

Energy-Aware Cooperative Traffic Offloading via Device-to-Device Cooperations: An Analytical Approach

Yuan Wu, Jiachao Chen, Li Ping Qian, Jianwei Huang, Xuemin (Sherman) Shen



Abstract—In this paper, we investigate the cooperative traffic offloading among mobile devices (MDs) which are interested in receiving a common content from a cellular base station (BS). For offloading traffic, the BS first sends the content to some selected MDs which then broadcast the received data to the other MDs, such that each MD can receive the entire content simultaneously. Due to each MD's limited transmit-power and energy budget, the transmission rate of the content should be properly designed, since it strongly influences whether and how long each MD can perform relaying. Therefore, different from most existing MDs cooperative schemes, we focus on a novel joint optimization of the content transmission rate and each MD's relay-duration, with the objective of minimizing the system cost accounting for the energy consumption and the cellular-link usage. To tackle with the technical challenge due to the coupling effect between the content transmission rate and each MD's relay-duration, we exploit the decomposable property of the joint optimization problem, based on which we characterize different possible cases for achieving the optimal solution. We then derive the optimal solution for each case analytically, and further propose an efficient algorithm for finding the globally optimal solution of the original joint optimization problem. Numerical results are provided to validate the proposed algorithm (including its accuracy and computational efficiency) and demonstrate that the optimal MDs' cooperative offloading can significantly reduce the system cost compared to some heuristic schemes. Several interesting insights about the cooperative offloading are also obtained.

Index Terms—Device-to-Device Communications, Mobile User Cooperation, Traffic Offloading, and Radio Resource Allocations.

1 INTRODUCTION

With the explosive growth of mobile applications, people are relying more heavily on mobile devices for sharing contents and watching video streaming, which yields a tremendously heavy traffic demand in cellular networks. As predicted by Cisco, the global mobile traffic will increase nearly tenfold between 2014 and 2019, and nearly three-fourths of the mobile traffic will be video streaming by 2019. Offloading traffic by exploiting mobile devices' (MDs') cooperations is widely considered as a promising approach for relieving such a traffic pressure. Given a group of MDs in close proximity and interested in downloading a common content, the cooperative offloading enables the cellular Base Station (BS) to first send part of the content to some

selected MDs, which will in turn relay the received data to their local neighbors¹. By exploiting close proximity among the MDs, the cooperative offloading improves the efficiency of content distribution, by lowering the energy consumption and reducing the traffic demand at the BS. The recent technology advances, e.g., WiFi-direct and LTE-Direct [1], [2], have made the device-to-device (D2D) communications implementable in practice [3] and have motivated a lot of academic studies [4]–[11] as well as industry standardization efforts (such as using D2D as an underlay to LTE-Advance networks) [12], [13].

A successful exploitation of the MDs' cooperative offloading requires a careful design of the cooperative scheme as well as the associated radio resource allocations. There is a large body of related studies devoted to this area, which roughly can be categorized into two groups: those focusing on the MDs' cooperative offloading for distributing real-time traffic and those focusing on distributing delay-tolerant traffic via cooperation. The studies in the first group mainly investigated how different MDs cooperate for distributing contents (e.g., which MDs relay which parts of the data) and how to motivate the MDs to cooperate. The studies in the second group, on the other hand, mainly investigated different opportunistic offloading (or forwarding) schemes that provided different tradeoffs between radio resource usage and delay performance in disseminating contents.

Our study here belongs to the first group of studies. In particular, we focus on the joint optimization of the transmission rate for content delivery as well as each MD's relay-duration. Although this issue has received little attention in the literature, it is an important issue in radio resource management for the following two reasons.

- *First*, due to each MD's limited transmit-power and energy budget, the transmission rate of the content influences the MDs' cooperations, i.e., which MDs can be selected for relaying and how long to relay. To better understand this point, consider a particular MD (labeled as MD 1) located at the center of a group of MDs. Such a location makes MD 1 an ideal candidate for relaying traffic to other MDs in the group. However, suppose that MD 1 has very limited transmit-power capacity and energy budget for relaying². In this case, MD 1 might be infeasible to perform relaying (or it can only perform relaying for a very short duration), if the transmission rate of the content is large, which requires a large transmit-power of MD 1 to perform relaying.

1. To facilitate the content relay, each MD is equipped with two radio interfaces of similar capabilities. This property has been conceived as a practice for most users' equipments in near future [1], [2], [12], [13].

2. The energy budget helps avoid the situation that an MD uses up its entire energy capacity to perform relaying.

Y. Wu and J. Chen are with College of Information Engineering, Zhejiang University of Technology, Hangzhou, China, (email: ieuw@zjut.edu.cn). Y. Wu is also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xian 710162, China.

L. Qian (corresponding author) is with College of Information Engineering, Zhejiang University of Technology, Hangzhou, China, (email: lpqian@zjut.edu.cn). J. Huang is with the Network Communications and Economics Lab, Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong (e-mail: jwhuang@ie.cuhk.edu.hk).

S. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: xshen@bbcr.uwaterloo.ca). This work is supported in part by the National Natural Science Foundation of China (61303235, 61379122, and 61572440), and the General Research Funds (Project Number CUHK 412713 and 14202814) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

- *Second*, the transmission rate of the content directly influences the usage of the cellular-link (under a fixed size of the content). Setting a smaller transmission rate of the content, although making more MDs eligible to perform relaying (as explained before), prolongs the transmission duration of the entire content. Thus, a longer use of cellular-link is required, which is unfavorable from network operator's point of view.

In this work, we are motivated to investigate an optimization framework that jointly controls the transmission rate of the content and the consequent relay-duration of each MD, with the objective of minimizing the total system cost accounting for both the energy consumption and the cellular-link usage.

The rest of this paper is organized as follows. In Section 2, we review the related studies and describe our key contributions in this study. In Section 3, we illustrate the system model and the cooperative traffic offloading scheme. In Section 4, we present the joint optimization framework, and decompose it into two subproblems. Sections 5 and 6 solve the two subproblems, respectively, by using backward induction. Numerical results are presented in Section 7, and we conclude this study in Section 8.

2 LITERATURE REVIEW AND CONTRIBUTIONS

The existing studies that investigated the resource managements for cooperative traffic offloading can be roughly categorized into the following two groups.

The first group of studies focused on the MDs' cooperative offloading for distributing real-time traffic. These results mainly investigated which MDs relay which parts of the contents, with the objective of optimizing different system-wide performances (e.g., saving the energy consumption or reducing the traffic at the BS). Specifically, Al-Kanj *et al.* in [7] investigated how to separate mobile users into different groups and select one leader of each group for relaying content, with the objective of minimizing the total power consumption. Further in [8], Al-Kanj *et al.* investigated the grouping and leader-selection problem for minimizing the bandwidth usage of the cellular link. Wang *et al.* in [14] considered cooperative traffic offloading in vehicular networks, and proposed a coalition game based approach for distributing contents within a group of vehicles. In [15], Cheng *et al.* took into account mobility of vehicles and studied the traffic offloading via WiFi networks. In [16], Kang *et al.* focused on optimizing network operator's revenue by selectively offloading users' traffic to third-party WiFi networks. Another related and important question is how to incentivize mobile users to cooperatively offload traffic. For instance, in [10], Gao *et al.* proposed a hybrid pricing-reimbursing policy for motivating the mobile users to play as WiFi-hosts and provide Internet connectivity for other users. In [17], Vu *et al.* proposed a heuristic tit-for-tat incentive mechanism to motivate users' cooperations. In [18], Niyato *et al.* proposed a sequential game model to analyze the cooperations between network operators and content providers for content delivery. In [19], Han *et al.* studied the cooperative traffic offloading from cellular operators to internet service providers (ISPs) which are usually closer to end users.

The second group of studies focused on distributing delay-tolerant traffic via the MDs' cooperative offloading. The main focus is to design different opportunistic offloading schemes for content distribution. In [23] and [24], different schemes were proposed to migrate delay-tolerant traffic from cellular networks to WiFi networks and D2D networks, respectively. In [25], Li *et*

al. proposed an energy-efficient opportunistic forwarding scheme for maximizing the message-delivery probability. In [26], Wang *et al.* proposed a hybrid pull-and-push scheme for opportunistic content delivery. In [27], Whitbeck *et al.* considered a push-based architecture for opportunistic content delivery and evaluated the influence of the number of content copies. In [28], Golrezaei *et al.* proposed a femto-caching scheme for video distribution³. In [29] and [30], Mavromoustakis *et al.* proposed a traffic-aware scheduling scheme and a social-aware process-offloading scheme for conserving energy consumptions of wireless devices (and thus prolonging their lifetimes), respectively.

Our study here belongs to the first group of studies. Different from the existing studies in Group 1 that modeled and analyzed the MDs' cooperations from a macro-view, we adopt a micro and analytical approach for modeling and optimizing the radio resource usage for a typical MDs' cooperation model. We focus on investigating the coupling effect between the transmission rate of the content and the MDs' relay-strategies. As we described before, such a coupling effect significantly influences the resource usage and the performance of the cooperative scheme. Moreover, incorporating the coupling effect in problem formulation leads to a challenging nonconvex problem, which requires a carefully designed solution methodology. Specifically, we focus on a scenario where a group of MDs are in close geographical proximity with each other, and are interested in downloading a common content from the BS. We consider a *typical cooperative model* [7], [8], [17], in which the BS first unicasts to some selected MDs, which at meantime broadcast their received data to the other MDs, such that each MD can obtain the entire content simultaneously⁴. Considering the MDs' limited energy budgets for receiving and relaying the content, we allow the BS to sequentially select different MDs to perform relaying.

Our key contributions in this study are summarized in the following three aspects.

First, we formulate an optimization framework that jointly controls the transmission rate of the content and the relay-duration of each MD, with the objective of minimizing a system-wide cost while guaranteeing a delay constraint for content delivery. The system-wide cost includes the total energy consumption of the BS and all MDs as well as the cost for cellular-link usage. In particular, our optimization framework considers both the transmit-power limit and the energy consumption budget of each MD to ensure that each MD's budget on its total energy consumption (for both receiving and relaying data) is respected.

Second, we characterize the optimal solution of the joint optimization problem and propose an efficient algorithm to compute the solution. We would like to emphasize that the joint optimization problem is difficult to solve, since the transmission rate of the content influences each MD's relay-duration in a complicated manner, which yields a difficult nonconvex optimization. To tackle with this difficulty, we identify the decomposition property of the joint optimization, based on which we characterize different

3. Besides the aforementioned two groups of studies, there also exist a huge body of research that investigated different aspects about content distributions in wireless networks. Interested readers please refer to [20] for a survey study.

4. Such a cooperative model has been widely adopted in the literature, e.g., for video streaming [21] and for distributing multimedia contents like the music-group-play [22]. Although the considered MDs' cooperative model shares a similar rationale as the peer-to-peer content sharing in wired networks, our proposed joint optimization framework that accounts for the coupling effect between the content transmission and the MDs' relay-strategies and the associated analysis of radio resource usages make our study a novel contribution to the state-of-the-art.

possible cases when achieving the optimum. We then analytically derive the corresponding optimal solution for each of these cases, and finally propose an efficient algorithm to find the optimal solution of the original joint optimization problem by using the derived analytical results.

Third, we perform extensive numerical simulations to validate the derived analytical results and the proposed algorithm to compute the optimal solution (i.e., the optimal transmission rate of the content and the optimal relay-duration of each MD). The results show that the proposed algorithm saves more than 90% of the computational time compared to an exhaustive search method, while guaranteeing to achieve the optimal solution. Besides, we show through simulations that the cooperative offloading with the jointly optimized content transmission rate and the MDs' relay-durations can significantly reduce the total system cost. Moreover, the optimal cooperative offloading can benefit the system more (i.e., saving a larger portion of the total system cost) when more MDs coexist for cooperations. The MDs' distribution also influences their consequent cooperations, i.e., more cooperations will be invoked when the MDs are further away from the BS and closer with each other.

3 SYSTEM MODEL

3.1 System Model and Cooperative Scheme

We consider a set $\mathcal{I} = \{1, 2, \dots, I\}$ of MDs who are in close proximity of each other and interested in downloading a common content from the BS. The size of the content is L bits, and the corresponding transmission rate is r bits/second. Hence, the total transmission duration of the content is $x = L/r$ (notice that the values of the decision variables r and x determine each other under a fixed L). Each MD has two radio interfaces, with one interface for receiving data from the BS (i.e., the cellular-link) and the other interface for local communications with the other MDs (i.e., the device-to-device link).

The cooperative offloading model works as follows. The BS first unicasts part of the content via the cellular-link to MD i at a rate r for a duration z_i . At the same time, MD i broadcasts its received data via the device-to-device link to all the other MDs at the same rate r . Followed by this procedure, the BS sequentially chooses different MDs for unicasting the data of the content, and the selected MDs then broadcast their received data to the other MDs to exploit the MDs' close proximity. Similar to [7], [8], [10], [38], we assume that the processing delay at each MD (when it performs relaying) is small enough and can be ignored, which simplifies our following quantitative modeling and analysis and enables us to derive clear analytical insights. Besides, we consider that the MDs are required to use the same transmission rate as the BS, which corresponds to a benchmark case that each MD uses the same coding rate/scheme (as that of the BS) for relaying the content, without invoking adaptive coding scheme.

Figure 1(a) illustrates the case that MD 1 performs the relay (i.e., local broadcasting) for duration z_1 . Specifically, the blue solid arrow represents the unicast transmission from the BS, and the two red dash-lines represent the broadcast-transmission from MD 1 for relaying its received data. Figure 1(b) illustrates the case that MD 2 performs the relay for duration z_2 . Notice that as different MDs relay the content at non-overlapping time periods, there is no interference among the MDs. Besides, we assume that when an MD is broadcasting content over its device-to-device link,

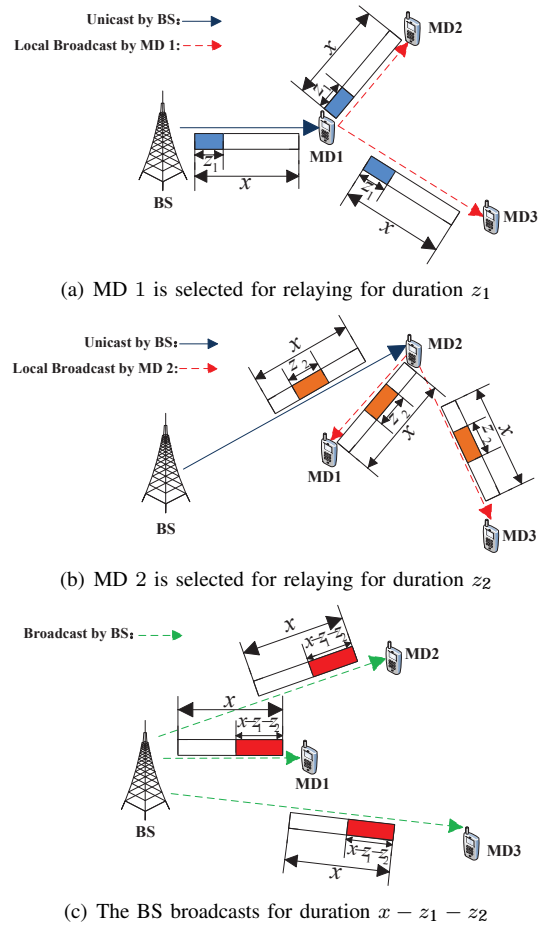


Fig. 1: Illustration of the considered cooperative scheme among three MDs, i.e., $\mathcal{I} = \{1, 2, 3\}$. The transmission-duration for whole content is x . MD 1 and MD 2 are selected for relaying the data for relay-durations z_1 and z_2 , respectively. The BS broadcasts to all MDs for the duration $x - z_1 - z_2$ to finish the delivery of the whole content.

it utilizes a frequency channel non-overlapping with that of the cellular-link, e.g., based on the LTE-Direct [12].

Due to the MDs' limited transmit-powers and energy capacities, the BS might need to broadcast some data to all MDs to finish the delivery of the whole content. The duration for the BS to broadcast, if needed, is $x - \sum_{i \in \mathcal{I}} z_i$. Figure 1(c) shows the case that the BS broadcasts to all MDs (i.e., the green dash-lines). However, the BS's broadcast-transmission is undesirable, since it consumes a significant transmit-power due to taking account of the MD with the worst channel condition from the BS.

Based on the above cooperative model, we aim at jointly optimizing the content transmission rate r (or equivalently its transmission duration x) and each MD i 's relay-duration z_i , in order to minimize the total system cost for the content delivery. The total system cost includes three parts: i) the energy consumption of the BS, ii) the energy consumption of each MD, and iii) the usage (occupancy) of the cellular-link. The details of the modeling are presented in the next two subsections. Since the optimal transmission duration for one content is usually very short (as we will show in the simulation section), we assume that the MDs' locations are relatively static (e.g., in indoor environment).

3.2 Energy Consumption of the BS

The energy consumption of the BS includes two parts, i.e., that for unicasting to MD i (when MD i is selected for relaying), and

that for broadcasting to all MDs. The details are as follows.

Energy Consumption of the BS for Unicasting: Suppose that MD i is selected by the BS for relaying with a duration z_i . During z_i , the BS unicasts to MD i at the transmission rate r . We use $F_{Bi}(r)$ to denote the required transmit-power by the BS to perform this unicasting (the subscript “B” stands for the BS), and such power depends on the choice of MD i (hence, the subscript i is included). Using the Shannon’s channel capacity formula, $F_{Bi}(r)$ can be expressed as $F_{Bi}(r) = (2^r - 1)n/g_{Bi}$, where for the sake of clear presentation, we assume an unit bandwidth of the channel. Parameter g_{Bi} denotes the channel power gain from the BS to MD i , and parameter n denotes the power of the background noise. In addition to the transmit-power, the BS also consumes a static circuit power dissipation q_B when it is transmitting data (due to the operations of the device electronics such as mixers, filters, and digital-to-analog converts). We assume that q_B is independent of the transmit-power. Taking into account the transmit power and the circuit power, the total energy consumption of the BS when selecting MD i for relaying with a duration z_i is given by⁵:

$$E_{Bi}(x, z_i) = (F_{Bi}(r) + q_B)z_i = (2^r - 1)\frac{n}{g_{Bi}}z_i + q_Bz_i. \quad (1)$$

Energy Consumption of the BS for Broadcasting: When $x - \sum_{i \in \mathcal{I}} z_i > 0$ (i.e., the total relay-duration of all MDs is less than the transmission duration x of the content), the BS needs to finish the content transmission by broadcasting to all MDs for a period of $x - \sum_{i \in \mathcal{I}} z_i$. In particular, we use function $F_{B0}(r) = (2^r - 1)\frac{n}{\min_{i \in \mathcal{I}}\{g_{Bi}\}}$ to denote the required transmit-power of the BS for successfully broadcasting to all the MDs at the transmission rate r . The $\min_{i \in \mathcal{I}}\{g_{Bi}\}$ in the denominator is due to the fact that the broadcasting of the BS should take into account the MD with the worst channel power gain. Then, the part of energy consumption for the BS to perform broadcasting is given by:

$$\begin{aligned} E_{B0}(x, \{z_i\}_{i \in \mathcal{I}}) &= (F_{B0}(r) + q_B)(x - \sum_{i \in \mathcal{I}} z_i) \\ &= \left((2^r - 1)\frac{n}{\min_{i \in \mathcal{I}}\{g_{Bi}\}} + q_B \right) (x - \sum_{i \in \mathcal{I}} z_i). \end{aligned} \quad (2)$$

Summarizing (1) and (2), the BS’s total energy consumption is

$$E_B^{\text{tot}}(x, \{z_i\}_{i \in \mathcal{I}}) = E_{B0}(x, \{z_i\}_{i \in \mathcal{I}}) + \sum_{i \in \mathcal{I}} E_{Bi}(x, z_i). \quad (3)$$

3.3 Energy Consumption of Each MD

The energy consumption of each MD i also includes two parts, i.e., that for data reception, and that for relaying its received data. The details are as follows.

Energy Consumption of each MD for Data Reception: The main operation of the MDs is data reception. According to [6] [8], the circuit power consumption of each MD when it is receiving data can be modeled as a constant, and we denote it by h_i for MD i . In particular, there are three possible scenarios in which MD i is receiving data, namely, i) when the BS unicasts the data to MD i (when MD i is selected as a relay), ii) when some other MD $i' \neq i$ broadcasts to MD i (when MD i' is selected as a relay), and iii) when the BS broadcasts to all MDs. Considering these

three scenarios, the energy consumption of MD i for receiving the whole content is given by:

$$E_i^{\text{rec}}(x) = h_i \left((x - \sum_{i \in \mathcal{I}} z_i) + z_i + \sum_{i' \neq i, i' \in \mathcal{I}} z_{i'} \right) = h_i x. \quad (4)$$

Energy Consumption of each MD for Relaying: Besides receiving data, if selected, MD i also relays its received data to the other MDs for a duration z_i . We use the following function

$$F_i(r) = (2^r - 1)\frac{n}{\min_{i' \neq i, i' \in \mathcal{I}}\{g_{ii'}\}} \quad (5)$$

to denote the required transmit-power of MD i for broadcasting to the other MDs (where $g_{ii'}$ is the channel gain from MD i to a different MD i'). Thus, the energy consumption of MD i for relaying its received data to all the other MDs is equal to $(F_i(r) + q_i)z_i$. In practice, the circuit power consumption of mobile device (when transmitting) is usually significantly smaller than that of the cellular base station. For example, according to [31]–[35], the circuit power consumption of cellular base stations when transmitting is around the order of 1-10W. In comparison, according to [7], [36], [37], the circuit power power of mobile devices when transmitting is around the order of 10mW, which is 1% or less of that of the BS. Therefore, for simplicity, we do not explicitly consider q_i in each MD’s energy consumption in the rest of this paper (we will show in Section 7 through numerical examples that the resulting relative error due to such an approximation is very marginal).

Summarizing the above two parts, the total energy consumption of MD i is given by:

$$E_i^{\text{tot}}(x, z_i) = E_i^{\text{rec}}(x) + F_i(r)z_i. \quad (6)$$

4 PROBLEM FORMULATION & DECOMPOSITION

4.1 Problem Formulation

We formulate an optimization problem that jointly controls the transmission rate r of the content, its transmission duration x , and the relay-duration z_i of each MD i . Our objective is to minimize the total system cost that includes the total energy consumption of the BS and all MDs as well as the cellular-link usage cost. Problem (P1) below gives the detailed problem formulation.

$$(P1): \min_{x, r, \{z_i\}_{i \in \mathcal{I}}} O(x, \{z_i\}_{i \in \mathcal{I}}) = \alpha E_B^{\text{tot}}(x, \{z_i\}_{i \in \mathcal{I}}) +$$

$$\sum_{i \in \mathcal{I}} \beta_i E_i^{\text{tot}}(x, z_i) + \gamma x$$

$$\text{subject to: } x = \frac{L}{r}, \quad (7)$$

$$x \leq T^{\text{max}}, \quad (8)$$

$$\sum_{i \in \mathcal{I}} z_i \leq x, \quad (9)$$

$$0 \leq z_i \leq x \mathbf{I}(F_i(r) \leq P_i^{\text{max}}), \forall i \in \mathcal{I}, \quad (10)$$

$$E_i^{\text{tot}}(x, z_i) \leq E_i^{\text{b}}, \forall i \in \mathcal{I}. \quad (11)$$

In Problem (P1), the first two terms in $O(x, \{z_i\}_{i \in \mathcal{I}})$ capture the energy consumption of the BS (weighted by α) and that of each MD i (weighted by β_i). The third term in $O(x, \{z_i\}_{i \in \mathcal{I}})$ accounts for the cost for the cellular-link usage (weighted by γ).

Constraint (7) explains the relationship between r and x , under the given file size L . Constraint (8) ensures that the transmission duration x cannot exceed a prefixed upper bound T^{max} , which corresponds to the strict deadline for delivering the content.

5. Similar model of the BS’s energy consumption, i.e., the one includes a dynamic part dependent on the served traffic rate and a fixed part accounting for the circuit processing, has been widely adopted [19], [31]–[33].

Constraint (9) ensures that the total relay-duration of all MDs cannot exceed the transmission duration x . Constraint (10) means that MD i is eligible for relaying, only if its required transmit-power for broadcasting $F_i(r)$ in (5) is below its transmit-power limit P_i^{\max} . Here, the indicator function $I(\mu) = 1$ if condition μ is satisfied, and $I(\mu) = 0$ otherwise. Constraint (11) ensures that MD i 's total energy consumption $E_i^{\text{tot}}(x, z_i)$ (as in (6), for both receiving and relaying the content) cannot exceed its energy consumption budget E_i^b (where the superscript "b" represents "budget"). Each MD i sets its own energy budget E_i^b based on its own interest in contributing in relaying⁶, and reports E_i^b to the BS truthfully. This means that we focus on the network performance optimization with complete network information (including energy budgets). We will consider the mechanism design problem that aims at inducing truthful telling behaviors with incomplete network information in our future work.

We notice that Problem (P1) is always feasible, since at least the BS can send the content to all MDs via broadcasting without invoking any MD's relaying. In this work, we use x^* (which leads to $r^* = L/x^*$) and $\{z_i^*\}_{i \in \mathcal{I}}$ to denote the optimal solution of Problem (P1). To derive x^* and $\{z_i^*\}_{i \in \mathcal{I}}$ analytically, we focus on the resource consumption for delivering one piece of content (the similar model also appeared in [26]). Our problem formulation can be further extended to investigate the case of multiple contents.

In Problem (P1), besides the linear constraint (9), the decision variable x (i.e., the transmission duration of the content) influences the decision variables $\{z_i\}_{i \in \mathcal{I}}$ (i.e., each MD's relay-duration) in a complicated manner. Specifically, x determines the required transmit-power of each MD for performing the consequent relaying, which thus influences i) whether MD i is eligible to be selected for relaying (according to (10)), and ii) how long MD i can perform relaying (according to (11)).

It can be verified that Problem (P1) is a nonconvex optimization problem with respect to x and $\{z_i\}_{i \in \mathcal{I}}$, since the objective function $O(x, \{z_i\}_{i \in \mathcal{I}})$ is not jointly convex in x and $\{z_i\}_{i \in \mathcal{I}}$ [39]. Thus, there does not exist a generic algorithm that can efficiently compute x^* , r^* , and $\{z_i^*\}_{i \in \mathcal{I}}$. This motivates us to solve Problem (P1) by exploiting its intrinsic decomposable structure as follows.

4.2 Decomposition of Problem (P1)

Function $O(x, \{z_i\}_{i \in \mathcal{I}})$ in Problem (P1), after making some manipulations, can be expressed as follows:

$$O(x, \{z_i\}_{i \in \mathcal{I}}) = \sum_{i \in \mathcal{I}} \left(\alpha F_{Bi} \left(\frac{L}{x} \right) - \alpha F_{B0} \left(\frac{L}{x} \right) + \beta_i F_i \left(\frac{L}{x} \right) \right) z_i + \alpha \left(F_{B0} \left(\frac{L}{x} \right) + q_B \right) x + \sum_{i \in \mathcal{I}} \beta_i h_i x + \gamma x, \quad (12)$$

in which only the first term depends on $\{z_i\}_{i \in \mathcal{I}}$. Hence, Problem (P1) can be decomposed into two subproblems as follows.

First, it is easy to see from (4), (6), and (11) that the transmission duration of the content x cannot exceed $\min_{i \in \mathcal{I}} \left\{ \frac{E_i^b}{h_i} \right\}$, otherwise some MD will violate its energy budget constraint even by just receiving. Together with (8), we can limit x in the interval of $[0, \min\{\min_{i \in \mathcal{I}} \{E_i^b/h_i\}, T^{\max}\}]$. In particular, if we fix the value of x , then we have the *bottom-layer subproblem* that

6. A small budget E_i^b implies that MD i is more interested in receiving the content, and a large budget E_i^b implies that MD i is also interested in helping other MDs by acting as a relay.

optimizes the relay-durations $\{z_i\}_{i \in \mathcal{I}}$ of each MD as follows:

(P1-Bottom): $O_{\text{bot}}(x) =$

$$\min_{\{z_i\}_{i \in \mathcal{I}}} \sum_{i \in \mathcal{I}} \left(\alpha F_{Bi} \left(\frac{L}{x} \right) - \alpha F_{B0} \left(\frac{L}{x} \right) + \beta_i F_i \left(\frac{L}{x} \right) \right) z_i$$

subject to: $\sum_{i \in \mathcal{I}} z_i \leq x,$ (13)

$$0 \leq z_i \leq x I \left(F_i \left(\frac{L}{x} \right) \leq P_i^{\max} \right), \forall i \in \mathcal{I}, \quad (14)$$

$$F_i \left(\frac{L}{x} \right) z_i \leq E_i^b - h_i x, \forall i \in \mathcal{I}. \quad (15)$$

In Problem (P1-Bottom), we have replaced r by x via using (7). Notice that the value of x is fixed in (13), (14), and (15), which are thus different from the original constraints (9), (10) and (11) in Problem (P1). We denote the optimal value of the bottom Problem (P1-Bottom) as $O_{\text{bot}}(x)$, which depends on x . We will analytically derive $O_{\text{bot}}(x)$ in Section 5.

After deriving $O_{\text{bot}}(x)$, we can substitute $O_{\text{bot}}(x)$ back into (12) and obtain the *top-layer subproblem* that optimizes the transmission duration x for the whole content as follows:

$$(P1-Top): \min_x O_{\text{bot}}(x) + \alpha \left(F_{B0} \left(\frac{L}{x} \right) + q_B \right) x + \sum_{i \in \mathcal{I}} \beta_i h_i x + \gamma x.$$

subject to: $0 \leq x \leq X^{\text{up}} = \min \left\{ \min_{i \in \mathcal{I}} \left\{ \frac{E_i^b}{h_i} \right\}, T^{\max} \right\}.$ (16)

By solving Problem (P1-Bottom) and Problem (P1-Top) in a way of backward induction, we can solve the original Problem (P1). The details are illustrated in the next two sections.

5 OPTIMAL SOLUTION OF (P1-BOTTOM)

In this section, we focus on solving Problem (P1-Bottom). Under a fixed value of x , the objective function and constraints (13), (14), and (15) of Problem (P1-Bottom) are linear with respect to the decision variables $\{z_i\}_{i \in \mathcal{I}}$. Therefore, Problem (P1-Bottom) is a linear programming problem.

To avoid confusion, we use $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ to denote the optimal solution of Problem (P1-Bottom), which depends on the given x . To derive $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$, we first introduce parameter M_i of each MD i as follows:

$$M_i = \alpha \frac{n}{g_{Bi}} + \beta_i \frac{n}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}} - \alpha \frac{n}{\min_{i' \in \mathcal{I}} \{g_{Bi'}\}}, \quad \forall i \in \mathcal{I}. \quad (17)$$

As we will illustrate soon, M_i is an important parameter that indicates how *helpful* MD i is in terms of performing relaying. For the sake of easy presentation, we make the following assumption in the rest of the paper.

Assumption 1: (An initial ordering of the MDs) In the rest of this paper, we assume that all MDs in \mathcal{I} have already been ordered according to an ascending order, i.e.,

$$M_1 \leq M_2 \leq \dots M_N < 0 \leq M_{N+1} \leq \dots \leq M_I, \quad (18)$$

always holds, where parameter N denotes the number of MDs whose $M_i < 0$. Recall that I denotes the total number of MDs.

Based on Assumption 1, we can derive $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ in the following proposition.

Proposition 1: The optimal solution of Problem (P1-Bottom) under a given value of x is as follows. For each MD i with $1 \leq i \leq N$, its unique optimal relay-duration is

$$z_i^{\text{bot}}(x) = \min \left\{ \eta_i \mathbb{I} \left(F_i \left(\frac{L}{x} \right) \leq P_i^{\text{max}} \right), \frac{E_i^{\text{b}} - h_i x}{F_i \left(\frac{L}{x} \right)} \right\}, \quad (19)$$

where η_i represents the *available relay-duration of MD i* , and it can be recursively computed as follows:

$$\eta_i = \max \left\{ x - \sum_{i'=1}^{i-1} z_{i'}^{\text{bot}}(x), 0 \right\}, \quad (20)$$

with the initial condition of $\eta_1 = x$. Besides, for each MD i with $N+1 \leq i \leq I$, its unique optimal relay-duration is $z_i^{\text{bot}}(x) = 0$.

Proof: Using M_i defined in (17), we first denote the objective function of Problem (P1-Bottom) under a fixed value of x by

$$H(\{z_i\}_{i \in \mathcal{I}}) = (2^{\frac{L}{x}} - 1) \sum_{i \in \mathcal{I}} M_i z_i. \quad (21)$$

Based on (21) and Assumption 1, it is easy to see that $z_i^{\text{bot}}(x) = 0$ holds if $i > N$ (otherwise, we can always decrease $H(\{z_i\}_{i \in \mathcal{I}})$ by setting $z_i^{\text{bot}}(x) = 0$ without violating any constraint).

We next prove (19) and (20) by showing contradiction. Without incurring any ambiguity, suppose that $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ is an optimal solution of Problem (P1-Bottom), but it does not satisfy (19) and (20). Our objective is to show via contradiction that $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ cannot be an optimal solution of Problem (P1-Bottom). Specifically, we consider the following three possible cases regarding $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$:

- *Case I:* suppose that i) there exists an MD i (with $i < N$), whose $z_i^{\text{bot}}(x) < \min \{ \eta_i \mathbb{I}(F_i(L/x) \leq P_i^{\text{max}}), (E_i^{\text{b}} - h_i x)/F_i(L/x) \}$, and ii) there at least exists another MD i' (with $i < i' \leq N$) whose $z_{i'}^{\text{bot}}(x) > 0$. Then, we can further reduce $H(\{z_i\}_{i \in \mathcal{I}})$ by replacing $z_i^{\text{bot}}(x)$ and $z_{i'}^{\text{bot}}(x)$ with $z_i^{\text{bot}}(x) + \epsilon$ and $z_{i'}^{\text{bot}}(x) - \epsilon$ (where ϵ is a very small positive number), respectively. Such an operation will not violate any constraint in Problem (P1-Bottom), which thus leads to a contradiction that $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ is optimal.
- *Case II:* suppose that i) there exists an MD i (with $i < N$), whose $z_i^{\text{bot}}(x) < \min \{ \eta_i \mathbb{I}(F_i(L/x) \leq P_i^{\text{max}}), (E_i^{\text{b}} - h_i x)/F_i(L/x) \}$, and ii) for each MD i' with $i < i' \leq N$, there exists $z_{i'}^{\text{bot}}(x) = 0$. Then, we can further reduce $H(\{z_i\}_{i \in \mathcal{I}})$ by directly replacing $z_i^{\text{bot}}(x)$ with $z_i^{\text{bot}}(x) + \epsilon$ (where ϵ is a very small positive number). Such an operation will not violate any constraint in Problem (P1-Bottom), which thus leads to a contradiction that $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ is optimal.
- *Case III:* if MD N 's $z_N^{\text{bot}}(x) < \min \{ \eta_N \mathbb{I}(F_N(L/x) \leq P_N^{\text{max}}), (E_N^{\text{b}} - h_N x)/F_N(L/x) \}$, then we can also reduce $H(\{z_i\}_{i \in \mathcal{I}})$ by directly replacing $z_N^{\text{bot}}(x)$ with $z_N^{\text{bot}}(x) + \epsilon$ (where ϵ is a very small positive number). Such an operation will not violate any constraint in Problem (P1-Bottom), which thus leads to a contradiction that $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ is optimal.

Through examining the above three cases of contradictions, we have finished the proof of Proposition 1. \square

Proposition 1 leads to the following corollary.

Corollary 1: Let MD i^{w} denote the MD with the worst channel power gain from the BS, i.e., $i^{\text{w}} = \arg \min_{i \in \mathcal{I}} \{g_{Bi}\}$. Then, $z_{i^{\text{w}}}^{\text{bot}}(x) = 0$ always holds.

Proof: According to (17), MD i^{w} always has its $M_{i^{\text{w}}} > 0$, which yields the above result. \square

Proposition 1 means that we do not need to consider those MDs with $N+1 \leq i \leq I$ when deriving the optimal transmission duration of the content, since they are *unhelpful for relaying the content*. Thus, we define the following subset of the MDs, denoted by $\tilde{\mathcal{I}}$, as follows:

$$\tilde{\mathcal{I}} = \{i | i = 1, 2, \dots, N\}. \quad (22)$$

Notice that $\tilde{\mathcal{I}} = \emptyset$ if $N = 0$. Specifically, the MDs in $\tilde{\mathcal{I}}$ are *potentially helpful in terms of relaying the content*. Nevertheless, we emphasize that at the optimal solution of Problem (P1), not necessarily all MDs in $\tilde{\mathcal{I}}$ will be selected for relaying.

Using Proposition 1, we can express the optimal objective value of Problem (P1-Bottom) as follows:

$$O_{\text{bot}}(x) = \sum_{i \in \tilde{\mathcal{I}}} M_i z_i^{\text{bot}}(x) (2^{\frac{L}{x}} - 1), \quad (23)$$

which will be used in the next section to solve Problem (P1-Top).

6 OPTIMAL SOLUTION OF PROBLEM (P1-TOP)

Using $O_{\text{bot}}(x)$, we continue to solve Problem (P1-Top) (repeated below) to determine the optimal transmission duration x^* .

$$\begin{aligned} \text{(P1-Top): } \max O_{\text{bot}}(x) + \alpha \left((2^{\frac{L}{x}} - 1) \frac{n}{\min_{i \in \mathcal{I}} \{g_{Bi}\}} + q_B \right) x \\ + x \sum_{i \in \mathcal{I}} \beta_i h_i + \gamma x \\ \text{subject to: } 0 \leq x \leq X^{\text{up}} = \left\{ \min_{i \in \mathcal{I}} \left\{ \frac{E_i^{\text{b}}}{h_i} \right\}, T^{\text{max}} \right\}. \end{aligned}$$

Deriving x^* analytically is difficult, since x influences $\{z_i^{\text{bot}}(x)\}$ in (19), and consequently influences $O_{\text{bot}}(x)$ in (23) in a complicated fashion. In particular, the transmission duration x influences $\{z_i^{\text{bot}}(x)\}$ in the following two aspects. *First, x influences whether MD i can relay or not according to whether the constraint that $F_i(\frac{L}{x}) \leq P_i^{\text{max}}$ is met or not.* For instance, suppose that x is so small such that the transmit-power required by MD i for relaying exceeds P_i^{max} . Then, MD i cannot be selected for relaying. *Second, x influences how long MD i can be selected for relaying (according to the constraint that $z_i F_i(\frac{L}{x}) \leq E_i^{\text{b}} - x h_i$).* Specifically, the smaller x , the larger transmit-power required by MD i for relaying, and thus the smaller relay-duration for MD i .

To derive x^* analytically, we characterize different subregions for x , such that we can obtain the analytical form of $O_{\text{bot}}(x)$. The key idea of characterizing different subregions is that we will further identify those MDs (in set $\tilde{\mathcal{I}}$, i.e., the set of potential helpful MDs for performing relaying) that are *not eligible* to perform relaying under a given value of x , due to their limited transmit-powers. The details are shown in the next subsections.

6.1 Characterizing Different Subregions for Variable x Thresholds for excluding MDs not eligible for relaying data:

We first consider the following threshold

$$\Gamma_i = \frac{L}{\log_2 \left(1 + \frac{P_i^{\text{max}} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}}{n} \right)}, \quad 1 \leq i \leq N, \quad (24)$$

regarding whether MD i is an *eligible candidate* to perform relaying or not. Specifically, MD i is eligible for relaying, only if the transmission duration x satisfies $x \geq \Gamma_i$ (which leads to $F_i(L/x) \leq P_i^{\text{max}}$). Otherwise, MD i is not eligible for relaying. Thus, starting from $x = \min \{ \min_{i \in \mathcal{I}} \{E_i^{\text{b}}/h_i\}, T^{\text{max}} \}$,

a decrease in x means that the transmission rate of the content increases, and less MDs are eligible for performing relaying.

Based on the above consideration, we further re-order the thresholds defined in (24) in an ascending order as follows:

$$\Gamma_1 \leq \Gamma_2 \leq \dots \leq \Gamma_{l-1} \leq \Gamma_l \leq \Gamma_{l+1} \leq \dots \leq \Gamma_N. \quad (25)$$

Different from (24), we now use subscript l as the index for thresholds $\{\Gamma_l\}_{1 \leq l \leq N}$ that follow the ordering in (25), i.e., $\Gamma_l \leq \Gamma_{l+1}$ always holds.

Moreover, given the index l of threshold Γ_l following (25), we define a mapping $T(l)$ to find the index of the MD that yields threshold Γ_l according to (24), i.e.,⁷

$$T(l) = \left\{ s \in \tilde{\mathcal{I}} \mid \frac{\min_{i' \neq s, i' \in \mathcal{I}} \{g_{si'}\} P_s^{\max}}{n} = 2^{\frac{l}{\Gamma_l}} - 1 \right\}. \quad (26)$$

Recall that due to the reordering in (25), Γ_l does not correspond to MD l , and that's why we need to define the mapping $T(l)$ in (26). To clarify the ambiguity, we emphasize that in the rest of this paper, subscript l is solely used as the index for thresholds $\{\Gamma_l\}_{1 \leq l \leq N}$.

Remark 1: (Illustration of the effect of ordering (25)): If $x > \Gamma_N$, then all MDs in $\tilde{\mathcal{I}}$ are eligible for performing relaying. On the other hand, if $x < \Gamma_1$, then none of the MDs is eligible (which results in that the BS needs to broadcast to all MDs directly). If x meets $\Gamma_l \leq x < \Gamma_{l+1}$, then we can define a set \mathcal{J}_l as follows:

$$\mathcal{J}_l = \{T(l+1), T(l+2), \dots, T(N)\}, \quad (27)$$

and each MD $i \in \mathcal{J}_l$ is not eligible for performing relaying (Figure 2 plots an example to show this point). Notice that $\mathcal{J}_l \subseteq \tilde{\mathcal{I}}$ always holds for $1 \leq l \leq N$. Thus, by considering different subregions of x in terms of $[\Gamma_l, \Gamma_{l+1}]$, $l = 0, 1, 2, \dots, N$, we can differentiate the influence of x on constraint (14). To facilitate discussions, we introduce $\Gamma_0 = 0$ and Γ_{N+1} to be a sufficiently large value.

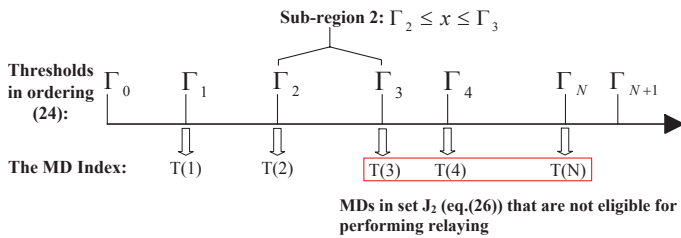


Fig. 2: Relationship between threshold Γ_l and MD $T(l)$. We consider the case of $\Gamma_2 \leq x \leq \Gamma_3$ as an example, and illustrate set $\mathcal{J}_2 = \{T(3), T(4), \dots, T(N)\}$ which are not eligible for performing relaying.

Special MD on the “Boundary”: Even if we focus on the subregion of $x \in [\Gamma_l, \Gamma_{l+1}]$, we still cannot analytically express the optimal solution of Problem (P1-Bottom) $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$. We need to further consider the influence of x in constraint (15). The key step is to characterize a special MD (let us say MD v), such that the energy budget E_v^b of MD v is not used up, while the available relay-duration η_{v+1} given by (20) is zero. In particular, if such an MD v exists, then the BS does not need to perform broadcasting, since the MDs’ capabilities have not been fully utilized. Otherwise, the BS needs to perform broadcasting

7. Given index l , if there exist several different MDs in $\tilde{\mathcal{I}}$ such that condition in (26) is met, then $T(l)$ just represents such a set of MDs. This will not influence our following analytical results and the proposed algorithm. However, because of the randomness in different MDs’ locations and the fading effect of the channel power gains, such case rarely happens.

to finish delivering the whole content. Therefore, we consider the following two different types of cases: i) Type-I cases in which the BS does not need to perform broadcasting, and ii) Type-II cases in which the BS needs to perform broadcasting. The details are as follows.

Type-I cases that do not require the BS to perform broadcasting:

The common property of the Type-I cases is as follows. Given x in the subregion $[\Gamma_l, \Gamma_{l+1}]$, there always exists a special MD $v \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$, such that the energy budget E_v^b of MD v is not used up, while the available relay-duration η_{v+1} given by (20) is zero. We denote this case by case (l, v) , whose definition is as follows.

Definition 1: (Case (l, v)): Given that x in the subregion $[\Gamma_l, \Gamma_{l+1}]$, MD $v \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$ has its energy budget not used up, i.e., $E_v^b - E_v(x, z_v^{\text{bot}}(x)) > 0$ (where $E_v(x, z_v^{\text{bot}}(x))$ is given in (6)), while the available relay-duration η_{v+1} given by (20) is zero, i.e., $x = \sum_{i' \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} z_{i'}^{\text{bot}}(x)$.

There exist at most $N(N+1)/2$ such cases of Type-I. Given case (l, v) , the optimal solution of Problem (P1-Bottom), which is given in Proposition 1 before, can be further detailed as follows.

Proposition 2: Given x and case (l, v) , the optimal solution of Problem (P1-Bottom) can be given by:

$$z_i^{\text{bot}}(x) = \frac{E_i^b - h_i x}{(2^{\frac{l}{x}} - 1) \frac{n}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}}}, \quad \text{when } 1 \leq i \leq v-1, \text{ and } i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l, \quad (28)$$

$$z_v^{\text{bot}}(x) = x - \sum_{s=1, s \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^{v-1} z_s^{\text{bot}}(x), \quad (29)$$

$$z_i^{\text{bot}}(x) = 0, \text{ when } v < i \leq I, \text{ or } i \in \mathcal{J}_l. \quad (30)$$

Proof: This proof is based on Proposition 1 and the definition of case (l, v) (i.e., Definition 1). According to Definition 1, for each MD $i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$ and $1 \leq i \leq v-1$, constraint (15) should be binding, which leads to (28). Besides, for MD v , constraint (13) should be binding, which leads to (29). Finally, (30) holds, because of the following three points: i) for each MD i with $N < i \leq I$, $z_i^{\text{bot}}(x) = 0$ holds based on Proposition 1, ii) for each MD i with $v < i \leq N$, $z_i^{\text{bot}}(x) = 0$ holds because of $\eta_i = 0$, and iii) for each MD $i \in \mathcal{J}_l$, $z_i^{\text{bot}}(x) = 0$ holds because of constraint (14). \square

Type-II cases that require the BS to perform broadcasting:

The common property of the Type-II cases is as follows. Given that x in the subregion $[\Gamma_l, \Gamma_{l+1}]$, each MD $i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$ has used up its energy budget, while there still exists a nonzero available relay-duration, i.e., $\sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} z_i^{\text{bot}}(x) < x$. This means that the BS needs to perform broadcasting to finish delivering the content. We denote this case by case (l, B) (where the capital letter “B” represents the BS), and its definition is as follows.

Definition 2: (Case (l, B)): Given x in subregion $[\Gamma_l, \Gamma_{l+1}]$, each MD $i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$ has used up its energy budget, i.e., $E_i(x, z_i^{\text{bot}}(x)) = E_i^b$, and the BS still needs to perform broadcasting to finish delivering the content, i.e., $x - \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} z_i^{\text{bot}}(x) > 0$.

There exist $N+1$ such cases of Type-II. Given case (l, B) , the optimal solution of Problem (P1-Bottom), which is given in Proposition 1 before, can be further detailed as follows.

Proposition 3: Given x and case (l, B) , the optimal solution of Problem (P1-Bottom) can be given by:

$$z_i^{\text{bot}}(x) = \frac{E_i^b - h_i x}{(2^{\frac{l}{x}} - 1) \frac{n}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}}}, \text{ when } i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l \quad (31)$$

$$z_i^{\text{bot}}(x) = 0, \text{ when } i \in \mathcal{I} \setminus \tilde{\mathcal{I}} \text{ or } i \in \mathcal{J}_l. \quad (32)$$

Correspondingly, in order to finish delivering the whole content, the BS broadcasts for the duration which is equal to

$$x - \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} \frac{E_i^b - h_i x}{(2^{\frac{x}{x}} - 1) \frac{n}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}}}. \quad (33)$$

Proof: This proof is based on Proposition 1 and the definition of case (l, B) (i.e., Definition 2). According to Definition 2, for each MD $i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l$, constraint (15) should be binding, which leads to (31). Meanwhile, for each MD $i \in \mathcal{J}_l$, $z_i^{\text{bot}}(x) = 0$ holds because of constraint (14), and for each MD i with $N < i \leq I$, $z_i^{\text{bot}}(x) = 0$ holds based on Proposition 1, which together lead to (32). Finally, the broadcasting duration of the BS (given in (33)) stems from (13) and Definition 2. \square

Until now, under a given x , we have analytically derived the optimal solution in (28)-(30) of Problem (P1-Bottom) in Proposition 2 by supposing that case (l, v) (of Type-I) holds. Meanwhile, we also derive the corresponding optimal solution in (31)-(33) of (P1-Bottom) in Proposition 3 by supposing that case (l, B) (of Type-II) holds. As a result, the optimal value of Problem (P1-Bottom), i.e., $O_{\text{bot}}(x)$ in (23), can be analytically detailed. We thus continue to solve Problem (P1-Top) in the next two subsections, in which we will also provide the conditions to verify whether case (l, v) (or case (l, B)) holds or not.

6.2 Analytical Solution for Each Case (l, v)

Given x and case (l, v) , we introduce function $W_{l,v}(x)$ to denote the objective function of Problem (P1-Top) under case (l, v) . By substituting (28) and (29) into (23), we can compactly express function $W_{l,v}(x)$ as follows:

$$W_{l,v}(x) = (2^{\frac{x}{x}} - 1)x \left(\alpha \frac{n}{g_{Bv}} + \beta_v \frac{n}{\min_{i \neq v, i \in \mathcal{I}} \{g_{vi}\}} \right) + x \left(\alpha q_B + \sum_{i \in \mathcal{I}} \beta_i h_i + \gamma + S_{l,v} \right) + Q_{l,v}, \quad (34)$$

where both $S_{l,v}$ and $Q_{l,v}$ are constant and depend on case (l, v) :

$$S_{l,v} = \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^{v-1} \left(\alpha \left(\frac{1}{g_{Bv}} - \frac{1}{g_{Bi}} \right) + \beta_v \frac{1}{\min_{i' \neq v, i' \in \mathcal{I}} \{g_{vi'}\}} - \beta_i \frac{1}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}} \right) h_i \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}, \quad (35)$$

$$Q_{l,v} = \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^{v-1} \left(\alpha \left(\frac{1}{g_{Bi}} - \frac{1}{g_{Bv}} \right) + \beta_i \frac{1}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}} - \beta_v \frac{1}{\min_{i' \neq v, i' \in \mathcal{I}} \{g_{vi'}\}} \right) E_i^b \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}. \quad (36)$$

Therefore, given case (l, v) , solving Problem (P1-Top) becomes equivalent to solving

$$\begin{aligned} \text{(P1-Top-(l,v))} \quad & \min_x W_{l,v}(x), \\ & \text{subject to: } \Gamma_l \leq x \leq \min \{X^{\text{up}}, \Gamma_{l+1}\}, \end{aligned}$$

where X^{up} has been defined in (16).

Let $x_{l,v}^*$ denote the optimal solution of Problem (P1-Top-(l,v)). Although $W_{l,v}(x)$ is complicated, we can analytically derive $x_{l,v}^*$ in the following proposition.

Proposition 4: Given case (l, v) , the optimal solution for Problem (P1-Top-(l,v)) can be given by:

$$x_{l,v}^* = \left[(\ln 2) \frac{L}{1 + \mathcal{W}(e^{-1} \frac{B_{l,v}}{A_{l,v}} - 1))} \right]_{\Gamma_l}^{\min \{X^{\text{up}}, \Gamma_{l+1}\}}, \quad (37)$$

where expression $[x]_a^b = \min \{ \max \{ a, x \}, b \}$, and $\mathcal{W}(\cdot)$ represents the Lambert W-function [40], i.e., the inverse function of $f(w) = w \exp(w)$. Meanwhile, parameters $A_{l,v}$ and $B_{l,v}$ are given by:

$$A_{l,v} = \alpha \frac{n}{g_{Bv}} + \beta_v \frac{n}{\min_{i \neq v, i \in \mathcal{I}} \{g_{vi}\}}, \quad (38)$$

$$B_{l,v} = \alpha q_B + \sum_{i \in \mathcal{I}} \beta_i h_i + \gamma + S_{l,v}. \quad (39)$$

Accordingly, $W_{l,v}^*(x_{l,v}^*) = A_{l,v} x_{l,v}^* (2^{\frac{x_{l,v}^*}{x_{l,v}^*}} - 1) + B_{l,v} x_{l,v}^* + Q_{l,v}$.

Proof: Before presenting the proof, we first give the following Lemma, which will be used for proving Proposition 4 later on.

Lemma 1: The following two results always hold:

- i) For each case (l, v) , we always have $S_{l,v} > 0$ and $Q_{l,v} < 0$, where $S_{l,v}$ and $Q_{l,v}$ are given in (35) and (36), respectively.
- ii) For each case (l, B) , we always have $S_{l,B} > 0$ and $Q_{l,B} < 0$, where $S_{l,B}$ and $Q_{l,B}$ are given in (46) and (47), respectively.

Proof of Lemma 1: We first prove result i). Based on the ordering in (18), $M_j \leq M_v < 0$ holds for $j < v < N$. By further using the definition of $\{M_i\}$ in (17), we have

$$\alpha \left(\frac{1}{g_{Bv}} - \frac{1}{g_{Bj}} \right) + \beta_v \frac{1}{\min_{i \neq v, i \in \mathcal{I}} \{g_{vi}\}} - \beta_j \frac{1}{\min_{i \neq j, i \in \mathcal{I}} \{g_{ji}\}} > 0, \quad \text{for } j < v < N.$$

As a result, $S_{l,v} > 0$ always holds. Similarly, we can show that $Q_{l,v} < 0$ always holds. We next prove result ii). According to ordering in (18), $M_j < 0$ holds for $j \leq N$. By further using the definition of $\{M_i\}$ in (17), we have $S_{l,B} > 0$ and $Q_{l,B} < 0$. \square

Now, we start to present the proof for Proposition 4. Under case (l, v) , $Q_{l,v}$ (defined in (36)) in $W_{l,v}(x)$ (defined (34)) is independent on x . Thus, solving Problem (P1-Top-(l,v)) is equivalent to solving the following Problem (P2):

$$\begin{aligned} \text{(P2):} \quad & \min_x V_{l,v}(x) = \min_x A_{l,v} (2^{\frac{x}{x}} - 1)x + B_{l,v} x, \\ & \text{subject to: } \Gamma_l \leq x \leq \min \{X^{\text{up}}, \Gamma_{l+1}\}. \end{aligned}$$

Notice that parameters $A_{l,v}$ and $B_{l,v}$ (defined in (38) and (39), respectively) are both positive constants, since $S_{l,v}$ (defined in (35)) is positive according to Lemma 1.

In particular, we can show that Problem (P2) is a *convex optimization problem* [39], since the second-order derivative of the objective function is always positive, i.e.,

$$\frac{d^2 V_{l,v}^2(x)}{dx^2} = \frac{L^2}{x^3} 2^{\frac{x}{x}} (\ln 2)^2 \geq 0, \quad \forall x \geq 0, \quad (40)$$

and the constraint in Problem (P2) is linear.

The convexity of Problem (P2) enables us to use the necessary and sufficient Karush-Kuhn-Tucker (KKT) condition to compute its optimal solution. By setting $\frac{dV_{l,v}(x)}{dx} = 0$, we obtain

$$\frac{dV_{l,v}(x)}{dx} = G_{l,v}(x) + \frac{B_{l,v}}{A_{l,v}} = 0, \quad (41)$$

where the auxiliary function $G_{l,v}(x)$ is defined as follows:

$$G_{l,v}(x) = \left(2^{\frac{x}{x}} - 1 \right) - \frac{L}{x} 2^{\frac{x}{x}} (\ln 2).$$

Notice that $G_{l,v}(x)$ is increasing in x , since (40) holds. Moreover, $G_{l,v}(x) < 0$ holds, since $\lim_{x \rightarrow \infty} G_{l,v}(x) = 0$. Therefore, the root of (41), if it exists, is unique. Moreover, if the root of (41) exists, then it corresponds to the optimal solution of Problem (P2). Although (41) is complicated, its root can be derived analytically. Specifically, (41) is equivalent to:

$$\left(1 - \frac{L}{x} \ln 2\right) 2^{\frac{L}{x}} = 1 - \frac{B_{l,v}}{A_{l,v}}.$$

By defining $y = 1 - \frac{L}{x} \ln 2$ and substituting x by the newly introduced variable y , we obtain

$$y = -\mathcal{W}\left(\frac{1}{e}\left(\frac{B_{l,v}}{A_{l,v}} - 1\right)\right),$$

where $\mathcal{W}(\cdot)$ denotes the Lambert W-function [40], which corresponds to the inverse function of $f(w) = w \exp(w)$. Consequently, we can obtain

$$x_{l,v}^* = (\ln 2) \frac{L}{1 + \mathcal{W}\left(\frac{1}{e}\left(\frac{B_{l,v}}{A_{l,v}} - 1\right)\right)}. \quad (42)$$

In addition, $x_{l,v}^*$ should be lower bounded by Γ_l and be upper bounded by $\min\{X^{\text{up}}, \Gamma_{l+1}\}$ (as required by case (l, v)). We thus need to consider the following two cases:

(Case i): If $x_{l,v}^*$ given in (42) is smaller than the lower bound Γ_l , then $\frac{dV_{l,v}}{dx}$ is positive when $x \in [\Gamma_l, \min\{X^{\text{up}}, \Gamma_{l+1}\}]$ (recall that the convexity of Problem (P2) implies that its first-order derivative $\frac{dV_{l,v}}{dx}$ is strictly increasing). Thus, to minimize the objective function, x should be set as the lower bound Γ_l .

(Case ii): If $x_{l,v}^*$ given in (42) is larger than the upper bound $\min\{X^{\text{up}}, \Gamma_{l+1}\}$, then $\frac{dV_{l,v}}{dx}$ is negative when $x \in [\Gamma_l, \min\{X^{\text{up}}, \Gamma_{l+1}\}]$. Thus, to minimize the objective function, x should be set as the upper bound $\min\{X^{\text{up}}, \Gamma_{l+1}\}$.

In summary, given case (l, v) , we obtain the optimal transmission duration $x_{l,v}^*$ in (37) for Problem (P1-Top-(l,v)). Further by using $x_{l,v}^*$, we can obtain $W_{l,v}(x_{l,v}^*) = V_{l,v}(x_{l,v}^*) + Q_{l,v}$ for case (l, v) . This finishes the proof of Proposition 4. \square

The result in Proposition 4 is based on the assumption that case (l, v) holds. We thus need to use the derived $x_{l,v}^*$ in Proposition 4 to verify whether case (l, v) holds or not. This leads to the following proposition.

Proposition 5: (Validation of case (l, v)): Case (l, v) holds, if the derived $x_{l,v}^*$ in (37) meets the following two conditions:

$$x_{l,v}^* (2^{\frac{L}{x_{l,v}^*}} - 1) + x_{l,v}^* \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^{v-1} \frac{h_i}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\} \geq \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^{v-1} \frac{E_i^b}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}, \quad (43)$$

$$x_{l,v}^* (2^{\frac{L}{x_{l,v}^*}} - 1) + x_{l,v}^* \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^v \frac{h_i}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\} \leq \sum_{i=1, i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l}^v \frac{E_i^b}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}. \quad (44)$$

Proof: According to the definition of case (l, v) , MD v is the ‘‘boundary’’ MD to perform relaying in set $\tilde{\mathcal{I}}$. Therefore, i) condition $x_{l,v}^* - \sum_{i=1}^{v-1} z_i^{\text{bot}}(x_{l,v}^*) > 0$ should hold, which yields (43) based on (28)-(30), and ii) condition $x_{l,v}^* - \sum_{i=1}^{v-1} z_i^{\text{bot}}(x_{l,v}^*) < (E_v^b - h_v x_{l,v}^*) (F_v(\frac{L}{x_{l,v}^*}))^{-1}$ should hold, which yields (44). \square

6.3 Analytical Solution for Each Case (l, B)

Given x and case (l, B) , we introduce $W_{l,B}(x)$ to denote the objective function of Problem (P1-Top). By substituting (31) and (33) into (23), we can compactly express function $W_{l,B}(x)$ as follows:

$$W_{l,B}(x) = (2^{\frac{L}{x}} - 1) x \alpha \frac{n}{\min_{i \in \mathcal{I}} \{g_{Bi}\}} + x(\alpha q_B + \sum_{i \in \mathcal{I}} \beta_i h_i + \gamma + S_{l,B}) + Q_{l,B}, \quad (45)$$

where both $S_{l,B}$ and $Q_{l,B}$ are constant and depend on index l :

$$S_{l,B} = - \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} \left(\alpha \frac{1}{g_{Bi}} + \beta_i \frac{1}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}} - \alpha \frac{1}{\min_{i' \in \mathcal{I}} \{g_{Bi'}\}} \right) h_i \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}, \quad (46)$$

$$Q_{l,B} = \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} \left(\alpha \frac{1}{g_{Bi}} + \beta_i \frac{1}{\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}} - \alpha \frac{1}{\min_{i' \in \mathcal{I}} \{g_{Bi'}\}} \right) E_i^b \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}. \quad (47)$$

Given case (l, B) , solving Problem (P1-Top) becomes equivalent to solving

$$\text{(P1-Top-(l,B))} \quad \min_x W_{l,B}(x),$$

subject to: $\Gamma_l \leq x \leq \min\{X^{\text{up}}, \Gamma_{l+1}\}$.

Let $x_{l,B}^*$ denote the optimal solution of Problem (P1-Top-(l,B)). We can analytically derive $x_{l,B}^*$ in the following proposition.

Proposition 6: Given case (l, B) , the optimal solution of Problem (P1-Top-(l,B)) can be given by:

$$x_{l,B}^* = \left[(\ln 2) \frac{L}{1 + \mathcal{W}(e^{-1}(\frac{B_{l,B}}{A_{l,B}} - 1))} \right]_{\Gamma_l}^{\min\{X^{\text{up}}, \Gamma_{l+1}\}}, \quad (48)$$

where parameters $A_{l,B}$ and $B_{l,B}$ are respectively given by:

$$A_{l,B} = \alpha \frac{n}{\min_{i \in \mathcal{I}} \{g_{Bi}\}}, \quad (49)$$

$$B_{l,B} = \alpha q_B + \sum_{i \in \mathcal{I}} \beta_i h_i + \gamma + S_{l,B}. \quad (50)$$

Accordingly, $W_{l,B}^*(x_{l,B}^*) = A_{l,B} x_{l,B}^* (2^{\frac{L}{x_{l,B}^*}} - 1) + B_{l,B} x_{l,B}^* + Q_{l,B}$.

Proof: The proof is similar as that for proving Proposition 4, in which we just need to use $A_{l,B}$ and $B_{l,B}$ to replace $A_{l,v}$ and $B_{l,v}$, respectively. Recall that both $A_{l,B}$ and $B_{l,B}$ are positive according to Lemma 1. \square

The result in Proposition 6 is based on the assumption that case (l, B) holds. We thus need to use the derived $x_{l,B}^*$ in Proposition 6 to verify whether case (l, B) holds or not. This leads to the following Proposition.

Proposition 7: (Validation of case (l, B)): Case (l, B) holds, if the derived $x_{l,B}^*$ in (48) meets the following conditions:

$$x_{l,B}^* (2^{\frac{L}{x_{l,B}^*}} - 1) + x_{l,B}^* \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} \frac{h_i}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\} \geq \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{J}_l} \frac{E_i^b}{n} \min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}. \quad (51)$$

Proof: According to the definition of case (l, B) , the BS needs to perform broadcasting for finishing the content delivery. Thus, $x_{l,B}^* > \sum_{i \in \tilde{\mathcal{I}} \setminus \mathcal{I}_l} z_i^{\text{bot}}(x_{l,B}^*)$ should hold, which yields (51) based on (31)-(33). \square

6.4 Efficient Algorithm for Finding the Global Optimum Solution of Problem (P1-Top)

Using the above analytical results, we propose a Joint Optimization of Transmission and Relay Durations (JOTRD) algorithm to solve Problem (P1-Top) and compute the optimal transmission duration for the whole content and each MD's relay-duration.

Algorithm JOTRD: to find the optimal solution of Problem (P1)

```

1: Initialize  $\phi$  as a very large positive number, e.g.,  $\phi = 10^8$ .
2: Set  $l = N$ , where  $N$  is obtained from the ordering (18).
3: while  $l \geq 0$  do
4:   Set  $v = 1$ .
5:   while  $v \leq N$  and  $v \notin \mathcal{I}_l$  do
6:     Derive  $x_{l,v}^*$  according to (37).
7:     if  $x_{l,v}^*$  meets (43) and (44) simultaneously then
8:       Evaluate  $W_{l,v}(x_{l,v}^*)$  according to (34).
9:       if  $W_{l,v}(x_{l,v}^*) < \phi$  then
10:        Derive  $\{z_j^{\text{bot}}(x_{l,v}^*)\}_{j \in \mathcal{I}}$  according to (28)-(30).
11:        Update  $\phi = W_{l,v}(x_{l,v}^*)$ , and record the currently
            best solution of Problem (P1) as:  $x^{*,c} = x_{l,v}^*$ ,  $r^{*,c} =$ 
             $\frac{L}{x_{l,v}^*}$ , and  $z_j^{*,c} = z_j^{\text{bot}}(x_{l,v}^*), \forall j \in \mathcal{I}$ 
12:        end if
13:      end if
14:      Set  $v = v + 1$ .
15:    end while
16:    Derive  $x_{l,B}^*$  according to (48).
17:    if  $x_{l,B}^*$  meets (51) then
18:      Evaluate  $W_{l,B}(x_{l,B}^*)$  according to (45).
19:      if  $W_{l,B}(x_{l,B}^*) < \phi$  then
20:        Derive  $\{z_j^{\text{bot}}(x_{l,B}^*)\}_{j \in \mathcal{I}}$  according to (31) and (32).
21:        Update  $\phi = W_{l,B}(x_{l,B}^*)$ , and record the currently best
            solution of Problem (P1) as:  $x^{*,c} = x_{l,B}^*$ ,  $r^{*,c} = \frac{L}{x_{l,B}^*}$ ,
            and  $z_j^{*,c} = z_j^{\text{bot}}(x_{l,B}^*), \forall j \in \mathcal{I}$ .
22:        end if
23:      end if
24:      Set  $l = l - 1$ .
25:    end while
26: Set the optimal solution of Problem (P1) as:  $x^* = x^{*,c}$ ,  $r^* =$ 
     $r^{*,c}$ , and  $z_j^* = z_j^{*,c}, \forall j \in \mathcal{I}$ .

```

Algorithm JOTRD enumerates all Type-I cases and Type-II cases. To this end, it consists of a two-layered loop, i.e., i) an *outer While-Loop* from Step 3 to Step 25 for enumerating all possible index l , and ii) given index l , an *inner While-Loop* from Step 5 to Step 15 for enumerating index v such that each possible case (l, v) is evaluated, and moreover, the *additional steps* from Step 16 to Step 23 for evaluating case (l, B) . Specifically, for each enumerated case (l, v) , we derive $x_{l,v}^*$ based on Proposition 4 in Step 6, and further verify whether case (l, v) holds in Step 7. If case (l, v) is valid and the obtained $W_{l,v}(x_{l,v}^*)$ can improve the currently best value ϕ , then we update ϕ and record the currently best solution of Problem (P1) in Step 11. Similarly, for

each enumerated case (l, B) , we derive $x_{l,B}^*$ based on Proposition 6 in Step 16, and further verify whether case (l, B) holds in Step 17. If case (l, B) is valid and the obtained $W_{l,B}(x_{l,B}^*)$ can improve the currently best value ϕ , then we update ϕ and record the currently best solution of Problem (P1) in Step 21. Finally, Algorithm JOTRD outputs the optimal solution of Problem (P1) in Step 26 based on the currently best solution⁸.

Proposition 8: Algorithm JOTRD is guaranteed to find the optimal solution of Problem (P1).

Proof: Notice that Algorithm JOTRD is designed to enumerate all possible Type-I cases and Type-II cases. Specifically, for each enumerated Type-I case (l, v) , Proposition 4 and Proposition 2 together give the unique optimal solution of Problem (P1). Meanwhile, for each enumerated Type-II case (l, B) , Proposition 6 and Proposition 3 together give the unique optimal solution of Problem (P1). Furthermore, based on Definition 1, there exist $N(N+1)/2$ different Type-I cases (where N is specified in the ordering (18)), and based on Definition 2, there exist $N+1$ different Type-II cases. Therefore, by enumerating and comparing with all these $(N+1)(N+2)/2$ cases, Algorithm JOTRD is guaranteed to find the optimal solution of Problem (P1). \square

The complexity of Algorithm JOTRD is analyzed as follows. *First*, Algorithm JOTRD requires a total of $O((N+1)(N+2)/2)$ rounds of iterations, because there exist $N(N+1)/2$ different Type-I cases and $(N+1)$ different Type-II cases. Recall that the value of N , which denotes the number of helpful MDs according to the ordering (18), is always smaller than I , i.e., the total number of the MDs. *Second*, within each iteration, for case (l, v) enumerated, Proposition 4 and Proposition 2 together give the optimal solutions x^* and $\{z_j^{\text{bot}}(x)\}_{j \in \mathcal{I}}$ analytically. Meanwhile, for case (l, B) enumerated, Proposition 6 and Proposition 3 together give the optimal solutions x^* and $\{z_j^{\text{bot}}(x)\}_{j \in \mathcal{I}}$ analytically. Thus, no additional iterative calculation is required within each round of iteration. In summary, Algorithm JOTRD is computationally efficient and is easy to be implemented at the BS.

Although the channel power gain information is required by the BS to perform Algorithm JOTRD, we notice that there is no need for the BS to collect all detailed information about each pair of two MDs. Instead, each MD i only needs to report the BS its worst channel gain involved to perform broadcasting to the other MDs (i.e., $\min_{i' \neq i, i' \in \mathcal{I}} \{g_{ii'}\}$), which can be estimated by MD i itself via the state-of-art channel estimation techniques.

7 NUMERICAL RESULTS

7.1 Setup of the Network Scenario

In this section, we perform numerical simulations to validate Algorithm JOTRD and the performance achieved by the MDs' optimal cooperations. We setup a scenario as shown in Figure 3, in which the BS is located at the origin $(0, 0)$. The group of MDs are randomly and independently located (according to a uniform distribution) within a circle. The central of the circle is $(D, 0)$, and its radius is R . We set $D = 50\text{m}$ and $R = 5\text{m}$ at the beginning (but will vary D and R later on). In particular, we assume that the MDs do not move during the period of interest, e.g., one period of T^{max}

⁸ We emphasize that the proposed two-step backward induction, in which we first derive the optimal $\{z_i^{\text{bot}}(x)\}_{i \in \mathcal{I}}$ as analytical functions of x in the bottom-problem and then optimize x in the top-problem by substituting each z_i with $z_i^{\text{bot}}(x)$, can optimally solve Problem (P1). However, backward induction with the alternative order (i.e., optimizing x first followed by $\{z_i\}_{i \in \mathcal{I}}$) fails to solve Problem (P1) optimally.

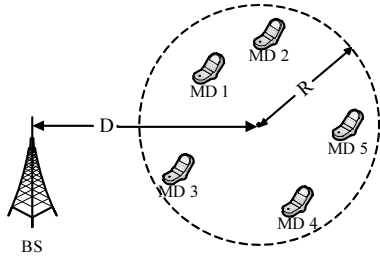


Fig. 3: Network scenario used for numerical experiments. We use $\mathcal{I} = \{1, 2, 3, 4, 5\}$ as an example. The MDs are randomly located within a circle. The central of the circle is $(D, 0)$. The radius of the circle is R .

(i.e., the delay bound for finishing delivery of the content). Thus, the channel power gain from the BS to each MD and that between the MDs remain unchanged (e.g., within one period of T^{\max}). In particular, we model the channel power gain from the BS to each MD i as $g_{Bi} = \frac{\xi_{Bi}}{l_{Bi}^\kappa}$, where parameter l_{Bi} denotes the distance between the BS and MD i , parameter κ denotes the power-scaling factor for the path-loss, and parameter ξ_{Bi} follows an exponential distribution with unit mean for capturing the fading.

Similar to [31], we set the static circuit power consumption of the BS during data transmission as $q_B = 1W$, the circuit power consumption of MD i during data reception as $h_i = 0.01W$ (i.e., 1% of the static circuit power consumption of the BS). Besides, the maximum transmit-power of each MD is $P_i^{\max} = 0.1W$, and the energy budget of each MD is $E_i^b = 0.1J$. The channel bandwidth is 1MHz for the cellular-link and the link between different MDs. The size of the content is $L = 1Mbits$. Besides, we set $\alpha = 1$, $\beta_i = 2, \forall i \in \mathcal{I}$, and $\gamma = 1$.

7.2 Performance of Algorithm JOTRD

Figure 4 validates the accuracy of Algorithm JOTRD in solving Problem (P1) optimally. We vary the number of the MDs $I = 10, 20, \dots, 50$ and the distance $D = 20, 40, 60$. For each tested case, we plot the average result (i.e., the total system cost) over 200 network scenarios which are randomly generated as described earlier. Figure 4 shows that Algorithm JOTRD achieves the optimal total system cost which is exactly same as the global optimum found by the exhaustive search method⁹, thus validating the accuracy of Algorithm JOTRD. Besides, it is observed that the total system cost increases in the number of the MDs.

Figure 5 validates the computational efficiency of Algorithm JOTRD. Specifically, we vary the distance $D = 30, 40, 50, 60$ and the number of the MDs $I = 10, 20, \dots, 50$. For each tested case, we plot the average result (i.e., the computational time) over 200 randomly generated network scenarios. Figure 5 shows that Algorithm JOTRD consumes a significantly less computational time than the exhaustive search method. Specifically, for each tested distance D , Algorithm JOTRD reduces the computational time by more than 90% on average. Furthermore, by comparing different subplots in Figure 5, we can observe that the computational time of Algorithm JOTRD increases mainly as the number of the MDs increases, but varies slightly as the distance D changes. This result is consistent with our earlier description about the computational complexity of Algorithm JOTRD (close to the end of Section 6).

9. The exhaustive search method enumerates the transmission duration x by using a very small step-size. For each enumerated x , we again use (19) to determine $\{z_j^{\text{opt}}(x)\}$ and thus evaluate $O_{\text{bot}}(x)$. Therefore, the exhaustive search method is guaranteed to achieve the global optimum for Problem (P1) with a negligible loss, as long as the chosen step-size is small enough. However, the downside of the exhaustive search method is that it consumes a significant computational time.

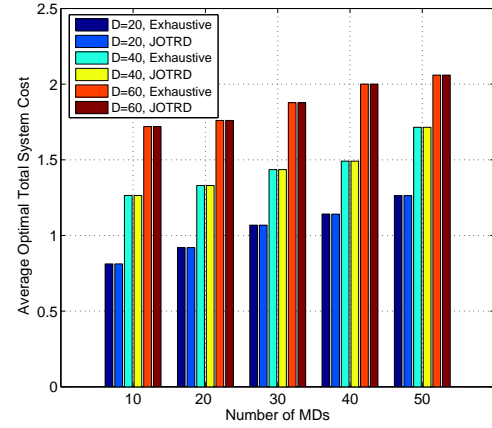


Fig. 4: Accuracy of Algorithm JOTRD in solving Problem (P1). We vary the number of the MDs $I = 10, 20, \dots, 50$ and the distance $D = 20, 40, 60$. For each tested case, the result (i.e., the total system cost) is averaged over 200 randomly generated network scenarios.

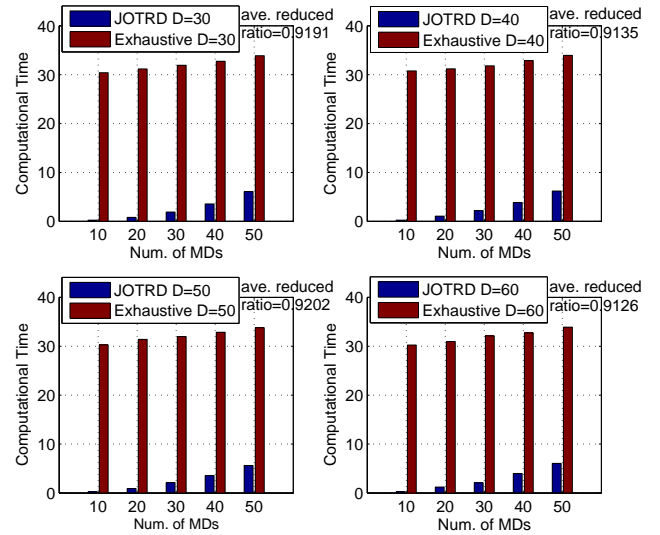


Fig. 5: Computational Efficiency of Algorithm JOTRD. We vary the distance $D = 30, 40, 50, 60$ and the number of the MDs $I = 10, 20, \dots, 50$. For each tested case, the result (i.e., the computational time) is averaged over 200 randomly generated network scenarios.

7.3 Performance Gain achieved by Cooperations

We present the advantage of reducing the total system cost by using the optimal MDs cooperation in Figures 6 and 7. To show this advantage, we compare the result of Algorithm JOTRD with those of two other heuristic approaches, namely, the *BS-only approach* and the *Equal-division approach*. In the BS-only approach, the BS directly broadcasts the whole content to all MDs and only optimizes its transmission-duration to minimize the total system cost. In the Equal-division approach, all the helpful MDs in \mathcal{I} equally share the transmission-duration for relaying the content, i.e., $z_i = x/N$, for all MDs in $\tilde{\mathcal{I}}^{10}$, and the BS optimizes its transmission-duration accordingly.

In Figure 6, we consider the distance between the BS and the central of circle $D = 50$ (in the left subplot) and $D = 60$ (in the

10. If some MD (let us say MD i) cannot afford the required transmit-power or the required relay-duration, then the BS takes over the job of MD i to deliver the content via broadcasting.

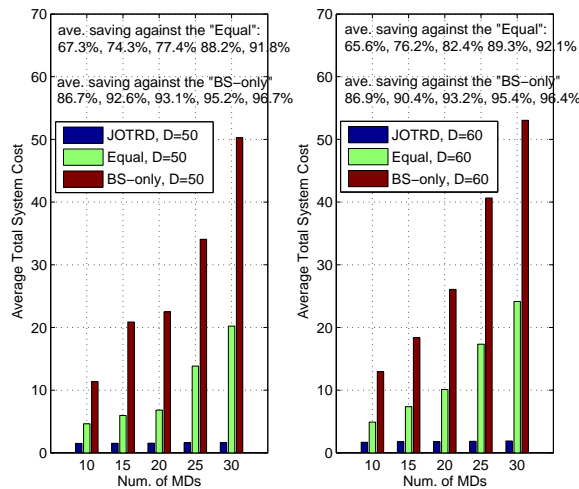


Fig. 6: Advantage of saving the total system cost by using the optimal MDs cooperations. We plot the average total system versus different number of the MDs $I = 10, 15, \dots, 30$. Left subplot: the distance between the BS and the central of the circle $D = 50$. Right subplot: $D = 60$. For each tested case, the result (i.e., the total system cost) is averaged over 200 randomly generated network scenarios. The numbers illustrated on the top of each subplot denote the average saving ratios of the total system cost by using Algorithm JOTRD against the Equal-distribution approach and the BS-only approach, respectively.

right subplot), and vary the number of the MDs $I = 10, 15, \dots, 30$. Figure 6 shows that Algorithm JOTRD significantly outperforms the BS-only approach and the Equal-distribution approach in terms of lowering the total system cost. For each tested case, we mark out the average saving ratios of the total system cost (i.e., the numbers listed on the top of subplot) by using Algorithm JOTRD against the Equal-distribution approach and the BS-only approach, respectively. Remarkably, the proposed optimal cooperation can save more than 60% of the system cost compared to the Equal-distribution approach, and saving more than 70% of the system cost compared to the BS-only approach. Moreover, the results show that the average saving ratio increases in the number of the MDs, i.e., a larger portion of the system cost is reduced. This is because a larger number of the MDs provides a larger freedom in performing cooperative relaying, which consequently yields a larger gain in terms of lowering the total system cost.

In Figure 7, we consider the number of MDs $I = 10$ (in the left subplot) and $I = 20$ (in the right subplot), and vary the distance $D = 20, 30, \dots, 60$. For each tested case, we mark out the average saving ratios of the total system cost (i.e., the numbers listed on the top of each subplot) by using Algorithm JOTRD. Figure 7 again shows that Algorithm JOTRD can significantly reduce the total system cost compared to the Equal-distribution approach and the BS-only approach. Meanwhile, the comparisons between the two subplots also verify that a larger saving ratio can be achieved when more MDs coexist for cooperation.

To further evaluate the advantage of the optimal cooperative scheme that jointly optimizes the content transmission rate and the MDs' relay-durations, we compare the optimal cooperative scheme with another heuristic scheme with fixed content transmission rate in Figure 8. Specifically, in the heuristic scheme, the content transmission duration x is heuristically fixed (which corresponds to a heuristically chosen transmission rate r), but the MDs' relay-durations are optimally given according to Proposition

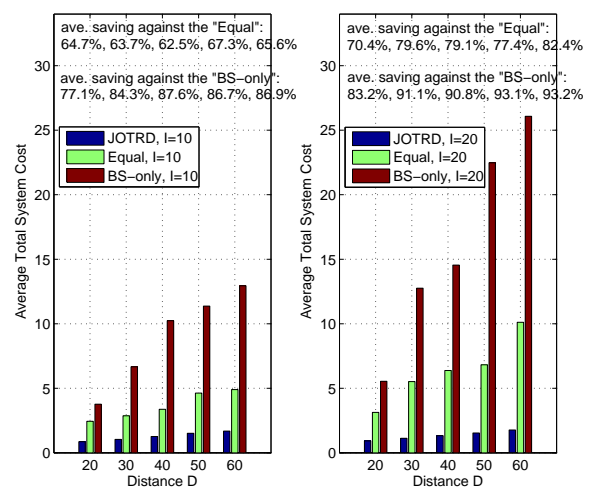


Fig. 7: Advantage of the MDs' cooperations in saving the total system cost. We plot the average total system versus different distance $D = 20, 30, \dots, 60$. Left subplot: We fix the number of MDs $I = 10$. Right subplot: $I = 20$. For each tested case, the result (i.e., the total system cost) is averaged over 200 randomly generated network scenarios. Besides, the numbers illustrated on the top of each subplot denote the average saving ratios of the total system cost by using Algorithm JOTRD against the Equal-distribution approach and the BS-only approach, respectively.

1. Each result in Figure 8 represents the average result for 200 randomly generated network scenarios.

In the left subplot of Figure 8 (with $D = 40m$), the right-most result labelled with x^* denotes the output of Algorithm JOTRD¹¹. Meanwhile, the other five results labelled with $x = 0.4, 0.6, 0.8, 1.0$, and 1.2 denote the output of the heuristic scheme with the content transmission duration fixed at $x = 0.4, 0.6, 0.8, 1.0$, and 1.2 second (which correspond to that the transmission rate $r = 2.5, 1.67, 1.25, 1.0, 0.83$ Mbps), respectively. As shown in the left subplot of Figure 8, Algorithm JOTRD can effectively reduce the total system cost as well as total cost for energy consumption, compared to the heuristic scheme with fixed transmission durations. In particular, as we have marked out in the left subplot, the *average* optimal transmission rate is 2.84Mbps. The right subplot of Figure 8 (with $D = 50m$) shows the similar advantage of Algorithm JOTRD, with the average optimal transmission rate equal to 2.57Mbps. The results in Figure 8 again verify the importance of jointly optimizing the content transmission rate and the MDs' relay-durations.

7.4 Impact of the MDs' Geographical Distribution

We show the impact of the MDs' geographical distribution on the system performance by varying the distance D and the radius R . Recall that the tuple of (D, R) locates the circle in which the MDs are randomly distributed. The left subplot of Figure 9 plots the total system cost (produced by Algorithm JOTRD) under different D and R , with the total number of MDs $I = 20$. The left subplot of Figure 9 shows that the total system cost increases in R , since a larger geographical distribution of the MDs necessitates a larger transmit-powers of the MDs for performing relaying and thus yields a greater system cost. For the similar reason, the total system cost also increases in D , since the BS

¹¹. For each randomly generated scenario, the corresponding x^* is different. That is why we label the result with x^* , instead of a particular numerical value.

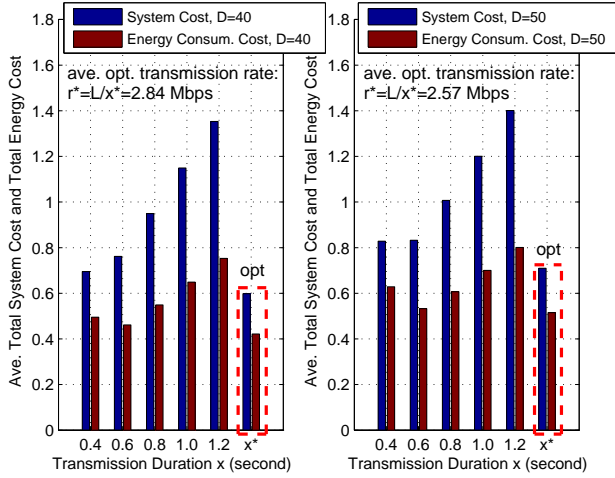


Fig. 8: Impact of the transmission duration x on the total system cost and total cost for energy consumption. Each result represents the average result over 200 randomly generated network scenarios. We fix the number of MDs $I = 10$ and $\alpha = \beta_i = \gamma = 0.5, \forall i \in \mathcal{I}$. Left subplot: $D = 40\text{m}$. Right subplot: $D = 50\text{m}$.

needs a larger transmit-power to transmit to some selected MDs (this point also has been reflected in Figure 4, Figure 6, and Figure 7). Moreover, in the left subplot, we use the optimal solution produced by Algorithm JOTRD to compute the ratio between the cellular-link usage cost and the total system cost, and mark out this ratio for each tested case. Interestingly, the results show that the ratio increases in R , which means that the cellular-link usage cost tends to be more significant in the total system cost. The trend is reflected in the right subplot of Figure 9 which shows the optimal transmission duration x^* versus different values of radius R , with the parameter-settings corresponding to the left subplot. The results show the optimal transmission duration x^* also increases in R . This is because a larger R means a larger transmit-power required by each MD to perform relaying, while prolonging transmission duration can reduce the content transmission rate and thus reduce the required transmit-powers.

We next show the impact of the MDs' distribution on their consequent optimal cooperation in Figure 10. To this end, we plot the ratio between the total relay-duration of all MDs and the total transmission duration, i.e. $\sum_{i \in \mathcal{I}} z_i^*/x^*$ (produced by Algorithm JOTRD) versus different (D, R) . Intuitively, $\sum_{i \in \mathcal{I}} z_i^*/x^* = 1$ means that the content delivery is completely performed via the MDs' cooperations. In comparison, $\sum_{i \in \mathcal{I}} z_i^*/x^* \ll 1$ means that the delivery is mainly performed by the BS's broadcasting, and little cooperation among the MDs is invoked. For easy comparison, we consider the scenario with a fixed distribution of the MDs, i.e., all the MDs are *evenly distributed* on the circle (whose central point is $(D, 0)$ and radius is R). Specifically, in the left subplot of Figure 10, we consider 5 MDs (i.e., $I = 5$) whose respective phases on the circle correspond to angles of $0, \frac{2\pi}{5}, \frac{4\pi}{5}, \frac{6\pi}{5},$ and $\frac{8\pi}{5}$. And in the right subplot, we consider 10 MDs (i.e., $I = 10$). Both subplots in Figure 10 show that the evaluated ratio increases in D , and decreases in R . This is because the larger D (which can be considered as a measure of the average distance between the BS and the MDs) encourages more cooperations among the MDs, and thus yields a greater ratio $\sum_{i \in \mathcal{I}} z_i^*/x^*$. On the other hand, the larger R (which can be considered as a measure of the average distance between different

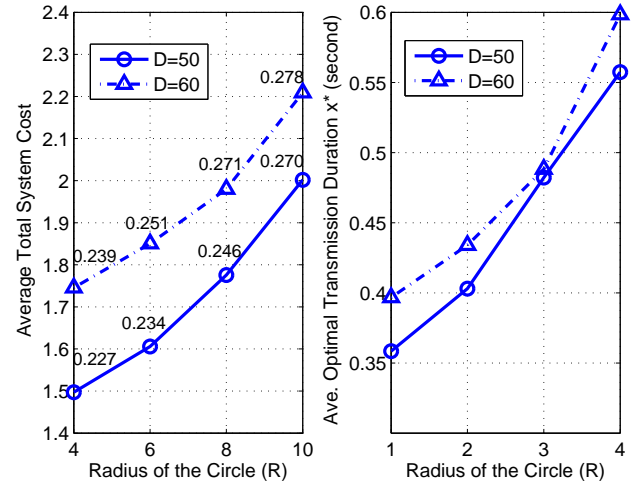


Fig. 9: Impact of the MDs' distribution on the optimal total system cost and the optimal transmission duration x^* (produced by Algorithm JOTRD). Left subplot: the total system cost versus different (D, R) . For each tested case, the result (i.e., the total system cost) is averaged over 200 randomly generated network scenarios. The number illustrated above each tested case represents the ratio between the cellular-link usage cost and the total system cost. Right subplot: the optimal transmission duration x^* versus different (D, R) .

MDs) discourages the cooperation among the MDs, and thus yields a smaller ratio $\sum_{i \in \mathcal{I}} z_i^*/x^*$.

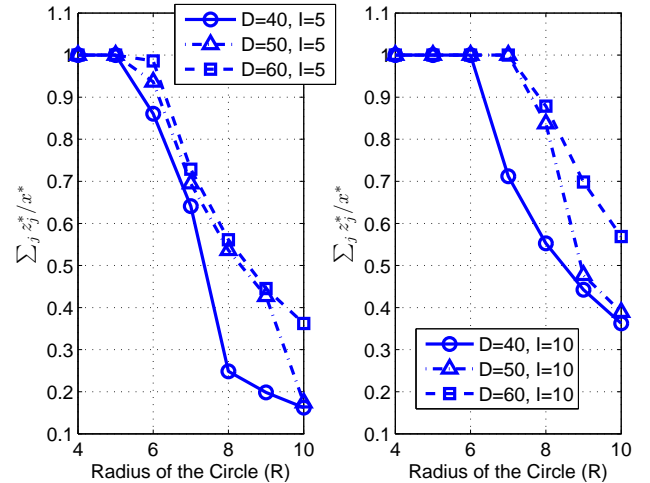


Fig. 10: Ratio between the total relay-duration of all MDs and the optimal transmission duration (produced by Algorithm JOTRD) versus different (D, R) . Left subplot: The total number of MDs $I = 5$ with $E_i^b = 0.01J$. Right subplot: $I = 10$ with $E_i^b = 0.01J$.

Figure 11 plots the ratio $\sum_{i \in \mathcal{I}} z_i^*/x^*$ versus different energy budget E_i^b of each MD. The similar trend, namely, $\sum_{i \in \mathcal{I}} z_i^*/x^*$ increases in D and decreases in R , is also reflected in Figure 11. Moreover, it is also observed that the degree of cooperation (represented by $\sum_{i \in \mathcal{I}} z_i^*/x^*$) is a non-decreasing function of each MD's energy budget E_i^b . When the MDs have higher energy budgets, they are more likely to take advantage of cooperative relaying in order to reduce the total system cost.

7.5 Tradeoff between Energy Consumption and Cellular-Link Usage

We next show the tradeoff between the energy consumption and the cellular-link usage. In particular, we consider $I = 10$ MDs, and

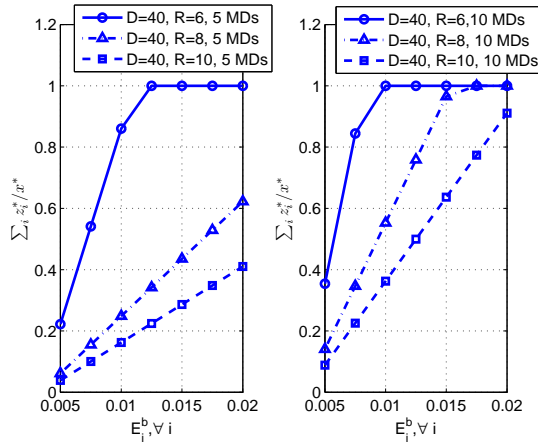


Fig. 11: Ratio between the total relay-duration of all MDs and the optimal transmission duration (produced by Algorithm JOTRD) versus different $\{E_i^b\}_{i \in \mathcal{I}}$. Left: The total number of MDs $I = 5$. Right: $I = 10$.

fix $\alpha = \beta_i = 0.5, \forall i \in \mathcal{I}$. Meanwhile, we vary γ from 0.25 to 1.5 to obtain different ratios γ/α . For each tested case, we plot the average result over 200 randomly generated network scenarios. The left subplot of Figure 12 shows that when the ratio γ/α increases, the average optimal transmission duration x^* decreases. This is because a larger weight on the cellular-link cost makes the BS more conservative in using the cellular-link for delivering the content, which leads to a shorter transmission duration x . However, reducing the cellular-link usage leads to a larger total energy consumption, which is reflected in the right subplot of Figure 12. Specifically, the right subplot of Figure 12 shows that when the ratio γ/α increases, the total energy consumption of the BS and all MDs (which is produced by Algorithm JOTRD) increases. This is because that a shorter transmission duration requires greater transmit-powers for both the BS and each MD to send data, which thus yields a larger total energy consumption. Meanwhile, as shown in Figure 12, the optimal transmission duration for the content (of 1Mbits) is usually very short (and the optimal transmission duration will be even shorter for a smaller content size). Such a short transmission duration allows us to assume that the MDs' locations are relatively static.

7.6 Marginal Error Due to Without Considering the MD's Circuit Power Consumption when Transmitting

Finally, to evaluate the error due to ignoring MD i 's circuit power consumption q_i during data transmission, we perform some numerical tests and show the results in Table 1¹². Specifically, we vary the topology-settings by varying the tuple of (D, R) , and for each setting, we test 200 randomly generated scenarios of the MDs' locations and compute the average relative error (with and without considering $\{q_i\}_{i \in \mathcal{I}}$). The results in Table 1 verify that the relative error due to without considering $\{q_i\}_{i \in \mathcal{I}}$ is very marginal, i.e., no greater than 3% for all the cases which we have tested. In particular, according to [7], [31]–[37], the value of q_i/q_B is usually even smaller than 1% (as used for testing in Table 1), and hence, the corresponding relative errors are believed to even

12. When considering $\{q_i\}_{i \in \mathcal{I}}$, we again adopt an exhaustive search method (similar to that in Footnote 9) to solve Problem (P1), and obtain the optimal system cost. In this method, we solve Problem (P1-Bottom) (which is a linear programming problem according to Proposition 1) in the bottom-layer for each given x . Based on the outcome from the bottom layer, we perform an exhaustive search (with the step-size small enough) in the top layer to solve Problem (P1-Top).

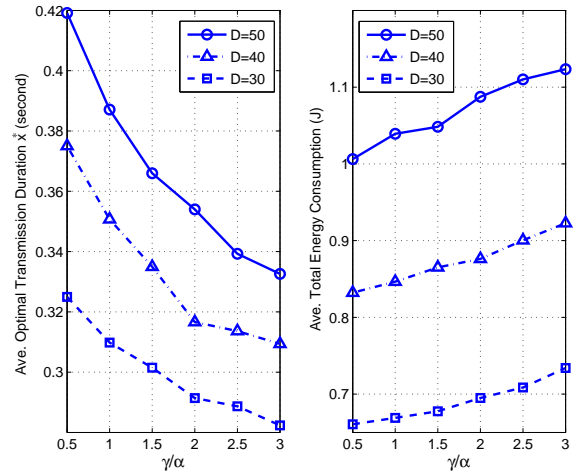


Fig. 12: Tradeoff between the total energy consumption and the cellular-link usage. We fix the number of the MDs $I = 10$ and $\alpha = \beta_i = 0.5, \forall i \in \mathcal{I}$. Left subplot: The average optimal transmission time versus different γ/α . Right subplot: The average total energy consumption versus different γ/α , corresponding to the left subplot.

smaller than those shown in Table 1. Therefore, the analytical results in this paper are of a sufficient accuracy. In other words, it is accurate enough to use Algorithm JOTRD in practice.

TABLE 1: Relative error of the total system cost ($q_i/q_B = 1\%$)

| Network topology | 5 MDs | 10 MDs | 15 MDs | 20 MDs | 25 MDs |
|------------------|--------|--------|--------|--------|--------|
| $D = 30, R = 3$ | 0.77% | 0.80 % | 0.73% | 0.75 % | 0.78 % |
| $D = 30, R = 4$ | 0.82% | 0.95 % | 1.01 % | 1.06 % | 1.18 % |
| $D = 30, R = 5$ | 1.27% | 1.45 % | 1.72 % | 1.74 % | 2.04% |
| $D = 40, R = 3$ | 0.63% | 0.64 % | 0.58 % | 0.63 % | 0.60 % |
| $D = 40, R = 4$ | 0.62% | 0.77 % | 0.87 % | 0.94 % | 0.93 % |
| $D = 40, R = 5$ | 0.77% | 1.06 % | 1.02 % | 1.31 % | 1.39 % |
| $D = 50, R = 3$ | 0.59 % | 0.57 % | 0.54 % | 0.54 % | 0.57 % |
| $D = 50, R = 4$ | 0.57 % | 0.55 % | 0.76 % | 0.78 % | 0.71 % |
| $D = 50, R = 5$ | 0.73 % | 0.78 % | 1.03 % | 1.14 % | 1.16 % |
| $D = 60, R = 3$ | 0.54% | 0.59 % | 0.55 % | 0.50 % | 0.49 % |
| $D = 60, R = 4$ | 0.58% | 0.56 % | 0.60 % | 0.61 % | 0.66 % |
| $D = 60, R = 5$ | 0.69% | 0.65 % | 0.85 % | 0.87 % | 0.94 % |

8 CONCLUSION

In this paper, we have investigated the MDs' cooperative traffic offloading for content distribution, by jointly optimizing the transmission rate of the content and the MDs' relay-durations. Due to the MDs' limited transmit-powers and energy budgets, the transmission rate of the content strongly influences the MDs' relay-durations, which makes the joint optimization problem difficult to solve. Our key idea to tackle this challenging problem is to exploit the decomposable structure of the joint optimization problem, based on which we characterized all possible cases for achieving the optimum. We then derived the optimal solution for each of these cases in an analytical manner, and further proposed an efficient algorithm to find the optimal solution of the original joint optimization problem based on the derived analytical results.

Extensive numerical results verify that the proposed Algorithm JOTRD can achieve the optimal solution of the joint optimization problem, while saving more than 90% of the computational time compared to the exhaustive search method. Meanwhile, numerical results also show that the optimal MDs' cooperation can significantly reduce the system cost. Moreover, we find that

the optimal MDs' cooperation can save a larger portion of the system cost when more MDs coexist for cooperation, and more cooperations are invoked when the MDs are further away from the BS while closer with each other (in which case the MDs' cooperation is more beneficial for reducing the system cost).

In this paper, we have focused on a cellular-controlled centralized approach, and the corresponding results can be considered as the performance benchmark for evaluating other relevant cooperative schemes. An interesting future direction is to design a distributed algorithm to implement this jointly optimal content transmission rate and the MDs' relay-durations, taking into account the impact of incomplete network information and the MDs' incentives for cooperations. Another important future direction is to further investigate the case of delivering multiple pieces of contents, and to design an efficient MDs' cooperative scheme that captures the coupling-effect among the MDs who are allowed to select different pieces of contents to relay.

REFERENCES

- [1] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "3GPP LTE Traffic Offloading onto WiFi Direct," in *Proc. of IEEE WCNCW'2013*, pp. 135-140, April 2013, Shanghai, China.
- [2] D. Camps-Mur, "Device-to-Device Communications with WiFi Direct: Overview and Experimentation," *IEEE Wireless Communications*, vol. 20, no. 3, pp. 96-104, June 2013.
- [3] P. Phunchongham, E. Hossain, and D.I. Kim, "Resource Allocation for Device-to-Device Communications Underlying LTE-advanced Networks," *IEEE Wireless Communications*, vol. 20, no. 4, pp. 91-100, Aug. 2013.
- [4] A. Aijaz, H. Aghvami, and M. Amani, "A Survey on Mobile Data Offloading: Technical and Business Perspectives," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104-112, April 2013.
- [5] S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina, and Y. Koucheryavy, "Cellular Traffic Offloading onto Network-Assisted Device-to-Device Connections," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 20-31, April 2014.
- [6] L. Al-Kanj, Z. Dawy, and E. Yaacoub, "Energy-Aware Cooperative Content Distribution over Wireless Networks: Design Alternatives and Implementation Aspects," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1736-1760, 2013.
- [7] L. Al-Kanj, Z. Dawy, W. Saad, E. Kutanoglu, "Energy-Aware Cooperative Content Distribution over Wireless Networks: Optimized and Distributed Approaches," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 3828-3847, Oct. 2013.
- [8] L. Al-Kanj, V. Poor, and Z. Dawy, "Optimal Cellular Offloading via Device-to-Device Communication Networks with Fairness Constraints," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8., pp. 4628-4643, 2014.
- [9] Y. Wu, J. Wang, L. Qian, and R. Schober, "Optimal Power Control for Energy Efficient D2D Communication and Its Distributed Implementation," *IEEE Communications Letters*, vol. 19, no. 5, pp. 815-818, Feb. 2015.
- [10] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas, "Hybrid Data Pricing for Network-Assisted User-Provided Connectivity," in *Proc. of IEEE INFOCOM'2014*, pp. 682-690, April 2014, Toronto, Canada.
- [11] Y. Wu, J. Wang, L. Qian, and R. Schober, "Energy-Aware Revenue Optimization for Cellular Networks via Device-to-Device Communication," in *Proc. of ICC'2015*, pp. 2085-2090, June 2015, London, United Kingdom.
- [12] Qualcomm Technologies, "LTE Advanced - Evolving and Expanding into New Frontiers," available online at <https://www.qualcomm.com/documents/lte-advanced-evolving-and-expanding-new-frontiers>, August, 2014.
- [13] C. Sankaran, "Data Offloading Techniques in 3GPP Rel-10 Networks: A Tutorial," *IEEE Communication Magazine*, vol. 50, no. 6, pp. 46-53, 2012.
- [14] T. Wang, L. Song, Z. Han, and B. Jiao, "Dynamic Popular Content Distribution in Vehicular Networks using Coalition Formulation Games," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 538-547, June 2013.
- [15] N. Cheng, N. Lu, N. Zhang, X. Shen, and J.W. Mark, "Vehicular WiFi Offloading: Challenges and Solutions," *Vehicular Communications (Elsevier)*, vol. 1, no. 1, pp. 13-21, Jan. 2014.
- [16] X. Kang, Y. Chia, S. Sun, and H. Chong, "Mobile Data Offloading Through A Third-Party WiFi Access Point: An Operator's Perspective," in *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5340-5351, Oct. 2014.
- [17] L. Vu, K. Nahrstedt, I. Rimac, V. Hilt, and M. Hofmann, "iShare: Exploiting opportunistic ad hoc connections for improving data download of cellular users," in *Proc. of GLOBECOM'2010*, pp. 1475-1480, Dec. 2010, Miami.
- [18] D. Niyato, P. Wang, W. Saad, and A. Hjørungnes, "Controlled Coalitional Games for Cooperative Mobile Social Networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1812-1824, May 2011.
- [19] T. Han, N. Ansari, "Enabling Mobile Traffic Offloading via Energy Spectrum Trading," *IEEE Transactions on Wireless Communications*, vol. 13, no. 6, pp. 3317-3328, June 2014.
- [20] T. Han, N. Ansari, "On Accelerating Content Delivery in Mobile Networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1314-1333, Third Quarter, 2013.
- [21] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, "MicroCast: Cooperative Video Streaming on Smartphones," in *Proc. of ACM MobiSys'2012*, pp. 57-70, June 2012, Low Wood Bay, United Kingdom.
- [22] Samsung, Music group play, available online at http://www.samsung.com/hk_en/galaxys4/
- [23] S. Dimatteo, P. Hui, B. Han, and V. O.K. Li, "Cellular Traffic Offloading through WiFi Networks," in *Proc. of IEEE MASS'2011*, pp. 192-201, Oct. 2011, Valencia, Spain.
- [24] Y. Li, G. Su, P. Hui, D. Jin, L. Su and L. Zeng, "Multiple Mobile Data Offloading Through Delay Tolerant Networks," in *Proc of ACM Mobicom'2011, CHANTS Workshop*, pp. 43-48, Sept. 2011, Las Vegas, NV.
- [25] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng, and D. Wu, "Energy-efficient Optimal Opportunistic Forwarding for Delay-Tolerant Networks," *IEEE Transactions Vehicular Technology*, vol. 59, no. 9, pp. 4500-4512, Aug. 2010.
- [26] X. Wang, M. Chen, Z. Han, T. Kwon, Y. Choi, "Content Dissemination by Pushing and Sharing in Mobile Cellular Networks: An Analytical Study," in *Proc. of IEEE MASS'2012*, pp. 353-361, Oct. 2012, Las Vegas, NV.
- [27] J. Whitbeck, M. Amorim, Y. Lopez, J. Leguay, V. Conan, "Relieving the Wireless Infrastructure: When Opportunistic Networks Meet Guaranteed Delays," in *Proc. of IEEE WoWMoM'2011*, pp. 1-10, June 2011, Lucca, Italy.
- [28] N. Golrezaei, A. Molisch, A. Dimakis, G. Caire, "Femtocaching and Device-to-Device Collaboration: A New Architecture for Wireless Video Distribution," *IEEE Commun. Magazine*, vol. 51, no. 4, pp. 142-149, April 2013.
- [29] C. X. Mavromoustakis, C.D. Dimitriou, G. Mastorakis, and E. Pallis, "Real-time Performance Evaluation of F-BTD Scheme for Optimized QoS Energy Conservation in Wireless Devices," in *Proc. of IEEE GLOBECOM Workshops, GLOBECOM'2013*, pp. 1151-1156, Dec. 2013, Atlanta, GA.
- [30] C. X. Mavromoustakis, G. Mastorakis, A. Bourdena, E. Pallis, G. Kormentzas, J. Rodrigues, "Context-oriented Opportunistic Cloud Offload Processing for Energy Conservation in Wireless Devices," in *Proc. of IEEE GLOBECOM Workshops, GLOBECOM'2014*, pp. 24-30, Dec. 2014, Austin, Texas.
- [31] D.W.K. Ng, E.S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in Multiuser OFDM Systems with Wireless Information and Power Transfer," in *Proc. of WCNC'2013*, pp. 3823-3828, April 2013, Shanghai, China.
- [32] J. Wu, Y. Bao, G. Miao, Z. Niu, "Base Station Sleeping and Power Control for Bursty Traffic in Cellular Networks," in *Proc. of ICC'2014*, pp. 837-841, June 2014, Sydney, Australia.
- [33] C. Li, J. Zhang, K.B. Letaief, "Energy Efficiency Analysis of Small Cell Networks," in *Proc. of ICC'2013*, pp. 4404-4408, June 2013, Budapest.
- [34] B. Debaillie, C. Desset, "Power Modeling of Base Stations," in *5GEE Summer school, Imec, Belgium*, Aug. 2014, available online at <http://wireless.kth.se/5green/wp-content/uploads/sites/19/2014/08/BDebaillie.pdf>.
- [35] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M.A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How Much Energy is Needed to Run a Wireless Network?" *IEEE Transactions Wireless Communications*, vol. 18, no. 5, pp. 40-49, Oct. 2011.
- [36] A. Carroll, G. Heiser, "An Analysis of Power Consumption in a Smartphone," in *Proc. of USENIX Annual Technical Conference'2010*, June 2010, Boston.
- [37] A. Gupta, P. Mohapatra, "Energy Consumption and Conservation in WiFi Based Phones: A Measurement-Based Study," in *Proc. of IEEE SEC-ON'2007*, pp. 122-131, June 2007, San Diego, CA.
- [38] Z. Zheng, L.X. Cai, R. Zhang, and X. Shen, "RNP-SA: Joint Relay Placement and Sub-carrier Allocation in Wireless Communication Networks with Sustainable Energy," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3818-3828, Sept. 2012.
- [39] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press: 2004
- [40] Lambert W-Function, Mathworld. [Online]. Available: <http://mathworld.wolfram.com/LambertW-Function.html>.



Yuan Wu (S'08-M'10) received the Ph.D degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2010. He is currently an Associate Professor in College of Information Engineering at Zhejiang University of Technology, Hangzhou, China. He was a Postdoctoral Research Associate at the Hong Kong University of Science and Technology during 2010-2011. He was visiting scholar at Princeton University, U.S. (Aug. 2009-Jan. 2010), Georgia State University, U.S. (Jan. 2013-March 2013), and University of

Waterloo, Canada (March 2016-Feb. 2017). His research interests focus on radio resource allocations for wireless communications and networks, cognitive radio networks, data offloading, device-to-device communications, and smart grids. Dr. Wu received the Distinguished Young Faculty Award by Zhejiang University of Technology in 2013, and the scholarship under the State Scholarship Fund of China Scholarship Council for Visiting Scholars in 2015. He is the co-recipient of the Second-class Outstanding Research Award for Zhejiang Provincial Universities in 2012.



Xuemin (Sherman) Shen (M'97-SM'02-F'09) is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc networks. He is an elected member of IEEE ComSoc Board of Governor, and the Chair of Distinguished Lecturers Selection Committee. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom16, Infocom14, IEEE VTC10 Fall,

and Globecom07, the Symposia Chair for IEEE ICC10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC08, the General Co-Chair for ACM Mobihoc15, the Chair for IEEE Communications Society Technical Committee on Wireless Communications. He also served/serves as the Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award from the University of Waterloo, and the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. Dr. Shen is a registered Professional Engineer of Ontario, Canada, a Fellow of IEEE, Engineering Institute of Canada, Canadian Academy of Engineering, and Royal Society of Canada, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.



Jiachao Chen is currently pursuing his M.S. degree in College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His research interest focuses on radio resource allocations for wireless communications and networks, and device-to-device communications.



Li Ping Qian (S'08-M'10) received her PhD degree in Information Engineering from the Chinese University of Hong Kong, Hong Kong, in 2010. She is currently an Associate Professor in the College of Information Engineering at Zhejiang University of Technology, Hangzhou, China. She was a Postdoctoral Research Associate at the Chinese University of Hong Kong, Hong Kong, during 2010-2011. Her research interests lie in the areas of wireless communication and networking and smart grids. She was a co-recipient of IEEE Marconi Prize Paper Award

in Wireless Communications (the Annual Best Paper Award of IEEE Transactions on Wireless Communications) in 2011, Second-class Outstanding Research Award for Zhejiang Provincial Universities in China in 2012, and Second-class Award of Science and Technology Given by Zhejiang Provincial Government in 2015. She was also a finalist to Hong Kong Young Scientist Award in 2011, the scholarship under the State Scholarship Fund of China Scholarship Council for Visiting Scholars in 2015, and Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars in 2015.



Jianwei Huang (S'01-M'06-SM'11-F'16) is an Associate Professor and Director of the Network Communications and Economics Lab (ncl.ee.cuhk.edu.hk), in the Department of Information Engineering at the Chinese University of Hong Kong. He received the Ph.D. degree from Northwestern University in 2005, and worked as a Postdoc Research Associate in Princeton University during 2005-2007. He is the co-recipient of 8 international Best Paper Awards, including IEEE Marconi Prize Paper Award in Wireless Communications in 2011. He has co-authored

four books: "Wireless Network Pricing," "Monotonic Optimization in Communication and Networking Systems," "Cognitive Mobile Virtual Network Operator Games," and "Social Cognitive Radio Networks". He has served as an Associate Editor of IEEE Transactions on Cognitive Communications and Networking, IEEE Transactions on Wireless Communications, and IEEE Journal on Selected Areas in Communications - Cognitive Radio Series. He is the Vice Chair of IEEE ComSoc Cognitive Network Technical Committee and the Past Chair of IEEE ComSoc Multimedia Communications Technical Committee. He is a Fellow of IEEE and a Distinguished Lecturer of IEEE Communications Society.