Reducing Cellular Signaling Traffic for Heartbeat Messages via Energy-Efficient D2D Forwarding

Yanqi Jin[1] Fangming Liu*[1] Xiaomeng Yi[1] Minghua Chen[2]
[1]Key Laboratory of Services Computing Technology and System, Ministry of Education,
School of Computer Science and Technology, Huazhong University of Science and Technology
[2]The Chinese University of Hong Kong

Abstract*—Mobile Instant Messaging (IM) apps, such as WhatsApp and WeChat, frequently send heartbeat messages to remote servers to maintain always-online status. Periodic heartbeat messages are small in size, but their transmissions incur heavy signaling traffic to frequently establish and release communication channels between base stations and smartphones, known as signaling storm. Meanwhile, smartphones also need to activate cellular data communication module frequently for transmitting short heartbeat messages, resulting in substantial energy consumption. To address these issues, we propose a Device-to-Device (D2D) based heartbeat relaying framework, in order to reduce signaling traffic and energy consumption in heartbeat transmission. The framework selects the 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 periods and the expiration time of the collected heartbeat messages, the framework schedules the transmissions of collected heartbeat messages to minimize signaling and energy consumption while satisfying time constrains. We implement and evaluate our solution on Android smartphones. The results from real-world experiments show that our solution achieves more than 50% signaling traffic reduction and up to 36% energy saving.

I. Introduction

Mobile Instant Messaging (IM) apps (e.g., WeChat, Line, WhatsApp) are dominating in smartphones. These IM apps periodically send short signaling messages, called heartbeat messages, to remote servers to maintain always-online status. Heartbeats are small in size but large in quantity. In Table. I, we illustrate the proportion of heartbeat messages in total number of transmitted messages in several popular apps, by summarizing the analysis in [1] and [2]. We can observe that nearly half of the messages are heartbeat messages.

TABLE I
Proportion of Heartbeats 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 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 establishment of Radio Resource Control (RRC) connections before data transmission. The establishment and release of RRC connections cause a lot of cellular signaling traffic. Frequent change of RRC state caused by heartbeat transmissions consumes a large amount of resource in control channel. 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 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 a 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 decrease the frequency of heartbeat transmissions. Some strategies, such as extending the period of the 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 decrease RRC transitions. These solutions might cause extra energy consumption and are hard to deploy [5]. It is also conceivable to use Device-to-Device (D2D) techniques to collect heartbeat messages from nearby devices and send them in aggregate fashion to the BS so as to reduce cellular signaling traffic. Besides, D2D communication, such as those using Bluetooth and Wi-Fi Direct, are more energy-efficient than cellular communication.

Based on the above analysis, in this paper, we propose a D2D based framework for heartbeat transmissions. D2D communication, a promising paradigm in next generation cellular technologies, refers to direct communication between two mobile clients. We set two roles, relay and User Equipment (UE), for smartphones. Mobile operators could select relays among the participating smartphone users to collect the heartbeat messages from nearby UE(s) via D2D communication and

*This work was supported in part by NSFC under Grant 61520106005, in part by the National 863 Hi-Tech Research and Development Program under grant 2015AA01A203, and in part by the National 973 Basic Research Program under Grant 2014CB347800. (Corresponding author: Fangming Liu)

schedule these messages to transmit them at once to BS. It is a win-win scheme for mobile network operator, relay, and UE. For mobile network operator, it can reduce the signaling traffic in cellular control channels at a low cost. For relays, they could get payback from mobile operator as they help the system reduce signaling traffic using its own connectivity and energy resource. For UEs, they can reduce the energy consumption in heartbeat transmissions since the D2D approach based on Bluetooth or Wi-Fi Direct is usually more energy-efficient in transmitting short heartbeat messages than the traditional cellular approach.

Although the idea of forwarding heartbeat messages via D2D approach reads simple, there are two key challenges we need to address. First, because of the inherit mobility of smartphones, relay has to deal with the unpredicted heartbeat messages from UE(s). Without proper scheduling, the relay would suffer from unnecessary excessive energy consumption. Second, improper D2D pairs might cause more energy consumption 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 minimize the delay raised by forwarding and reduce the energy consumption. As for the second issue, we design a mechanism for UEs to determine when to use relay to forward heartbeat messages and when to send the message directly via cellular network to optimize energy consumption.

We design a prototype, implement it on Android platform, and evaluate its performance in reducing cellular signaling traffic for mobile operators and saving energy for smartphones. The prototype consists of three main components: D2D Detector, Message Monitor and Message Scheduler. It uses Wi-Fi Direct as the D2D communication mechanism to collect heartbeat messages. Then, the relay determines a specific timing to send these collected heartbeat messages to achieve minimum energy consumption and maximum cellular signaling traffic reduction.

The remainder of this paper is organized as follows. In Section II, we introduce the background of the signaling storm problem and D2D technologies, 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 based framework. In Section IV, we present the implementation of the prototype. Section V evaluates the performance of the D2D based framework. Section VI reviews related work and Section VII concludes the paper.

II. Background and Motivation

A. Heartbeat Message

Heartbeat messages are used to support real-time communication or push notification services for IM apps. IM servers set expiration timers to determine a client is online or not [3]. In order to maintain 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. 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 are based on XMPP protocol [6]; and Facebook Messenger uses MQTT protocol [7]. There are also many IM apps choosing to implement the heartbeat mechanism using proprietary solutions.

B. Signaling Storm

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

The mobile network operators get profit based on the users' cellular data traffic, not the signaling traffic. Moreover, the massive signaling traffic greatly deteriorates user experience on cellular network, such as higher rate of paging failure. Overall, the mobile network operator has great motivation to reduce signaling traffic at low cost.

C. D2D Technique

D2D techniques, which support direct communication without BS intervention, are developing rapidly. Currently, popular D2D techniques, including Bluetooth, Wi-Fi Direct, NFC, ZigBee, are implemented in common apps, like PayPal, WeChat, etc. 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 under research. These techniques achieve good performance in neighbor discovery and energy saving. For example, LTE Direct enables the discovery of thousands of devices in proximity of approximately 500 meters. Inspired by the advantages of D2D communication mentioned above and its promising development in the future, we develop the D2D based scheme for solving the signaling storm and energy consumption issues studied in this paper.

D. D2D Communication for Remitting Signaling Storm

D2D communication 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 D2D approach instead of cellular network. In this way, the heartbeat transmissions will not bring any traffic to cellular control channels. The number of RRC connection decreases because the collected heartbeat messages from UEs are transmitted once by the relay in an aggregate fashion. Thus, if the relay mechanism is applied widely, the signaling traffic on the cellular channel will decrease significantly.

A previous work [4] reveals that cellular communication incurs higher power consumption than the common D2D communication over Bluetooth and Wi-Fi. Therefore, D2D techniques can potentially improve energy efficiency when using D2D for heartbeat transmission. 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 communication is also perfectly suitable for the proposed scenario. The signaling storm problem usually occurs in the region with high-density crowd. In D2D communication, more peers and more opportunities of a high-quality D2D communication could be found in such high-density crowd scenario.

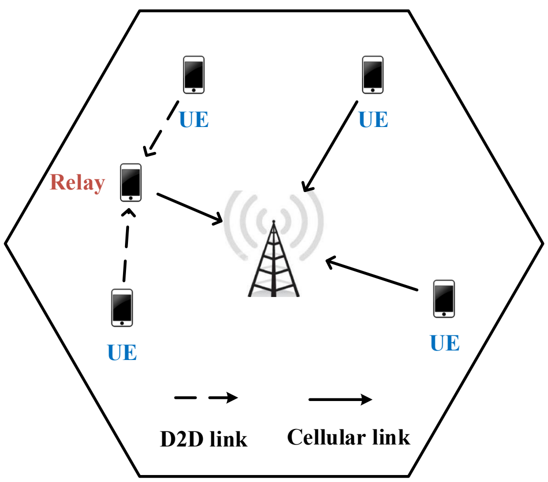

Fig. 1. An Illustrative Example of Forwarding Messages via Relay

III. System Design

Our system aims to address heartbeat-induced signaling storm in cellular control channel, while tries to reduce the energy consumption in heartbeat transmissions. To do so, we need to decrease the frequency of heartbeat transmissions in smartphones. The lower frequent heartbeat transmissions lead to not only less signaling traffic, but also less energy consumption for smartphones. To achieve less heartbeat transmissions, it is essential to reduce the frequency of heartbeat messages in app design. However, the reduction will impact the instantaneity of these IM apps. For app developers standing, they are unwilling to solve the problem for mobile network operators at the expense of worsening user experience. In addition, with the rapid growth of heartbeat messages, it is hard to address the problem of excessive cellular signaling traffic simply with such elementary strategies. There are also many other schemes to lower the frequency of heartbeat transmissions. Some of these change RRC mechanisms, which vary in different cellular networks, to reduce cellular signaling traffic. Consequently, these schemes are usually hard to be compatible with all the cellular networks. Moreover, they would be dropped with the development of cellular networks. Some of these bring in more energy consumption. The flaw might keep these schemes away from smartphone users for limited battery of smartphones.

A. Design Insights

Is there a scheme to reduce the frequency of heartbeat transmissions while avoiding the flaws mentioned above? Initially, we consider the feature of independence to cellular network. D2D techniques are taken into our consideration for its unique feature, supporting direct communication without BS intervention. Our basic idea is to use some specific smartphones to collect the heartbeat messages from nearby smartphones and send them in an aggregate fashion. In order to distinguish senders and receivers, we assign two roles, relay and UE. As illustrated in Fig. 1, when a UE intends to send heartbeat messages, it could activate the service provided by our framework. Then, the service will discover the relays nearby. Meanwhile, the matching mechanism of this service would match a proper relay or choose to send the heartbeat messages via cellular network directly. When the group of a relay and UE(s) is formed, UE(s) could send the heartbeat messages to the relay. Finally, the message-scheduling algorithm of this service would schedule these heartbeat messages and determine a sending time with least cellular signaling traffic.

Although the feature of direct communication guarantees the effectiveness of cellular signaling traffic reduction, there are still many challenges before the D2D based scheme coming into reality.

Primarily, the D2D based scheme might cause more energy consumption than the traditional cellular approach. Admittedly, D2D techniques are more energy-efficient in data transmission. However, the D2D scheme has the potential of more energy consumption than the original cellular scheme in heartbeat transmission, because the energy consumption in discovering and connecting devices is inevitable for all the D2D techniques. Thus, we need to optimize the energy consumption through minimizing the times of discovering and connecting and maximizing the period of more energy-efficient D2D transmission. We will discuss the design of energy saving strategies in detail in Sec. III. C.

Then, the D2D based scheme might bring in extra security issues. Admittedly, forwarding heartbeat messages might bring in some worries of leaking private information of users, such as users' ID. In fact, our framework nearly does not bring any comprise in security even when the relays are malicious. Firstly, the forwarded data has already been encrypted via the protocols offered by IM apps before it sends to relay. For example, heartbeat messages based on

MQ Telemetry Transport (MQTT) will be equipped with lightweight cryptography, which can be handled with Secure Sockets Layer (SSL). Thus, even if the relay obtains the forwarded messages, it would not get the encrypted data in it. Furthermore, there is another concern about the negative impacts caused by forwarding. For example, the relay might be attacked or some accidents occur on the relay, which results in failure of message transmission. Actually, in the original cellular network, the BS, which receives smartphones' messages and "relays" them to the server, also has similar security concern. Although it may be harder to attack BS as compared to the relaying smartphone in our solution, the fact is that relay actively accept the forwarding and the transmission time is short for heartbeat messages are small in size. Certainly, the security concern cannot be eliminated. The detailed security measures applied in this scenario could be further discussed in the future.

Besides, the data transmission should not suffer from higher failure rate resulting from the D2D based scheme. In the proposed scheme, UEs send their heartbeat messages to the relay nearby instead of the BS. The accidents, which lead to transmission failure, might occur because of exceptions of the relay(s) involved. For example, the relay has ran out of its battery or lost connection to cellular network before all the collected heartbeat messages are sent to BS. Besides, during the data transmission via D2D approach, the physical distance between involved smartphones might exceed the maximum communication distance of the chosen D2D technique while smartphones movement. To avoid these accidents, we design a feedback mechanism for UEs. Once the matched relay transmitting the collected heartbeat messages successfully, the proposed framework will notify the connected UE through feedback information. In case that the UE does not receive the feedback information after a certain interval, it will send the heartbeat messages via cellular network.

Except for the independence feature, the scheme also needs to be easy to deploy for mobile network operator. D2D techniques, as a key paradigm in the next generation of cellular networks, avoid extra deployment for mobile network operator. Compared with upgrading the existing cellular networks, the D2D based scheme inherently has the potential to be more economical and easier in deployment.

However, in our scheme, relays might suffer excessive cellular traffic and energy consumption resulting from more data communication. **What are the incentives to be a relay for mobile users?** The incentives of relay in D2D offloading has been studied in many previous researches, such as [12], [13]. The strategy, using micro-payment scheme, is reasonable and acceptable in practice [14]. Facing the incentive problem in this scenario, we offer a feasible solution. Considering a similar scenario, Uber, a famous transportation app around the world, provides a service, named uberPOOL [15]. The service matches the riders heading in the same direction to saving money for all the riders involved. For the reason that relays contribute a lot in cellular signaling traffic reduction, mentioned in Section II, mobile operators could offer some rewards, such as offering some free cellular data, or reducing the cost for their service. In the real world, there is a commercial product, named Karma Go [16], equipped with the similar incentive mechanism. Karma Go could grab a cellular connection and turn it into a personal Wi-Fi signal, which could be shared with the people nearby. When the people nearby get online through Karma Go, the owner of Karma Go earns rewards, either $1 in credits or 100 MB of free data. Through this easy and effective incentive mechanism, mobile network operator could achieve our scheme at low expense. The discussion about the deployment scheme in detail for mobile network operator is beyond the scope of this paper.

B. System Overview

Let us take an overview to the whole system. Fig. 2 depicts the overall architecture of the D2D based framework for heartbeat messages. It consists of three main components: Message Monitor, D2D Detector and Message Scheduler. Messages Monitor detects the heartbeat transmission, and then obtains their messages and transmission-related data. D2D Detector discovers the peers and establish D2D connections. Completed D2D connection, Message Scheduler runs the scheduling algorithm to schedule the collected messages, which will output a timing to send these messages.

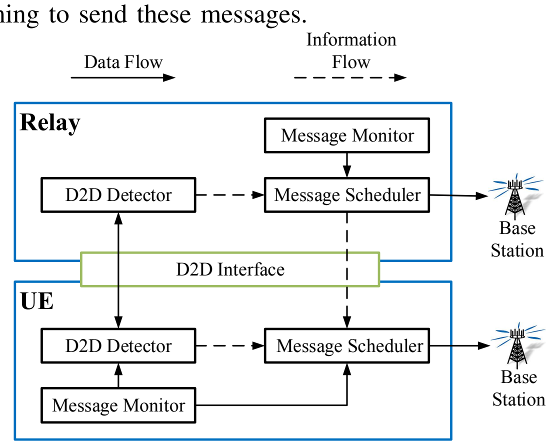

Fig. 2. Software Architecture

C. Energy-Saving Feature

Battery life of smartphones is an explosive issue. While processing speed doubles around every 18 months, battery capacity takes almost a decade to improve to the same degree. That gap is starting to cause problems. Equipped with faster CPU, larger screen and thinner design, smartphone users have to face the severe energy shortage. Thus, we must avoid excessive energy consumption and try to save some energy further. In order to achieve this goal, we take energy-saving consideration into the system design in several aspects.

In the D2D discovery phase, we attempt to make a pre-judgment before establishing D2D connection, which aims to reduce the chances of short-duration D2D connection. The short-duration D2D connection causes an energy-inefficient situation that the energy spent in D2D discovery and connection is accounted excessive proportion of the entire D2D communication. Besides, the transmission through D2D approach is more energy-efficient than that through traditional cellular approach. Thus, we prefer longer-duration D2D connection for more heartbeat transmissions completed during the connection in terms of energy saving. In order to make an effective prejudgment, we set two parameters, i.e., distance between the UE and the relay involved, capacity of the relay. We can obtain the relative distances between the UE and the discovered relays through signal strength in D2D discovery. The relative distance, to some degree, represents the possibility of disconnection between the two devices. Because of the limited communication distance of D2D technique and the inherent mobility of smartphones, disconnection is more likely to occur when the two devices with longer distance. As for capacity of the relay, it refers to the maximum number of collected heartbeat messages, which is set by users. The users, as relays, could adjust the value according their situations in reality, such as their battery usage. In sum, the proposed system tries to match the available relay, with the shortest distance.

In terms of data transmission, we design a scheduling strategy to achieve as less cellular connections as possible. According to the forwarding mechanism discussed before, relays arrange the forwarded messages based on the scheduling strategy. The scheduling strategy determine which messages could be sent together and when these messages should be sent. More forwarded messages are sent simultaneously, more energy and signaling could be saved for cellular connection causes abundant energy consumption and signaling traffic. Without the scheduling strategy, the proposed framework would consume more energy than the original system and lose the signaling-saving feature. Thus, the scheduling strategy, as the core of the proposed framework, directly affect the performance in energy saving and signaling reduction.

TABLE II
NOTATIONS OF THE SCHEDULING ALGORITHM

Notation	Description
T	The period of heartbeat of relay
M	The maximization number of collected heartbeats
k	Number of the forwarded heartbeats from UE(s)
T_k	The expiration time of the No.k forwarded heartbeats
t_k	The arriving time of the No.k heartbeat
t	The sending time of relay's heartbeat

In order to achieve best performance in energy saving and signaling reduction, the scheduling strategy tries to minimize the times of cellular connection. However, excessive delay caused by the proposed framework might make the heartbeat messages expired, leading to transmission failure. Thus, the proposed scheduling strategy aims to find a solution for the tradeoff.

Message Scheduling Algorithm outlines the process of scheduling the collected heartbeat messages for relay. We try to reduce signaling traffic through delaying the heartbeat messages of relay and sending them with the forwarded heartbeat messages from UE(s) in a cellular connection. The algorithm is designed to find the proper sending timing under the condition of unpredicted arriving time of the forwarded heartbeat messages from UE(s).

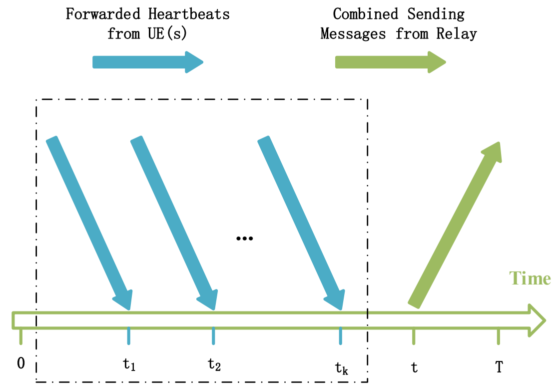

Fig. 3. An Illustrative Example of Scheduling the Collected Messages

There exists a mature algorithm, Nagle's algorithm [17], to solve a similar problem. Nagle's algorithm has been proposed to avoid frequent small data packet leading to congestion in TCP/IP networks. In Nagle's algorithm, the number of small data packets is reduced by combining a number of small outgoing messages, and sending them all at once. As long as there is a sent packet for which the sender has received no acknowledgement, the sender should keep buffering its output until it has a full packet's worth of output, so that output can be sent all at once [18]. Admittedly, there are some differences between the "small packet problem" in TCP/IP networks and this scenario. For example, the buffer size need to vary according to the expiration time of heartbeat messages in this scenario; periods of the relay's heartbeat messages constrain the longest delay. However, we could gain a lot of benefits through this algorithm. Firstly, Nagle's algorithm works in reduction of signaling traffic in this scenario. Through delaying the heartbeat messages of relay and sending them with the forwarded heartbeat messages from UE(s) in a cellular connection, we could effectively cut the number of cellular connection caused by the heartbeat messages from UE(s). In addition, Nagle's algorithm is suitable for the proposed framework. The algorithm is easy to implement and bring in a small amount of computation, which is apposite to smartphones because of their limited battery and computing capacity. Therefore, we modify Nagle's algorithm in some details and apply it in the proposed framework.

To better illustrate the main idea of the scheduling algorithm, we take a heartbeat period of relay as an example shown in Fig. 3. Notations are defined in Table II. We set the longest delayed time as T, although it is usually set as $3T$ for commercial apps, such as WeChat. When the delay exceeds T, the received effective heartbeat messages would be less than the system without modification. To make less effect on the system, we constrain $t \in [0, T]$. Because t_k is unpredicted, it is obvious that the longer time delayed, the more possibility of the most heartbeat messages collected it would be. Besides, number of the collected heartbeat messages should be less than M (we offer a default value based on the experiments and the users could adjust the value according to practice) because of the limited capacity of relay. Furthermore, to guarantee the forwarding success, we also need to constrain the delayed time $t - t_k$ less than the expiration time of heartbeat T_k. Once the heartbeat sent, the relay won't collect forwarded heartbeat messages from UE(s) until the next heartbeat period. The message scheduling algorithm is presented in Algorithm 1.

Algorithm 1 Message Scheduling Algorithm

Input: $M, T, T_1, T_2, ..., T_k$
Initialize $k \leftarrow 0$;
When forwarded heartbeat arrives
Get t_k;
$k++$;
if $k < M$ && $t - t_k < T_k$ && $t < T$ **then**
  pending;
else
  send data now;
end if

D. UI Design

In order to realize the interaction, such as applying for relay and exchanging payback, between relay user and mobile network operator, we design an interface platform, as shown in Fig. 4, for relays and mobile network operators. The smartphone users could apply as a relay through the application. Besides, the users, who is set as relays, are able to activate our framework through the UI, which also provides the information about the amount of collected heartbeat messages and the reward from mobile network operator.

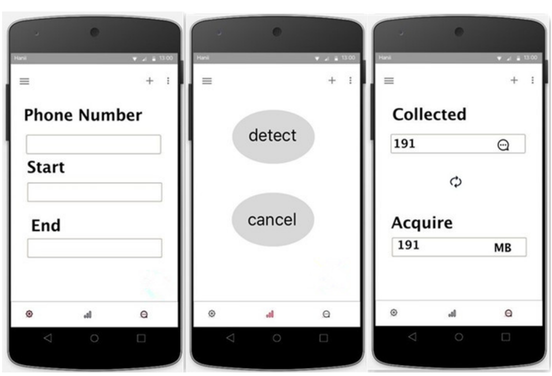

Fig. 4. UI for Relay

IV. Implementation

We now present the design of the prototype that implements the D2D based framework on Android smartphones. The prototype allows users to discover D2D peers via Wi-Fi Direct. Then, the system would select the discovered relay(s) based on the matching information. If a proper relay is found, the UE will build D2D connection with it. The UEs could send its heartbeat messages to the connected relay through Wi-Fi Direct. The prototype also provides a user interface for relays to exchange its reward from mobile network operators based on the messages it collected.

A. The Choice of D2D Techniques

Nowadays, there exists several popular D2D techniques, which have already been applied widely. For example, we could share the music in one smartphone with other smartphones via Bluetooth. In addition, we use Near Field Communication (NFC) to recharge IC card. We choose Wi-Fi Direct among these popular D2D techniques to be implemented in our framework, due to its advantage over other D2D techniques in the scenario studied in this paper. For example, while Bluetooth indeed has the potential to complete D2D communication with low energy, its communication range is typically less than 10 m, too limited to meet our need. Besides, LTE Direct as an innovative D2D technology enabling the discovery of thousands of devices in the proximity of approximately 500 meters [19]. Nonetheless, many countries, such as China, have not deployed the technique mostly. Because of its limitation of deployment, we have to abandon this alternative for generality consideration. Wi-Fi Direct has ideal communication distance and generality, which makes it perfect in this scenario. It is equipped 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, which enables smartphones to connect with others from different manufacturers without any extra hardware deployment.

B. Message Monitor

In Message Monitor module, one major challenge for us is to detect heartbeat messages. Message Monitor is required to obtain the messages to be forwarded and their related parameters, such as their period and expired time. 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 order to achieve Message Monitor, we design a set of APIs for app developers to integrate the proposed D2D based framework into their existing apps.

What is the motivation for app developers to apply the D2D based framework through the proposed APIs? Admittedly, there is a little bit engineering workload while modifying the code of existing apps to complete the integration. However, it is worthy for the modification. Firstly, the framework offers great effect in signaling traffic. Controlling signaling messages are regarded as the responsibility for the developers, according to Global System for Mobile Communications Association (GSMA) guidelines. Furthermore, according to [10], developers lack the necessary information and support infrastructure about green software engineering, although they do care and think about the energy when they build applications. This framework provides an easy and feasible scheme for developers to achieve energy-efficient feature.

C. D2D Discovery and Connection

The D2D Detector aims to discover the suitable peers and determine whether to establish 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, broadcast by Android system. There are four intents, `WIFI_P2P_PEERS_CHANGED_ACTION`, `WIFI_P2P_STATE_CHANGED_ACTION`, `WIFI_P2P_CONNECTION_CHANGED_ACTION`, `WIFI_P2P_THIS_DEVICE_CHANGED_ACTION`, needed to respond [20]. To discover peers that are available to build a D2D connection in range, the class `WifiP2pManager` offers a set of methods. We implemented the `ActionListener` interface offered by `WifiP2pManager`, the system could invoke for successful and unsuccessful discovery. In unsuccessful cases, we used `Toast` to show the failure message to notify users. After fetching the list of available peers, we set `groupOwnerIntend`, which represents the inclination to be a group owner as 15, the maximum value, for relays initially. Moreover, the message scheduling algorithm would reduce `groupOwnerIntend` proportionally until 0 while relay collects heartbeat messages from connected UE(s). For UE(s), we set `groupOwnerIntend` as 0, the minimum value, and select the nearest available relay.

V. Performance Evaluation

To evaluate our D2D based framework, we design a prototype on Android smartphones. We start with characterizing the framework performance in terms of cellular signaling traffic reduction and energy saving, and then, 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, as an equipment, could provide instant current with the constant voltage 3.7V, which offers energy consumption under help of the power tool software in PC [21].

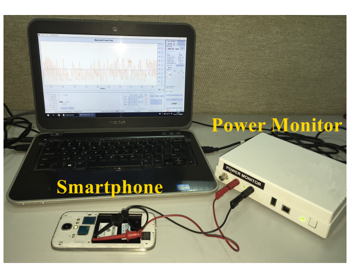

Fig. 5. Experiment Setup for Energy Measurement

Firstly, we attempt to discover the energy consumption in message transmission. We measured the instant current while a 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 D2D transmission, spurts at the moment of the data transmission, and then it descends rapidly. In contrast, the instant current, in cellular transmission, spurts and lasts for a longer period. Thus, as to data transmission, Wi-Fi Direct consumes less energy than cellular network, which is the preliminary evidence for the energy-saving feature of our D2D based framework.

Furthermore, a lot of energy have to be spent in D2D discovery and connection when we apply the D2D approach. How about the energy consumption in D2D discovery and connection, especially when compared with the energy consumption in data transmission? To figure out this question, we analyze the energy consumption in different phases of the D2D based framework for relay and UE separately, as shown in Table III. In D2D discovery, UE initiates the D2D discovery, which consumes a little bit more energy than relay does. In data transmission, UE only needs to send its messages to relay through D2D approach. While relay collects the forwarded messages through D2D approach and send the messages and its own messages through cellular approach. Based on the results in Table III, we can conclude that the energy consumed in discovery and connection are relatively close; the energy used in discovery and connection is accounted for a large proportion in the whole communication. The conclusion might induce the possibility that the proposed D2D based framework consumes more energy than the traditional cellular approach.

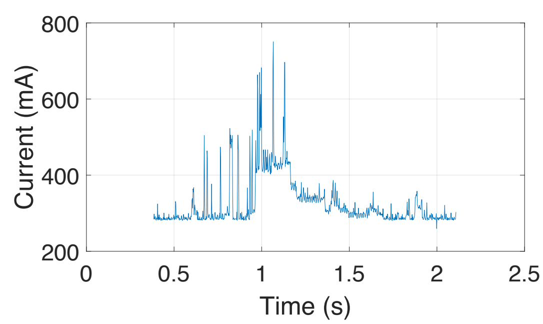

Fig. 6. Energy Consumption in D2D Transfer

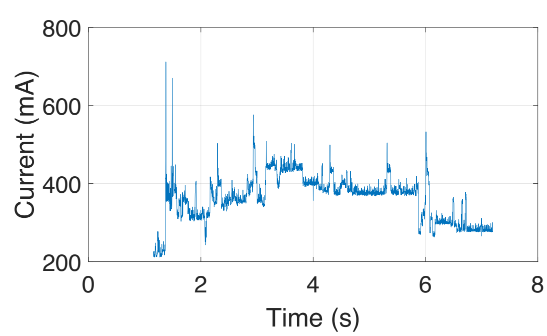

Fig. 7. Energy Consumption in Cellular Transfer

To explore the situation of energy efficiency, we measured our prototype in different scenarios. Firstly, we set different D2D connection time, which reflects the number of forwarded heartbeat messages. We use the transmission times of the forwarded heartbeat messages in standard size, 54 Bytes, during the D2D connection as a variable. In addition, we use a relay connected with one UE, 1 meter away from it. The system without any modification is set as the original system, which would be regarded as the baseline in experiments. Fig. 8 shows that the energy consumption of relay and the original system increases with the D2D connection time increasing, but the increased range of the UE largely falls behind the relay and the original system. Moreover, the difference in energy consumption between the relay and the original system stays almost constant, although the energy consumption of the relay is always slightly higher than that of original system. As a result, as the D2D connection time grows, the saved energy of the UE will exceed considerably to the wasted energy of the relay. Fig. 9 shows that on the period of first message forwarded, the D2D approach reaches nearly the same energy consumption as the original system. In the aspect of UE, it saves 55% energy consumption as compared to the energy consumption of cellular network. When seven heartbeat messages are forwarded, the whole system based on D2D communication could save 36% energy consumption. In brief, if D2D connection lasts longer, our D2D based framework will save more energy. For UEs, it always achieves tremendous energy saving. For the whole system, the performance in energy saving will improve with the D2D connection time growing.

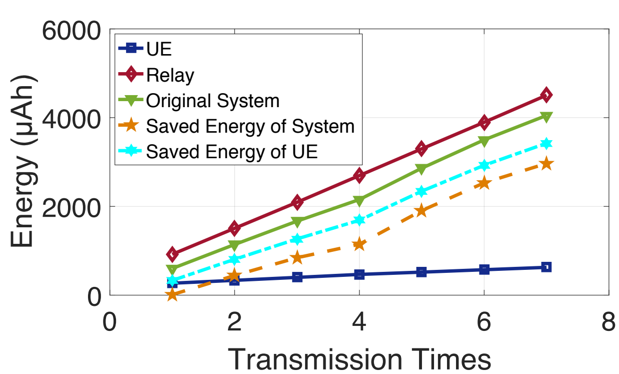

Fig. 8. Energy Consumption Comparison for the Whole System, UE and Relay

To disclose the energy consumption in this scenario with more details, we analyze the difference in energy consumption in different phases of our D2D based framework for relay and UE separately, as shown in Table III. From the result of Table III, the energy consumed in discovery and connection are relatively close; the major difference in energy consumption appears in the D2D communication part. A relay spends much more energy in receiving messages from UE than the energy consumed, for UE, to send messages to a relay.

Because a relay might serve multiple UEs in real-world scenario, we then conducted the experiment on the condition

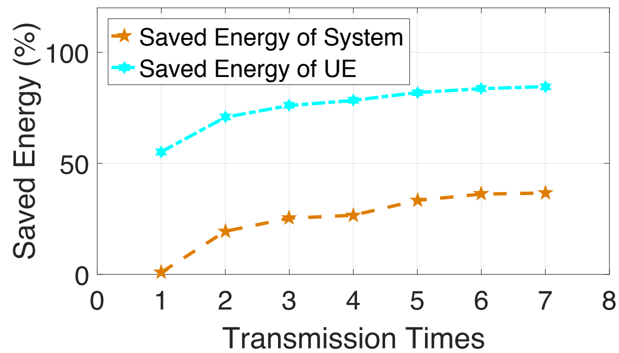

Fig. 9. Saved Energy of the Whole System and UE

TABLE III
ENERGY CONSUMPTION IN DIFFERENT PHASE

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

that a relay was connected with multiple UEs. The size of heartbeat messages and the communication distance stay as the previous one. Fig. 10 gives a comparison of energy consumption when a relay is connected with different numbers of UEs with the D2D connection time growing. It is evident that the increased energy consumption caused by more UEs connected is large in amount for the relay when only few heartbeat messages forwarded. However, when the connection time lasts long enough, the impact of the multiple connected UEs can be neglected for its little proportion. Fig. 11 shows the relationship between the wasted energy caused by the relay and the energy saved by the UE. With more UEs connected with a relay and longer D2D connection time, ratio of the wasted energy caused by the relay and the energy saved by the UE drops from around 97% to around 5%. To find out the reason, we measured the energy consumption in discovery phase, connection phase, and forwarding phase, respectively. According to the composition of energy consumption, we find that more times awaking, caused by the increasing UEs, to receive messages lead to the major difference in energy consumption. We present the detailed data of receiving phase in Table IV, which concludes an approximate linear relationship between the energy consumption of receiving data and the number of connected UEs. In addition, the excessive energy caused by multiple connected UEs is significantly less than the energy consumption used to send the collected messages. Consequently, the influence of more connected UEs drops gradually as more heartbeat messages sent.

TABLE IV
ENERGY CONSUMPTION IN D2D COMMUNICATION

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	

Communication distance is also a necessary factor needed to be taken into consideration in real-world scenario, for

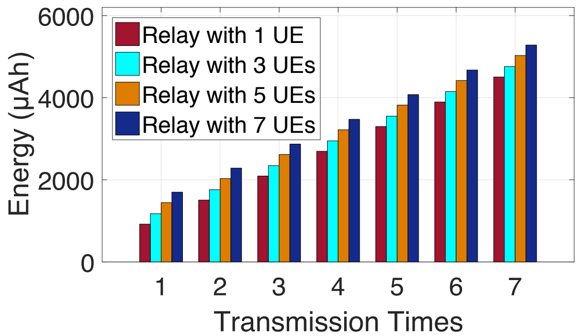

Fig. 10. Energy Consumption of a Relay with Multiple UEs

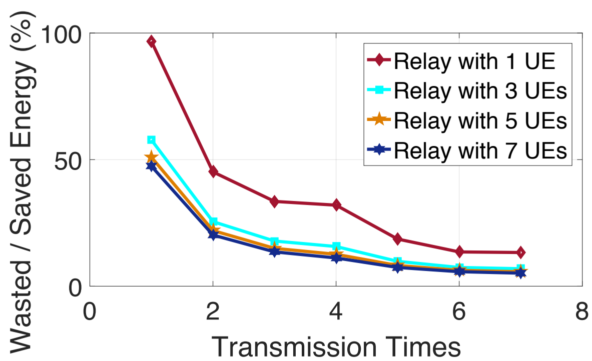

Fig. 11. Ratio of Wasted Energy to Saved Energy

communication distance varies in a large range in practice. We set three different communication distances in experiments, attempting to reveal the tendency of the energy consumption. Fig. 12 exposes that with the communication distance increased, Wi-Fi Direct consumes more energy apparently. We could predict that UE might consume more energy than original system when the communication distance beyond a certain value. Hence, we try to match a relay with the UE as close as possible for lower energy consumption.

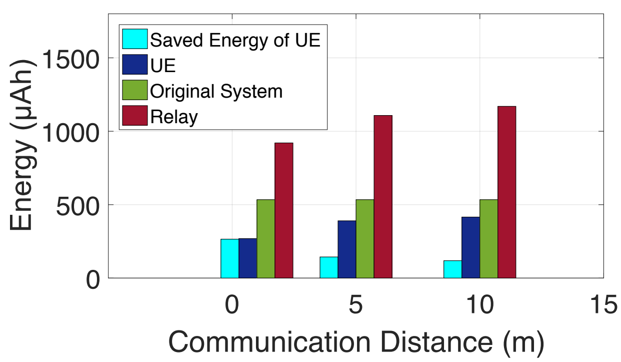

Fig. 12. Energy Consumption in Different Communication Distance

Because different IM apps always send heartbeat messages with different sizes, we evaluate the prototype with different-sized message transmissions. According to the range of the common heartbeat sizes, we set 54 Bytes as the standard of message size, and gradually scale it up until around 300 Bytes. Fig. 13 shows that the energy consumption stays almost constant, which is appropriate for small-sized messages. The result of energy consumption might be different if the size of message largely exceeds the standard, which conflicts with the system constrain of the small-sized messages.

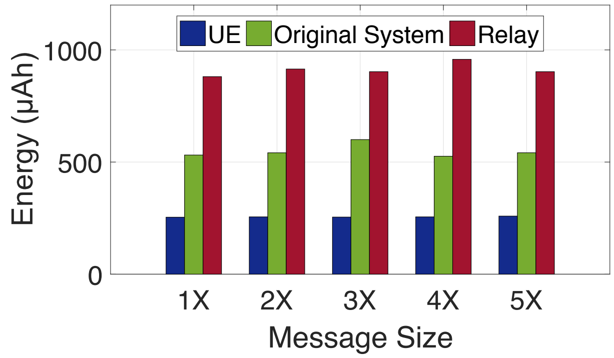

Fig. 13. Energy Consumption in Different Message Sizes

B. Performance in Signaling Consumption

Setup. To measure the cellular signaling traffic in heartbeat transmissions, we use NetOptiMaster to capture the layer 3 messages of our prototype and a heartbeat transmission emulator in WCDMA network separately. The layer 3 messages of signaling reveal the operation of cellular network. 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 the cellular network approach in heartbeat transmissions.

Similar to the energy consumption evaluation, we conducted the measurement in different D2D connection time via proposed D2D approach and traditional cellular approach. In practice, the connected UE would send more heartbeat messages via D2D approach with longer D2D connection time, leading to less RRC connection between the BS and the UE. Besides, we also evaluate the impacts of different numbers of the connected UEs for a relay.

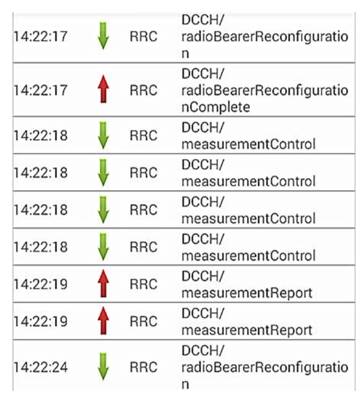

Fig. 14. Experiment Setup for Signaling Measurement

Results. Fig. 15 shows the cellular signaling traffic for the UE, the relay with different numbers of connected UEs, and the original system. Based on the result, the UE brings in no extra cellular signaling traffic via D2D communication because of the direct communication between smartphones. As for the relay and the original system, the cellular signaling traffic of the relay is nearly the same as the original system. With the number of the connected UE increasing, the cellular signaling traffic of the relay nearly stay invariable. Thus, our D2D based framework achieves more than 50% cellular signaling traffic saving through D2D forwarding and the performance in cellular signaling traffic saving will improve if more UEs are connected with a relay.

With the transmission times increasing, we can observe that the relay with more UEs connected incurs slightly more cellular signaling traffic. The reason of this phenomenon might be that more data in once transmission incurs more cellular traffic. When the size of transmitted messages exceeds some value, the cellular signaling traffic would slightly increase correspondingly. In addition, we could conclude from this figure that the increased cellular signaling traffic is minor for the total cellular signaling traffic. Thus, it proves that our scheme, cutting down the times of heartbeat messages transmissions and sending them aggregately, is effective in reducing cellular signaling traffic.

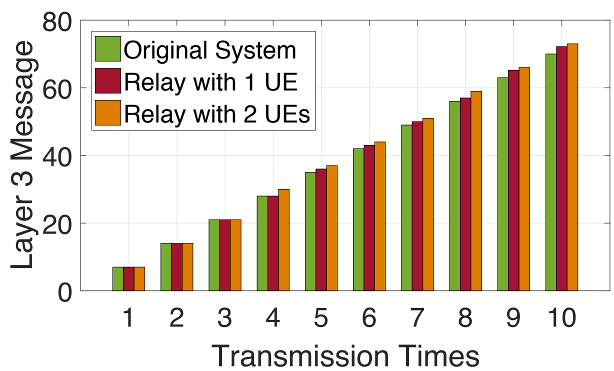

Fig. 15. Layer 3 Message Consumption

C. Limitations

We also found several limitations of our framework during the experiments.

Our framework based on Wi-Fi Direct is not easy to use in pratice. The users have to bear redundant process during discovering neighbors and establishing D2D connections. For example, a relay has to confirm the each change of its connected UE(s) manually. However, we originally hope relays could complete heartbeat messages forwarding without much interruption to their daily use. In the next generation of cellular networks, D2D techniques would be embedded in cellular networks. It might refine the redundant process. Thus, we think that our framework would be friendlier to users with the development of D2D technology.

Our framework has some negative impacts. Firstly, heartbeat messages transmitted via Wi-Fi Direct incur some extra traffic to Wi-Fi, although it is minor. Besides, UE is likely to suffer from more energy consumption because it has to send its heartbeat messages again if some exceptions occur in D2D forwarding. Our mechanism for UEs to select the proper relay, even if effectively reduce the chances of these situations, cannot eliminate it.

VI. Related Work

Many researchers have already studied signaling storm because of its unneglectable impact on mobile network, especially with the rapid development of smartphone network application. [3] gives a detailed analysis about the origin of signaling storm and its negative influence. There are also many solutions proposed to relieve such problem. For example, [2] presents an insightful view on the impact of periodic messages on the mobile wireless network. [22], [11], [23] propose some schemes of modification on communication protocol to reduce the signaling traffic. In the industry, push notification server, such as Apple Push Server, is designed to reduce the number of servers connected with UE, resulting in less periodic messages for keeping the connection between smartphones and remote servers alive.

Energy-efficient approaches for smartphone are common in recent studies, such as [24], [25]. [26] employs fast dormancy to save energy with higher signaling overhead, which aggravates signaling storm while reducing energy consumption. Many works, like [1], [27], schedule the delay-tolerant data to achieve energy saving, which is also applied in this work for energy saving and lower signaling overhead.

D2D technique, as an important component of next generation cellular communication, is popular in academic researches. [28], [29] applied D2D communication in content sharing scenario, and design the mechanism to fit the scenario. [30], [31] combines the D2D techniques with traditional cellular networks to reduce data traffic. To the best of our knowledge, there is no existing work, which applies D2D techniques to reduce signaling traffic and achieves amount of energy saving in common scenario simultaneously.

VII. Conclusion

In this paper, we presented a D2D based framework 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 the framework, we utilize voluntary smartphones, i.e., relays, to collect heartbeat messages from nearby smartphones, i.e., UEs, via 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. To illustrate the advantage of D2D communication for heartbeat transmissions, we implemented a prototype of our framework in Android smartphones, which, instead of using cellular networks, conduct heartbeat transmissions via energy-efficient Wi-Fi Direct. Through real-world experiments, we demonstrate that (1) the feasibility of the proposed framework; (2) in the worst situation where there is only one UE connected to the relay, our framework can still reduce about 50% cellular signaling traffic of heartbeat transmissions; (3) the proposed framework can save at most 36% energy for the whole system (relays and UEs) in heartbeat transmissions. For UEs only, it can achieve up to 55% energy saving. Our framework could be further applied in other periodic message, such as advertisements and diagnostic messages of apps, transmission with the following constrains to reduce cellular signaling traffic. The messages (1) are small in size and short in duration, (2) don't need to reply, (3) are delay-torrent.

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 *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on*. IEEE, 2015, pp. 113–122.

[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 *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 51–60.

[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, 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, 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 *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 177–188.

[6] XMPP protocol. [Online]. Available: http://xmpp.org/

[7] MQTT protocol. [Online]. Available: http://mqtt.org/

[8] T. Stoner, X. Wei, J. Knight, and L. Guo, "Experience: Rethinking rrc state machine optimization in light of recent advancements," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 477–485.

[9] H. Zhang, K. Xue, and P. Hong, "Facing the signaling storm: A method with stochastic concept," in *Wireless Communications and Signal Processing (WCSP), 2014 Sixth International Conference on*. IEEE, 2014, pp. 1–6.

[10] Signaling. [Online]. Available: https://en.wikipedia.org/wiki/Signal

[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, 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," vol. PP, no. 99, pp. 1–1, 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 *IEEE GLOBECOM Workshops*, 2013, pp. 163–169.

[14] Y. Li, D. Jin, P. Hui, and Z. Han, "Optimal base station scheduling for device-to-device communication underlaying cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 1–1, 2015.

[15] uberPOOL. [Online]. Available: https://www.uber.com/ride/uberpool

[16] Karma Go. [Online]. Available: https://yourkarma.com/how-it-works

[17] J. Nagle, "Congestion control in ip/tcp internetworks," *Acm Sigcomm Computer Communication Review*, vol. 14, no. 4, pp. 11–17, 1984.

[18] Nagle's algorithm. [Online]. Available: https://en.wikipedia.org/wiki/Nagle%27s_algorithm

[19] LTE Direct. [Online]. Available: https://www.qualcomm.com/invention/technologies/lte/direct

[20] Android Developers. [Online]. Available: https://developer.android.com/training/connect-devices-wirelessly/wifi-direct.html

[21] Monsoon Solutions Inc. Power Monitor. [Online]. Available: http://www.msoon.com/LabEquipment/PowerMonitor/

[22] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "Top: Tail optimization protocol for cellular radio resource allocation," in *Network Protocols (ICNP), 2010 18th IEEE International Conference on*. IEEE, 2010, pp. 285–294.

[23] 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 *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014, pp. 3499–3504.

[24] 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 *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 195–199.

[25] Y. Cui, S. Xiao, X. Wang, M. Li, H. Wang, and Z. Lai, "Performance-aware energy optimization on mobile devices in cellular network," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 1123–1131.

[26] 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 *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 101–112.

[27] 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 *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013, p. 7.

[28] 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, 2016.

[29] 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, 2016.

[30] T. Bansal, K. Sundaresan, S. Rangarajan, and P. Sinha, "R2d2: Embracing device-to-device communication in next generation cellular networks," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 2014, pp. 1563–1571.

[31] 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 *Mobile Ad Hoc and Sensor Systems (MASS), 2015 IEEE 12th International Conference on*. IEEE, 2015, pp. 19–27.