This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

Low-Latency Multi-flow Cooperative Broadcast in Fading Wireless Networks

Chenxi Qiu, Haiying Shen*, *IEEE Member*, Lei Yu, and Sohraab Soltani

Abstract—Though a cooperative broadcast scheme has been proposed for fading environments, it has two defects: First, it only handles a packet flow from a single source node in the network, but does not consider the scenario of multiple packet flows simultaneously broadcasted from different source nodes. Second, it only allows a single relay node to forward a packet in each time slot, though multiple relay nodes forwarding in a time slot can significantly reduce broadcast latency. In this paper, we aim achieve low-latency multi-flow broadcast in wireless multi-hop networks with fading channels. To describe the interference among the transmission in different flows, we incorporate the Rayleigh fading model to the Signal to Noise Ratio (SNR) model. Then, we introduce a cooperative diversity scheme which allows multiple relays forwarding in a time slot to reduce broadcast latency. We then formulate an interesting problem: In a fading environment, what is the optimal relay allocation schedule to minimize the broadcast latency? We propose a warm up heuristic algorithm for single-flow cooperative broadcast, based on which, we further propose a heuristic algorithm for multi-flow cooperative broadcast. Simulation results demonstrate that the two algorithms achieve lower broadcast latency than a previous method.

Index Terms—cooperative broadcast, cooperative diversity, low latency.

1 Introduction

In wireless networks, broadcasting is a particularly important mechanism for disseminating a message from one source to all other nodes. Blind flooding, in which each node forwards the broadcast packet exactly once is the simplest way to implement broadcasting. However, its high cost and excessive redundant transmissions lead to severe contention and collisions. For efficient broadcasting, many approaches have been proposed [1]–[10], among which the cooperative broadcast [1], [5]–[10] has gained increasing attention.

In cooperative broadcast, a packet receiver cooperatively combines received weak signal power from different relays to recover the original packet in broadcasting. Due to the broadcast nature of the wireless channel, a packet transmitted by a sender can be heard by all of its neighbor nodes. Thus, a node can receive multiple copies of a specific packet from multiple relays in broadcasting and cooperatively combine the signal power in an additive fashion using a cooperative diversity technique (e.g., maximal ratio combining (MRC)) [6] to recover this packet. The efficiency of broadcasting is improved by combining weak signals rather than discarding them. However, most previous cooperative broadcast solutions [6]–[10] did not take into account the fading environments, where the transmissions between relay nodes are susceptible to random fluctuations in signal strength due to node mobility in a multi-path propagation environment. Therefore, these approaches lack robustness and may not guarantee high delivery ratio (i.e., percent of nodes successfully receiving the packet) in fading environments.

Recently, a cooperative broadcast scheme [1] has been proposed to particularly address the fading environments. It aims to improve delivery ratio while minimizing the broadcast latency. Like [1], we define "broadcast latency" as the number of time slots required for one packet to be distributed throughout the entire network, because the broadcast process is finished only if all the nodes receive the packet. The broadcast scheme in [1] incorporates the Rayleigh fading model into broadcast tree construction, and exploits cooperative diversity to reduce the size of the tree. However, it has two defects which limit its applicability. First, it only handles a packet flow from a single source node in the network, but does not consider the scenario of multiple packet flows simultaneously broadcasted from different source nodes (i.e., *multi-flow broadcast*). Second, it only allows a single relay node to forward the packet in each time slot, though multiple relay nodes forwarding in a time slot can significantly reduce broadcast latency. Indeed, in many distributed network applications, e.g., high-quality multimedia applications [11], broadcasts can be initiated by multiple source nodes simultaneously. In such a case, efficient allocation of relay nodes for different data flow broadcasts must be considered. Directly using the relay allocation method for "single-flow" will degrade the multi-flow broadcast performance, because when a node is a relay node for multiple flows, allocating

- * *Corresponding Author. Email: shenh@clemson.edu. Telephone number: (864) 6565931*
- *C. Qiu, L. Yu and H. Shen are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, 29634. E-mail: {chenxiq, shenh, leiy}@clemson.edu*
- *S. Soltani is with Intelligent Automation, Inc., 15400 Calhoun Drive, Suite 400, Rockville, MD 20855. E-mail: ssoltani@i-a-i.com*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

this node to serve only one of the flows inevitably increases the broadcast latency of the other flows.

We use an example to illustrate this problem. In Fig. 1, there are six sensor nodes v_1-v_6, and two packet flows f_1 and f_2 initiated from v_1 and v_2, respectively. Both flows will traverse v_3. If v_3 is selected as the relay for f_1, then the broadcast latency of f_2 will be increased; similarly, allocating v_3 only to f_2 increases the broadcast latency of f_1. Two challenges here are how to allocate the nodes for each flow and how a node selects flow among multiple flows to forward in order to achieve high delivery ratio and minimum broadcast latency.

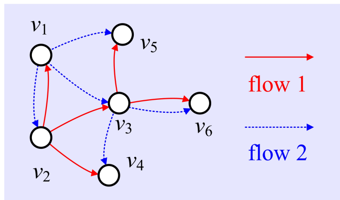


Fig. 1. Multi-flow broadcast

To address the challenges, in this paper, we introduce a cooperative diversity scheme which allows multiple nodes forwarding in a time slot to reduce broadcast latency; thus, a group of nodes, termed a *cooperative relay set*, can forward a packet in the same time slot. We also introduce a probabilistic relay allocation mechanism, where a node can be assigned to forward packets for multiple flows, and it probabilistically selects a packet in different flows to forward at each time slot. Thus, our broadcast model distinguishes from the previous single-flow broadcast scheme [1], where only one node in the network relays the packet in a time slot and each node serves only a single flow, and the interference between the transmissions does not need to be taken into account.

Based on our broadcast model, we then formulate and analyze an interesting problem: In a fading environment, what is the optimal relay allocation schedule (i.e., which node to forward which flow's packet at each time slot) to minimize the broadcast latency? We call this relay allocation optimization problem *Minimum Slotted Delay Cooperative Broadcast (MSDCB)* problem. We prove that the MSDCB problem is NP-hard and $o(\log N)$ inapproximable under some restrictions. We find an optimal solution for this problem which has high time complexity. We then propose a warm up heuristic algorithm for single-flow cooperative broadcast, based on which, we further propose a heuristic algorithm for multi-flow cooperative broadcast.

The simulation results demonstrate that our algorithms perform better than the previous PCDB scheme [1] and achieve nearly "the lowest broadcast latency". Also, interestingly, packet forwarding between cooperative relay sets can further reduce the broadcast latency without increasing the number of relay nodes in the network.

The list of this paper's contributions is as follows:

- We mathematically model and analyze our multi-flow cooperative broadcast scheme with our introduced *cooperative relay set* forwarding and probabilistic relay allocation mechanism in fading wireless networks. We derive a closed-form expression of the probability of successful packet reception.
- Based on our model, we formulate the optimization problem of minimizing the broadcast latency and prove that the problem is NP-hard and $o(\log N)$ inapproximable given some restrictions. We find an optimal solution for this problem, which has a high time complexity.
- A warmup scheme, named *Probabilistic Cooperative Broadcast Heuristic algorithm for single-flow* (PCBH-S) is introduced, and based on this scheme, a heuristic algorithm, named *Probabilistic Cooperative Broadcast Heuristic algorithm for multi-flow* (PCBH-M) is proposed to solve our multi-flow optimization problem.
- The performances of PCBH-S and PCBH-M are evaluated by simulation experiments. The results demonstrate that PCBH-S and PCBH-M outperform previous typical approaches. Also, PCBH-S is comparable to the optimal solution in achieving the "the lowest broadcast latency".

The remainder of this paper is organized as follows. Section 2 builds the mathematical model. Section 3 defines the MSDCB problem and Section 4 introduces the properties of MSDCB. Guided by these properties, in Section 5 and Section 6, we propose an optimal algorithm and two heuristic algorithms (PCBH-S and PCBH-M) for MSDCB. Section 7 presents the simulation results for PCBH-S and PCBH-M. Section 8 presents related work. Section 9 concludes this paper and future work.

2 System Model

We consider a wireless network consisting of a set of nodes $\mathcal{V} = \{v_1, ..., v_N\}$, and a set of packet flows $\mathcal{F} = \{f_1, ..., f_M\}$, where each flow, say f_j ($j = 1, ..., M$), is broadcasted from a source node s_j to all other nodes denoted by set $\mathcal{D}_j$, i.e., $D_j = \mathcal{V} \backslash s_j$. To broadcast a flow to all other nodes, we need to select a set of relay nodes to forward the packet, where the relay nodes can only be selected from the nodes that have received the packet. As in [1], we assume that time is discretized into fixed duration time slots and each node cannot complete a packet reception and a packet transmission in the same time slot, namely *half-duplex* mode.

We assume that the relays in different flows use different frequency bands. Hence, we do not consider the co-channel interference among the relays from different flows. To determine a set of cooperative relay nodes for a flow, we need to consider the fading channels between these relay nodes and other nodes, as well as how cooperative communication can improve the packet delivery ratio under fading environment. Therefore, the selections of fading channel model and cooperative communication model are very important. In the following, we will describe the fading channel model (Section 2.1) and the cooperative communication model (Section 2.2) in detail.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

2.1 Fading Channel Model

We assume a frequency-flat time-varying wireless channel. For the transmitted signal from sender v_i received by receiver v_j, the channel effect can be modeled by a single, complex, random channel coefficient $h_{i,j}$. We consider a Rayleigh fading channel in which all $|h_{i,j}|^2$ are independent and exponentially distributed with a mean value

$$\sigma_{i,j}^2 = P_i d_{i,j}^{-\alpha} \tag{1}$$

where P_i, $d_{i,j}$ and α represent the transmission power of v_i, the distance between v_i and v_j, and the path loss exponent, respectively. The instantaneous signal power $P_{i,j}$ received by v_j from v_i is a random variable with Cumulative Distribution Function (CDF)

$$F_{P_{i,j}}(x) = 1 - e^{-x/\sigma_{i,j}^2}. \tag{2}$$

Typically, there are two types of channel models used to judge whether a packet is correctly decoded by the receiver. The first type model is called the Signal to Noise ratio (SNR) model. Suppose there is relay node v_i sending a packet to a node v_j, then the SNR v_j receives is given by

$$SNR_{i,j} = \frac{P_{i,j}}{N_0}, \tag{3}$$

where N_0 is the noise power density [1]. Node v_j can successfully receive the packet from v_i iff $SNR_{i,j} \geq \gamma_{\text{th}}$ [1], where γ_{th} is the fixed decoding threshold.

2.2 Cooperative Communication Model

The reliability of a packet signal can be improved by diversity schemes which use two or more stochastically independent communication channels to transmit copies of a packet to one receiver [12]. Diversity schemes can exploit independent channels in time, frequency and space to obtain a decrease in error probability, which is called a *diversity gain*. In this paper, we will exploit spatial diversity by using channels from different senders.

The multiple copies of a packet have to be combined in some way at the receiver into a single packet to combat transmission errors. Here we assume the Maximum Ratio Combining (MRC) filter commonly used in diversity receivers [13]. If the sum of all the received instantaneous SNRs is above the decoding threshold γ_{th}, the original packet can be successfully decoded from the packet copies. For example, suppose that G is the set of nodes sending the packet to v_j and the received SNRs from all nodes in G, i.e., $X_{i,j}$ for $\forall v_i \in G$, are independent, then the sum SNR that v_j receives follows a hypoexponential distribution [14] that has Probability Density Function (PDF):

$$f_{\sum_{v_i \in G} X_{i,j}}(x) = \sum_{v_i \in G} \beta_{i,j} e^{-\beta_{i,j} x} \prod_{v_k \in G, k \neq i} \frac{\beta_{k,j}}{\beta_{k,j} - \beta_{i,j}} \tag{4}$$

where $\beta_{i,j} = N_0/\sigma_{i,j}^2$. v_j can successfully decode the original message if the sum SNR $\sum_{v_i \in G} X_{i,j}$ is above the decoding threshold γ_{th}, then the probability that v_j cannot correctly receive the packet can be calculated as:

$$\Pr\left[\sum_{v_i \in G} X_{i,j} < \gamma_{\text{th}}\right] = \int_0^{\gamma_{\text{th}}} f_{\sum_{v_i \in G} X_{i,j}} \mathrm{d}x \tag{5}$$

Using MRC to achieve diversity gain requires multiple relays to broadcast the same packets at the same time, which is possible only if all the forwarders are perfectly synchronized. Like previous works [1], [6], [15], in this paper we assume that all the nodes in the system can be perfectly synchronized. Previous work by Jagannathan *et al.* [16] has investigated the effect of time synchronization error on MRC diversity techniques and Alamouti coding. Mei *et al.* [17] has also analyzed this asynchronous effect on time-reverse space-time coding system and space-time OFDM system. All these works have shown that cooperative system has a good tolerance for small synchronization error.

3 Problem Formulation

In this section, we formulate the broadcast latency minimization problem in multi-flow and fading environment, called *Minimum Slotted Delay Cooperative Broadcast problem* (*MSDCB*). We first introduce the concept of cooperative relay set (Section 3.1). Unlike the scheme in [1] that only uses one single relay to broadcast packet in each time slot, we consider a group of nodes, termed a cooperative relay set, that forwards packets in each time slot in order to decrease the broadcast latency. Then, we introduce the probabilistic allocation mechanism (Section 3.2), in which a node can be shared by more than one flow, and if it is required to forward packets for different flows simultaneously, it probabilistically selects one flow to forward its packet. Finally, we formulate the MSDCB problem.

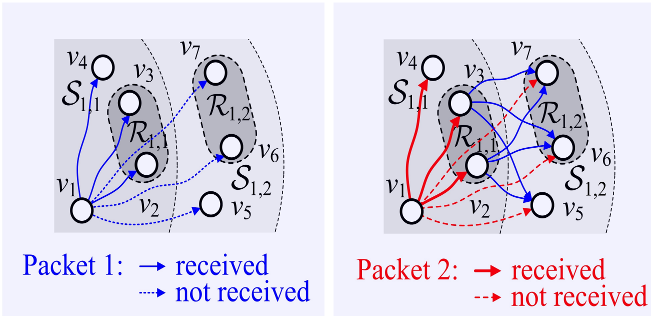
3.1 Cooperative Relay Set Forwarding

In this section, we introduce the concept of cooperative relay set and its properties, and other related definitions. In cooperative broadcast based on cooperative relay set, a packet is transmitted from a relay set to another relay set consecutively.

Definition 3.1: (*Broadcast latency*) The packet delay of node v_i for flow f_j is defined as the minimum number of time slots a packet of flow f_j needs to be transmitted from source v_j to v_i. We assume that every packet from the same flow has the same packet delay for a destination node. The broadcast latency of flow f_j from source s_j , denoted by L_j, is defined as the maximum of packet delays of f_j at all other nodes in the network.

Definition 3.2: (*Cooperative relay set*) For a packet generated at time slot t_0, its age at a specified time slot t is $t - t_0 + 1$. Then, the k^{th} cooperative relay set of a flow f_j is defined as the set of nodes responsible for forwarding the packets of f_j at age k ($1 \leq k \leq L_j$). We use $\mathcal{R}_{j,k}$ to denote the k^{th} cooperative relay set of f_j and specify its size constraint: $|\mathcal{R}_{j,k}| \leq K$. For f_j, we say $\mathcal{R}_{j,l}$ is $\mathcal{R}_{j,k}$'s *previous set* if $l < k$. We use $\mathcal{B}_{j,k}$ to represent $\cup_{l=1}^{k} \mathcal{R}_{j,l}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers



(a) At time slot 1 (b) At time slots 2,3

Fig. 3. An example of serial transmission

Property 3.1: Given a flow f_j, let $\mathcal{S}_{j,k}$ be the set of nodes at which the packet delay equals k. Let $\mathcal{A}_{j,k} = \cup_{l=1}^{k} \mathcal{S}_{j,l}$, then $\mathcal{R}_{j,k} \subset \mathcal{B}_{j,k} \subseteq \mathcal{A}_{j,k}$.

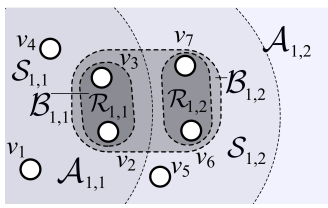


Fig. 2. Example of Def. 3.2

Fig. 2 gives an example for these definitions. v_1 is the source node of flow f_1, and it broadcasts a packet at time slot 1. At the first time slot, the packet's age is 1. $\{v_2, v_3, v_4\}$ correctly receive the packet and $\{v_2, v_3\}$ are selected as the relay nodes to forward the packet. By definition, the packet delay of $\{v_2, v_3, v_4\}$ in f_1 is 1, $\mathcal{S}_{1,1} = \{v_2, v_3, v_4\}$, and the 1$^{\text{st}}$ cooperative relay set of f_1 is $\mathcal{R}_{1,1} = \{v_2, v_3\}$. Similarly, at the second time slot, the packet's age is 2. Suppose $\{v_5, v_6, v_7\}$ correctly receive the packet and $\{v_5, v_6\}$ are selected as the relay nodes to forward the packet. Hence, the packet delay of $\{v_5, v_6, v_7\}$ for f_1 is 2, $\mathcal{S}_{1,2} = \{v_5, v_6, v_7\}$, and $\mathcal{R}_{1,2} = \{v_5, v_6\}$. Notice that the each relay set is not fixed due to channel variation. For instance, in Fig. 2, v_4 could receive packets from v_1 but v_2 could not, then $\mathcal{R}_{1,2} = \{v_3, v_4\}$. Because the system we study is not deterministic, we cannot guarantee that each relay node can successfully receive the packet. But using probability theory, we can make sure each relay node can receive packets with a high probability.

Definition 3.3: (*Serial transmission*) For any packet flow, say f_j, its cooperative relay sets form a sequence $\{\mathcal{R}_{j,1}, \mathcal{R}_{j,2}, \mathcal{R}_{j,3}, ..., \mathcal{R}_{j,L_j}\}$ such that the nodes in $\mathcal{R}_{j,l}$ $(1 \le l \le L_j)$ finish transmitting a packet before the nodes in $\mathcal{R}_{j,l+1}$ start transmitting the same packet. To guarantee *serial transmission*, the cooperative relay sets must satisfy the following two properties:

Property 3.2: $\mathcal{R}_{j,k}$ cannot be empty $(1 \le k \le L_j)$, and also its size cannot exceed K: $0 < |\mathcal{R}_{j,k}| \le K$. We say a cooperative relay set $R_{i,j}$ is *full* if $|R_{i,j}| = K$.

Fig. 3 gives an example of *serial transmission* of *cooperative relay sets*, where the source node $s_1 = v_1$ and two cooperative relay sets $\mathcal{R}_{1,1} = \{v_2, v_3\}$ and $\mathcal{R}_{1,2} = \{v_6, v_7\}$. At time slot 1 (in Fig. 3 (a)), v_1 broadcasts Packet 1 to the whole network, and only nodes v_2, v_3 and v_4 successfully receive it. Then, look at Fig. 3 (b) shows the transmission in time slot 2 and 3: at time slot 2, v_2 and v_3 broadcast and v_5, v_6 and v_7 receive the packet. Then, at time slot 3, v_6 and v_7 broadcast Packet 1, and at the same time, v_1 broadcasts Packet 2 and nodes v_2, v_3 and v_4 receive it. Such serial transmission will never stop until all the nodes in $D_1 = \{v_i | 2 \le i \le 7\}$ successfully receive the packets sent from v_1.

3.2 Probabilistic Relay Allocation Mechanism

When multiple flows are transmitted simultaneously, the flows may compete for the relay nodes. Fig. 4 gives an example, in which $\mathcal{R}_{1,1} = \{v_1, v_2\}$ and $\mathcal{R}_{1,2} = \{v_5, v_6\}$ are relay sets of flow f_1, and $\mathcal{R}_{2,1} = \{v_2, v_3\}$ and $\mathcal{R}_{2,2} = \{v_7, v_8\}$ are relay sets of flow f_2. $\mathcal{R}_{1,1}$ and $\mathcal{R}_{2,1}$, which intersect at v_2, are responsible for forwarding the packets of f_1 and f_2, respectively. The competition occurs at v_2 because v_2 can only transmit one packet in a time slot. One possible solution is to let v_2 serve only one of the flows as cooperative relay. However, it would increase the broadcast latency of the other flows.

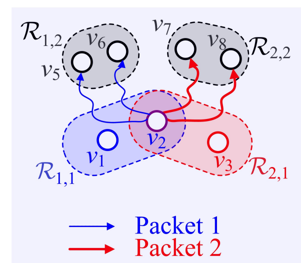


Fig. 4. Relay allocation

In order to efficiently solve the relay competition problem among multiple flows, we propose a *probabilistic allocation mechanism* to allocate cooperative relay nodes, in which a node can be shared by more than one flow, and if it is required to forward packets for different flows simultaneously, it probabilistically selects one flow to forward its packet. A random variable $Y_{i,j,k}$ $(\mathrm{E}(Y_{i,j,k}) = p_{i,j,k})$ which follows Bernoulli distribution is used to represent whether v_i needs to forward the packet for $\mathcal{R}_{j,k}$. $Y_{i,j,k}$ is 1 if v_i is the cooperative relay node for $\mathcal{R}_{j,k}$, and 0 otherwise. The probability that v_i serves relay set $\mathcal{R}_{j,k}$ is $p_{i,j,k}$.

Definition 3.4: (*Cooperative relay set schedule*): The cooperative relay set schedule (schedule in short) is defined as the schedule that determines the probability of each node serving a cooperative relay set. In a schedule, we call the probability of a node v_i serving a flow f_j v_i's *allocation probability* for f_j.

We use a matrix $\mathbf{P} = \{p_{i,j,k}\}_{N \times M \times L}$ to represent a cooperative relay set schedule, where $L \ge \max\{L_1, ..., L_M\}$. Let $p_{i,j,k} = 0$ when v_i is not in $\mathcal{R}_{j,k}$ or $k > L_j$. A *schedule* is *optimal* if $\max\{L_1, ..., L_M\}$ is minimal. Notice that the schedule matrix needs to be updated if some parameters of the system are changed. For example, when the number of flows is changed from 3 to 2, then the system needs to re-compute the schedule matrix according to the setup of each flow.

Suppose that $v_i \in \mathcal{R}_{j,k}$. When previous $k-1$ cooperative relay sets complete the transmission of a packet in f_j, the sum SNR that v_i receives can be represented by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

$$Z_{i,j,k} = \sum_{l=1}^{k-1} \sum_{v_r \in \mathcal{R}_{j,l}} X_{r,i} Y_{r,j,l}. \quad (6)$$

Theorem 3.1: In the SNR model, suppose all $X_{r,i}$ and $Y_{r,j,l}$ in Equ. (6) are independent, then the closed form of $Z_{i,j,k}$'s PDF is given by

$$f_{Z_{i,j,k}}(x) = \sum_{l<k, v_r \in \mathcal{R}_{j,l}} E_{r,i,j} e^{-\beta_{r,i} x} u(x) + E_{j,k} \delta(x) \quad (7)$$

where $E_{j,k} = \prod_{l<k, v_r \in \mathcal{R}_{j,l}} (1 - p_{r,j,l})$ and $E_{r,i,j}$ is given by

$$E_{r,i,j} = \frac{\prod_{l<k, v_t \in \mathcal{R}_{j,l}} (\beta_{r,i} p_{t,j,l} + \beta_{t,i} - \beta_{r,i})}{\prod_{l<k, v_t \in \mathcal{R}_{j,l}, t \neq r} (\beta_{t,i} - \beta_{r,i})}$$

The proof of Theorem 3.1 is provided in Appendix (Section 10.1). Accordingly, in the SNR model, the probability that node v_i cannot correctly decode the packet before the time slot for the packet forwarding of f_j is given by

$$\Pr[Z_{i,j,k} < \gamma_{\text{th}}] = \int_0^{\gamma_{\text{th}}} f_{Z_{i,j,k}} \mathrm{d}x$$
$$= \sum_{l<k, v_r \in \mathcal{R}_{j,l}} \frac{E_{r,i,j} \left(1 - e^{-\beta_{r,i} \gamma_{\text{th}}}\right)}{\beta_{r,i}} + E_{j,k} \quad (8)$$

We use the constant ε to represent the acceptable error probability for the network. We say a node can *successfully receive* a packet if its received sum SNR, denoted as Z, satisfies $\Pr[Z < \gamma_{\text{th}}] < \varepsilon$.

Definition 3.5: (*Informed node* and *Candidate*) For any flow f_j $(j = 1, 2, 3, ..., M)$, we say a node is informed of f_j if the probability it cannot successfully receive every packet of f_j is smaller than ε; otherwise we say the node is *uninformed* of f_j. Informed nodes of f_j are the relay candidates of f_j.

Therefore, in the probabilistic allocation mechanism, nodes refer to $\mathbf{P} = \{p_{i,j,k}\}_{N \times M \times L}$ to decide whether it needs to forward a packet of a specific flow at a specific time slot. The challenge here is how to determine the schedule so that the broadcast latency of each flow is minimized. We formulate this problem below and study this problem in Section 4.

3.3 Problem Statement

A schedule must satisfy the following three conditions in order to successfully broadcast packets in each flow. Here, ε denotes the acceptable error probability of the network, $\mathcal{V}$ denotes the set of nodes, $\mathcal{F}$ denotes the set of flows, L is the max broadcast latency among $\{L_1, L_2, L_3, ..., L_M\}$ and K denotes the size constraint of cooperative relay sets.

Condition 1 (*Connectivity condition*): Like [1], we require that: for any node $v_i \in \mathcal{R}_{j,k}$, the probability that v_i cannot successfully receive a packet in $f_j \in \mathcal{F}$ before the time of forwarding the packet should be smaller than ε: $\Pr[Z_{i,j,k} < \gamma_{\text{th}}] < \varepsilon$.

Under fading channels, there is no guarantee for a node to be informed before a specific time slot due to the fluctuation of channel conditions, and each node has a probability to be informed at different time slots. So condition 1 actually makes sure that the nodes served as relay nodes in a time slot has high probability to be informed before that time slot (e.g., the probability equals 0.999 when ε is set to be 0.001). Another metric, called *flow delivery probability*, which is defined as

$$\eta_j = \frac{\text{\# of nodes that decode the packet in } f_j}{N} \quad (9)$$

can more directly reflect the performance of the system in aspect of packet delivery ratio. Property 3.3 provides a lower bound of each flow delivery probability in the form of ε:

Property 3.3: When ε is small enough, i.e., $\varepsilon \leq \frac{1}{N}$

$$\eta_j \geq 1 - N\varepsilon, \ \forall j. \quad (10)$$

According to Property 3.3, a given threshold η for flow delivery probability, if ε satisfies $\varepsilon \leq \frac{1-\eta}{N}$, then each η_j will be higher than η. For example, when $\eta = 0.99$ and $N = 100$, the flow delivery probability will be higher than 0.99 if ε is smaller than 0.0001.

Condition 2 (*Successful delivery condition*): After all the nodes in $\mathcal{B}_{j,L_j}$ forward the packet in f_j, for any node $v_i \in \mathcal{D}_j = \mathcal{V} \setminus v_j$, the probability that v_i cannot successfully receive the packet should be smaller than ε: $\Pr\left[Z_{i,j,L_j+1} < \gamma_{\text{th}}\right] < \varepsilon$.

Condition 3 (*Probability condition*): The elements $p_{i,j,k}$ in the schedule matrix $\mathbf{P} = \{p_{i,j,k}\}_{N \times M \times L}$ are subject to the following constraints:

- $\sum_{j=1}^{M} \sum_{k=1}^{L} p_{i,j,k} \leq 1$ $(1 \leq i \leq N)$, i.e., the sum of the probabilities that v_i is allocated in all cooperative relay sets cannot exceed 1 (we say a node v_i is *fully used* iff $\sum_{j=1}^{M} \sum_{k=1}^{L} p_{i,j,k} = 1$);
- $\sum_{i=1}^{N} p_{i,j,k} > 0$ $(1 \leq j \leq M$ and $1 \leq k \leq L_j)$, i.e., any cooperative relay set cannot be empty (according to *Property 3.2*);
- $\sum_{i=1}^{N} \lceil p_{i,j,k} \rceil \leq K$ $(1 \leq j \leq M$ and $1 \leq k \leq L_j)$, i.e., the size of the cooperative relay cannot exceed K.

Note that our problem does not require the flow-rate conservation condition. That is, for each flow, the incoming rate at a node can be either larger or smaller than the outgoing rate. For a flow broadcasted from a source, we have two following cases: 1) If a node has been selected as a node in the relay set of the flow, it still has a probability not to forward the packet for this flow. In this case the node's incoming rate is possibly larger than its outgoing rate; 2) A node may forward the same packet of the flow multiple times when it is put into multiple relay sets for this flow. In this case the node's incoming rate can be smaller than its outgoing rate. Also, a node might receive multiple copies of a flow when it hears the same packet with SNR higher than γ_{th} in different time slots. In this case, the node will drop the duplicated packets.

Our objective is to find the minimized broadcast latency for each flow. The problem can be formulated as follows:

Definition 3.6: (*Minimum Slotted Delay Cooperative Broadcast problem (MSDCB)*):

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

Instance: *A finite set of nodes $\mathcal{V}$, a set of flows $\mathcal{F}$, constant ε $(0 < \varepsilon < 1)$, integers L and K. (The instance is denoted by $I(\varepsilon, \mathcal{V}, \mathcal{F}, L, K)$).*

Question: *Existence of a broadcast relay schedule (or schedule matrix) such that* Condition 1, Condition 2 *and* Condition 3 *are satisfied.*

4 Problem analysis

In this section, we analyze the problem presented in Section 3.3.

Property 4.1: In SNR model, increasing the value of any element in schedule matrix, without violating *Condition 3*, does not decrease $\Pr[Z_{i,j,k} > \gamma_{\text{th}}]$ for any $Z_{i,j,k}$

Theorem 4.1: In SNR model, if MSDCB has a solution, there always exists an optimal schedule $\mathbf{P}$ for which $\forall v_i \in \mathcal{V}$ satisfies either $\sum_{j=1}^{M}\sum_{k=1}^{L} p_{i,j,k} = 1$ or $\sum_{j=1}^{M}\sum_{k=1}^{L} p_{i,j,k} = 0$.

$$\sum_{j=1}^{M}\sum_{k=1}^{L} p_{i,j,k} = 1 \text{ or } \sum_{j=1}^{M}\sum_{k=1}^{L} p_{i,j,k} = 0. \tag{11}$$

Corollary 4.1: If MSDCB instance $I(\varepsilon, \mathcal{V}, \mathcal{F}, L, K)$ with $\mathcal{F} = \{f_1\}$ (single-flow case) has a solution, there exists an optimal schedule $\mathbf{P}$ where $p_{i,1,k}$ can be only 1 or 0.

Corollary 4.2: If $I(\varepsilon, \mathcal{V}, \mathcal{F}, L, K)$ has a solution, there exists an optimal schedule such that any cooperative relay set $\mathcal{R}_{j,k}$ satisfies at least one of the following three conditions:

- $\mathcal{R}_{j,k}$ is full: $|\mathcal{R}_{j,k}| = K$;
- All the nodes informed are in $\mathcal{B}_{j,k} = \cup_{l=1}^{k}\mathcal{R}_{j,l}$;
- All the nodes fully used are in $\mathcal{R}_{j,k}$.

Proof: For the sake of contradiction, suppose that there is no optimal schedule such that every cooperative relay set satisfies at least one of the above three conditions. Then for any optimal schedule $\mathbf{P}$ which has cooperative relay sets satisfying none of the above three conditions, we can increase the probability of every node in these cooperative relay sets to make them fully used without violating *Condition 3*. In this way, according to *Property 4.1*, we can get another optimal schedule that satisfies one of the above three conditions, which contradicts with the assumption that there is no such optimal schedule. □

Theorem 4.2: MSDCB is NP-hard under the following constraints:

- the magnitude of fading for any signal follows a Rayleigh distribution and is independent of others;
- The cooperative relay set size constraint is K (constant value);
- $F = \{f_1\}$ (single-flow case).

Theorem 4.3: MSDCB remains NP-hard for the multi-flow case given the following constraints:

- The magnitude of any signal fading follows an independent Rayleigh distribution;
- The cooperative relay set size is upper bounded by K (constant value);
- The number of flows M has constant complexity.

Corollary 4.3: MSDCB is $o(\log N)$ inapproximable given the constraints in *Theorem 4.2* or *Theorem 4.3*:

5 The Optimal Broadcast

Dynamic programming (DP) is typically applied to optimization problems. It first breaks down the original problem into a set of subproblems, and then finds the problem solution by combining the subproblem solutions. Using DP, we break down MSDCB into a series of subproblems called *instantaneous relay allocation problems*. Each subproblem is to find the optimal allocation of relay nodes to inform all the destinations with the minimum number of time slots, given a set of relay node candidates and their destinations. Then, the minimum broadcast latency can be achieved based on the optimal relay allocation at each time slot. destination selection.

Let $C(\mathcal{G}, \mathcal{R})$ denote the minimum number of time slots required to inform $\mathcal{G}$ with the set of relay nodes $\mathcal{R}$ ($\mathcal{R} \subseteq \mathcal{G} \subseteq \mathcal{V}$). Initially, $C(\{s_1\}, \{s_1\}) = 0$, and $C(\mathcal{G}, \{v_i\}) = \infty$ when $v_i \neq s_1$. Then, the minimum time to broadcast to all the nodes in $\mathcal{G}$ can be calculated using the following equation:

$$C(\mathcal{G}, \mathcal{R}) = \min\{C(\mathcal{G}', \mathcal{R}') + 1\}. \tag{12}$$

where min is taken over all possible sets for $\mathcal{G}'$ and $\mathcal{R}'$ such that: 1) $\mathcal{R}'$ can inform each node in $\mathcal{G}'$ 2) $\mathcal{R}$ can inform each node in $\mathcal{G}$, 3) the number of newly selected nodes in each time slot is no larger than K, i.e., $|\mathcal{R} \backslash \mathcal{R}'| \leq K$. Based on Equ. (12), we present an optimal algorithm for MSDCB in single-flow case as shown in Algorithm 1, which returns the minimum broadcast latency of the optimal broadcast schedule.

Algorithm 1: DP($\mathcal{V}, s_1$): output the minimum latency.

```
1  begin
2      Initiate C({s1}, {s1}) by 0 and all C(G, R) s.t.
       ((G, R) ≠ ({s1}, {s1})) by ∞;
3      for n ← 1 to N do
4          for each subset G ⊆ V s.t. G has size n do
5              for each R ⊆ G do
6                  for each R' ⊆ R, G' ⊆ G s.t.
                   |R\R'| ≤ K and G' can be informed by
                   R' do
7                      if R can inform each node in G then
8                          C(G, R) ←
                           min{C(G, R), C(G', R') + 1};
9      latency ← min_{R⊆V} C(V, R);
10     return latency;
```

As Algorithm 1 shows, we first initiate the values of all $C(\mathcal{G}, \mathcal{R})$ s.t. $\mathcal{R} \subseteq \mathcal{G} \subseteq \mathcal{V}$ by ∞, except $C(\{s_1\}, \{s_1\}) = 0$. Then, we iteratively calculate the value of $C(\mathcal{G}, \mathcal{R})$ by increasing the size of $\mathcal{G}$ from 1 to N, with 1 in each step. As Line 6 – 8 show, when we have obtained the value of $C(\mathcal{G}, \mathcal{R})$ for all $|\mathcal{G}| \leq k$, we can derive the value of $C(\mathcal{G}, \mathcal{R})$ for all $|\mathcal{G}| = k + 1$ according to Equ. (12). Here, we use Equ. (8) to calculate the probability of each node to be informed. DP invokes $O(N^{K+1}2^{3N})$ calls to Equ. (8), which takes $O(N^2)$ time. Thus, the algorithm has totally $O(N^{K+3}2^{3N})$ time, which is extremely high and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

impractical for large-scale networks. Hence, in Section 6, we will propose simpler and more time-efficient heuristic algorithms for MSDCB in general cases.

6 Heuristic Algorithms

We have shown that the MSDCB problem is NP-hard and $o(\log N)$ inapproximable, so we need to propose time efficient heuristic algorithms to obtain suboptimal solutions. According to the corollaries and properties described in Section 4 (e.g., the second term in *Property 4.1* implies that moving relay nodes to their *previous cooperative relay sets* will increase the probability of successful transmission), we develop two Probabilistic Cooperative Broadcast Heuristic (PCBH) algorithms for the single-flow and multi-flow cases. The algorithm for single-flow (PCBH-S) and the algorithm for multi-flow (PCBH-M) are described as follows.

6.1 Probabilistic Cooperative Broadcast Heuristic algorithm for Single-flow (PCBH-S)

In the case of single-flow ($\mathcal{F} = \{f_1\}$), any node in $\mathcal{V}$ must be in either an *informed set* or an *uninformed set*. We use $\mathcal{S}_{1,k}$ ($k = 1, 2, 3, ..., L$) to represent the informed set in the k^{th} time slot and $\mathcal{B}_{1,k}$ to represent the set of relay nodes that have been selected, i.e., $\mathcal{B}_{1,k} = \cup_{l=1}^{k} \mathcal{R}_{1,l}$. In addition, let $\mathcal{S}_{1,0} = \mathcal{B}_{1,0} = \{s_1\}$, and let $\mathbf{G} = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, ..., \mathcal{G}_C\}$ denote the set of $\mathcal{V}$'s subsets with cardinality no larger than K, where $C = \sum_{l=1}^{K} \frac{N!}{l!(N-l)!}$.

Algorithm 2: Probabilistic Cooperative Broadcast Heuristic algorithm for Single-flow (PCBH-S).

1 **begin**
2 Initiate $\mathcal{B}_{1,0}$ by $\{s_1\}$;
3 Initiate $\mathcal{S}_{1,0}$ by the set of nodes informed by $\mathcal{B}_{1,0}$;
4 $\mathbf{P} \leftarrow \mathbf{0}$; // $\mathbf{P} = \{p_{i,1,k}\}_{N \times 1 \times T}$ is the schedule matrix
5 Initiate k by 0;
6 **while** $\mathcal{S}_{1,k} \subsetneq \mathcal{V}$ **do**
7   Find $\mathcal{R} \subseteq \mathcal{S}_{1,k}$ s.t. $|\mathcal{R}| \leq K$ that informs the largest number of nodes in $\mathcal{V} \backslash \mathcal{S}_{1,k}$;
8   $k \leftarrow k + 1$;
9   **for** *each node* $v_i \in \mathcal{G}$ **do**
10    $p_{i,1,k} \leftarrow 1$;
11   $\mathcal{B}_{1,k} \leftarrow \mathcal{R}$;
12   $\mathcal{S}_{1,k}$ is given by the set of nodes informed by $\mathcal{B}_{1,k}$;

Algorithm 2 shows the pseudocode of PCBH-S. In each iteration PCBH-S always selects as many nodes in relay set as possible from the candidates (indicated by *Corollary 4.2*), as line 7 shows, and each node in relay set has probability 1 to forward the packet (indicated by *Corollary 4.1*), as line 10 shows. If the number of candidates is larger than the constraint of relay set size K, the algorithm selects K candidates as relay nodes that can inform the most number of uninformed nodes; otherwise, all the candidates are selected as relay nodes.

Property 6.1: Given that $F = \{f_1\}$ and MSDCB has feasible solutions, Algorithm 2 can always find a feasible solution for MSDCB .

Property 6.2: Algorithm 2 can achieve the optimal schedule for MSDCB given the constraints that $\mathcal{F} = \{f_1\}$ and $\forall k : 1 \leq k \leq L$, $\mathcal{S}_{1,k}$ and $\mathcal{B}_{1,k}$ satisfies $|\mathcal{S}_{1,k}/\mathcal{B}_{1,k}| \leq K$.

6.2 Probabilistic Cooperative Broadcast Heuristic algorithm for Multi-flow (PCBH-M)

Also, in the multi-flow case, any node in $\mathcal{V}$ must be in either an *informed set* or an *uninformed set* in each flow. Accordingly, similar to the single-flow case, we use $\mathcal{S}_{j,k}$ to represent the informed set for flow f_j in the k^{th} iteration and use $\mathcal{B}_{j,k}$ to represent the nodes that have been selected in relay set.

In the k^{th} iteration, there are two cases when selecting new relay nodes for flow f_j: (1) $|\mathcal{S}_{j,k} \backslash \mathcal{B}_{j,k}| \leq K$, i.e., the number of candidates in f_j is less than or equals K; (2) $|\mathcal{S}_{j,k}/\mathcal{B}_{j,k}| > K$, i.e., the number of candidates in f_j is larger than K. For case (1), all the candidates should be put into relay set (indicated by *Property 4.1*); while for case (2) only K nodes are put into relay set from the candidates (indicated by *Corollary 4.2*). Any K nodes in the set of candidate relay nodes for f_j, i.e., $\mathcal{S}_{j,k} \backslash \mathcal{B}_{j,k}$, constitute a candidate of relay set for f_j (denoted as $\mathcal{C}_{l,j,k}$), so there are totally $A_{j,k}$ different relay set candidates $\mathbf{C}_{j,k} = \{\mathcal{C}_{1,j,k}, \mathcal{C}_{2,j,k}, ..., \mathcal{C}_{A_{j,k},j,k}\}$ for f_j, where $A_{j,k} = \frac{|\mathcal{S}_{j,k}/\mathcal{B}_{j,k}|!}{(|\mathcal{S}_{j,k}/\mathcal{B}_{j,k}|-K)! \times K!}$. Note that if $|\mathcal{S}_{j,k}/\mathcal{B}_{j,k}| < K$, all the nodes in $|\mathcal{S}_{j,k}/\mathcal{B}_{j,k}|$ constitute one relay set candidate. The collection of M relay sets for M flows (one for each flow) in the k^{th} iteration is denoted as $\mathbf{C}_{u,k} = \{\mathcal{C}_{u_1,k}, \mathcal{C}_{u_2,k}, ..., \mathcal{C}_{u_M,k}\}$ ($\mathcal{C}_{u_j,k} \in \mathbf{C}_{j,k}$). Then, there are total $D_k = \prod_{j=1}^{M} A_{j,k}$ number of possible collections in the k^{th} iteration.

Given a collection $\mathbf{C}_{u,k}$ of relay sets for M flows, we need to determine the matrix $\mathbf{P}(k) = \{p_{i,j,k}\}_{N \times M}$ to make the total reception failure probability of all the uninformed nodes as small as possible, which can be formulated as the following *non-linear programming (NLP)* problem:

$$\min \sum_{j=1}^{M} \sum_{v_i \notin \mathcal{S}_{j,k}} \Pr\left[Z_{i,j,k} < \gamma_{\text{th}}\right] \tag{13}$$

$$\begin{aligned} \text{s.t.} \sum_{j=1}^{M} p_{i,j,k} &\leq 1 - \sum_{j=1}^{M} \sum_{l=1}^{k-1} p_{i,j,l} \qquad (14) \\ \sum_{i=1}^{N} p_{i,j,k} &> 0 \\ p_{i,j,k} &= 0 \text{ if } v_i \text{ is not in any set of } \mathbf{C}_{u,k} \end{aligned}$$

where $p_{i,j,l}$ ($l = 1, ..., k-1$) in Equ. (14) is pre-calculated. We use the notation $[\mathbf{P}(k)\ Result] = \text{NLP}(\mathbf{P}, \mathbf{C}_{u,k}, \gamma_{\text{th}})$ to denote the above non-linear programming with the return values $\mathbf{P}(k)$ and $Result$, where $Result$ represents the minimum value of the objective function (13) and $\mathbf{P}(k)$ is the solution of the NLP. Algorithm 3 shows the pseudo code of PCBH-M. In PCBH-M, first we initiate

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

$\mathcal{B}_{j,0}$ and $\mathcal{S}_{j,0}$ by s_j and the set of nodes informed by s_j, respectively (line 2-3). Then, we calculate $[\mathcal{S}_{1,k}, ..., \mathcal{S}_{M,k}]$ with k from 1 to L. For each $[\mathcal{S}_{1,k}, ..., \mathcal{S}_{M,k}]$, we enumerate all possible combination of relay sets, and for each possible combination, we first determine the schedule $\mathbf{P}(k)$ by solving the NLP problem defined by Equ. (13) and Equ. (14) (line 10), and then find the schedule the minimum $Result$ value (line 11 – 14) shows.

Property 6.3: Algorithm 3 can always find a feasible solution for MSDCB with $F = \{f_1, f_2, f_3, ..., f_M\}$ within N iterations if in each iteration $\mathrm{NLP}(\mathbf{P}, \mathcal{C}_u, \gamma_{\mathrm{th}})$ has a solution.

Algorithm 3: Probabilistic Cooperative Broadcast Heuristic algorithm for Multi-flow (PCBH-M).

1 **begin**
2 Initiate $\mathcal{B}_{j,0}$ by $\{s_j\}$;
3 Initiate $\mathcal{S}_{j,0}$ by the set of nodes informed by s_j;
4 $\mathbf{P} \leftarrow \mathbf{0}$; // $\mathbf{P} = \{p_{i,1,k}\}_{N \times 1 \times T}$ is the schedule matrix
5 Initiate k by 0;
6 **while** $\exists \mathcal{S}_{j,k} \subsetneq \mathcal{V}$ $(j = 1, 2, 3, ..., M)$ **do**
7 $flag \leftarrow \infty$;
8 **for** $j \leftarrow 1$ *to* M **do**
9 **for** $u \leftarrow 1$ *to* $A_{j,k}$ **do**
10 $[\mathbf{P}'(k)\ Result] \leftarrow \mathrm{NLP}(\mathbf{P}, \mathbf{C}_{u,k}, \gamma_{\mathrm{th}})$;
11 **if** $Result < flag$ **then**
12 $flag \leftarrow Result$;
13 $\mathbf{P}(k) \leftarrow \mathbf{P}'(k)$;
14 $\mathbf{C} \leftarrow \mathbf{C}_{u,k}$;
15 **for** $j \leftarrow 1$ *to* M **do**
16 $\mathcal{B}_{j,k} \leftarrow \mathbf{C}(j) \cup \mathcal{B}_{j,k}$;
17 $S_{j,k}$ equals the set of nodes informed by $\mathbf{P}$;
18 Increase k by 1;

7 Performance Evaluation

In this section, we evaluate the performance of PCBH-S and PCBH-M through discrete-event simulation implemented in MATLAB. We compared these PCBH algorithms with three algorithms. Recall that in PCBH, a node can be allocated to multiple flows and probabilistically forwards the packet for each flow. Two compared algorithms are the variants of PCBH. One is a deterministic allocation algorithm, denoted by CBH, that exclusively allocates every node v_i to the flow f_j for which v_i has the highest probability $p_{i,j,k}$ in time slot k. The other algorithm is a random allocation algorithm, denoted by RAND, that randomly and exclusively assigns a set of nodes with size K to each flow. The third algorithm is PCDB [1] that also considers fading channels and cooperative diversity in single flow and uses the probability of successful packet receipt as a metric to construct the broadcast backbone. However, it serially selects a single relay node in every time slot. Since each node in PCDB cannot be shared by multiple flows at the same time, to make fair comparison, all other methods are performed with the same bandwidth and the same transmission power as PCDB. For example, suppose a node in PCDB is required to broadcast a single flow, say f_1, while PCBH allows the node to broadcast 2 flows, say f_2 and f_3, then the bandwidth and transmission power allocated to f_1 equals the sum of those allocated to f_2 and f_3.

In the simulation, we generated a random deployment of nodes in the region with $1000\mathrm{m} \times 1000\mathrm{m}$ for the single-flow case and $700\mathrm{m} \times 700\mathrm{m}$ for the multi-flow case, and randomly chose a source node for each flow. The parameters were set as follows [1]: $\alpha = 4$ (path loss exponent), $P_i = 20\mathrm{dBm}$ (transmission power), $\gamma_{\mathrm{th}} = 25.8\mathrm{dB}$ (decoding threshold), $N_0 = 4.32 \times 10^{-18}\mathrm{W/Hz}$ (Noise power density) and $R = 1\mathrm{Mbit/s}$ (data rate).

Our evaluation focuses on the performance metric of *broadcast latency*, which is the number of slots a packet requirs to inform all nodes in the network.

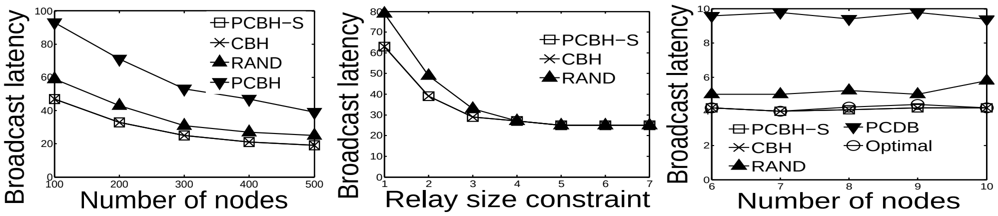
7.1 Single-flow Cooperative Broadcast

Fig. 5(a) and Fig. 5(b) compare the broadcast latency of PCBH-S, CBH, RAND and PCDB in the single-flow case. In Fig. 5(a), the number of nodes is increased from 100 to 400. From the figure, we observe that the average broadcast latency follows PCDB $>$ RAND $>$ CBH $\approx$ PCBH-S. The broadcast latency of PCDB is the highest because it only uses one relay node in each iteration but other approaches utilize the multiple relay nodes to decrease delay. The broadcast latency of RAND is higher than that of CBH and PCBH-S because both PCBH-S and CBH allocate relay nodes with the goal to inform as many nodes as possible in each iteration, while RAND just randomly selects relay nodes. The performances of PCBH-S and CBH are almost the same because in the single-flow case the schedules calculated by PCBH-S and CBH are the same according to *Corollary 4.1*.

In Fig. 5(b), the cooperative relay set size constraint K is changed from 1 to 7. We do not show the results of PCDB in this figure because PCDB only selects one relay node in each iteration, which means K=1. From the figure, we find that the broadcast latency follows RAND $>$ CBH $\approx$ PCBH, which is consistent with the result in Fig. 5(a). It is intriguing to observe that for each algorithm, as K increases, the broadcast latency decreases when $K \leq 4$, and it remains nearly constant when $K > 4$. This is because when $K > 4$, the size of a relay set becomes bounded by the number of candidates in each iteration which is smaller than K, so increasing K does not decrease the broadcast latency anymore.

Fig. 5(c) compares the performance of PCBH, CBH, RAND and PCDB with the optimal solution calculated by DP. Because the time complexity of DP algorithm is too high, we only simulate a small scale network. From the figure, we see that the broadcast latency follows

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers



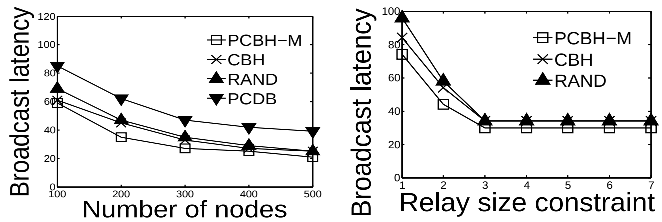
(a) Different number of nodes (b) Different relay size constraint (c) Comparison with DP

Fig. 5. Broadcast latency for single flow

Optimal ≈ PCBH-S ≈ CBH < RAND < PCDB. It verifies that as the optimal method, PCBH-S and CBH can achieve the lowest broadcast delay.

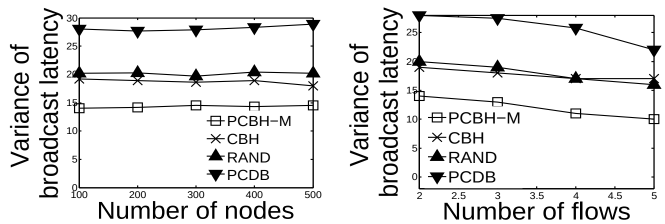
7.2 Multi-flow Cooperative Broadcast

Fig. 6(a) and Fig. 6(b) compare the average broadcast latency per flow of PCBH-M, CBH, RAND and PCDB when there are two flows in the network, with different number of nodes and different relay size constraints K, respectively. Fig. 6(a) shows that the average broadcast latency decreases as the number of nodes increases. It is because that higher node density causes smaller average distance between any two nodes and thus higher received signal energy on average, enabling more nodes to be informed in each time slot. Fig. 6(b) shows that as K increases, the broadcast latency decreases when $K \leq 3$, and it remains approximately the same when $K > 3$. In both Fig. 6(a) and Fig. 6(b), the broadcast latency flows PCDB > RAND > CBH > PCBH-M (no PCDB in Fig. 6(b)), which are consistent with the results shown in Fig. 5(a) and Fig. 5(b), except that CBH and PCBH-S has the same performance in the single-flow case while PCBH-M outperforms CBH in the multi-flow case. This is because in the single-flow case, no competition exists among different flows, thus the results for selecting cooperative relays are the same between PCBH-S and CBH (indicated by *Corollary 4.1*). However, in the multi-flow case, due to exclusive allocation in CBH, arbitrarily allocating a relay node to any flow would increase the broadcast latency of other flows. Since PCBH-M uses probabilistic mechanism for allocating relay nodes, which provides efficient relay resource sharing, it achieves lower broadcast latency than that of CBH. We also compare the variance of broadcast latency of different methods in Fig. 7(a) (with different number nodes) and Fig. 7(b) (with different number of flows). Both figures demonstrate that PCBH-M PCBH-M can best balance the broadcast latency of different flows due to its probabilistic allocating relay mechanism.



(a) Different number of nodes (b) Different relay size constraint

Fig. 6. Broadcast latency in multi-flow case with different number of nodes and different relay size constraint



(a) Different number of nodes (b) Different number of flows

Fig. 7. Variance of broadcast latency in multi-flow case

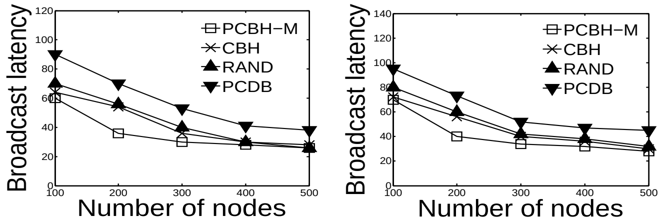
In Fig. 8(a) and Fig. 8(b), we compare the schemes PCBH-M, CBH, RAND and PCDB when there are two flows in the network, with different channel models including Weibull fading model and Rician fading model [14]. In Weibull fading model, we set the *shape parameter* by 3. In Rician fading model, we set the *Rician factor* by 0.5. From these two figures we can find that PCBH-M has the lowest broadcast latency among the four schemes in both Weibull fading model and Rician fading model, which demonstrates the robustness of PCBH-M against different channel models.

Fig. 9(a) and Fig. 9(b) compare the broadcast latencies of PCBH under different number of flows, with different path loss exponent α and different number of nodes N, respectively. From these two figures we can see: (1) the broadcast latency increases with the number of flows increases; (2) the broadcast latency becomes higher when the path loss exponent increases, i.e., when the channel fading is more severe. For observation (1), it is because that when the number of flows increases, the number of relay nodes assigned to each flow in a single time slot becomes smaller (since each node can serve only one flow in a single time slot). This further leads to more time slots to finish the broadcast. For observation (2), it is because that when fading is more severe, the power strength of received signal decreases more rapidly as the distance increases (according to Formula (1)). Thus, the probability of success is lower for each transmission, which results in more transmissions and more time slots to complete the broadcast.

8 Related Work

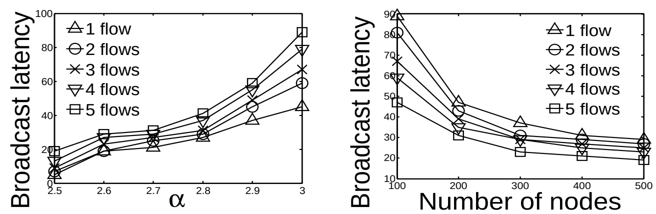
Broadcast latency, as an important metric to measure the efficiency of broadcast protocols in applications that have

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers



(a) Weibull distribution (b) Rician distribution

Fig. 8. Broadcast latency in multi-flow case under different fading models



(a) Different path lose exponent (b) Different number of nodes

Fig. 9. Broadcast latency of PCBH in multi flow case with different number of flows

stringent end-to-end delay requirements, has received significant attention in wireless networks. Schein and Gallager [18] firstly introduced the multiple-access relay channel (MARC), and also derived the channel's capacity outer and inner bounds. Schein *et al.* [19] also introduced a real, discrete-time Gaussian parallel relay network in, which is theoretically important in the context of network information theory. Chlamtac and Kutten [20] studied the complexity of minimum delay broadcast scheduling with the consideration of interference and proved that the problem is NP-hard for general graphs. Gandhi *et al.* [21] showed that the problem remains NP-hard for disk graphs and presented a simple, distributed, collision-free broadcasting algorithm with broadcast latency and the number of retransmissions within O(1) times of their optimal values. Chou *et al.* [15] considered the multi-rate minimum delay broadcast problem, in which nodes may transmit several times at strictly decreasing rates, and exploited both the wireless multicast advantage and the multirate nature of the network to achieve low-delay broadcast. They proved that this problem is NP-hard. Dai *et al.* [22] investigated a linear network coding (NC) construction at the relay in the multiple access relay channel (MARC) system to minimize the average packet loss rate at the destination. In particular, they investigated efficient algorithms for the cases where each relay is allowed to transmit in a single time slot and multiple time slots, respectively. Dai [23] *et al.* also proposed a single source-destination pair communicating via a layer of parallel relay nodes under slow fading environment.

Cooperative communication has been widely exploited for broadcasting [1], [3], [5], [9], [10], [24]–[28] to reduce energy consumption and broadcast delay. One class of schemes [3] make use of MCDS heuristics [4]. The difference between computing MCDS problem and computing broadcast tree is that the broadcast over an MCDS can start from any node while a broadcast trees is routed at a particular node. To apply MCDS heuristics, the wireless network is modeled as a graph in which nodes are connected based on a link quality metric. A broadcast backbone is constructed such that all remaining nodes in the network can be reached through it. Comparing to the algorithm based on broadcast tree, the algorithm using MCDS heuristic may have higher broadcast latency because a MCDS may have more nodes than a minimum latency broadcast tree.

Wu *et al.* [3] proposed an Extended Minimum CDS (E-MCDS) approach that extends the graph by quasi-neighbors, which can reach other nodes using cooperative transmissions. Maric *et al.* [9] and Hong *et al.* [10] considered energy consumption in cooperative broadcasts. Maric *et al.* [9] and Hong *et al.* [10] aimed to find a transmission schedule and power assignment such that every transmitting node can accumulate enough energy from previous transmissions and the total power consumed is minimal. Mergen *et al.* [5] studied the effect of the source/relay transmission powers and the decoding threshold (the minimum SNR required to decode a transmission) on the number of nodes reached by cooperative broadcast. Lu *et al.* [7] studied how to coordinate relay stations in a wireless relay network to cooperatively serve mobile stations such that the network performance can be optimized. Sharma *et al.* [8] considered a problem for concurrent sessions, where in each session, a packet from a source node needs to traverse multiple hops before reaching its destination node. They solved this problem by dividing it into two subproblems: how to optimally allocate relay nodes and how to select paths after relay nodes' allocation. Gentian *et al.* [24] designed a distributed network-wide broadcasting protocol. The protocol takes into account the physical layer dependencies that arise with cooperative diversity. It is showed that computing the optimal solution to the cooperative broadcast problem is NP-hard, and centralized algorithms are proposed with approximation ratio $O(N^\epsilon)$. To improve the efficiency and scalability of broadcast service, Zhang *et al.* [25] proposed a protocol called Chorus, which can effectively reconstruct the packets with the same data through collision resolution achieved by using symbol-level iterative decoding and combining the resolved symbols. Subramanian *et al.* [26] proposed a distributed heuristic flooding method, namely UFlood, which aims to reduce latency by dynamically choosing the relays likely to lead to all nodes informed in the least time. In UFlood each receiver has a feedback for determining which relay's transmission will have the most benefit to receivers.

Recently, Baghaie *et al.* [6] studied delay constrained energy-efficient broadcast in cooperative multihop wireless networks and showed that this problem is NP-hard. They further considered the problem in multi-flow case and formulated the problem as a combinatorial optimization problem, which is proved to be NP hard and $o(n^{1/7-\epsilon})$ inapproxmiable [27]. In addition, Baghaie *et al.* proposed a heuristic for the problem and evaluated the performance of this heuristic under different channel conditions and against the analytical upper and lower bounds in [28]. However, their problem formulation did not take into account the time-varying fading environments, where the transmissions between relay nodes are susceptible to random fluctuations in signal strength due to node mobility in a multi-path propagation environment. Therefore, the optimal solution of DMECB lacks robustness and may not guarantee high delivery ratio in fading environments. Indeed, previous studies showed that the optimization solutions are sufficient in non-fading environments but may suffer a low delivery ratio under wireless channel fading [1]. Obviously, a schedule minimizing energy consumption will not necessarily minimize broadcast latency and vice versa.

The work most closely related to our work is by Lichte *et al.* [1]. They presented cooperative broadcast approach for wireless multihop networks with low latency by constructing small broadcast trees. They incorporated a Rayleigh fading model directly into tree construction to re-obtain complete distribution with high probability and showed that finding minimum latency cooperative broadcasts is NP-hard. However, in this proposed approach, relay nodes are selected serially and are not allowed to broadcast the packet in the same slot. Furthermore, the authors failed to consider the multi-flow case. The deterministic scheme may lead to unbalanced flow latency for different flows in the multi-flow case.

9 Conclusion

In this paper, we have proposed a cooperative cooperative broadcast mechanism with probabilistic relay node allocation and relay-set based forwarding for multi-flow broadcast in fading wireless networks to minimize the broadcast latency. We mathematically modeled our mechanism and formulated the MSDCB problem based on the model. Specifically, we incorporated the Rayleigh fading model into the SNR model to reflect the interference among the transmissions in different flows. We have proved the MSDCB problem is NP-hard and $o(\log N)$ inapproximable given some constraints and derived a number of properties of the problem. Based on these properties, we proposed an optimal algorithm and developed two heuristic algorithms named PCBH-S and PCBH-M for the single-flow and multi-flow cooperative broadcasts, respectively. The simulation results demonstrate that PCBH-S and PCBH-M outperforms a typical previous approach and PCBH-S can achieve almost the optimal performance. In our future work, we aim to develop a continuous time model instead of discrete time model and take into account interference among different flows. In our current design, the schedule matrix needs to be recomputed once the parameters of any flow change, which may lead to high computation cost. Hence, we also aim to design a more time-efficient strategy, which is suitable for dynamic environment.

Acknowledgements

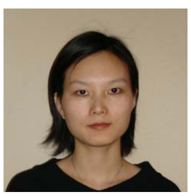
This research was supported in part by U.S. NSF grants NSF-1404981, IIS-1354123, CNS-1254006, CNS-1249603, Microsoft Research Faculty Fellowship 8300751.

References

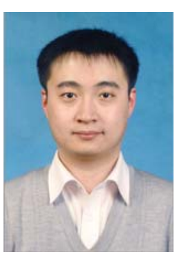
[1] H. S. Lichte, H. Frey, and H. Karl, "Fading-resistant low-latency broadcasts in wireless multihop networks: The probabilistic cooperation diversity approach," in *Proc. of MobiHoc*, 2010.
[2] P. Sinha, R. Sivakumar, and V. Bharghavan, "Enhancing ad hoc routing with dynamic virtual infrastructures," in *IEEE Conf. Comput. Commun*, 2001.
[3] J. Wu, M. Cardei, F. Dai, and S. Yang, "Extended dominating set and its applications in ad hoc networks using cooperative communication," *IEEE Trans. Parallel Distrib. Syst.*, 2006.
[4] S. Guha and S. Khuller, "Approximation algorithms for connected dominating sets," *Algorithmica*, 1998.
[5] B. S. Mergen, A. Scaglione, and G. Mergen, "Asymptotic analysis of multi-stage cooperative broadcast in wireless networks," *Joint special issue of the IEEE Transactions on Information Theory and IEEE/ACM Trans. On Networking*, 2006.
[6] M. Baghaie and B. Krishnamachari, "Delay constrained minimum energy broadcast in cooperative wireless networks," in *Proc. of Infocom*, 2011.
[7] H.-C. Lu and W. Liao, "On cooperative strategies in wireless relay networks," in *Proc. of Infocom*, 2011.
[8] S. Sharma, Y. Shi, Y. T. Hou, H. D. Sherali, and S. Kompella, "Cooperative communications in multi-hop wireless networks: joint flow routing and relay node assignment," in *Proc. of Infocom*, 2010.
[9] I. Maric and R. D. Yates, "Cooperative multicast for maximum network lifetime," *IEEE J. Sel. Areas Commun*, 2005.
[10] Y.-W. Hong and A. Scaglione., "Energy-efficient broadcasting with cooperative transmissions in wireless sensor networks.," *IEEE Trans. Wireless Communications*, 2006.
[11] W. Tu, "Efficient resource utilization for multi-flow wireless multicasting transmissions," *IEEE J. Sel. Areas Commun.*, 2012.
[12] J. N. Laneman, D. N. C. Tse, and G.W.Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior.," *IEEE Trans. Inf. Theory*, 2004.
[13] J. D. Proakis, *Digital Communications*. McGraw-Hill Science, 2000.
[14] S. M. Ross, *Introduction to Probability Models, 8th Edition*. Amsterdam: Academic Press, 2003.
[15] C. T. Chou, A. Misra, and J. Qadir, "Low-latency broadcast in multirate wireless mesh networks," *IEEE J. Sel. Areas Commun.*, 2006.
[16] S. Jagannathan, H. Aghajan, and A. Goldsmith, "The effect of time synchronization errors on the performance of cooperative miso systems," in *Proc. of GLOBECOM*, 2011.
[17] Y. Mei, Y. Hua, A. Swami, and B. Daneshrad, "Combating synchronization errors in cooperative relays.," in *Proc. of ICASSP*, 2005.
[18] G. Kramer and A. J. van Wijngaarden, "On the white gaussian multiple-access relay channel," *IEEE ISIT*, 2000.
[19] B. Schein and R. Gallager, "The gaussian parallel relay network," *IEEE ISIT*, 2000.
[20] I. Chlamtac and S. Kutten, "On broadcasting in radio networks: problem analysis and protocol design," *IEEE ISIT*, 2000.
[21] R. Gandhi, S. Parthasarathy, and A. Mishra, "Minimizing broadcast latency and redundancy in ad hoc networks," in *ACM Intl. Symp. Mobile Ad Hoc Networking & Computing*, 2003.
[22] M. Dai, H. Y. Kwan, and C. W. Sung, "Linear network coding strategies for the multiple access relay channel with packet erasures," *IEEE Transactions on Wireless Communications*, 2011.
[23] M. Dai, C. W. Sung, and Y. Wang, "Distributed on-off power control for amplify-and-forward relays with orthogonal space-time block code," *IEEE Transactions on Wireless Communications*, 2011.
[24] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, and P. V. Krishnamurthy, "On broadcasting with cooperative diversity in multi-hop wireless networks," *JSAC*, 2007.
[25] X. Zhang and K. G. Shin, "Chorus: Collision resolution for efficient wireless broadcast," in *Proc. of Infocom*, 2010.

[26] J. Subramanian, R. Morris, and H. Balakrishnan, "Uflood: High-throughput flooding over wireless mesh networks," in *Proc. of Infocom*, 2012.
[27] M. Baghaie, "On hardness of multiflow transmission in delay constrained cooperative wireless networks," in *Proc. of GLOBECOM*, 2004.
[28] M. Baghaie, D.Hochbaum, and B. Krishnamachari, "Multiflow transmission in delay constrained cooperative wireless networks," *CoRR*, 2012.
[29] Oppenheim and Willsky, *Signals and Systems, 2nd Edition*. Prentice Hall, 1997.
[30] R. E. Miller and J. W. Thatcher, "Reducibility among combinatorial problems," *Complexity of Computer Computations*, 1972.

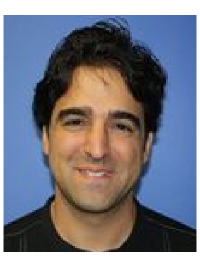
Chenxi Qiu Chenxi Qiu received the BS degree in Telecommunication Engineering from Xidian University, China, in 2009. He currently is a Ph.D student in the Department of Electrical and Computer Engineering at Clemson University, SC, United States. His research interests include sensor networks and wireless networks.



Haiying Shen Haiying Shen received the BS degree in Computer Science and Engineering from Tongji University, China in 2000, and the MS and Ph.D. degrees in Computer Engineering from Wayne State University in 2004 and 2006, respectively. She is currently an Associate Professor in the Department of Electrical and Computer Engineering at Clemson University. Her research interests include distributed computer systems and computer networks, with an emphasis on P2P and content delivery networks, mobile computing, wireless sensor networks, and grid and cloud computing. She was the Program Co-Chair for a number of international conferences and member of the Program Committees of many leading conferences. She is a Microsoft Faculty Fellow of 2010 and a member of the IEEE and ACM.



Lei Yu Mr. Lei Yu received his PhD degree in computer science from Harbin Institute of Technology, China. He was postdoc research fellow in ECE in Clemson University. His research interests include cloud computing, sensor networks, wireless networks, and network security.



Sohraab Soltani Dr. Sohraab Soltani received his B.E. degree (Hons.) from the Shiraz University, Shiraz, Iran, in 2003 and M.S and Ph.D. degree in Computer Science from Michigan State University (MSU) in 2009. Due to his outstanding research effort, he has received a prestigious dissertation completion award from MSU. Dr. Soltani's research interests include computer security, stochastic models, data mining, statistical information inference, wireless video and cognitive mobile communications, ad-hoc and sensor networks and protocol design. He has extensive knowledge in computer security, anomaly and intrusion detection, protocol design, statistical data analysis, computer networking and wireless multimedia communication and has published several papers in competitive conferences and journals in these areas. Dr. Soltani has been a PI or Co-PI in several NSF, SBIR and STTR projects.

10 Appendix

10.1 Proof of Theorem 3.1

Proof: First, we calculate the PDF of $Q_{r,i,j,k} = X_{r,i}Y_{r,j,k}$: when $x \geq 0$

$$\begin{aligned} & \Pr\left[Q_{r,i,j,k} \leq x\right] \\ = & \Pr\left[Q_{r,i,j,k} \leq x | Y_{r,j,k} = 0\right] \Pr\left[Y_{r,j,k} = 0\right] \\ + & \Pr\left[Q_{r,i,j,k} \leq x | Y_{r,j,k} = 1\right] \Pr\left[Y_{r,j,k} = 1\right] \\ = & 1 - p_{r,j,k} + p_{r,j,k}(1 - e^{-\beta_{r,i}x}); \end{aligned} \tag{15}$$

when $x < 0$, $\Pr\left[Q_{r,i,j,k} \leq x\right] = 0$. By calculating the derivation of $\Pr\left[Q_{r,i,j,k} \leq x\right]$, we can obtain the PDF of $Q_{r,i,j,k}$:

$$\begin{aligned} f_{Q_{r,i,j,k}}(x) &= \frac{\mathrm{d}\Pr\left[Q_{r,i,j,k} \leq x\right]}{\mathrm{d}x} \\ &= (1 - p_{r,j,k})\delta(x) + p_{r,j,k}\beta_{r,i}e^{-\beta_{r,i}x}u(x) \end{aligned} \tag{16}$$

where $\delta(x)$ is impulse function and $u(x)$ is Heaviside step function [29]. For any node in $\mathcal{R}_{j,k}$, say v_i, the PDF of $Z_{i,j,k}$ equals the convolution of the PDF of all $Q_{r,i,j,l}$ that $v_r \in \mathcal{B}_{j,k}$. We use $H_{Z_{i,j,k}}(s)$ and $H_{Q_{r,i,j,l}}(s)$ to represent the Laplace transform of $f_{Z_{i,j,k}}$ and $f_{Q_{r,i,j,l}}$ respectively, where $H_{Q_{r,i,j,l}}(s) = \frac{(1-p_{r,j,l})s+\beta_{r,i}}{s+\beta_{r,i}}$. Since the characteristic function of the sum of independent random variables is the product of the characteristic functions of all the random variables, then

$$\begin{aligned} H_{Z_{i,j,k}}(s) &= \prod_{l<k, v_r \in \mathcal{R}_{j,l}} H_{Q_{r,i,j,l}}(s) \\ &= \prod_{l<k, v_r \in \mathcal{R}_{j,l}} \frac{(1 - p_{r,j,l})s + \beta_{r,i}}{s + \beta_{r,i}} \end{aligned} \tag{17}$$

Notice that the values of all $\beta_{r,i}$ are distinct, so we can expand $H_{Z_{i,j,k}}$ to the following form

$$H_{Z_{i,j,k}}(s) = \sum_{v_r \in \mathcal{B}_{j,k}} \frac{E_{r,i,j}}{s + \beta_{r,i}} + E_{j,k} \tag{18}$$

where $E_{j,k} = \prod_{l<k, v_r \in \mathcal{R}_{j,l}} (1 - p_{r,j,l})$ and $E_{r,i,j}$ is given by

$$\begin{aligned} E_{r,i,j} &= (s + \beta_{r,i})\, H_{Z_{i,j,k}}(s) | s = -\beta_{r,i} \\ &= \frac{\prod_{l<k, v_t \in \mathcal{R}_{j,l}} (\beta_{r,i}p_{t,j,l} + \beta_{t,i} - \beta_{r,i})}{\prod_{l<k, v_t \in \mathcal{R}_{j,l}, t \neq r} (\beta_{t,i} - \beta_{r,i})} \end{aligned} \tag{19}$$

Finally, the PDF of $f_{Z_{i,j,k}}$ is obtained by the inverse Fourier transform of $H_{Z_{i,j,k}}(\omega j)$ as follows

$$f_{Z_{i,j,k}} = \sum_{v_r \in \mathcal{B}_{j,k}} E_{r,i,j}e^{-\beta_{r,i}x}u(x) + E_{j,k}\delta(x)$$

□

10.2 Proof of Property 3.3

Proof: For any flow f_j, denote the number of relay nodes by M_j. Then, when $\varepsilon \leq \frac{1}{N}$,

$$\begin{aligned} \eta_j &= \Pr\{\text{All } M_j \text{ relay nodes decode the packet}\} \\ &= (1 - \varepsilon)^{M_j} \geq (1 - \varepsilon)^N \geq 1 - N\varepsilon. \end{aligned}$$

□

10.3 Proof of Theorem 4.1

Proof: We use $\mathbf{P}$ to represent the original schedule matrix. Suppose that $p_{i,j,k}$ in $\mathbf{P}$ is increased and the new schedule matrix is $\mathbf{P}'$. Let $Z_{r,q,l}$ and $Z'_{r,q,l}$ represent the sum SNR that a node v_r receives for packets from flow f_q after the previous $l-1$ cooperative relay sets complete the transmission under two different schedule matrix $\mathbf{P}$ and $\mathbf{P}'$ respectively. According to Equ. (8) and Equ. (19), in SNR model, for any $v_r \in \mathcal{V}$, $q = j$, and $l < k$

$$\Pr\left[Z_{r,q,l} > \gamma_{\text{th}}\right] = \Pr\left[Z'_{r,q,l} > \gamma_{\text{th}}\right] \tag{20}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

for any $v_r \in \mathcal{V}$, $q = j$, and $l \geq k$

$$\Pr\left[Z_{r,q,l} > \gamma_{\text{th}}\right] < \Pr\left[Z'_{r,q,l} > \gamma_{\text{th}}\right] \tag{21}$$

for any $v_r \in \mathcal{V}$, $q \neq j$

$$\Pr\left[Z_{r,q,l} > \gamma_{\text{th}}\right] = \Pr\left[Z'_{r,q,l} > \gamma_{\text{th}}\right] \tag{22}$$

Hence, for any $Z_{r,q,l}$ and $Z'_{r,q,l}$

$$\Pr\left[Z_{r,q,l} > \gamma_{\text{th}}\right] \leq \Pr\left[Z'_{r,q,l} > \gamma_{\text{th}}\right] \tag{23}$$

□

10.4 Proof of Theorem 4.2

Proof: The proof is similar to the proof of NP-hardness of the *Minimum Slotted Cooperative Broadcast (MSCB)* problem in [1], which requires finding a series of nodes $\{v_1, v_2, v_3, ..., v_H\}$ for broadcasting such that all the remaining nodes can successfully receive the packets. In [1], the authors prove the NP-hardness of MSCB by constructing a polynomial time reduction of the NP-hard problem *Set Cover*: a collection $\mathcal{C} = \{C_1, C_2, C_3, ..., C_m\}$ of a finite set $\mathcal{U} = \{u_1, u_2, u_3, ..., u_n\}$ and an integer H are given, and the problem is whether there exists a subset $\mathcal{C}' \subseteq \mathcal{C}$ with $|\mathcal{C}'| \leq H$ such that every element of $\mathcal{U}$ belongs to at least one element of $\mathcal{C}'$. Here, we also construct a polynomial time reduction of the Set Cover problem by constructing the following instance (see Fig. 10 for an example): $\mathcal{V} = s \cup \mathcal{R}' \cup \mathcal{D}'$, where $\mathcal{R}' = \{r_1, r_2, r_3, ..., r_m\}$ (r_i corresponds to $C_i \in \mathcal{C}$) and $\mathcal{D}' = \{d_1, d_2, d_3, ..., d_n\}$ (d_i corresponds to $u_i \in \mathcal{U}$). For each pair of distinct nodes $x, y \in \mathcal{V}$, select independent exponentially distributed random variables $X_{x,y}$ and $X_{y,x}$ ($X_{x,y} \sim X_{x,y}$, i.e., they are independent and identically distributed) such that if $\{x, y\} = \{s, r_i\}$ or $\{x, y\} = \{r_i, d_j\} \wedge u_j \in C_i$, $\Pr[X_{xy} > \gamma_{\text{th}}] = \Pr[X_{yx} > \gamma_{\text{th}}] < \varepsilon$; Otherwise, $E[X_{xy}] = E[X_{yx}] \leq \frac{1-\varepsilon}{H+1}$. The question is that given source node s and time T where $T = \frac{H}{K}$ ($H \bmod K = 0$), whether a schedule $\mathbf{P}$ exists such that *Condition 1*, *Condition 2* and *Condition 3* are met? Hermann [1] has proved that the existence of solution $\{r_1, r_2, r_3, ..., r_H\}$ of MSCB in the constructed network with source node S is a necessary and sufficient condition for the existence of solution of the set cover problem. Accordingly, we need to prove that, in the constructed network, the solution of MSDCB exists iff the solution of MSCB exists.

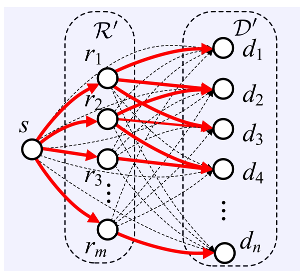


Fig. 10. Proof of *Theorem 4.2*

⇒: Assume there exists a solution for MSCB, we then can also find a solution for MSDCB with cooperative relay sets $\{\mathcal{R}_{1,1}, \mathcal{R}_{1,2}, \mathcal{R}_{1,3}, ..., \mathcal{R}_{1,T}\}$, where $\mathcal{R}_{1,i} = \{r_{(i-1)K+1}, r_{(i-1)K+2}, r_{(i-1)K+3}, ..., r_{iK}\}$ ($i = 1, 2, 3, ..., T$). Each relay node forwards the packet with probability 1. Obviously, this schedule satisfies *Condition 3*. It also satisfies *Condition 1* and *Condition 2* according to the proof in [1].

⇐: Assume there is no solution for MSCB in the constructed instance, which indicates that there is no schedule that can finish broadcasting only using H nodes; each node has probability 1 in $\mathcal{R}'$. Then, according to *Corollary 4.1*, no schedule can finish broadcasting using only H nodes. Because the size of cooperative relay sets is bounded by K, and at least $\lceil \frac{H+1}{K} \rceil = T + 1$ cooperative relay sets are required. Thus, there is no schedule that can finish broadcasting within T slots. □

10.5 Proof of Theorem 4.3

Proof: Similar to the proof of Theorem 4.2, we consider a Set Cover decision problem with a collection $\mathcal{C} = \{C_1, C_2, C_3, ..., C_m\}$ of a finite set $\mathcal{U} = \{u_1, u_2, u_3, ..., u_n\}$ and a given integer H, and reduce it to the following instance of MSDCB with M flows(e.g., see Fig. 11): $T = \frac{H+1}{K} + 1 \geq M + 1$, $(H+1) \bmod K = 0$, $\mathcal{F} = \{f_1, f_2, f_3, ..., f_M\}$ where each flow f_j has a distinct source node s_j, the node set $\mathcal{V} = \{\mathcal{S}, \mathcal{R}'_1, \mathcal{R}'_2, ..., \mathcal{R}'_M, \mathcal{D}', w\}$ where $\mathcal{S} = \{s_1, s_2, s_3, ..., s_M\}$, $\mathcal{R}'_j = \{r_{j,1}, r_{j,2}, r_{j,3}, ..., r_{j,m}\}$ and $\mathcal{D}' = \{d_1, d_2, d_3, ..., d_n\}$. Here $r_{j,i}$ corresponds to $C_i \in \mathcal{C}$ and d_k corresponds to $u_k \in \mathcal{U}$. For each pair of distinct nodes $x, y \in \mathcal{V}$, we select independent exponentially distributed random variables $X_{x,y}$ and $X_{y,x}$ ($X_{x,y} \sim X_{y,x}$) satisfying $\Pr[X_{x,y} > \gamma_{\text{th}}] = \Pr[X_{y,x} > \gamma_{\text{th}}] < \varepsilon$ if (1) $x \in \mathcal{R}'_i$ and $y \in \mathcal{R}'_j$ for any pair of node sets $\mathcal{R}'_i$ and $\mathcal{R}'_j$; (2) $x = s_j$ and $y \in \mathcal{R}'_j$ ($1 \leq j \leq M$); (3) $\{x, y\} = \{r_{j,i}, d_k\}$ and $u_k \in C_i$ ($1 \leq j \leq M$) and (4) $\{x, y\} = \{w, s_k\}$, for $\forall s_k \in \mathcal{S}$. Otherwise, it follows $E[X_{x,y}] = E[X_{y,x}] \leq \frac{1-\varepsilon}{H+1}$. The question is that given source nodes $\{s_1, s_2, s_3, ..., s_M\}$ and time T, does a schedule exist such that *condition 1*, *condition 2* and *condition 3* are met? Now, we prove that, in this specific instance, the solution of MSDCB with M flows exists iff the solution of Set Cover exists.

⇒: Assume there exists a solution $\mathcal{C}' = \{C_{i_1}, C_{i_2}, ..., C_{i_m}\}$ ($m \leq H$) for Set Cover. Then, we can construct a feasible schedule for MSDCB with M-flow instance that satisfies Conditions 1-3. That is, for each flow f_j, in the 1^{st} time slot s_j broadcasts the packet, and from the 2^{nd} time slot to $\lceil \frac{m}{K} \rceil^{\text{th}}$ time slot, nodes $\{r_{j,i_1}, r_{j,i_2}, ..., r_{j,i_m}, w\}$ are selected as relay nodes. Since $m \leq H$, $\lceil \frac{m+1}{K} \rceil \leq T - 1$, which implies these relay nodes can be arranged within $T-1$ time slots with relay set size constraint K. To avoid w being used by different flows in the same time slot, let w be selected by f_j in the $(j+1)^{\text{th}}$ time slot, which is feasible because $T - 1 \geq M$. Obviously, Condition 3 is satisfied because for each flow f_j, every relay node serves at most one flow in each time slot and the size of each relay set is upper bounded by K. Also, Condition 1 is satisfied because all the nodes in $\mathcal{R}'_j$ ($1 \leq j \leq M$) have been informed by s_j in the 1^{st} time slot. Because of the existence of a solution for Set Cover, $\forall u_k \in \mathcal{U}$, $\exists C_{i_l} \in \mathcal{C}'$ such that $u_k \in C_{i_l}$. Then, for $\forall f_j \in \mathcal{F}$ and $d_k \in \mathcal{D}'$, we select $\{r_{j,i_l} | 1 \leq l \leq m\}$ as relay nodes such that

$$\Pr[\sum_{r=1}^{m} X_{r_{j,i_l},d_k} > \gamma_{\text{th}}] < \Pr[X_{r_{j,i_l},d_k} > \gamma_{\text{th}}] < \varepsilon. \tag{24}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2015.2456011, IEEE Transactions on Computers

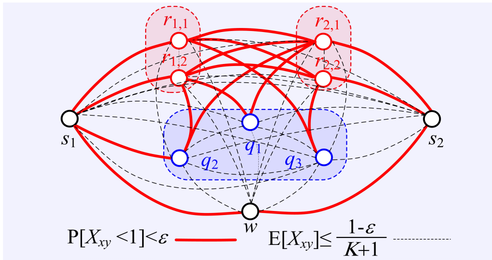


Fig. 11. Proof of *Theorem 4.3*

Thus, every node in $\mathcal{D}'$ is informed. In addition, $\forall s_k \in \mathcal{S}$,

$$\Pr[X_{w,s_k} > \gamma_{\text{th}}] < \varepsilon. \quad (25)$$

Thus, every node in $\mathcal{S}$ is informed by the relay node w and hence Condition 2 is satisfied.

$\Leftarrow$: Assume there exists a feasible schedule for MSDCB with M flows: for each flow f_j, in the 1$^{\text{st}}$ time slot, s_j broadcasts the packet and from the 2$^{\text{nd}}$ time slot to T^{th} time slot, nodes $\{r_{j,i_1}, r_{j,i_2}, ..., r_{j,i_m}, w\}$ are selected as relay nodes. Here, m must be no larger than $KT-K-1 = H$ due to the constraint of set size K. All the nodes in $\mathcal{D}'$ can be informed. Then we prove that we can also find a solution for the Set Cover decision problem, which is $\mathcal{C}' = \{C_{i_1}, C_{i_2}, ..., C_{i_m}\}$. We first assume that this solution cannot solve the Set Cover decision problem; that is, there exists $u_k \notin C_{i_l}$ for all $C_{i_l} \in \mathcal{C}'$. Hence, $\forall r_{j,i_l} \in \mathcal{R}'_j$, $E\left[X_{r_{j,i_l},d_k}\right] \le \frac{(1-\varepsilon)\gamma_{\text{th}}}{H+1}$, then by MarKov's inequality, we can derive that

$$\Pr[\sum_{r_{j,i_l}\in\mathcal{R}'_j} X_{r_{j,i_l},d_k} > \gamma_{\text{th}}] \le \frac{E(\sum_{r_{j,i_l}\in\mathcal{R}'_j} X_{r_{j,i_l},d_k})}{\gamma_{\text{th}}}$$
$$\le \frac{E(\sum_{r_{j,i_l}\in\mathcal{R}'_j} X_{r_{j,i_l},d_k})}{\gamma_{\text{th}}} < 1-\varepsilon \quad (26)$$

which implies that $\Pr[\sum_{r_{j,i_l}\in\mathcal{R}'_j} X_{r_{j,i_l},d_k} < \gamma_{\text{th}}] > \varepsilon$. Thus, d_k is not informed after the 2$^{\text{nd}}$ time slot, which contradicts with the result that all the nodes in $\mathcal{D}'$ can be informed. Therefore, $\mathcal{C}' = \{C_{i_1}, C_{i_2}, ..., C_{i_m}\}$ is a solution for the Set Cover problem. □

10.6 Proof of Corollary 4.3

Proof: The reduction used in the construction of the instance in *Theorem 4.2* preserves the approximation factor. That is, if one can find an α-approximation for MSDCB given the above constraints, there must exist an α-approximation for set cover. Based on [30], we know that the Set Cover problem is $o(\log N)$ inapproximable, thus MSDCB must be $o(\log N)$ inapproximable. Similarly, this result holds given the constraints in *Theorem 4.3*. □

10.7 Proof of Property 6.1

Proof: Assume Algorithm 2 cannot find any solution. Thus, under Algorithm 2, there must exist an iteration, say the k^{th} iteration, in which $\texttt{SetInform}(\mathcal{S}_{1,k}) = \mathcal{S}_{1,k}$, i.e., all candidates have been used but no new node informed; otherwise, the algorithm will never stop until all the nodes are informed. Then, we have that $\forall \mathcal{J} \subseteq \mathcal{S}_{1,k}$, $\mathcal{J}$ satisfies $\texttt{SetInform}(\mathcal{J}) \subseteq \mathcal{S}_{1,k}$, because $\texttt{SetInform}(\mathcal{J}) \subseteq \texttt{SetInform}(\mathcal{S}_{1,k}) = \mathcal{S}_{1,k}$. However, since MSDCB has feasible solutions, we can prove that there must exists a solution that has an iteration in that the nodes from $\mathcal{S}_{1,k}$ can successfully inform the nodes not in $\mathcal{S}_{1,k}$, and then the property is proved by such contradiction. Consider the set partition of $\mathcal{V}$: $\mathcal{S}_{1,k}$ and $\mathcal{V} \setminus \mathcal{S}_{1,k}$. Suppose u is the first node being informed in $\mathcal{V}\setminus\mathcal{S}_{1,k}$, then it must be informed by the nodes from $\mathcal{S}_{1,k}$, because if it is informed by the nodes from $\mathcal{V}\setminus\mathcal{S}_{1,k}$, then u is not the first node being informed in $\mathcal{V} \setminus \mathcal{S}_{1,k}$. □

10.8 Proof of Property 6.2

Proof: We use $\texttt{SetInform}(\mathcal{S})$ to denote the set of nodes informed by the relay set $\mathcal{S}$. Clearly, for any two sets $\mathcal{B}_1 \subseteq \mathcal{V}, \mathcal{B}_2 \subseteq \mathcal{V}$, if $\mathcal{B}_1 \subseteq \mathcal{B}_2$, $\texttt{SetInform}(\mathcal{B}_1) \subseteq \texttt{SetInform}(\mathcal{B}_2) \subseteq \mathcal{V}$ and $|\texttt{SetInform}(\mathcal{B}_1)| \le |\texttt{SetInform}(\mathcal{B}_2)| \le N$. Thus, in the k^{th} iteration, Algorithm 2 always selects all the nodes in $\mathcal{S}_{1,k}/\mathcal{B}_{1,k}$ as relay nodes because $|\texttt{SetInform}(\mathcal{S}_{1,k}/\mathcal{B}_{1,k})| \le K$ and $|\texttt{SetInform}((\mathcal{S}_{1,k}/\mathcal{B}_{1,k}) \cup \mathcal{B}_{1,k})| = \max\{|\texttt{SetInform}(\mathcal{G}_i \cup \mathcal{B}_{1,k})| : \mathcal{G}_i \subseteq \mathcal{S}_{1,k}/\mathcal{B}_{1,k}\}$. We prove the theorem by contradiction. Assume that Algorithm 2 cannot achieve the optimal schedule and its schedule result is denoted by $\mathcal{P}'$. Let $\mathcal{P}$ be the optimal schedule which has the same cooperative relay set as in $\mathcal{P}'$ in each of the first $l-1$ iterations and $\mathcal{P}$ has maximum $l-1$ among all optimal schedules. In other words, no other optimal schedules have all the same relay sets as in $\mathcal{P}'$ in the first h ($h > l-1$) iterations. $l = 1$ means that no optimal schedules have Let $\{\mathcal{R}_{1,1}, \mathcal{R}_{1,2}, \mathcal{R}_{1,3}, ..., \mathcal{R}_{1,L}\}$ be the cooperative relay sets in $\mathcal{P}$. Now we can construct a new schedule $\mathcal{P}''$ (its cooperative relay sets are $\{\mathcal{R}''_{1,1}, \mathcal{R}''_{1,2}, ..., \mathcal{R}''_{1,L_1}\}$) by changing $\mathcal{R}_{1,l}$ to $\mathcal{S}_{1,l}/\mathcal{B}_{1,l}$ via moving some nodes from $\mathcal{R}_{1,t}$ to $\mathcal{R}_{1,l}$ ($t > l$) or adding some candidate nodes from outside of relay sets into $\mathcal{R}_{1,l}$. According to *Property 4.1*, such operation does not decrease the probability of successful reception of packets for any node. Thus, $\mathcal{P}''$ is also an optimal schedule. However, $\mathcal{P}''$ has the same cooperative relay sets with $\mathcal{P}'$ in the first l iterations. A contradiction with that $\mathcal{P}$ has maximum $l-1$ among all optimal schedules. □

10.9 Proof of Property 6.3

Proof: Obviously the schedule calculated by Algorithm 3 satisfies *Condition 1* (*Serial transmission condition*) and *Condition 3* (*Probability condition*). Then, we need to prove that the schedule satisfies *Condition 2* (*successfully delivery*) and show that it is finished within N iterations. Because in each iteration $\text{NLP}(\mathbf{P}, \mathbf{C}_{u,k}, \gamma_{\text{th}})$ has a solution, the algorithm selects at least one node as relay node for the flows of which the packets cannot be received by all nodes. Since the total number of nodes is N, the algorithm runs at most N iterations. The algorithm ends in the k^{th} iteration only if $\forall j$, either $\mathcal{B}_{j,k} = \mathcal{V}$ or $\mathcal{S}_{j,k} = \mathcal{V}$, where in both cases the algorithm satisfies *Condition 2*. □