduce the signaling storm, while still maintaining healthy data traffic, since they charge users based on their data traffic but not signaling traffic.

C. D2D Technique

D2D techniques, which support direct communication without BS intervention, are in rapid development and deployment. Currently, D2D techniques, including Bluetooth, Wi-Fi Direct, NFC, and ZigBee, have already been implemented in popular apps, such as PayPal and WeChat. In addition, as a promising and important paradigm in the next generation of cellular technologies, more advanced D2D techniques, such as LTE Direct developed by Qualcomm, are becoming mature. These techniques achieve better performance in neighbor discovery and energy saving than existing techniques. For illustration, LTE Direct enables D2D discovery of thousands of devices in proximity of approximately 500 meters. Inspired by the advantages of D2D techniques mentioned above and their promising development in the future, we propose a D2D-assisted framework for solving the signaling storm and energy consumption issues studied in this paper.

D. D2D Communication for Remitting Signaling Storm

D2D communications could be used to remit signaling storm, as it incurs no or little signaling traffic. The UEs send their heartbeat messages to the other smartphone via the D2D approach instead of cellular networks. In this way, the heartbeat transmissions will not bring any traffic to cellular control channels. The number of RRC connections decreases because the collected heartbeat messages from UEs are transmitted once by the relay in an aggregated fashion. Thus, if the relay mechanism is applied widely, the signaling traffic of cellular channels will decrease significantly.

A previous work [4] reveals that cellular communications incur higher energy consumption than common D2D communications over Bluetooth and Wi-Fi. Therefore, D2D techniques can potentially improve energy efficiency when utilizing D2D communication in heartbeat transmissions. There have been already many works focusing on the D2D-enabled cellular networks for energy efficiency enhancement, e.g., [11]. The technique of switching between cellular and D2D modes has been successfully applied in many existing studies. Therefore, the feasibility of applying D2D techniques cooperating with the cellular data transmission has been proved.

Furthermore, D2D communications are also perfectly suitable for the studied scenarios in this paper. The signaling storm problem usually occurs in the crowded regions. In these regions, more D2D peers and opportunities of high-quality D2D communication could be found.

III. System Design

eDirect aims to address the heartbeat-induced signaling storm in cellular control channels, while reducing the energy consumption in heartbeat transmissions as much as possible. To do so, we need to reduce the interactions between smartphones and BSs, by collecting heartbeat messages from nearby smartphones and transmitting them in an aggregated manner.

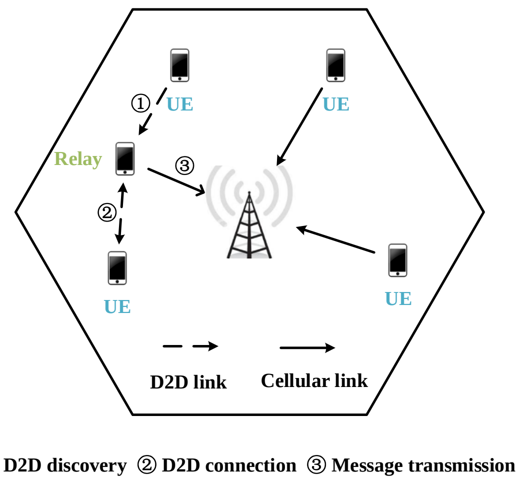


Fig. 1. An Illustrative Example of Forwarding Messages via a Relay.

A. System Overview

Our basic idea is to use relays to collect the heartbeat messages from nearby UEs and send them in an aggregated fashion. As illustrated in Fig. 1, our framework breaks down a heart transmission into three processing phases. **D2D discovery.** When a UE intends to send heartbeat messages by our framework, it activates the D2D discovering service in eDirect. **D2D connection.** If there are relays nearby discovered, based on the matching information of the discovered relays, the matching mechanism of eDirect determines whether using D2D relaying for heartbeat transmissions before establishing the D2D connection. **Message transmission.** After the D2D group formed, the UE sends its heartbeat messages to the relay by D2D communications. These forwarded heartbeat messages are scheduled by the relay to determine a sending time with the least cellular signaling traffic.

Fig. 2 depicts the overall architecture of eDirect. It consists of three main modules: Message Monitor, D2D Detector and Message Scheduler. Messages Monitor detects heartbeat transmissions, and then obtains their messages and transmission-related data. D2D Detector discovers D2D peers and establishes D2D connections. We use data flow and control flow to denote data transmission of heartbeat messages and scheduling related control messages respectively. In particular, once D2D connection is established, the UE can send its heartbeat messages to the relay. Along with the messages, the UE also notifies Message Scheduler in the relay with basic information of heartbeat messages such as the corresponding UE's ID, heartbeat period, deadline, and destination server. According to these information, the scheduler decides when to send all the received messages in a batch. For Message Scheduler in UE, it decides whether to send heartbeat messages through cellular network according to control information from its D2D Detector and relay's Message Scheduler. The D2D Detector notifies the Message Scheduler whether the UE is connected

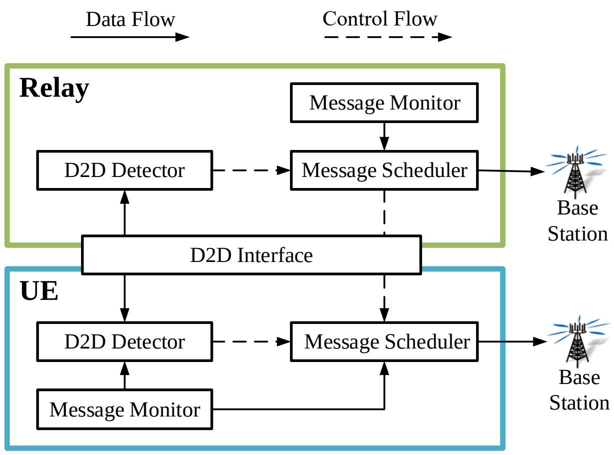


Fig. 2. eDirect: Software Architecture for Heartbeat Relaying.

to a relay and information from relay's Message Scheduler indicates whether a heartbeat message is sent out by the relay in time.

B. Design Insights

We discuss several design considerations in the following.

To be a relay. In this part, we discuss two issues, i.e., how to be a relay and what are the incentives for mobile users to be a relay. For the first issue, every smartphone user with network connected can apply to be a relay through our framework. It is worth noting that, unlike peer-to-peer systems, in eDirect there is no synchronization mechanism to elect relays from a bunch of UEs. Instead, relay selection is performed in a centralized fashion. Based on the crowd density of the region, the network operator can estimate the number of UEs that could be transmitted to alleviate signaling storm. Before applying to be a relay, a mobile user can perform D2D peer discovery, so as to record the ID of UEs it can potentially connect. According to the list of UEs reported by each relay candidate, the network operator can properly select relays by approving or rejecting the applications. In eDirect, we utilize a simple relay selection method that greedily selects the relay with the largest number of discovered D2D peers until its limited number of connected UEs is reached.

To deal with the incentive issue in eDirect, we offer a feasible bonus mechanism for relays. For the reason that relays contribute in cellular signaling traffic reduction but suffer from extra cellular traffic and energy consumption, it is reasonable for the operator to give them some bonuses, such as offering some free cellular data, or reducing their service cost. Through the following bonus mechanism, the operator could implement our proposed solution without incurring a high cost. In practice, according to the load of base station, the network operator can calculate its benefit if a mobile device is relayed, correspondingly it can set the bonus to relays for relaying a UE. According to the bonus, each user can decide whether to become a relay. This can be modeled as an auction problem where the network operator bids for using users' devices as relays. How to design an auction like incentive mechanism with ideal properties such as truthfulness, fairness and social ware maximization is an interesting research topic, and has already been well-studied in D2D offloading studies, e.g., [12] and [13]. The discussion of the incentive mechanism in detail for operator is beyond the scope of this paper.

To illustrate the feasibility of this bonus mechanism, we take a commercial product, named Karma Go [14], as an example, which equips a similar bonus mechanism. Karma Go could grab a cellular connection and turn it into a personal Wi-Fi signal, which could be shared with people nearby. When people nearby get online through Karma Go, the owner of Karma Go earns rewards, either \$1 of credit or 100 MB of free data.

Privacy concerns. Admittedly, forwarding heartbeat messages might introduce concerns of leaking users' private information, such as users' ID. In fact, our framework nearly does not bring any problems in security even when the relays are malicious. Firstly, D2D techniques and heartbeat protocols have already been armed with multiple security mechanisms. For example, Wi-Fi Direct, a common D2D technique, offers Wi-Fi Protected Access 2 (WPA2) technique. Besides, the forwarded heartbeat messages based on MQ Telemetry Transport (MQTT) will be encrypted via Secure Sockets Layer (SSL) before they are sent to the relay. Thus, even if the relay obtains the forwarded messages, it would not get the encrypted data within.

Mobile devices vs. fixed Access Points (APs). Although it is convenient and economical to utilize the deployed fixed APs to forward heartbeat messages, we adopt smartphones as relays in our scheme for the following reasons. Firstly, these fixed APs are beyond control of mobile network operators. The fixed APs deployed by different units are hard to be regarded as steady and reliable relays. Furthermore, the scalability will be constrained by the number and distribution of these APs. Because smartphones' mobility, the random demands are hard to be satisfied by the fixed APs. Based on the analysis above, we choose mobile users as relays. The D2D technique, as a key paradigm in the next generation of cellular networks, avoids extra deployment cost for mobile network operators. Compared with upgrading the existing cellular networks, the D2D-based scheme inherently has the potential to be more economical and easier in deployment.

Incentives for developers to integrate our framework. Admittedly, the extra engineering workload of modifying the code of existing apps is required to integrate our framework. However, developers still have motivations for the modification. Firstly, the framework is effective in reducing signaling traffic. Controlling signaling traffic is regarded as the responsibility for developers, according to Global System for Mobile Communications Association (GSMA) guidelines. Furthermore, [15] concludes that developers do care and think about the energy consumption when they build applications, but lack the necessary information and support infrastructure about green software engineering. This framework provides an easy and feasible scheme for developers to achieve energy-efficient feature in heartbeat transmissions.

Feedback mechanism. When the transmission failure happens in our framework, we need a mechanism to discover it and deal with it. Thus, we design a feedback mechanism to guarantee successful heartbeat transmissions. We

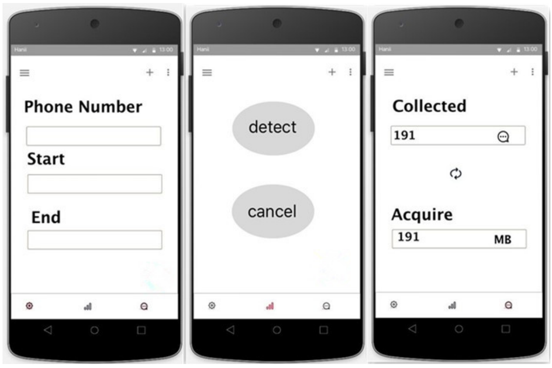


Fig. 3. eDirect: UI for Relays.

set a boolean value named `transmissionResult`, with the `false` value initially, to indicate if the heartbeat message is sent successfully. Once the matched relay sending the collected heartbeat messages successfully, it will set `transmissionResult` as `true` and feed back to the requesting UE. In case that the UE does not receive the feedback information after a heartbeat period, it will send the heartbeat message via cellular networks. For the reason that deadlines of heartbeat messages are always a few times of their cycles, the latency caused by the feedback mechanism would not impact the keep-alive function.

C. UI Design

For convenience of interacting with our framework, such as applying for relays and exchanging the bonus between relay users and mobile network operators, we design an interface, as shown in Fig. 3, for mobile users and mobile network operators. The smartphone users could apply to be a relay through the application. Besides, relays are able to activate our framework through the UI, which also presents the amount of the collected heartbeat messages and the bonus from mobile network operators.

IV. D2D Peer Matching

A. Challenges in D2D Connection Processing

eDirect might consume more energy than the cellular approach in heartbeat transmissions. The extra energy consumption occurs in two situations. The first one is caused by the short D2D connection, which means the established D2D connection only sends few messages during the short D2D duration. Intuitively, D2D techniques are more energy-efficient in data transmissions. However, the energy consumption in discovering and connecting devices for D2D techniques is considerable in short D2D connections. Thus, we need to optimize the energy consumption by minimizing the times of D2D discovery and connection, and maximizing the period of the more energy-efficient D2D transmission. The second one is caused by heartbeat transmission failures in our framework. The transmission failure can be ascribed to the transmission exceptions of the relays in our framework. For example, when a relay runs out of its battery before all the forwarded heartbeat messages are sent to the BS, some heartbeat messages might not meet their deadlines. Besides, during the data transmission via D2D communications, when the physical distance between the involved smartphones exceeds the maximum communication distance of the chosen D2D technique, the heartbeat messages fail to be forwarded to the relay. Thus, we need a matching mechanism to make a prejudgment before establishing the D2D connection to avoid these unsatisfactory situations.

B. D2D Peer Matching Mechanism

The main purpose of the D2D peer matching mechanism is to constrain the energy consumption caused by transmission accidents in eDirect. In our framework, the D2D peer matching mechanism is designed to choose a proper relay for UEs. We set two parameters: the physical distance between the requesting UE and the discovered relay candidate, the intention to be connected with the discovered relay. We use Received Signal Strength Indication (RSSI), able to be obtained during the D2D discovery phase, to indicate the physical distance between the requesting UE and discovered relay candidate. The longer physical distance between the two mobile devices means higher possibility of disconnection between them, because of the limited communication distance of D2D techniques and the inherent mobility of smartphones. Thus, the D2D peer matching mechanism always prefers the closer relay candidate.

In practice, letting each UE connect to the relay with the largest RSSI is a greedy strategy, which might not give globally optimal performance. If we consider relays as knapsacks and UEs as items, when a UE is assigned to different relays we can acquire different values. Searching for the optimal assignment of UEs is a multi-knapsack problem, which is NP-hard in general. Moreover, in D2D communication the global information is usually not available to individual. Therefore, the problem should be solved in a peer-to-peer manner, which makes it even harder to get the optimal solution. In Sec. VII, we investigate the impact of different peer matching strategies, and demonstrate that the greedy strategy can achieve near-optimal energy consumption.

To allow relays to indicate their intention to forward messages, we use `groupOwnerIntend`, a parameter defined in `WifiP2pManager`, offered by Android API. This parameter, ranging from 0 to 15, refers to the current available capacity of the relay. It is set by relay users in practice. The relay users could adjust the value of this parameter based on their willingness to be a relay and smartphone status, such as their battery usage and current message relay frequency. For instance, a relay with sufficient battery power and low message relay frequency might be more willing to forward messages from other UEs, therefore it can set `groupOwnerIntend` with a higher value.

The system contains a parameter mapping mechanism to relay's `groupOwnerIntend` value and UE's demand for transmission. In D2D discovery phase, along with the D2D peer discovery messages, the relay sends the value of `groupOwnerIntend` parameter to UEs. According to the received `groupOwnerIntend` value and the heartbeat period of IM apps, a UE picks out relay candidates, from all discovered relays, that can satisfy its transmission demand. Then it chooses the relay candidate with the largest RSSI.

When the D2D peer matching mechanism fails to find a proper relay for the UE, it notifies the UE to send heartbeat messages via cellular networks directly.

V. HEARTBEAT MESSAGE SCHEDULING

The heartbeat message scheduling processing achieves two main functions in eDirect. The first one is to maximize the benefit of D2D relaying through scheduling the transmission time of the collected heartbeat messages from UEs and the local heartbeat messages. The second one is to determine the transmission approach, the D2D approach or cellular approach, to transmit heartbeat messages for UEs. We describe the Message Scheduler in relays and UEs respectively, corresponding to the two purposes.

TABLE II
NOTATIONS OF THE SCHEDULING ALGORITHMS.

Notation	Description
T	The smallest cycle of the collected heartbeat messages
M	The maximum of the collected heartbeat messages
n	The number of the forwarded heartbeat messages from UEs
T_n	The deadline of the No.n forwarded heartbeat message
a_n	The arrival time of the No.n heartbeat message
t_n	The sending time of relay's heartbeat message
y	The cost of the heartbeat transmission via the D2D approach
z	The cost of the heartbeat transmission via the D2D approach for relays
β	The weighted cost of the latency in heartbeat transmissions
Δt	The period from the arriving time of the last received message to the present

A. Message Scheduler in Relays

This component is designed to reduce signaling traffic through delaying the local and forwarded heartbeat messages in relay, and sending them in a cellular connection, instead of each heartbeat message establishing a cellular connection. Message scheduling algorithm in relays dispatches the local and forwarded heartbeat messages at right time, when the peak benefit of heartbeat forwarding is reached.

1) Heartbeat Scheduling Problem: When an unpredictable stream of heartbeat messages arrives at a relay, the relay needs to send these heartbeat messages together with its local heartbeat messages. Delaying these heartbeat messages longer, the possibility of collecting new heartbeat messages grows, which would reduce more signaling traffic and energy consumption. However, delaying them too long would interfere with the heartbeat mechanism and relays cannot hold unlimited forwarded heartbeat messages for their limited capacity. Thus, we aim to find a right time to achieve minimal cost in heartbeat transmissions.

We give the definition of our problem. We list the main notations in Table II. The inputs are the sequence of arrival time of the forwarded heartbeat messages I, the maximum of the collected heartbeat M, and the smallest cycle of the sending heartbeat messages T. The smallest cycle of the sending heartbeat messages requires to be updated once the new forwarded heartbeat messages arrive. The output is a set of time $t_1, ..., t_k$. The cost of the heartbeat transmission for relays is defined as

$$z = k\beta + \sum_{i=1}^{k} latency(i). \tag{1}$$

The cost consists of two parts. The first part is the energy consumption in heartbeat transmissions. The second part indicates the cost in delaying heartbeat messages.

2) Solutions: The heartbeat scheduling problem is similar to TCP acknowledgment problem, defined by Daniel [16]. It is origin from the classic ski-rental problem. We refer to several researches about this problem. Nagle [17] offered a batching example. To reduce the network congestion caused by small data packets, it combines a number of small data packets with a buffer constrained, and sends them all at once. Seidan [18] proposed a randomized online algorithm with a competitive ratio of $e/(e-1)$. Based on the design insights of them, we present our solution.

Heartbeat scheduling algorithm for delay-tolerant applications. For some applications, such as advertisement and diagnosis information, their performance will not be impacted by their delayed heartbeat message in a heartbeat cycle. In this scenario, ignoring the cost of latency, we propose a message scheduling algorithm for these applications. When Message Scheduler delays the local heartbeat message, relay is likely to receive some forwarded heartbeat messages during the waiting time, which leads to lower energy cost. However, the number of the collected heartbeat messages must be less than M, because of the limited capacity of relays. Besides, to guarantee the forwarding success, we also need to constrain the waiting time $t - t_n$ of the No. n heartbeat message less than the smallest heartbeat cycle T. In sum, the Algorithm 1 delays the transmission as long as possible to get the most forwarded messages before reaching the constraints. Once the new heartbeat messages arriving, the constraints, M and T, will be updated according to the parameters of the new heartbeat messages.

Algorithm 1 Heartbeat Scheduling Algorithm for Delay-Tolerant Applications.

Input: M, T
Initialize $n \leftarrow 0$;
When forwarded message arrives
Get a_n;
Update T;
$n++$;
if $n < M$ && $latency < T$ **then**
  Pending;
else
  Send data now;
end if

Online scheduling algorithm for delay-sensitive applications. For the applications requiring instant notifications, the longer latency of the heartbeat messages increases the probability of delaying push services. In this part, we extend

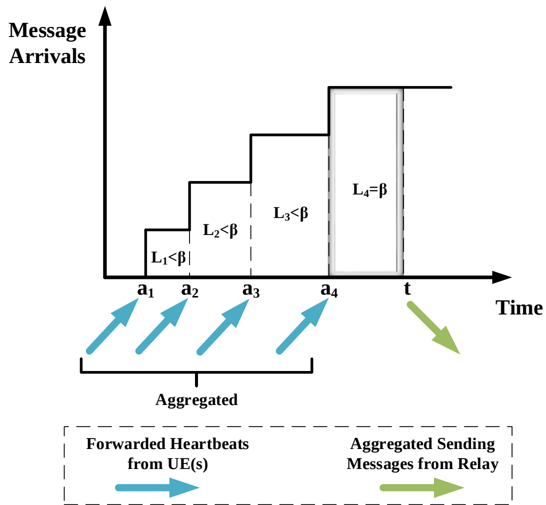


Fig. 4. An Illustrative Example of Algorithm 2.

our message scheduling algorithm, Algorithm 1, to adapt to these applications. In this solution, we take the impacts of the delayed messages into consideration. We take a cycle of the local heartbeat message as the scheduling period. Thus, the total cost in the heartbeat transmission is defined as

$$y = \beta + latency(n),$$

and our objective is to minimize the cost.

Let $M(t, t')$ be the number of the received heartbeat messages between the time t and t'. The latency cost is:

$$latency(n) = \sum_{i=1}^{n} M(a_{i-1}, a_i)(t - a_i)$$
$$a_0 = 0,$$

and there is no heartbeat message received in the period $[a_n, t]$. As illustrated in Fig. 4, the area $L_1 + L_2$ represents the $latency(2)$. For one more heartbeat message received, a β can be saved. We set

$$\Delta latency = \beta \tag{2}$$

as the threshold to send the aggregated heartbeat messages. In the example, when L_4 reaches β, our framework sends the aggregated heartbeat messages.

As shown in Fig. 4, we take a cycle of the local heartbeat message as the scheduling period. The original sending time of the local heartbeat message is set as 0. The Algorithm 2 always checks $\Delta latency$, i.e. area $L_i (i \in [1, n])$ in the figure. As time goes, if $L_i < \beta$ when a_{i+1} arrives, the algorithm checks L_{i+1} until it equals to β. The process is presented in Algorithm 2.

3) Algorithm Analysis: In the heartbeat scheduling problem, the forwarded heartbeat messages are unpredictable, but the decision depends on the future information. Thus, we take the online design for this algorithm. The Algorithm 2 is a deterministic online algorithm. We always utilize competitive ratio to analyze an online algorithm. The competitive ratio is defined as the ratio between the performance of the online algorithm in the worst case and the performance of the optimal offline algorithm. The online algorithm with a smaller

Algorithm 2 Deterministic Online Algorithm for Delay-Sensitive Applications.

$t \in [0, T)$
Input: M, T, β
Initialize $n \leftarrow 0$;
When forwarded message arrives
Get a_n;
Update T;
$n{+}{+}$;
Ensure: $n < M$ && $latency < T$
if $M(0, a_n)(t - a_n) = \beta$ **then**
  Send data at time t;
else
  Pending;
end if

competitive ratio has better performance when unpredictable events occur. Thus, even without any assumption of the future information, the online algorithm with a small competitive ratio is still robust.

Theorem 1. *Algorithm 2 is 2-competitive.*

Proof. Consider an arbitrary input sequence I. Based on the Eq. (2), we can obtain

$$L_i \leqslant \beta.$$

And the offline optimal solution, noted as OPT, sends heartbeat messages between any pair of successive sending time of Algorithm 2, proved in [19]. Thus,

$$\begin{aligned} z &\leqslant E_{OPT} + latency_{OPT} + \beta \\ &\leqslant 2(E_{OPT} + latency_{OPT}) = 2z_{OPT}. \end{aligned}$$

4) Discussion: The heartbeat scheduling problem in heartbeat forwarding is similar to the classic ski-rental problem. In the ski-rental problem, the decision is whether to rent skis or to buy them with unpredictable future trips when go skiing. In the heartbeat problem, the decision is whether to send the current heartbeat messages or to wait for the possible incoming heartbeat messages in the future. In the first cycle of relay's local heartbeat message, the heartbeat scheduling problem is the same as the classic ski-rental problem. However, the difference between them occurs in the next cycle. In the next cycle, the forwarded heartbeat messages from the connected UEs are predictable for the heartbeat messages are periodic. When the connected UEs do not change, the heartbeat problem becomes an offline scheduling problem. Therefore, the Message Scheduler of eDirect will not run the Algorithm 2, and set $t = a_k$. With this feature, the message scheduling algorithm achieves less computation and closer to the offline optimal algorithm than the deterministic online algorithm in the ski-rental problem.

B. Message Scheduler in UEs

In terms of the Message Scheduler in UEs, its main purpose is to determine the transmission approach of heartbeat messages. It relies on the results of the D2D peer matching mechanism discussed in Sec. IV and the message scheduling algorithms in relays.

The heartbeat messages in UEs will be sent through the D2D approach only if it successfully matches a proper peer and the received `transmissionResult` equals `true`. Otherwise, UEs will send their heartbeat messages through cellular transmission.

VI. Implementation

We now present the implementation of the prototype system based on eDirect on Android smartphones. The prototype system allows users to discover D2D peers via Wi-Fi Direct. Then, the system selects the discovered relay(s) based on the D2D peer matching mechanism. If a proper relay is found, the UE will build a D2D connection with it. The connected UE can send its heartbeat messages to the relay through Wi-Fi Direct. The relay schedules the forwarded heartbeat messages via the message scheduling algorithms presented in Sec. V.

A. The Choice of D2D Techniques

Nowadays, there exist several popular D2D techniques, which have already been applied widely. For example, we share music across smartphones via Bluetooth. And, we utilize Near Field Communication (NFC) to recharge IC cards. Although Bluetooth has the potential to complete D2D communications with lower energy consumption, its communication range is typically less than 10 meters, which is hard to satisfy our requirements. Besides, LTE Direct as an innovative D2D technique enables us to discover thousands of devices in the proximity of approximately 500 meters [20]. Nonetheless, many countries, such as China, have not deployed the technique yet. Because of its limited deployment, we have to abandon this alternative for generality consideration. Wi-Fi Direct has an ideal communication distance and generality, which makes it perfect in these scenarios. It is equipped with longer communication distance and higher transmission speed than Bluetooth. In terms of generality, the smartphones with Android OS newer than 4.0 version support Wi-Fi Direct technique, which enables smartphones to connect to others from different manufacturers without any extra hardware deployment. Thus, we choose Wi-Fi Direct among these popular D2D techniques in eDirect, due to its advantages over other D2D techniques in heartbeat transmission scenarios.

B. Message Monitor

In the Message Monitor module, the major challenge for us is to detect heartbeat messages. Message Monitor is required to obtain the forwarding messages and their related parameters, i.e., their beat frequencies and deadlines. However, it is hard to obtain the data across applications in Android application layer without the permission of app developers and users. Android offers `Content Provider`, as a basic component of Android, to realize data exchange across applications. In eDirect, we offer a set of APIs for app developers to integrate it into their existing apps. The developers need to submit the forwarding messages and their related parameters through the interfaces.

C. D2D Discovery and Connection

The D2D Detector aims to discover suitable D2D peers for UEs and determines whether to establish the D2D connection through Wi-Fi Direct. A `Broadcast Receiver`, which responds to the events relevant to Wi-Fi Direct, receives the Wi-Fi Direct intents. There are four intents broadcasted by Android system, i.e., `WIFI_P2P_PEERS_CHANGED_ACTION`, `WIFI_P2P_STATE_CHANGED_ACTION`, `WIFI_P2P_CONNECTION_CHANGED_ACTION`, and `WIFI_P2P_THIS_DEVICE_CHANGED_ACTION`, required to respond [21]. To discover D2D peers that are available to build a D2D connection in the range, `WifiP2pManager` class offers a set of methods. We implement the `ActionListener` interface offered by `WifiP2pManager` class, the system is able to be invoked by successful and unsuccessful discoveries. In unsuccessful cases, we utilize `Toast` to show the failure message to notify users. After fetching the list of available D2D peers, we set the parameter `groupOwnerIntend`, which represents the inclination to be a group owner, as 15, the maximum value for relays initially. Moreover, the D2D peer matching mechanism would reduce `groupOwnerIntend` of the relay proportionally until 0 while it collects heartbeat messages from the connected UE(s). For UEs, we set `groupOwnerIntend` as 0, the minimum value.

VII. Performance Evaluation

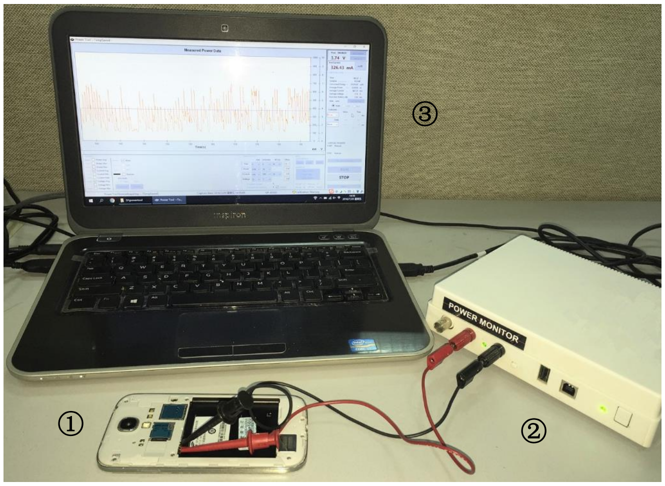
To evaluate eDirect, we implement the prototype system on Android smartphones. We start with characterizing the performance of eDirect in terms of cellular signaling traffic reduction and energy saving. Then, we analyze the possible negative impacts caused by our framework in different scenarios.

A. Performance in Energy Saving

To comprehensively evaluate the performance in energy saving, we conducted several experiments. In the experiments, we ran the prototype on Samsung Galaxy S4 smartphone with Android OS and captured the instant current every 0.1 seconds through Power Monitor, as shown in Fig. 5. Power Monitor captures the instant current under the constant voltage 3.7V, which is able to obtain instant energy consumption of the smartphone with the help of the power tool software in PC [22].

First, we attempt to discover energy consumption in data transmissions via different approaches. We measured the instant current while the smartphone was sending the same message using our D2D approach and the cellular approach, respectively. The results, shown in Fig. 6 and Fig. 7, reveal that the instant current, in the D2D transmission, spurts at the beginning of the data transmission, and then it descends rapidly. As for the cellular transmission, the instant current spurts and lasts for a longer period. Thus, in terms of data transmissions, Wi-Fi Direct has higher energy efficiency than cellular networks, which is the preliminary evidence for the energy-saving feature of our D2D based framework.

Furthermore, the D2D approach may consume substantial energy. Compared with cellular transmission, extra phases like peer discovery and connection are required before data transmission. Moreover, a relay should keep the WiFi interface active to accept new WiFi Direct connections. Is



① Smartphone equipped with eDirect
② Power Monitor capturing instant current
③ Power Tool software running in PC

Fig. 5. Experiment Setup for Energy Measurement.

the energy consumption overhead in D2D communication significant, compared with the energy consumption in the data transmission? To figure out this question, we analyze the energy consumption in different phases of the D2D relaying framework for relays and UEs respectively, as shown in Table III. In the D2D discovery, the UEs initiate the D2D discovery service, which leads to more energy consumption than relays. In the data transmission, UEs only need to send their messages to relays through the D2D approach, while relays have two phases in the data transmission, the receiving phase and sending phase. In the receiving phase, relays collect the forwarded messages through the D2D approach. As for the sending phase, they send the forwarded messages combined with their own messages through the cellular approach. Based on the results in Table III, we can conclude that the energy consumption gaps between relays and UEs in the D2D discovery and connection phases are narrow; the energy consumption in the D2D discovery and connection are accounted for a large proportion in the whole transmission. We also evaluate the energy consumption of keeping WiFi interface active, so as to analyze the energy overhead for relays. On average, keeping WiFi active would cost 281.6 μAh per minute, which contributes to a large portion of relay energy consumption. The above analysis explains why the proposed D2D relaying framework might consume more energy than the traditional cellular approach in some cases. We show in our paper that our proposed framework incurs minor energy (about 36% less) than the traditional approaches in typical scenarios.

We investigate the energy consumed to send heartbeat messages during daily usage of IM applications. We analyze the energy consumption of WeChat and QQ, two popular IM applications in China, during a period of 90 minutes. As a comparison, we also evaluated the consumed energy when the APPs send messages through the D2D approach. From Fig. 8, we can observe that cellular transmission contributes to around 30% of overall energy consumption for both WeChat and QQ. By simply using eDirect to send heartbeat messages via D2D approach, the energy consumption can reduce by up to 20%.

To explore the energy efficiency of our prototype system, we measured its energy consumption in different scenarios. Firstly, we set different D2D connection durations, which reflect the number of forwarded heartbeat messages, in the experiments. We use transmission time of the forwarded heartbeat messages during the D2D connection as a variable. The messages are in standard size, 54 Bytes. In addition, we use a relay connected with one UE, 1 meter away from it. In the following analysis, if not specified, we use “Cellular Transmission” to denote the baseline energy consumption for a single device that sends heartbeat messages through cellular network. “UE” and “Relay” denote the energy consumption of one UE and one relay respectively. We use “eDirect” to denote the energy of all devices (i.e., both UEs and relays) in the evaluation. Therefore, “Saved energy of eDirect” as the saved energy in terms of all devices, is calculated as the energy consumption of “Cellular Transmission” multiplies the total number of devices subtracted by the energy consumption of “eDirect”.

Fig. 9 shows that the energy consumption of the relay and cellular transmission increases with the number of forwarded messages increasing, but the increased range of the UE largely falls behind the relay and cellular transmission. Moreover, the difference in energy consumption between the relay and cellular transmission nearly stays constant, although the energy consumption of the relay is always slightly higher than that of cellular transmission. This observation indicates that the major extra energy consumption of relay comes from the need to keep WiFi interface active. As a result, as the number of forwarded messages grows, the saved energy of the UE will considerably exceed the extra energy spent by the relay. Fig. 10 shows that on the period of the first message forwarded, the energy consumption of the D2D approach reaches nearly the same as cellular transmission. As for the UE, it saves 55% of the energy consumption as compared to the energy consumption in the cellular approach. When seven heartbeat messages are forwarded, the proposed system based on the D2D communication could save 36% of the energy consumption. In brief, if the D2D connection lasts longer, our D2D relaying framework will save more energy. For the UE, it always saves tremendous energy.

To disclose more details about the energy consumption in this scenario, we analyze the difference of energy consumption between different phases of the D2D transmission for relays and UEs respectively, as shown in Table III. From the results of Table III, the energy consumption in the D2D discovery and connection are similar; the major difference in energy consumption occurs in the D2D forwarding phase. The relay spends much more energy in receiving messages than the consumed energy of the UE to send messages.

TABLE III
ENERGY CONSUMPTION IN DIFFERENT PHASES.

	Discovery	Connection	Forwarding
UE (μAh)	132.24	63.74	73.09
Relay (μAh)	122.50	60.29	132.45

Considering that a relay might serve multiple UEs in real-world scenario, we then conduct the experiment to evaluate

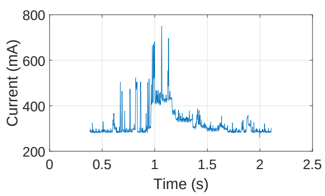


Fig. 6. Energy Consumption in the D2D Transmission.

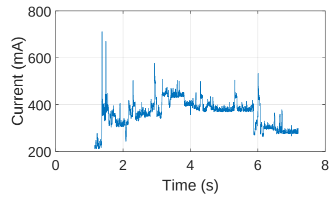


Fig. 7. Energy Consumption in the Cellular Transmission.

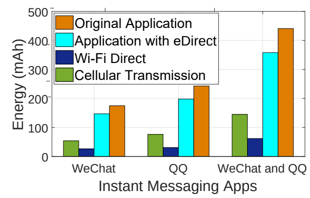


Fig. 8. Energy Consumption compared with Instant Messaging Apps.

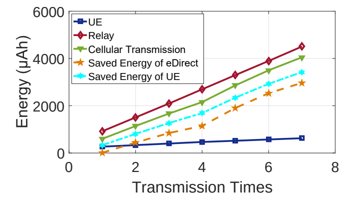


Fig. 9. Energy Consumption of Cellular Transmission, UE, and Relay.

the energy consumption in this situation. We configure the initial value of parameter `groupOwnerIntend` to be 15 for the relay and reduce the value by 2 whenever a UE is connected. Therefore, a relay can connect at most 7 UEs. The size of heartbeat messages and the communication distance keep the previous settings. Fig. 11 gives a comparison of the energy consumption in the case that a relay is connected with different numbers of UEs. It is evident that the increased energy consumption caused by more UEs connected is large in amount for the relay when only a small amount of heartbeat messages forwarded. However, when the D2D connection lasts long enough, the impact of the multiple connected UEs can be neglected for its little proportion. Fig. 12 shows the relationship between the extra energy wasted by the relay and the energy saved by the UE. With more UEs connected with a relay and longer D2D connection duration, the ratio between the extra energy wasted by the relay and the energy saved by the UE drops from around 97% to 5%. To find out the reason, we measured the energy consumption in D2D discovery phase, D2D connection phase, and D2D forwarding phase, respectively. According to the composition of the energy consumption, we find that the relay has to be waked more frequently to receive messages because of the connected UEs increasing, which leads to the significant difference in the energy consumption. We present the detailed data of the energy consumption in receiving messages in Table IV, which concludes an approximate linear relationship between the energy consumption in receiving data and the number of the connected UEs. In addition, the excessive energy caused by multiple connected UEs is significantly less than the energy consumption in sending the collected messages. Consequently, the influence of multiple connected UEs drops gradually as the number of the sent heartbeat messages increasing.

TABLE IV
ENERGY CONSUMPTION IN THE DATA TRANSMISSION.

Times	1	2	3	4
Receiving (μAh)	123.22	252.40	386.106	517.97
Times	5	6	7	
Receiving (μAh)	655.82	791.178	911.196	

It is also necessary to take communication distance between relays and UEs into consideration, for it varies in a large range in practice. We set three typical different communication distances in experiments to reveal variation tendency of the energy consumption. Fig. 13 exposes that with the communication distance increasing, Wi-Fi Direct technique consumes more energy in message transmission apparently. Although the UE still saves energy in message transmissions when the communication distance is 10 meters, according to the increasing tendency, we can predict that the UE will consume more energy in message transmissions than cellular transmission when the communication distance between the two involved devices beyond a certain value. Table V presents the relationship between distance and RSSI in our measurement. The RSSI is negatively correlated with distance. In eDirect, we try to match the UE with a relay with the largest RSSI to lower energy consumption in D2D communication.

TABLE V
THE RELATIONSHIP BETWEEN DISTANCE AND RSSI.

Distance (m)	0	1	5	10
RSSI (dBm)	-9	-33	-42	-68

We further evaluate the impact of peer matching strategy on energy saving. We consider ten devices in total among which two devices are randomly selected as relays. For each relay, we configure it as being able to connect at most 5 UEs. We perform evaluations in three scenarios denoted as A, B and C with distance between any two devices ranging from 0 to 5 meters, 0 to 10 meters, and 5 to 10 meters respectively. We compared the energy consumption of three strategies and present the results in Fig. 15. The Greedy strategy chooses the relay with the largest RSSI. The Random strategy randomly chooses a relay for each UE. The Optimal strategy calculates the optimal solution with a brutal-force search, assuming the global information can be acquired. It can be observed that the greedy strategy that eDirect adopts can achieve similar energy consumption compared to the optimal strategy. In addition, the difference among the performance of different peer match strategies becomes smaller as the distance between devices grows.

Because the heartbeat messages are always with different sizes in practice, we evaluate the prototype system with different-sized message transmissions. According to the range of common heartbeat sizes, we set 54 Bytes as the standard of message size, and gradually scale it up until around 300 Bytes. Fig. 16 shows that the energy consumption stays almost constant, which applies to small-sized messages only. The conclusion might be invalid if the size of message largely exceeds the standard, however, such cases conflict with the fact that heartbeat messages are small in size. Thus, the size of heartbeat messages has little impact on the energy consumption in the message transmission.

B. Performance in Signaling Reduction

Setup: To measure the cellular signaling traffic in heartbeat transmissions, we use NetOptiMaster to capture the layer 3 messages of the smartphones with WCDMA network connection. The layer 3 messages of signaling reveal the interactions

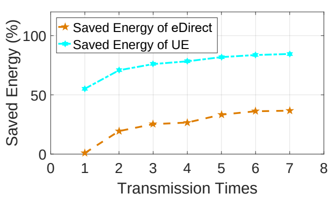


Fig. 10. Saved Energy of eDirect and UE.

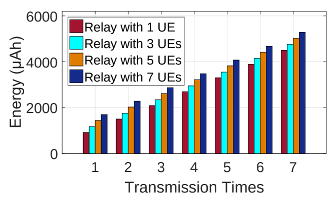


Fig. 11. Energy Consumption of a Relay with Multiple UEs.

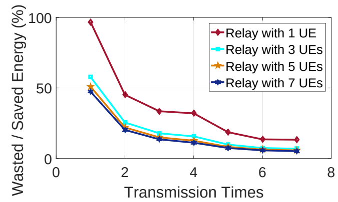


Fig. 12. Ratio of the Wasted Energy of the Relay to the Saved Energy of the UE.

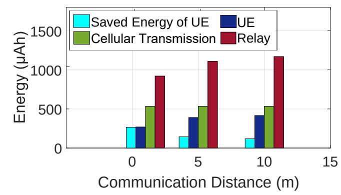


Fig. 13. Energy Consumption in Different Communication Distances.

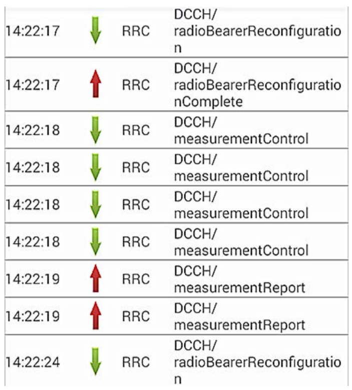


Fig. 14. Part of the Captured Signaling Traffic.

between the cellular network and smartphones. Fig. 14 shows a part of the captured cellular signaling traffic. Through contrasting the captured cellular signaling traffic, we try to figure out the differences between the D2D approach and cellular network approach in the heartbeat transmission.

Similar to the energy consumption evaluation, we conducted the measurement in different D2D connection durations via the proposed D2D approach and traditional cellular approach respectively. In practice, when the connected UE sends more heartbeat messages via the D2D connection, it leads to less RRC connections between the BS and the UE. Besides, we evaluate the signaling consumption in the case that a relay connects multiple UEs.

Results: Fig. 17 shows the cellular signaling traffic for the UE, relay with multiple connected UEs, and cellular transmission. Based on the results, the UE brings in no extra cellular signaling traffic via the D2D transmission because of its direct communication between smartphones. As for the relay and cellular transmission, the cellular signaling traffic of the relay is nearly the same as that of cellular transmission. With the number of the connected UE increasing, the cellular signaling traffic of the relay nearly stays constant. Thus, our D2D relaying framework can reduce more than 50% cellular signaling traffic in the heartbeat transmission through D2D forwarding and its performance will improve if more UEs are connected with a relay.

With the transmission time increasing, we can observe that the relay with more UEs connected incurs little extra cellular signaling traffic. The reason of this phenomenon might be that more data in once transmission incurs more cellular signaling traffic. When the size of the transmitted messages exceeds the threshold, the cellular signaling traffic would slightly increase correspondingly. In addition, we can conclude from this figure that the increased cellular signaling traffic is minor for the total cellular signaling traffic. Thus, it proves that our scheme, through D2D forwarding and sending heartbeat messages in an aggregated fashion, is effective in reducing the cellular signaling traffic.

TABLE VI
THE PERCENTAGES OF DIFFERENT TRAFFIC IN FREQUENCY AND DATA USAGE.

	Flash Traffic	Durable Traffic
Frequency	73%	27%
Data Usage	59%	41%

C. Interference from Wi-Fi Direct

Wi-Fi Direct, the applied D2D technique in the proposed framework, shares the 2.4 GHz band with commercial APs. Wi-Fi Direct always utilizes Channel 1, 6 and 11 in 2.4 GHz as the listen channels, which are the same as that of commercial APs. Based on this fact, it is possible for the wireless co-channel interference between the devices and commercial APs to occur. The co-channel interference might degrade the performance of AP. To discover the impacts of the interference, we evaluate the performance of AP with its peak throughput in practice at different bandwidth settings and communication distances respectively.

Before the interference evaluation, we classify the traffic of the AP based on the traffic feature. The AP in the experiments is deployed in a laboratory, with over 10 devices connected daily. We collected its traffic data in three weeks. Based on the traffic feature, we divide the traffic into two classes, flash traffic and durable traffic. Flash traffic, which always occurs in web browsing, instant messenger and e-mail, represents the intermittent and short network requests. Durable traffic, e.g., downloading, indicates the continuous and long network connections. Their frequency and data usage distribution are presented in Table VI.

According to Table VI, the frequency of the durable traffic is much lower than that of the flash traffic. However, the durable traffic lasts considerably longer than the flash traffic. The flash traffic features a short duration, always finished in seconds, and a small amount of data usage. These features make the possibility of the interference lower for the heartbeat transmission is also short in time over the D2D forwarding. Thus, we focus on the durable traffic in the following interference evaluations.

1) Co-channel interference in different bandwidth settings: In this part, we evaluate the co-channel interference between the AP deployed in the laboratory and two smartphones, Samsung Galaxy S4 and Moto XT1040, equipped with Wi-Fi Direct. The communication channels of the AP and Wi-Fi Direct are set as Channel 1. And, there are no other detectable APs in Channel 1. In this circumstance, we reach the peak throughput via a PC connected with the AP and set

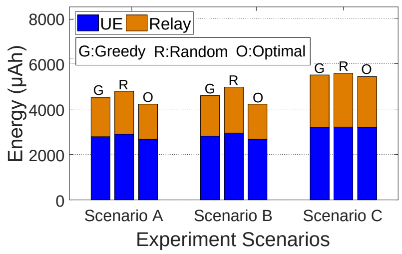


Fig. 15. Energy Consumption in Different Scenarios.

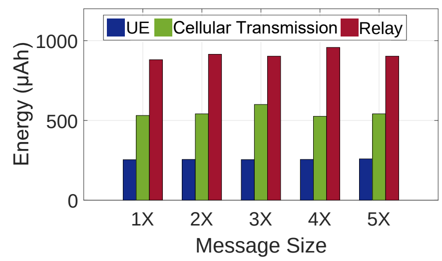


Fig. 16. Energy Consumption in Different Message Sizes.

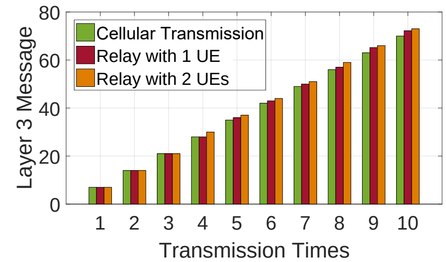


Fig. 17. Layer 3 Message Consumption.

TABLE VII
PROBABILITY OF CO-CHANNEL INTERFERENCE.

Bandwidth	5 Mbps	10 Mbps	15 Mbps	25 Mbps
Probability	4%	12%	28%	33%

TABLE VIII
PROBABILITY OF CO-CHANNEL INTERFERENCE.

RSSI	-75 dBm	-50 dBm	-25 dBm
Probability	0%	4%	43%

the bandwidth limitation as 25 Mbps (the maximal available bandwidth) and 5 Mbps (the common available bandwidth) sequentially. At the same time, the two smartphones run our prototype system to send heartbeat messages once per minute through Wi-Fi Direct forwarding with the distance of -50 dBm RSSI. To analyze the impacts caused by the co-channel interference, we utilize Wireshark and Iperf to capture the instant traffic during the experiments. Through repeating the experiments at the same time in three weeks, we select a part of the captured traffic to reveal the co-channel interference in Fig. 18 and compile the probability statistics of this interference in Table VII.

Because of the unstable wireless condition, the throughput fluctuation in a small range is common and normal. To find out the impacts caused by the interference, we record the amplitude when the Wi-Fi Direct communication occurs and its adjacent average fluctuation. When the co-channel interference occurs, the instant throughput degrades, while the gap between the two values is obvious.

As shown in Fig. 19, under the condition of 25 Mbps bandwidth, the fluctuation in throughput is small. However, the co-channel interference is obvious and leads to the severe throughput degradation. On the contrary, the original fluctuation in the setting of 5 Mbps bandwidth makes the interference inconspicuous. To discover the relationship between the probability of the interference and bandwidth, we conclude the results from the collected data, presented in Table VII. With the public AP connected with multiple devices, the unstable wireless bandwidth is hard to reach such large value in daily use. For the common available bandwidth in practice, the probability of the co-channel interference is acceptable.

2) Co-channel interference in different distances: The distance between the AP and smartphones makes the RSSI of Wi-Fi different, which also impacts the co-channel interference significantly. In this experiment, we set two smartphones with the distance of -50 dBm and -25 dBm respectively. The other settings stay as before.

TABLE IX
IMPACTS ON COMMON FUNCTIONS.

Functions	Calling	Messaging	Gaming
Work Normally?	√	√	√
Connections	Cellular	WLAN	Bluetooth
Work Normally?	√	×	√

Fig. 20 and Fig. 21 depict the difference of the throughput fluctuation under the condition of different Wi-Fi RSSI. It is obvious that the amplitude of throughput fluctuation and the frequency of the co-channel interference soar in -25 dBm RSSI, compared with that in -50 dBm RSSI. The difference of the fluctuation caused by the Wi-Fi Direct interference reaches around 80%, which indicates a serious throughput degradation. We also count the frequency of the co-channel interference and present its probability in Table VIII. Based on the table, the co-channel interference impacts the normal work of the AP apparently in the case of -25 dBm RSSI. However, -25 dBm RSSI indicates the distance between the AP and smartphones is less than 1 meter, which is rare in daily scenarios.

In conclusion, the co-channel interference caused by the D2D forwarding of heartbeat messages can be neglected in most daily cases.

D. Impacts on Common Functions of Smartphones

In this subsection, we test whether the common functions of smartphone work normally while the framework is forwarding messages. We conducted several tests on Samsung Galaxy S4, equipped with Android 4.4, with LTE-TDD, a widely used 4G telecommunication technology. The results presented in Table. IX show that only the Wi-Fi connection will be interrupted when forwarding messages. This is because the Wi-Fi interface does not support Wi-Fi Direct and Wi-Fi connection to the AP to simultaneously exist.

VIII. RELATED WORK

Many researchers have already studied the signaling storm problem because of its considerable impacts on mobile cellular networks, with the rapid development of smartphone network applications. [3] gives a detailed analysis about the origins of signaling storm and its negative influences on cellular networks. [2] presents a reason of frequent interactions between the BS and smartphones through analyzing the impacts of periodic messages on the mobile wireless network. Based on these studies, the conclusion shows heartbeat transmissions are the major reason of triggering signaling storm. There are also many solutions proposed to solve this problem. In the

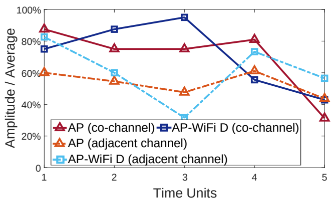


Fig. 18. Fluctuation in Different Channel Settings.

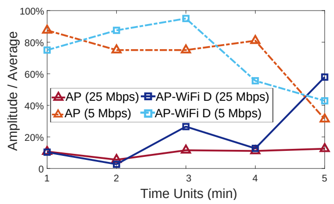


Fig. 19. Fluctuation in Different Bandwidth Settings.

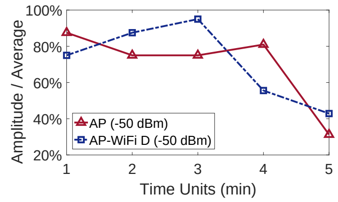


Fig. 20. Fluctuation in -50 dBm RSSI.

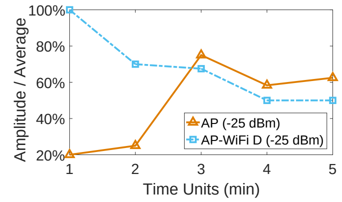


Fig. 21. Fluctuation in -25 dBm RSSI.

academia, most solutions are to modify the RRC mechanism of the cellular networks. For example, [23] proposes a scheme to eliminate tail time when releasing radio resource to optimize the radio resource allocation. [24] reduces the interactions between the BS and smartphones through extending the RRC connected state. In the industry, the push notification server, such as Apple Push Server, is designed to reduce the number of servers connected with UEs. It achieves less heartbeat messages to keep the connection between smartphones and remote servers.

Both the academia and industry are concerned by the short battery life of smartphones [25]. Many studies focus on reducing the energy consumption for smartphones. [26] discovers the unnecessary energy consumption caused by foreground network traffic persisting after switching from foreground to the background. [27] studies the tail energy in data transmissions, which presents an evidence that heartbeat transmissions waste a lot of energy. [28] employs fast dormancy to save energy with higher signaling overhead, which aggravates the signaling storm problem while reducing the energy consumption. Many works, such as [1] and [29], schedule the delay-tolerant data to achieve energy saving, which is also applied in this work for energy saving and lower signaling overhead. To satisfy the performance requirements while saving energy, [30] proposes a metric to capture the performance cost due to energy-saving approaches and scheduling them online.

D2D communication, as an important component of the next generation cellular networks, is popular in academic researches. [31] and [32] apply the D2D communication in content sharing scenarios, and design mechanisms to reduce cellular data traffic. [33] and [34] combine D2D techniques with the cellular networks to reduce cellular data traffic. To the best of our knowledge, there is no work utilizing D2D techniques to achieve reducing the signaling traffic of BSs and the energy consumption of smartphones simultaneously.

This work significantly extends the preliminary work [35]. We take the impact of delaying heartbeat messages into consideration and propose a deterministic online algorithm for delay-sensitive applications, which extends application scenarios of the scheduling mechanism in Message Scheduler. The new proposed algorithm, with competitive ratio of 2, is robust even without any information about the future incoming heartbeat messages, which is suitable for mobile devices.

IX. Conclusion and Future work

In this paper, we presented an energy-efficient D2D-assisted relaying framework, eDirect, to deal with the side-effects brought by small-size and numerous heartbeat messages in smartphone apps, i.e., serious cellular signaling overload for BSs and huge energy consumption for smartphones. In eDirect, we utilize voluntary smartphones, i.e., relays, to collect heartbeat messages from nearby smartphones, i.e., UEs, via the D2D approach, and aggregating the heartbeat messages to transmit over one single cellular channel, saving not only energy consumption for smartphones but also cellular signaling traffic for BSs. We implemented a prototype system with Wi-Fi Direct technique. The real-world experiments with our prototype implementation demonstrate the feasibility and effectiveness of eDirect. In the worst situation, our framework can still reduce about 50% of the cellular signaling traffic and save up to 36% of the energy for relays and UEs in heartbeat transmissions. In terms of UEs, it can achieve up to 55% energy saving.

We believe that the more advanced D2D technique, embedded in the next generation of cellular networks would further improve the convenience of eDirect. With this technique, eDirect can get rid of the operating system requirements and eliminate the impacts on Wi-Fi connecting when forwarding messages.

References

[1] T. Zhang, X. Zhang, F. Liu, H. Leng, Q. Yu, and G. Liang, "etrain: Making wasted energy useful by utilizing heartbeats for mobile data transmissions," in *Proc. IEEE ICDCS*, Columbus, Ohio, USA, Jun. 29 - Jul. 2, 2015.

[2] F. Qian, Z. Wang, Y. Gao, J. Huang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Periodic transfers in mobile applications: network-wide origin, impact, and optimization," in *Proc. ACM WWW*, Lyon, France, Apr. 16 - 20, 2012.

[3] Y. Choi, C.-h. Yoon, Y.-s. Kim, S. W. Heo, and J. A. Silvester, "The impact of application signaling traffic on public land mobile networks," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 166-172, Jan. 2014.

[4] F. Liu, P. Shu, and J. C. Lui, "Appatp: An energy conserving adaptive mobile-cloud transmission protocol," *IEEE Transactions on Computers*, vol. 64, no. 11, pp. 3051-3063, Nov. 2015.

[5] S. Rosen, H. Luo, Q. A. Chen, Z. M. Mao, J. Hui, A. Drake, and K. Lau, "Discovering fine-grained rrc state dynamics and performance impacts in cellular networks," in *Proc. ACM MobiCom*, Maui, Hawaii, USA, Sept. 7 - 11, 2014.

[6] "XMPP protocol," Internet: xmpp.org/.

[7] "MQTT protocol," Internet: mqtt.org/.

[8] "Signaling," Internet: en.wikipedia.org/wiki/Signal.

[9] T. Stoner, X. Wei, J. Knight, and L. Guo, "Experience: Rethinking rrc state machine optimization in light of recent advancements," in *Proc. ACM MobiCom*, Paris, France, Sept. 7 - 9, 2015.

[10] H. Zhang, K. Xue, and P. Hong, "Facing the signaling storm: A method with stochastic concept," in *Proc. IEEE WCSP*, Hefei, China, Oct. 23 - 25, 2014.

[11] M. Sheng, Y. Li, X. Wang, J. Li, and Y. Shi, "Energy efficiency and delay tradeoff in device-to-device communications underlaying cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 92-106, Jan. 2016.

[12] L. Pu, X. Chen, J. Xu, and X. Fu, "D2d fogging: An energy-efficient and incentive-aware task offloading framework via network-assisted d2d collaboration," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3887-3901, Dec. 2016.

[13] N. Mastronarde, V. Patel, J. Xu, and M. Van, der Schaar, "Learning relaying strategies in cellular d2d networks with token-based incentives," in *Proc. IEEE GLOBECOM Workshops*, Atlanta, GA, USA, Dec. 9 - 13, 2013.

[14] "Karma Go," Internet: yourkarma.com/how-it-works.

[15] I. Manotas, C. Bird, R. Zhang, D. Shepherd, C. Jaspan, C. Sadowski, L. Pollock, and J. Clause, "An empirical study of practitioners' perspectives on green software engineering," in *Proc. ACM ICSE*, Austin, Texas, USA, May 14 - 22, 2016.

[16] D. R. Dooly, S. A. Goldman, and S. D. Scott, "Tcp dynamic acknowledgment delay: theory and practice," in *Proc. ACM STOC*, Dallas, Texas, USA, May 24 - 26, 1998.

[17] J. Nagle, "Congestion control in ip/tcp internetworks," *Acm Sigcomm Computer Communication Review*, vol. 25, no. 1, pp. 61-65, Jan. 1995.

[18] S. S. Seiden, "A guessing game and randomized online algorithms," in *Proc. ACM STOC*, Portland, OR, USA, May 21 - 23, 2000.

[19] A. R. Karlin, C. Kenyon, and D. Randall, "Dynamic tcp acknowledgement and other stories about e/(e-1)," in *Proc. ACM STOC*, Crete, Greece, Jul. 4 - 6, 2001.

[20] "LTE Direct," Internet: www.qualcomm.com/invention/technologies/lte/direct.

[21] "Android developers," Internet: developer.android.com/training/connect-devices-wirelessly/wifi-direct.html.

[22] "Monsoon Solutions Inc. Power Monitor," Internet: www.msoon.com/LabEquipment/PowerMonitor/.

[23] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "Top: Tail optimization protocol for cellular radio resource allocation," in *Proc. IEEE ICNP*, Kyoto, Japan, Oct. 5 - 8, 2010.

[24] J. Wu, S. Zhou, Z. Niu, C. Liu, P. Yang, and G. Miao, "Traffic-aware data and signaling resource management for green cellular networks," in *Proc. IEEE ICC*, Sydney, Australia, Jun. 10 - 14, 2014.

[25] F. Liu, P. Shu, H. Jin, L. Ding, J. Yu, D. Niu, and B. Li, "Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 1–0, 2013.

[26] S. Rosen, A. Nikravesh, Y. Guo, Z. M. Mao, F. Qian, and S. Sen, "Revisiting network energy efficiency of mobile apps: Performance in the wild," in *Proceedings of the 2015 ACM Internet Measurement Conference, IMC 2015, Tokyo, Japan, October 28-30, 2015*.

[27] P. Shu, F. Liu, H. Jin, M. Chen, F. Wen, Y. Qu, and B. Li, "etime: energy-efficient transmission between cloud and mobile devices," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 14 - 19, 2013.

[28] P. K. Athivarapu, R. Bhagwan, S. Guha, V. Navda, R. Ramjee, D. Arora, V. N. Padmanabhan, and G. Varghese, "Radiojockey: mining program execution to optimize cellular radio usage," in *Proc. ACM MobiCom*, Istanbul, Turkey, Aug. 22 - 26, 2012.

[29] N. D. Lane, Y. Chon, L. Zhou, Y. Zhang, F. Li, D. Kim, G. Ding, F. Zhao, and H. Cha, "Piggyback crowdsensing (pcs): energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities," in *Proc. ACM SenSys*, Rome, Italy, Nov. 11 - 14, 2013.

[30] Y. Cui, S. Xiao, X. Wang, M. Li, H. Wang, and Z. Lai, "Performance-aware energy optimization on mobile devices in cellular network," in *Proc. IEEE INFOCOM*, Toronto, Canada, Apr. 27 - May 2, 2014.

[31] J. Jiang, S. Zhang, B. Li, and B. Li, "Maximized cellular traffic offloading via device-to-device content sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 82-91, Jan. 2016.

[32] V. Sciancalepore, D. Giustiniano, A. Banchs, and A. Hossmann-Picu, "Offloading cellular traffic through opportunistic communications: Analysis and optimization," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 122-137, Jan. 2016.

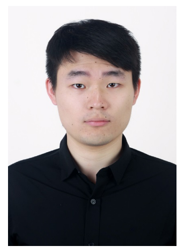
[33] T. Bansal, K. Sundaresan, S. Rangarajan, and P. Sinha, "R2d2: Embracing device-to-device communication in next generation cellular networks," in *Proc. IEEE INFOCOM*, Toronto, Canada, Apr. 27 - May 2, 2014.

[34] L. Deng, Y. Zhang, M. Chen, Z. Li, J. Y. Lee, Y. J. Zhang, and L. Song, "Device-to-device load balancing for cellular networks," in *Proc. IEEE MASS*, Dallas, TX, USA, Oct. 19 - 22, 2015.

[35] Y. Jin, F. Liu, X. Yi, and M. Chen, "Reducing cellular signaling traffic for heartbeat messages via energy-efficient D2D forwarding," in *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, GA, USA, June 5-8, 2017*, 2017, pp. 1301–1311.

Xiaomeng Yi received his Ph.D. degrees in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2017. His current research interests focus on modeling and performance optimization of GPU accelerated data processing systems.

Li Pan received his B.Eng. degree in the School of Computer Science and Technology, Huazhong University of Science and Technology, China. He is currently a M.Eng. student in the School of Computer Science and Technology, Huazhong University of Science and Technology. His current research interests focus on edge computing and mobile computing.



Yanqi Jin received his B.Eng. degree in School of Automation, Huazhong University of Science and Technology, China. He is currently a M.Eng. student in School of Computer Science and Technology, Huazhong University of Science and Technology. His research interests include mobile computing and green computing.

Fangming Liu (S'08, M'11, SM'16) received the B.Eng. degree from the Tsinghua University, Beijing, and the Ph.D. degree from the Hong Kong University of Science and Technology, Hong Kong. He is currently a Full Professor with the Huazhong University of Science and Technology, Wuhan, China. His research interests include cloud computing and edge computing, datacenter and green computing, SDN/NFV/5G and applied ML/AI. He received the National Natural Science Fund (NSFC) for Excellent Young Scholars, and the National Program Special Support for Top-Notch Young Professionals. He is a recipient of the Best Paper Award of IEEE/ACM IWQoS 2019, ACM e-Energy 2018 and IEEE GLOBECOM 2011, as well as the First Class Prize of Natural Science of Ministry of Education in China.

Minghua Chen (S04 M06 SM 13) received his B.Eng. and M.S. degrees from the Department of Electronic Engineering at Tsinghua University in 1999 and 2001, respectively. He received his Ph.D. degree from the Department of Electrical Engineering and Computer Sciences at University of California at Berkeley in 2006. He spent one year visiting Microsoft Research Redmond as a Postdoc Researcher. He joined the Department of Information Engineering, the Chinese University of Hong Kong, in 2007, where he is now an Associate Professor. He is also currently an Adjunct Associate Professor in Tsinghua University, Institute of Interdisciplinary Information Sciences. He received the Eli Jury award from UC Berkeley in 2007 (presented to a graduate student or recent alumnus for outstanding achievement in the area of Systems, Communications, Control, or Signal Processing) and The Chinese University of Hong Kong Young Researcher Award in 2013. He also received several best paper awards, including the IEEE ICME Best Paper Award in 2009, the IEEE Transactions on Multimedia Prize Paper Award in 2009, and the ACM Multimedia Best Paper Award in 2012. He is currently serving on the Steering Committee of ACM e-Energy. He serves as an Associate Editor of IEEE/ACM Transactions on Networking in 2014 - 2018. He serves as TPC Co-Chair of ACM e-Energy 2016 and General Chair of ACM e-Energy 2017. He receives the ACM Recognition of Service Award in 2017 for service contribution to the research community. His recent research interests include energy systems (e.g., smart power grids and energy-efficient data centers), intelligent transportation systems, distributed optimization, multimedia networking, wireless networking, delay-constrained network coding, and characterizing the benefit of data-driven prediction in algorithm/system design.