On Link Scheduling in Dual-Hop 60-GHz mmWave Networks

Zhifeng He, *Student Member, IEEE*, Shiwen Mao, *Senior Member, IEEE*, Sastry Kompella, *Senior Member, IEEE*, and Ananthram Swami, *Fellow, IEEE*

***Abstract*—We tackle the problem of minimizing the maximum expected delivery time of all transmission pairs in 60 GHz mmWave networks. The network is dual-hop with multiple transmitter-receiver pairs, relays, and a centralized controller (termed PicoNet coordinator). We jointly optimize relay and link selection for the transmitter–receiver pairs to minimize delivery time, while the reflected non-line-of-sight transmission links are exploited to get around obstacles. To reduce computational complexity, we develop a decomposition principle to transform the joint-optimization problem into a link selection subproblem and a relay assignment subproblem when there are a sufficient amount of relays. A tight performance bound for the proposed algorithm is proved. We also develop a heuristic scheme to handle the case when there are no enough relays. The superior performance of our proposed algorithms is validated with extensive simulations and comparisons with benchmark schemes.**

***Index Terms*—5G wireless, mmWave networks, optimization, scheduling, 60 GHz.**

I. Introduction

As a core technology for future 5G Wireless systems [1]–[3], 60 GHz millimeter wave (mmWave) communications has attracted significant standardization efforts recently [4], [5]. Although the huge, license-free bandwidth (up to 7GHz in many countries) is a great advantage, signals propagating in 60 GHz channel suffers from serious attenuation, which is significantly higher than that in 5 GHz channels [6], [7]. Beamforming has been used as an essential technique to overcome such attenuation. The signal attenuation and beamforming together results in a reduced interference to neighboring links, which can be exploited for significantly enhanced spatial reuse. According to [3], [8]–[11], the highly directional links, especially in the outdoor environment, can be regarded as *pseudo-wired* with negligible collision probabilities. Therefore, it is possible to schedule concurrent transmissions at multiple links without interfering each other, so as to improve network capacity.

In addition, mmWave signals in the 60 GHz band usually do not diffract around or penetrate obstacles [12]. A line-of-sight (LOS) path between the transmitter and receiver is usually required for a successful transmission. The IEEE 802.15.3c activities consider the use of beamforming technique to switch the LOS link to a Non-LOS (NLOS) link when the direct link is blocked (e.g., by a human body), where relay nodes can be used to forward data for a hidden receiver [12] or wall reflections can be utilized, which forms an NLOS link, so that network connectivity can be enhanced. The authors of [13] proposed a flexible link model that considers both LOS and NLOS for extending the coverage of mmWave networks, where static reflectors were used to produce NLOS links. The authors in [14] investigated how to select a backup NLOS link to support high performance beam switching when the LOS link was blocked. It may require a higher transmission power to transmit via NLOS links using reflections, however, the network coverage and throughput can be improved.

In this work, we consider both LOS and multiple reflected NLOS links between source, relay, and destination nodes, in order to exploit at maximum the potential of mmWave networks, which provides an advantage over prior works on scheduling in 60 GHz mmWave networks [10], [11], [15]–[21]. Furthermore, the 60 GHz transmissions are highly directional, in order to overcome the high attenuation, making it susceptible to blockage of the LOS path by obstacles or pedestrians, while the blockage may appear or disappear occasionally due to the movement of objects in between or the movement of transmitter or receiver themselves [22], [23]. Therefore, we capture the link state dynamics in this work with a Markov chain model. Specifically, a joint link and relay selection problem is formulated as a Nonlinear Integer Programming (NIP) problem. We develop effective heuristic algorithms to solve the NP-hard problem. When there are enough relays in the network to support the source destination pairs that have their LOS path blocked or are out of range, a Decomposition Principle is proposed to break down the formulated problem into a link selection sub-problem and a relay selection sub-problem. We prove that the two sub-problems together provide a sub-optimal solution to the formulated problem with a tight performance bound and greatly reduced

Manuscript received July 26, 2016; revised March 23, 2017 and June 1, 2017; accepted June 14, 2017. Date of publication June 21, 2017; date of current version December 14, 2017. This work was supported by the National Science Foundation under Grant CNS-1320664. This paper was presented in part at the IEEE INFOCOM 2016, San Francisco, CA, USA, April 2016. The review of this paper was coordinated by Dr. X. Huang. *(Corresponding author: Shiwen Mao.)*

Z. He and S. Mao are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA (e-mail: zzh0008@tigermail.auburn.edu; smao@ieee.org).

S. Kompella is with the Information Technology Division, Naval Research Laboratory, Washington, DC 20375, USA (e-mail: sk@ieee.org).

A. Swami is with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: a.swami@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2017.2717840

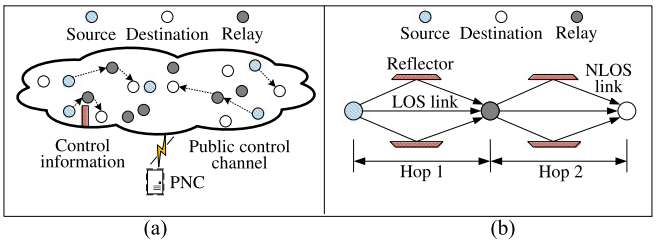


Fig. 1. (a) Network model. (b) Two-hop relay path model.

complexity. Another heuristic algorithm is proposed to deal with the case when there are no enough relays. The performance of our proposed algorithms is validated with extensive simulation studies, with various performance metrics including delay, MEDT, throughput, and fairness, and comparisons with benchmark schemes.

In the remainder of this paper, we present the system model and problem formulation in Section II-A. We develop the Decomposition Principle and the heuristic algorithms in Section III and evaluate their performance in Section IV. Related works are reviewed in Section V and Section VI concludes the paper.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1(a), we consider a 60 GHz ad hoc network consisting of multiple nodes and one PNC. Each node can be either a source node (S), a destination node (D), or a potential relay node (R). When the source and destination nodes are unable to directly communicate with each other (e.g., permanently blocked by an obstacle/wall, or out of range), a relay is used to forward their traffic. Due to point-to-point 60 GHz links (unlike traditional broadcast-based relay networks), we assume that each SD pair can choose only one relay at a time. However, a relay may serve multiple SD pairs at different time slots (but not at the same time slot).

To overcome the deafness problem, which makes it highly challenging for coordination of the highly directional links, we assume a lower frequency common control channel (e.g., a WiFi channel) for all nodes and the PNC [24]. Due to omnidirectional transmissions, better propagation, and larger coverage, the nodes on the control channel actually form a single-hop network. Network state and control information of the dual-hop 60 GHz network can be effectively exchanged among the nodes on the control channel [10], [11], [19]–[21], [24], including the following information at each time slot.

1) The PNC collects network state information, such as traffic requests, link status, etc., from each node
2) The PNC broadcasts link transmission schedules to all the nodes if it makes a new schedule at this time slot
3) Receiver nodes inform the PNC whether the reception is successful, so that the PNC can infer the link status.

As shown in Fig. 1(b), there may be multiple links between a pair of nodes within range of each other: there may be one LOS link, as well as other NLOS links formed by exploiting reflections from walls and other surfaces [8]. Due to moving obstacles, the state of a link is either blocked or unblocked at each time slot. We assume the link state follows a discrete-time Markov process [18], [19], while the nodes learn the transition probabilities of their links from history data and inform the PNC these parameters. If the LOS link is more likely to be blocked, an NLOS link may be a better choice. A successful transmission on a link requires the link being unblocked.

Without loss of generality, we assume each node is equipped with an electronically steerable antenna array to beamform in the transmitting or receiving directions; so each node works in the *half-duplex* mode [25]. Both transmission and reception are directional with a very narrow beamwidth. The beamforming weights learned when receiving from a given node can then be used to transmit back to that node, assuming *channel reciprocity*. As shown in [26], "point-to-point 60 GHz devices typically have a beamwidth of less than 5 degrees." For instance, the Airlinx Communications GE60 series (a commercial 60 GHz point-to-point product) have a beamwidth of 1.4° with a gain of 40 dBi and the Airlinx Communications GE60X series has a beamwidth of 0.6° with a gain of 46 dBi [27]. A probabilistic analysis is presented in [8] on the interference caused by uncoordinated transmissions in such highly directional 60 GHz networks. The analysis shows that "interference can essentially be ignored in the MAC design" and the links can be regarded as *pseudo-wired* [8]. We adopt such a pseudo-wired link model in this paper, as in prior works [8]–[11].[1]

A relay can be in one of the three states at each time slot: idle, transmitting, or receiving. If a relay is selected for a source, it receives from the source in the first hop. Once finishing the reception, the relay transmits the received packet to the specified destination in the second hop. Until the packet is successfully transmitted to the destination, the relay cannot receive more data from this or other sources due to the half-duplex operation. If a relay is not selected for any source, it stays in the idle state. This model is as that in prior works [10], [11], [28].

B. Dynamic Link Blockage Model

For a link l, denote C_l^t as the event that link l is unblocked at time slot t, and $\bar{C}_l^t$ the opposite case. Recall that link state follows a discrete-time two-state Markov chain. Let $0 \le p_l, q_l \le 1$ be the one-step transition probability from blocked to unblocked, and from unblocked to blocked, respectively. The one-step transition probability matrix of link l is

$$\mathbf{P}_l(1) = \begin{pmatrix} P(\bar{C}_l^{t+1}|\bar{C}_l^t) & P(C_l^{t+1}|\bar{C}_l^t) \\ P(\bar{C}_l^{t+1}|C_l^t) & P(C_l^{t+1}|C_l^t) \end{pmatrix} = \begin{pmatrix} 1-p_l & p_l \\ q_l & 1-q_l \end{pmatrix}.$$

[1]Note that the network model considered in [8] is an outdoor mesh network. Although the network model considered in this paper includes a centralized controller, the network is actually an ad hoc network since all the transmissions are between transmitter-receiver pairs, rather than through the centralized controller. The centralized controller only collects information from the network, through a common control channel (e.g., a larger range WiFi channel), and computes the transmission schedule. It is not involved in data transmissions. Furthermore, the analysis and conclusion in [8] is actually based on the narrow beamwidth assumption, which can be satisfied in a practical 60 GHz networks [26]. If interference is not negligible, the interference link model considered in our prior work [19]–[21] can be used, which does not change the structure of the problem and solution, but will make the problem more complicated.

The n-step transition probability matrix of link l is [29]

$$\begin{aligned}\mathbf{P}_l(n) &= \begin{pmatrix} P(\bar{C}_l^{t+n}|\bar{C}_l^t) & P(C_l^{t+n}|\bar{C}_l^t) \\ P(\bar{C}_l^{t+n}|C_l^t) & P(C_l^{t+n}|C_l^t) \end{pmatrix} \\ &= \frac{1}{p_l+q_l}\begin{pmatrix} q_l & p_l \\ q_l & p_l \end{pmatrix} + \frac{(1-p_l-q_l)^n}{p_l+q_l}\begin{pmatrix} p_l & -p_l \\ -q_l & q_l \end{pmatrix} \\ &= \begin{pmatrix} 1-p_l(n) & p_l(n) \\ q_l(n) & 1-q_l(n) \end{pmatrix}, \quad n = 1, 2, \ldots. \end{aligned} \tag{1}$$

C. Expected Delivery Time (EDT)

We consider two types of SD pairs. The first type, denoted as $\mathcal{S}_i$, is that the source and destination are within one-hop distance with each other and are not permanently blocked (e.g., by a wall). Hence the SD pair can either communication with each other directly, or use a relay if the direct link is poor. The second type, denoted as $\mathcal{S}_j$, is that the SD pair are either out of range or blocked by a permanent obstacle between them. Thus a relay is needed for them to communicate with each other. Define $\mathcal{S}_{i\cup j} = \mathcal{S}_i \cup \mathcal{S}_j$. We next derive the expected delivery time (EDT) for the relay-assisted and direct transmission cases.

1) EDT via Relay: Let s denote a source with destination $d(s)$, and r be a relay that can communicate directly with both s and $d(s)$. Denote a link between s and r as l_{sr}, and the set of all s–r links as $\mathcal{L}_{\mathrm{sr}}$. Similarly, we define $l_{\mathrm{rd}(s)}$ as a link between r and $d(s)$ and $\mathcal{L}_{\mathrm{rd}(s)}$ as the set of these links.

Let $T_{l_{\mathrm{sr}}}$ be the delivery time from s to r when link l_{sr} is used in the first hop for a block of data no greater than the channel capacity (which is normalized to a time slot), and T be the current time slot, $T \geq 1$. The expectation $\mathbb{E}(T_{l_{\mathrm{sr}}})$ is the average number of trials until the first successful transmission happens on l_{sr}, which can be expressed as

$$\begin{aligned}\mathbb{E}(T_{l_{\mathrm{sr}}}) &= \sum_{t=1}^{\infty} tP(T_{l_{\mathrm{sr}}} = t) \\ &= 1 \cdot P(C_{l_{\mathrm{sr}}}^T) + \sum_{t=2}^{\infty} tP(\bar{C}_{l_{\mathrm{sr}}}^T)(1-p_{l_{\mathrm{sr}}})^{(t-2)}p_{l_{\mathrm{sr}}} \\ &= 1 + \frac{1-P(C_{l_{\mathrm{sr}}}^T)}{p_{l_{\mathrm{sr}}}}, \end{aligned} \tag{2}$$

where $P(C_{l_{\mathrm{sr}}}^T)$ is the probability that l_{sr} is unblocked at T.

Now let $T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}$ be the delivery time from r to $d(s)$ when link l_{sr} is chosen in the first hop and link $l_{\mathrm{rd}(s)}$ is used in the second hop. To derive $\mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}})$, we first note that

$$\begin{aligned}\mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}|T_{l_{\mathrm{sr}}} = t) &= \sum_{t'=1}^{\infty} t'P(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}} = t'|T_{l_{\mathrm{sr}}} = t) \\ &= 1 + \frac{1-P(C_{l_{\mathrm{rd}(s)}}^{T+t})}{p_{l_{\mathrm{rd}(s)}}}. \end{aligned} \tag{3}$$

According to the *law of total expectation*, we have

$$\begin{aligned}\mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}) &= \mathbb{E}(\mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}|T_{l_{\mathrm{sr}}} = t)) \\ &= \sum_{t=1}^{\infty} \mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}|T_{l_{\mathrm{sr}}} = t)P(T_{l_{\mathrm{sr}}} = t) \\ &= 1 + \frac{1}{p_{l_{\mathrm{rd}(s)}}} - \frac{1}{p_{l_{\mathrm{rd}(s)}}}\sum_{t=1}^{\infty} P(C_{l_{\mathrm{rd}(s)}}^{T+t})P(T_{l_{\mathrm{sr}}} = t). \end{aligned} \tag{4}$$

To calculate $\mathbb{E}(T_{l_{\mathrm{sr}}})$ in (2) and $\mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}})$ in (4), we need to derive $P(C_{l_{\mathrm{sr}}}^T)$ and $P(C_{l_{\mathrm{rd}(s)}}^{T+t})$. Let t_l be the last time (before the current time T) that PNC knew the state of link l (being either blocked or unblocked). We have

$$\begin{aligned}P(C_l^T) &= P(\bar{C}_l^{t_l})p_l(T-t_l) + P(C_l^{t_l})(1-q_l(T-t_l)) \\ &= \frac{p_l}{p_l+q_l} + \frac{(1-p_l-q_l)^{T-t_l}}{p_l+q_l}\left(q_l + P(C_l^{t_l})(p_l+q_l)\right). \end{aligned} \tag{5}$$

Furthermore, we derive the summation term in (4), which is given below.

$$\begin{aligned}&\sum_{t=1}^{\infty} P\left(C_{l_{\mathrm{rd}(s)}}^{T+t}\right)P(T_{l_{\mathrm{sr}}} = t) = \sum_{t=1}^{\infty}\Bigg(P\left(\bar{C}_{l_{\mathrm{rd}(s)}}^{T_{l_{\mathrm{rd}(s)}}}\right)p_{l_{\mathrm{rd}(s)}}(T+t \\ &- T_{l_{\mathrm{rd}(s)}}) + P\left(C_{l_{\mathrm{rd}(s)}}^{T_{l_{\mathrm{rd}(s)}}}\right)(1-q_{l_{\mathrm{rd}(s)}}(T+t-T_{l_{\mathrm{rd}(s)}}))\Bigg)P(T_{l_{\mathrm{sr}}} = t) \\ &= \frac{P(\bar{C}_{l_{\mathrm{sr}}}^T)}{p_{l_{rd}} + q_{l_{\mathrm{rd}(s)}}} + \frac{P(\bar{C}_{l_{\mathrm{sr}}}^T)(p_{l_{\mathrm{rd}(s)}} + q_{l_{\mathrm{rd}(s)}})p_{l_{\mathrm{sr}}}}{p_{l_{\mathrm{rd}(s)}} + q_{l_{\mathrm{rd}(s)}}} \\ &\cdot \frac{(1-p_{l_{\mathrm{sr}}} - q_{l_{\mathrm{rd}(s)}})^{T-T_{l_{\mathrm{rd}(s)}}+2}}{(1-(1-p_{l_{\mathrm{rd}(s)}} - q_{l_{\mathrm{rd}(s)}})(1-p_{l_{\mathrm{sr}}}))}. \end{aligned} \tag{6}$$

Substituting (5) and (6) into (2) and (4), we thus derive the closed-form expression for the EDT when link l_{sr} and link $l_{\mathrm{rd}(s)}$ are chosen for the two-hop relay path, denoted as $\mathbb{E}(T_{\mathrm{srd}(s)})$, which is given by

$$\mathbb{E}(T_{\mathrm{srd}(s)}) = \mathbb{E}(T_{l_{\mathrm{sr}}}) + \mathbb{E}(T_{l_{\mathrm{sr}},\, l_{\mathrm{rd}(s)}}). \tag{7}$$

2) EDT via Direct Link: Consider the case when s and $d(s)$ use a direct link $l_{\mathrm{sd}(s)}$ between them to communicate without using a relay, where $s \in \mathcal{S}_i$. The EDT from s to $d(s)$ via link $l_{\mathrm{sd}(s)}$, denoted as $\mathbb{E}(T_{\mathrm{sd}(s)})$, can be derived as $\mathbb{E}(T_{\mathrm{sd}(s)}) = 1 + \frac{1-P(C_{l_{\mathrm{sd}(s)}}^T)}{p_{l_{\mathrm{sd}(s)}}}$.

D. Problem Formulation

Let $\mathcal{R}(s)$ be the set of relays that can communicate directly with both source s and its destination $d(s)$, and $\mathcal{S}(r)$ be the set of sources that can communicate directly with relay r. Denote all the relays and sources as $\mathcal{R}$ and $\mathcal{S}_{i\cup j}$, respectively. Let $\mathcal{L}_{\mathrm{sd}(s)}$ be the set of all s-$d(s)$ links. We then define the following decision

variables.

$$x_{l_{\rm sr}} = \begin{cases} 1, & \text{source } s \text{ transmits on link } l_{\rm sr} \text{ in hop 1} \\ 0, & \text{otherwise,} \end{cases}$$

$$\forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), l_{\rm sr} \in \mathcal{L}_{\rm sr} \tag{8}$$

$$x_{l_{{\rm rd}(s)}} = \begin{cases} 1, & \text{relay } r \text{ transmits on link } l_{{\rm rd}(s)} \text{ in hop 2} \\ 0, & \text{otherwise} \end{cases}$$

$$\forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), l_{{\rm rd}(s)} \in \mathcal{L}_{{\rm rd}(s)} \tag{9}$$

$$x_{l_{{\rm sd}(s)}} = \begin{cases} 1, & \text{source } s \text{ transmits to its destination } d(s) \\ & \text{via direct link } l_{{\rm sd}(s)} \\ 0, & \text{otherwise} \end{cases}$$

$$\forall\, s \in \mathcal{S}_{i\cup j}, l_{{\rm sd}(s)} \in \mathcal{L}_{{\rm sd}(s)}. \tag{10}$$

Since each relay r can be selected by at most one SD pair and only one link can be selected at each hop, we have

$$\sum_{s\in\mathcal{S}(r)} \sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} x_{l_{\rm sr}} \le 1, \forall r \in \mathcal{R}. \tag{11}$$

Note that if relay r is selected by s in hop 1, then r must also be selected by $d(s)$ in hop 2, i.e.,

$$\sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} x_{l_{\rm sr}} = \sum_{l_{{\rm rd}(s)}\in\mathcal{L}_{{\rm rd}(s)}} x_{l_{{\rm rd}(s)}}, \forall s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s). \tag{12}$$

An SD pair can use either a relay or a direct link to communicate directly, but not both. So we have the following constraint.

$$\sum_{r\in\mathcal{R}(s)} \sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} x_{l_{\rm sr}} + \sum_{l_{{\rm sd}(s)}\in\mathcal{L}_{{\rm sd}(s)}} x_{l_{{\rm sd}(s)}} = 1, \forall\, s \in \mathcal{S}_{i\cup j}. \tag{13}$$

Furthermore, a type $\mathcal{S}_j$ SD pair has to use a relay, i.e.,

$$\sum_{r\in\mathcal{R}(s)} \sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} x_{l_{\rm sr}} = 1, \forall\, s \in \mathcal{S}_j. \tag{14}$$

If a relay is selected for an SD pair with source s and destination $d(s)$, the EDT from s to $d(s)$, denoted as g_s, is

$$g_s = \sum_{r\in\mathcal{R}(s)} \sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} \mathbb{E}(T_{l_{\rm sr}}) x_{l_{\rm sr}} + \sum_{r\in\mathcal{R}(s)} \sum_{l_{\rm sr}\in\mathcal{L}_{\rm sr}} x_{l_{\rm sr}} \sum_{l_{{\rm rd}(s)}\in\mathcal{L}_{{\rm rd}(s)}} \mathbb{E}(T_{l_{\rm sr}, l_{{\rm rd}(s)}}) x_{l_{{\rm rd}(s)}}. \tag{15}$$

If a relay is not selected, the EDT from s to $d(s)$, denoted as u_s, is

$$u_s = \sum_{l_{{\rm sd}(s)}\in\mathcal{L}_{{\rm sd}(s)}} \mathbb{E}(T_{l_{{\rm sd}(s)}}) x_{l_{{\rm sd}(s)}}. \tag{16}$$

Before our problem formulation, we first need to determine whether each SD pair in $\mathcal{S}_j$ can have a relay. Recall that an $\mathcal{S}_j$ type SD pair must use a relay to communicate. Let index variable $y_{\rm rs} = 1$ denote that relay r is assigned to SD pair s, and $y_{\rm rs} = 0$ otherwise. This problem can be formulated as follows.

$$\textbf{P0}: \max: \sum_{s\in\mathcal{S}_j} \sum_{r\in\mathcal{R}(s)} y_{\rm rs} \tag{17}$$

$$\text{s.t.} \sum_{s\in\mathcal{S}(r)\cap\mathcal{S}_j} y_{\rm rs} \le 1,\ \forall\, r \in \mathcal{R} \tag{18}$$

$$\sum_{r\in\mathcal{R}(s)} y_{\rm rs} \le 1,\ \forall\, s \in \mathcal{S}_j \tag{19}$$

$$y_{\rm rs} = 0, \forall\, r \in \mathcal{R}, s \in \mathcal{S}_j, \text{and } r \notin \mathcal{R}(s). \tag{20}$$

To show how problem **P0** determines whether there are enough relays for the network, let's denote $\|\mathcal{R}\|$ as the number of relays in the network, and Y as the Objective Function Value (OFV) of problem **P0**. The objective of problem **P0** is to allocate relays to as many type $\mathcal{S}_j$ SD pairs as possible, under the constraints that:

1) each relay can serve at most one SD pair;
2) each SD pair can be served by at most one relay;
3) a relay r can serve an SD pair with source s only if $r \in \mathcal{R}(s)$.

Then Y will always be no greater than $\|\mathcal{R}\|$. If $Y \ge \|\mathcal{S}_j\|$, then each SD pair in $\mathcal{S}_j$ can be served by a relay; otherwise, there are some SD pairs in $\mathcal{S}_j$ that cannot have a relay, since the maximum number of type $\mathcal{S}_j$ SD pairs that can have a relay in the optimal solution is less than the total number of type $\mathcal{S}_j$ SD pairs. We next examine two cases.

1) Case 1: When $Y \ge \|\mathcal{S}_j\|$: This is the case when each SD pair in $\mathcal{S}_j$ can have a relay. In this case, our objective is to minimize the **MEDT** among all the SD pairs. We thus have the following problem formulation.

$$\textbf{P1}: \min:\ \max_{s\in\mathcal{S}_{i\cup j}} \{g_s + u_s\}$$
$$\text{s.t.}\ (8)-(14). \tag{21}$$

When $Y \ge \|\mathcal{S}_j\|$, problem **P1** must have a solution. Although the constraints are linear, the objective function is not. Therefore problem **P1** is a nonlinear integer programming problem (NIP), which is generally NP-hard. In the next section, we propose a Decomposition Principle to solve this problem.

2) Case 2: When $Y < \|\mathcal{S}_j\|$: In this case, problem **P1** is not applicable since a type $\mathcal{S}_j$ SD pair may not have a relay to forward its data, if all relays within range are assigned to other SD pairs. We develop a heuristic algorithm to address this case in Section III-G.

III. Problem Decomposition and Solution

In this section, we present the Decomposition Principle for the case when each SD pair in $\mathcal{S}_j$ can have a relay, which breaks down problem **P1** into a subproblem of link selection and another subproblem of relay selection. The basic idea is to determine the link selection for each relay first, and then determine the relay selection based on the result of link selection. Moreover, the link selection sub-problem can be further decomposed into three sub-problems: (i) link selection in hop 1, (ii) link selection in hop 2, and (iii) direct link selection. We develop effective algorithms to solve the decomposed problems,

and more important, prove a tight bound on the optimality gap for the decomposition principle solution. In the case that there are no enough relays for the SD pairs in $\mathcal{S}_j$, we develop a heuristic algorithm that can still produce highly competitive solutions in Section III-G.

A. Optimal Choice and Greedy Choice

We first define an optimal choice, Optimal Choice 1 (OC1), and a greedy choice, Greedy Choice 1 (GC1), as follows.

1) *Optimal Choice 1 (OC1):* Given a link $l_{\mathrm{rd}(s)}$ in hop 2, choose the hop 1 link as

$$l_{\mathrm{sr}}^{*} = \arg\min_{l_{\mathrm{sr}} \in \mathcal{L}_{\mathrm{sr}}} \{\mathbb{E}(T_{l_{\mathrm{sr}}}) + \mathbb{E}(T_{l_{\mathrm{sr}}, l_{\mathrm{rd}(s)}})\}. \tag{22}$$

That is, choose the hop 1 link that minimizes the EDT from s to $d(s)$ for a given hop 2 link.

2) *Greedy Choice 1 (GC1):* Given a link $l_{\mathrm{rd}(s)}$ in hop 2, choose the hop 1 link as

$$l_{\mathrm{sr}}^{+} = \arg\min_{l_{\mathrm{sr}} \in \mathcal{L}_{\mathrm{sr}}} \mathbb{E}(T_{l_{\mathrm{sr}}}). \tag{23}$$

That is, choose the hop 1 link that minimizes the EDT from s to r for a chosen hop 2 link.

Obviously, the choice of l_{sr}^{*} depends on $l_{\mathrm{rd}(s)}$ but that of l_{sr}^{+} does not. We have the following theorem for OC1 and GC1.

Theorem 1: For a given relay r and a hop 2 link, GC1 can achieve an EDT from s to $d(s)$ via r that is at most one time slot greater than that achieved by OC1.

Proof: Let the hop 2 link be l, and recall that l_{sr}^{+} and l_{sr}^{*} are the links chosen by GC1 and OC1, respectively. For two time slots t_1 and t_2, denote $\Delta_t = t_2 - t_1$, which is an integer.

We consider the following four cases.

1) *Case 1: $p_l + q_l \le 1$ and $\Delta_t \ge 0$:* From (3) and (5), we have (24) given on Page 13. Since $0 < (1 - p_l - q_l)^{\Delta_t} < 1$, then $-1 < (1 - p_l - q_l)^{\Delta_t} - 1 < 0$. And since $-p_l \le q_l - P(C_l^{t_l})(p_l + q_l) \le q_l$, we have Eqn (24) as shown at the bottom of this page

$$-1 \le -\frac{p_l}{p_l + q_l} \le \frac{q_l - P(C_l^{t_l})(p_l + q_l)}{p_l + q_l} \le \frac{q_l}{p_l + q_l} \le 1.$$

Since $(1 - p_l - q_l)^{T + t_2 - t_l} < 1$, it follows (24) that

$$\left|\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l} | T_{l_{\mathrm{sr}}^{+}} = t_1) - \mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l} | T_{l_{\mathrm{sr}}^{*}} = t_2)\right| \le 1, \tag{26}$$

where $|\cdot|$ denotes the absolute value.

(2) *Case 2: $p_l + q_l \le 1$ and $\Delta_t < 0$:* A similar reasoning as case 1 yields the following inequality.

$$\left|\mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l} | T_{l_{\mathrm{sr}}^{*}} = t_2) - \mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l} | T_{l_{\mathrm{sr}}^{+}} = t_1)\right| \le 1$$
$$\Rightarrow \text{Inequality (26)}. \tag{27}$$

(3) *Case 3: $1 < p_l + q_l \le 2$ and $\Delta_t \ge 0$:* It follows that

$$|(1 - p_l - q_l)^{\Delta_t} - 1| \le |(1 - p_l - q_l)^{\Delta_t}| + 1 \le 1$$
$$+ |1 - p_l - q_l| = 1 - (1 - p_l - q_l) = p_l + q_l. \tag{28}$$

Then we have (25) as shown at the bottom of this page which implies inequality (26).

(4) *Case 4: $1 < p_l + q_l \le 2$ and $\Delta_t < 0$:* Again, a similar reasoning as the above yields (27).

From Cases 1, 2, 3 and 4, we conclude that for all $t_1 \ge 1, t_2 \ge 1$, and $0 \le p_l + q_l \le 2$, inequality (26) holds true.

According to (4), we can compute $\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l}) - \mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l})$ as inequality (29) as shown at the bottom of this page.

$$\mathbb{E}\left(T_{l_{\mathrm{sr}}^{+}, l} | T_{l_{\mathrm{sr}}^{+}} = t_1\right) - \mathbb{E}\left(T_{l_{\mathrm{sr}}^{*}, l} | T_{l_{\mathrm{sr}}^{*}} = t_2\right) = \frac{P(C_{rd}^{T+t_2}) - P(C_{rd}^{T+t_1})}{p_{l_{\mathrm{rd}(s)}}} = \frac{(1 - P(C_l^{t_l}))q_l - P(C_l^{t_l})p_l}{p_l + q_l}((1 - p_l - q_l)^{T+t_2-t_l}$$
$$- (1 - p_l - q_l)^{T+t_1-t_l}) = \frac{q_l - P(C_l^{t_l})(p_l + q_l)}{p_l + q_l}((1 - p_l - q_l)^{T+t_2-t_l})((1 - p_l - q_l)^{\Delta_t} - 1). \tag{24}$$

$$\left|\frac{q_l - P(C_l^{t_l})(p_l + q_l)}{p_l + q_l}((1 - p_l - q_l)^{T+t_2-t_l})((1 - p_l - q_l)^{\Delta_t} - 1)\right| = \left|\frac{q_l - P(C_l^{t_l})(p_l + q_l)}{p_l + q_l}\right| \cdot \left|(1 - p_l - q_l)^{T+t_2-t_l}\right|$$
$$\cdot \left|(1 - p_l - q_l)^{\Delta_t} - 1\right| \le \left|q_l - P(C_l^{t_l})(p_l + q_l)\right| \cdot \frac{1}{p_l + q_l} \cdot \left|(1 - p_l - q_l)^{T+t_2-t_l}\right| (p_l + q_l)$$
$$= \left|q_l - P(C_l^{t_l})(p_l + q_l)\right| \cdot \left|(1 - p_l - q_l)^{T+t_2-t_l}\right| \le 1. \tag{25}$$

$$\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l}) - \mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l}) = \mathbb{E}\left(\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l} - T_{l_{\mathrm{sr}}^{*}, l} \left| T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right.)\right) = \sum_{t_1, t_2} \mathbb{E}\left(T_{l_{\mathrm{sr}}^{+}, l} - T_{l_{\mathrm{sr}}^{*}, l} \left| T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right.\right)$$
$$\cdot P\left(T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right) = \sum_{t_1, t_2} \left(\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l} \left| T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right.) - \mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l} \left| T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right.)\right) \cdot P\left(T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2\right)$$
$$= \sum_{t_1, t_2} \left(\mathbb{E}(T_{l_{\mathrm{sr}}^{+}, l} \left| T_{l_{\mathrm{sr}}^{+}} = t_1\right.) - \mathbb{E}(T_{l_{\mathrm{sr}}^{*}, l} \left| T_{l_{\mathrm{sr}}^{*}} = t_2\right.)\right) \cdot P(T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2) \le \sum_{t_1, t_2} 1 \cdot P(T_{l_{\mathrm{sr}}^{+}} = t_1, T_{l_{\mathrm{sr}}^{*}} = t_2) = 1. \tag{29}$$

Recall $\mathbb{E}(T_{l_{\rm sr}^+}) = \min_{l_{\rm sr} \in \mathcal{L}_{\rm sr}} \mathbb{E}(T_{l_{\rm sr}})$. Assume $\mathbb{E}(T_{l_{\rm sr}^+}) - \mathbb{E}(T_{l_{\rm sr}^*}) = P(C^T_{l_{\rm sr}^*})/p_{l_{\rm sr}^*} - P(C^T_{l_{\rm sr}^+})/p_{l_{\rm sr}^+} = \alpha \le 0$. By (29) we have

$$\begin{aligned} -1+\alpha \le\ & \mathbb{E}(T_{l_{\rm sr}^+,l}) + \mathbb{E}(T_{l_{\rm sr}^+}) - \mathbb{E}(T_{l_{\rm sr}^*,l}) - \mathbb{E}(T_{l_{\rm sr}^*}) \\ \le\ & 1+\alpha \le 1. \end{aligned} \tag{30}$$

Moreover, if $\alpha \le -1$, then we have $\mathbb{E}(T_{l_{\rm sr}^+,l}) + \mathbb{E}(T_{l_{\rm sr}^+}) - \mathbb{E}(T_{l_{\rm sr}^*,l}) - \mathbb{E}(T_{l_{\rm sr}^*}) \le 0$. Recall that $\mathbb{E}(T_{l_{\rm sr}^*}) + \mathbb{E}(T_{l_{\rm sr}^*,l}) = \min_{l_{\rm sr} \in \mathcal{L}_{\rm sr}} \{\mathbb{E}(T_{l_{\rm sr}}) + \mathbb{E}(T_{l_{\rm sr},l})\}$. Therefore $\mathbb{E}(T_{l_{\rm sr}^+,l}) + \mathbb{E}(T_{l_{\rm sr}^+}) - \mathbb{E}(T_{l_{\rm sr}^*,l}) - \mathbb{E}(T_{l_{\rm sr}^*}) = 0$, which means GC1 equals to OC1 in terms of EDT from s to $d(s)$.

Thus we conclude that Theorem 1 holds true. ■

In the following, we show how to use GC1 to reduce problem **P1** into a simpler problem.

B. Link Selection in Hop 1

The problem is to minimize the MEDT among all the SD pairs while there are more than one relays. With GC1, we consider links $l_{\rm sr}^+$, for all $s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s)$ in hop 1 of problem **P1**, as

$$\begin{aligned} & x_{l_{\rm sr}^+} \in \{0,1\}, x_{l_{\rm sr}} = 0, \forall\, l_{\rm sr} \ne l_{\rm sr}^+, l_{\rm sr} \in \mathcal{L}_{\rm sr}, s \in \mathcal{S}_{i\cup j}, \\ & r \in \mathcal{R}(s). \end{aligned} \tag{31}$$

Substituting constraint (31) into problem **P1**, then we have a reduced problem, termed problem **P2**, as follows.

$$\begin{aligned} \textbf{P2}:\ \min:\ & \max_{s\in\mathcal{S}_{i\cup j}} \Bigg\{ \sum_{r\in\mathcal{R}(s)} \mathbb{E}(T_{l_{\rm sr}^+}) x_{l_{\rm sr}^+} \\ & + \sum_{r\in\mathcal{R}(s)} x_{l_{\rm sr}^+} \sum_{l_{{\rm rd}(s)}\in\mathcal{L}_{{\rm rd}(s)}} \mathbb{E}(T_{l_{\rm sr}^+, l_{{\rm rd}(s)}}) x_{l_{{\rm rd}(s)}} \\ & + \sum_{l_{{\rm sd}(s)}\in\mathcal{L}_{{\rm sd}(s)}} \mathbb{E}(T_{l_{{\rm sd}(s)}}) x_{l_{{\rm sd}(s)}} \Bigg\} \end{aligned} \tag{32}$$

$$\text{s.t.} \sum_{s\in\mathcal{S}(r)} x_{l_{\rm sr}^+} \le 1,\ \forall\, r \in \mathcal{R}, \tag{33}$$

$$x_{l_{\rm sr}^+} = \sum_{l_{{\rm rd}(s)}\in\mathcal{L}_{{\rm rd}(s)}} x_{l_{{\rm rd}(s)}},\ \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), \tag{34}$$

$$\sum_{r\in\mathcal{R}(s)} x_{l_{\rm sr}^+} + \sum_{l_{{\rm sd}(s)}\in\mathcal{L}_{{\rm sd}(s)}} x_{l_{{\rm sd}(s)}} = 1,\ \forall\, s \in \mathcal{S}_{i\cup j} \tag{35}$$

$$\sum_{r\in\mathcal{R}(s)} x_{l_{\rm sr}^+} = 1, \forall\, s \in \mathcal{S}_j \tag{36}$$

$$\begin{aligned} & x_{l_{\rm sr}^+} = \{0,1\},\ \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s) \\ & \text{Constraints (9) and (10).} \end{aligned} \tag{37}$$

The number of decision variables of problem **P2** is much less than that of problem **P1**. We will prove below that the difference between the OFV of problem **P2** and that of problem **P1** is at most one time slot. We first introduce a lemma as a basis of the proof. For ease of presentation, let $\mathcal{S}_1$ denote the set of sources that are assigned with relays in the optimal solution to problem **P1**, i.e., $z_s = \mathbb{E}(T_{l_{\rm sr}}) + \mathbb{E}(T_{l_{\rm sr}, l_{{\rm rd}(s)}})$, for all $s \in \mathcal{S}_1$, and $\mathcal{S}_2$ be the set of sources that are not assigned with relays and communicate with their destinations using a direct link, i.e., $z_s = \mathbb{E}(T_{l_{{\rm sd}(s)}})$, for all $s \in \mathcal{S}_2$. Also denote $\mathcal{S}_{1\cup 2} = \mathcal{S}_1 \cup \mathcal{S}_2$. Note that $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}_i \cup \mathcal{S}_j$.

Lemma 1: Denote $\phi^* = \{x_{l_{\rm sr}^*} = 1, x_{l^*_{{\rm rd}(s)}} = 1,\ \forall\, s \in \mathcal{S}_1$, and $x_{l^*_{{\rm sd}(s)}} = 1, \forall\, s \in \mathcal{S}_2\}$ as the optimal solution to problem **P1**. For all $s \in \mathcal{S}_1$, set $x_{l_{\rm sr}^+} = 1$ and then set $x_{l_{\rm sr}^*} = 0$. Then $\phi = \{x_{l_{\rm sr}^+} = 1, x_{l^*_{{\rm rd}(s)}} = 1, \forall\, s \in \mathcal{S}_1$, and $x_{l^*_{{\rm sd}(s)}} = 1, \forall\, s \in \mathcal{S}_2\}$ is a feasible solution to problem **P2**.

Proof: Comparing ϕ with ϕ^*, only the hop 1 link choice is different. Since for all s, we set $x_{l_{\rm sr}^+} = 1$ and then set $x_{l_{\rm sr}^*} = 0$, the link choice of hop 1 still satisfies all the constraints in problem **P2**. Hence ϕ is a feasible solution to problem **P2**. ■

Theorem 2: The OFV of problem **P2** is at most one time slot greater than that of problem **P1**.

Proof: If ϕ^* is also the optimal solution to problem **P2**, denote the corresponding EDT of each SD pair as z_s^*, for all $s \in \mathcal{S}_{1\cup 2}$, and the OFV of problem **P1** as $\mathbf{z} = \max_{s\in\mathcal{S}_{1\cup 2}}\{z_s^*\}$. if ϕ is a feasible solution to problem **P2**, denote the corresponding EDT of each SD pair as z_s^+, for all $s \in \mathcal{S}_{1\cup 2}$. Then the difference between the OFV of problem **P2** and that of problem **P1** can be written as

$$\max_{s\in\mathcal{S}_{1\cup 2}} z_s^+ - \mathbf{z} = \max_{s\in\mathcal{S}_{1\cup 2}} \{z_s^+ - \mathbf{z}\}. \tag{38}$$

Since links $\{l_{\rm sr}^+, \forall\, s, r \in \mathcal{R}(s)\}$ are chosen by GC1, it follows Theorem 1 that $z_s^+ - z_s^* \le 1$, for all $s \in \mathcal{S}_1$. Besides, we have $z_s^+ = z_s^* = \mathbb{E}(T_{l^*_{{\rm sd}(s)}})$, for all $s \in \mathcal{S}_2$. Thus we have

$$z_s^+ - z_s^* \le 1,\ \forall\, s \in \mathcal{S}_{1\cup 2}. \tag{39}$$

Since $\mathbf{z} \ge z_s^*$, for all $s \in \mathcal{S}_{1\cup 2}$, we have

$$\begin{aligned} & z_s^+ \le z_s^* + 1 \le \mathbf{z} + 1 \Rightarrow z_s^+ - \mathbf{z} \le 1, \forall s \in \mathcal{S}_{1\cup 2}, \\ & \Rightarrow \max_{s\in\mathcal{S}_{1\cup 2}} \{z_s^+ - \mathbf{z}\} \le 1 \Rightarrow \max_{s\in\mathcal{S}_{1\cup 2}} \{z_s^+\} - \mathbf{z} \le 1. \end{aligned}$$

Thus we conclude that Theorem 2 holds true. ■

C. Link Selection in Hop 2

Lemma 1 indicates that $\phi = \{x_{l_{\rm sr}^+} = 1, x_{l^*_{{\rm rd}(s)}} = 1, \forall\, s \in \mathcal{S}_1$, and $x_{l^*_{{\rm sd}(s)}} = 1, \forall\, s \in \mathcal{S}_2\}$ is a feasible, but not necessarily optimal solution to problem **P2**. Furthermore, $l^*_{{\rm rd}(s)}, \forall\, s \in \mathcal{S}_1$ is hard to obtain because it requires computing the EDT of all possible links in hops 1 and 2. To obtain the optimal solution to problem **P2**, we first define another greedy choice, termed Greedy Choice 2 (GC2), as follows.

1) *Greedy Choice 2 (GC2):* Given a link $l_{\rm sr}^+$ obtained by GC1 in hop 1, choose the hop 2 link $l^+_{{\rm rd}(s)}$ as

$$l^+_{{\rm rd}(s)} = \arg\min_{l_{{\rm rd}(s)}\in\mathcal{L}_{{\rm rd}(s)}} \mathbb{E}(T_{l_{\rm sr}^+, l_{{\rm rd}(s)}}). \tag{40}$$

That is, choose the hop 2 link that minimizes the EDT from r to $d(s)$ for given hop 1 link $l_{\rm sr}^+$.

With GC2, we only consider links $l^+_{\mathrm{rd}(s)}$, for all $s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s)$ in hop 2 for problem **P2**, which means

$$x_{l^+_{\mathrm{rd}(s)}} \in \{0,1\}, x_{l_{\mathrm{rd}(s)}} = 0,\ \forall\, l_{\mathrm{rd}(s)} \neq l^+_{\mathrm{rd}(s)}, l_{\mathrm{rd}(s)} \in \mathcal{L}_{\mathrm{rd}(s)},\ \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s). \tag{41}$$

Then we have the following claims for the optimal solution to problem **P2**.

Lemma 2: Denote $\hat{\phi}^* = \{x_{l^+_{\mathrm{sr}}} = 1, x_{l^*_{\mathrm{rd}(s)}} = 1,\ \forall\, s \in \mathcal{S}_1, x_{l^*_{\mathrm{sd}(s)}} = 1, \forall\, s \in \mathcal{S}_2\}$ as the optimal solution to problem **P2**. For all $s \in \mathcal{S}_1$, set $x_{l^+_{\mathrm{rd}(s)}} = 1$ and $x_{l^*_{\mathrm{rd}(s)}} = 0$. Then $\hat{\phi} = \{x_{l^+_{\mathrm{sr}}} = 1, x_{l^+_{\mathrm{rd}(s)}} = 1, \forall\, s \in \mathcal{S}_1, x_{l^*_{\mathrm{sd}(s)}} = 1, \forall\, s \in \mathcal{S}_2\}$ is a feasible solution to problem **P2**.

Proof: Comparing $\hat{\phi}$ with $\hat{\phi}^*$, only the hop 2 link choice is different. Since for all s, we set $x_{l^+_{\mathrm{rd}(s)}} = 1$ and then set $x_{l^*_{\mathrm{rd}(s)}} = 0$, the link choice of hop 2 still satisfies all the constraints in problem **P2**. Hence $\hat{\phi}$ is a feasible solution to problem **P2**. ■

Lemma 3: With the optimal solution to problem **P2**, the link selection problem in hop 2 is $\{x_{l^+_{\mathrm{rd}(s)}} \in \{0,1\}, x_{l_{\mathrm{rd}(s)}} = 0, \forall\, l_{\mathrm{rd}(s)} \neq l^+_{\mathrm{rd}(s)}, l_{\mathrm{rd}(s)} \in \mathcal{L}_{\mathrm{rd}(s)}, \forall\, s \in \mathcal{S}_1, r \in \mathcal{R}(s)\}$.

Proof: Recall $\hat{\phi}$ is a feasible solution to problem **P2**. With this solution, define $z_s^{++} = \mathbb{E}(T_{l^+_{\mathrm{sr}}}) + \mathbb{E}(T_{l^+_{\mathrm{sr}}, l^+_{\mathrm{rd}(s)}})$, for all $s \in \mathcal{S}_1$, and $z'_s = \mathbb{E}(T_{l^+_{\mathrm{sr}}}) + \mathbb{E}(T_{l^+_{\mathrm{sr}}, l_{\mathrm{rd}(s)}})$, for all $l_{\mathrm{rd}(s)} \neq l^+_{\mathrm{rd}(s)}, l_{\mathrm{rd}(s)} \in \mathcal{L}_{\mathrm{rd}(s)}$, for all $s \in \mathcal{S}_1$.

Since $\mathbb{E}(T_{l^+_{\mathrm{sr}}, l^+_{\mathrm{rd}(s)}}) = \min_{l_{\mathrm{rd}(s)} \in \mathcal{L}_{\mathrm{rd}(s)}} \mathbb{E}(T_{l^+_{\mathrm{sr}}, l_{\mathrm{rd}(s)}})$, we have $z_s^{++} \leq z'_s$, for all $s \in \mathcal{S}_1$. We also have $z_s^{++} = z'_s = \mathbb{E}(T_{l^*_{\mathrm{sd}(s)}})$, for all $s \in \mathcal{S}_2$.

Let $\mathbf{z}' = \max_{s \in \mathcal{S}_{1\cup 2}}\{z'_s\}$. It follows that

$$z_s^{++} \leq z'_s \leq \mathbf{z}',\ \forall s\ \in \mathcal{S}_{1\cup 2}. \tag{42}$$

We thus have

$$\begin{aligned} &\max_{s \in \mathcal{S}_{1\cup 2}}\{z_s^{++}\} - \max_{s \in \mathcal{S}_{1\cup 2}}\{z'_s\} = \max_{s \in \mathcal{S}_{1\cup 2}}\{z_s^{++}\} - \mathbf{z}' \\ &= \max_{s \in \mathcal{S}_{1\cup 2}}\{z_s^{++} - \mathbf{z}'\} \leq 0 \\ &\Rightarrow \max_{s \in \mathcal{S}_{1\cup 2}}\{z_s^{++}\} \leq \max_{s \in \mathcal{S}_{1\cup 2}}\{z'_s\}, \end{aligned} \tag{43}$$

where the inequality in (43) is due to Inequality (42).

Note that the objective of problem **P2** is to minimize the MEDT among all SD pairs. So $\hat{\phi}$ is optimal to problem **P2** and Lemma 3 holds true. ■

D. Link Selection in the Direct Path

Link selection when the SD pair communicate directly, in the optimal solution to problem **P2**, can also be obtained with a greedy approach.

For all $s \in \mathcal{S}_j$ and $l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}$, set $\mathbb{E}(T_{l_{\mathrm{sd}(s)}})$ to an arbitrary constant. because the constraints will ensure that for all $s \in \mathcal{S}_j$ the direct link will not be selected. Define a greedy choice, termed Greedy Choice 3 (GC3), as follows.

1) *Greedy Choice 3 (GC3):* Choose the link in the direct SD path as $l^+_{\mathrm{sd}(s)} = \arg\min_{l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}} \mathbb{E}(T_{l_{\mathrm{sd}(s)}})$. That is, choose the link in the direct SD path that minimizes the EDT from s to $d(s)$.

Lemma 4: With the optimal solution to problem **P2**, the link selection in the direct path is $\{x_{l^+_{\mathrm{sd}(s)}} \in \{0,1\}, x_{l_{\mathrm{sd}(s)}} = 0, \forall\, l_{\mathrm{sd}(s)} \neq l^+_{\mathrm{sd}(s)}, l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}, \forall s \in \mathcal{S}_{i\cup j}\}$.

Proof: Define $\hat{z}_s = \mathbb{E}(T_{l^+_{\mathrm{sd}(s)}})$, for all $s \in \mathcal{S}_2$, and $\tilde{z}_s = \mathbb{E}(T_{l_{\mathrm{sd}(s)}})$, for all $l_{\mathrm{sd}(s)} \neq l^+_{\mathrm{sd}(s)}, l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}, s \in \mathcal{S}_2$. Then we have $\hat{z}_s \leq \tilde{z}_s$. We also have $\hat{z}_s = \tilde{z}_s$, for all $s \in \mathcal{S}_1$.

Let $\bar{z} = \max_{s \in \mathcal{S}_{1\cup 2}}\{\tilde{z}_s\}$. It follows that

$$\hat{z}_s \leq \tilde{z}_s \leq \bar{z},\ \forall s\ \in \mathcal{S}_{1\cup 2}. \tag{44}$$

We thus have

$$\begin{aligned} &\max_{s \in \mathcal{S}_{1\cup 2}}\{\hat{z}_s\} - \max_{s \in \mathcal{S}_{1\cup 2}}\{\tilde{z}_s\} = \max_{s \in \mathcal{S}_{1\cup 2}}\{\hat{z}_s\} - \bar{z} \\ &= \max_{s \in \mathcal{S}_{1\cup 2}}\{\hat{z}_s - \bar{z}\} \leq 0 \\ &\Rightarrow \max_{s \in \mathcal{S}_{1\cup 2}}\{\hat{z}_s\} \leq \max_{s \in \mathcal{S}_{1\cup 2}}\{\tilde{z}_s\}, \end{aligned} \tag{45}$$

where the inequality in (45) is due to Inequality (44).

Note that the objective of problem **P2** is to minimize the MEDT among all SD pairs. So $\hat{z}_s$ is the optimal solution for direct link selection to problem **P2**, and Lemma 4 holds true. ■

E. Relay Assignment

Now that the hop 1, hop 2, and direct link selection subproblems having been solved with GC1, GC2, and GC3, respectively, we next solve the remaining problem of relay assignment. Substituting the following into problem **P2**,

$$\begin{aligned} &x_{l^+_{\mathrm{rd}(s)}} = x_{l^+_{\mathrm{sr}}}, x_{l_{\mathrm{sr}}} = 0, x_{l_{\mathrm{rd}(s)}} = 0, x_{l_{\mathrm{sd}(s)}} = 0, \forall\, l_{\mathrm{sr}} \neq l^+_{\mathrm{sr}}, \\ &\forall\, l_{\mathrm{rd}(s)} \neq l^+_{\mathrm{rd}(s)}, \forall\, l_{\mathrm{sd}(s)} \neq l^+_{\mathrm{sd}(s)}, l_{\mathrm{sr}} \in \mathcal{L}_{\mathrm{sr}}, l_{\mathrm{rd}(s)} \in \mathcal{L}_{\mathrm{rd}(s)}, \\ &l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}, \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s), \end{aligned} \tag{46}$$

we obtain a reduced problem, termed **SP2**, as follows.

$$\begin{aligned} \textbf{SP2:}\ \min:\ &\max_{s \in \mathcal{S}_{i\cup j}} \left\{ \sum_{r \in \mathcal{R}(s)} \left(\mathbb{E}(T_{l^+_{\mathrm{sr}}}) + \mathbb{E}(T_{l^+_{\mathrm{sr}}, l^+_{\mathrm{rd}(s)}}) \right) x_{l^+_{\mathrm{sr}}} + \right. \\ &\left. \sum_{l_{\mathrm{sd}(s)} \in \mathcal{L}_{\mathrm{sd}(s)}} \mathbb{E}(T_{l^+_{\mathrm{sd}(s)}}) x_{l^+_{\mathrm{rd}(s)}} \right\} \\ \text{s.t.}\ &\sum_{r \in \mathcal{R}(s)} x_{l^+_{\mathrm{sr}}} + x_{l^+_{\mathrm{sd}(s)}} = 1, \forall s \in \mathcal{S}_{i\cup j} \end{aligned} \tag{47}$$

$$x_{l^+_{\mathrm{sd}(s)}} = \{0,1\}, \forall s \in \mathcal{S}_{i\cup j}$$

$$\text{Constraints (33), (36), and (37).} \tag{48}$$

Also the OFV of problem **SP2** equals to that of problem **P2**. According to Theorem 2, we have Theorem 3 as follows.

Theorem 3: The OFV of problem **SP2** is at most one time slot greater than that of problem **P1**.

F. Decomposition Principle and Problem Reformulation

With the analysis in Sections III-A to III-E, we are now able to present the following theorem on the Decomposition Principle.

Theorem 4: Problem **P1** can be solved with the following four-step procedure, and the OFV of the solution is at most one time slot larger than that of the optimal solution.

1) *Step 1:* Choose the set of links in hop 1, i.e., $\{l^+_{\rm sr}\}$, as

$$l^+_{\rm sr} = \arg\min_{l_{\rm sr} \in \mathcal{L}_{\rm sr}} \mathbb{E}(T_{l_{\rm sr}}), \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s).$$

2) *Step 2:* With $\{l^+_{\rm sr}\}$, choose the set of links in hop 2, i.e., $\{l^+_{{\rm rd}(s)}\}$, as

$$l^+_{{\rm rd}(s)} = \arg\min_{l_{{\rm rd}(s)} \in \mathcal{L}_{{\rm rd}(s)}} \mathbb{E}(T_{l^+_{\rm sr}, l_{{\rm rd}(s)}}), \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s).$$

3) *Step 3:* Choose the set of links in the direct path, i.e., $\{l^+_{{\rm sd}(s)}\}$, as

$$l^+_{{\rm sd}(s)} = \arg\min_{l_{{\rm sd}(s)} \in \mathcal{L}_{{\rm sd}(s)}} \mathbb{E}(T_{l_{{\rm sd}(s)}}), \forall\, s \in \mathcal{S}_{i\cup j}, r \in \mathcal{R}(s).$$

4) *Step 4:* With (46) derived, solve problem **SP2**.

Let the problem in Step 1, Step 2, and Step 3 of Theorem 4 be termed **SP1**. Note that problem **SP2** is not in the general Integer Linear Programming (ILP) form. To solve problem **SP2**, we reformulate it into a linear programming (LP) problem. Introducing a new variable $w = \max_{s\in\mathcal{S}_{i\cup j}}\{\mathbb{E}(T_{l^+_{{\rm sd}(s)}})x_{l^+_{{\rm sd}(s)}} + \sum_{r\in\mathcal{R}(s)}(\mathbb{E}(T_{l^+_{\rm sr}}) + \mathbb{E}(T_{l^+_{\rm sr}, l^+_{{\rm rd}(s)}}))x_{l^+_{\rm sr}}\}$, we have

$$w \geq \mathbb{E}(T_{l^+_{{\rm sd}(s)}})x_{l^+_{{\rm sd}(s)}} + \sum_{r\in\mathcal{R}(s)}(\mathbb{E}(T_{l^+_{\rm sr}}) + \mathbb{E}(T_{l^+_{\rm sr}, l^+_{{\rm rd}(s)}}))x_{l^+_{\rm sr}},\ \forall\, s \in \mathcal{S}_{i\cup j}. \tag{49}$$

Then **SP2** can be rewritten as

$$\mathbf{SP2}' : \min : w$$
$$\text{s.t. Constraints, (33), (36), (37), (47), (48), and (49).}$$

Problem **SP2'** is a mixed integer linear programming problem (MILP) and can be solved with an existing effective solver. Once the relay and link selection are completed, the PNC will inform the nodes to start transmission as scheduled. If and only if at least one of the following events happens, the PNC will reschedule the link selection and relay assignment for all the SD pairs based on feedback.

1) *Case 1:* If a source had no traffic in the previous time slot but has new traffic to send in the current time slot.
2) *Case 2:* Whenever a relay finishes transmission to a destination and thus becomes available for source(s).

G. When $Y < \|\mathcal{S}_j\|$

If a type $\mathcal{S}_j$ SD pair cannot be served by a relay, its EDT cannot be defined as in (15) or (16). Thus we cannot directly employ the Decomposition Principle to solve the link and relay assignment problem in this case. We then propose a heuristic algorithm to solve the problem. The basic idea is to maximize the number of SD pairs that can transmit concurrently by relay assignment. We let each type $\mathcal{S}_i$ SD pair transmit via its direct link, and then assign relays to type $\mathcal{S}_j$ SD pairs to maximize the number of $\mathcal{S}_j$ SD pairs that can transmit concurrently. The more concurrent transmissions, the smaller the MEDT.

The Heuristic algorithm is presented in Algorithm 1.

Algorithm 1: Heuristic Algorithm for Link and Relay Assignment When Some $\mathcal{S}_j$ SD Pairs Do Not Have Relays.

```
Solve problem P0 ;
if Y ≥ ||S_j|| then
    Apply the Decomposition Principle to solve problem P1 ;
else
    for ∀s ∈ S_i do
        Choose direct link l^+_{sd(s)} to communicate with d(s) ;
    end
    Assign relays to type S_j SD pairs according to the solution
    to P0 ;
    Denote the set of type S_j SD pairs that have a relay as S'_j ;
    Find l^+_{sr}, l^+_{rd(s)}, for all s ∈ S'_j, r ∈ R(s) ;
end
```

H. Complexity Analysis

1) When $Y \geq \|\mathcal{S}_j\|$: Since problem **SP1** is easy to solve, we just compare the complexity of problem **P1** and problem **SP2'** from the following aspects.

1) Problem **P1** is an NIP, while problem **SP2'** is an MILP. Currently there are existing efficient solvers for MILP, such as the Gorubi MIP solver and the Matlab Intlinprog function (implementing the Branch and Bound algorithm). Such kind of problems have been solved effectively in prior works [19], [28], especially when the solution space is relatively small.
2) The number of decision variables of problem **P1** is $\sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{\rm sr}\| + \sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{{\rm rd}(s)}\| + \sum_{s\in\mathcal{S}_i}\|\mathcal{L}_{{\rm sd}(s)}\|$. The number of decision variables of **SP2'** is $\sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}\|l^+_{\rm sr}\| + \sum_{s\in\mathcal{S}_{i\cup j}}\|l^+_{{\rm sd}(s)}\| = \sum_{s\in\mathcal{S}_{i\cup j}}\sum_{r\in\mathcal{R}(s)}1 + \sum_{s\in\mathcal{S}_{i\cup j}}1 = \sum_{s\in\mathcal{S}_{i\cup j}}\|\mathcal{R}(s)\| + \|\mathcal{S}_{i\cup j}\|$, which is considerably smaller than that of problem **P1**.

2) When $Y < \|\mathcal{S}_j\|$: In this case, we first need to solve problem **P0** in order to decide if there are enough relays in the network. In problem **P0**, since each SD pair in $\mathcal{S}_j$ can use at most one relay and each relay can be used by at most one SD pair in $\mathcal{S}_j$, it can be seen that problem **P0** becomes a maximum weight matching problem on a bipartite graph that matches relays to SD pairs. Only one edge is allowed for a relay r and an SD pair s, and the edge weights are defined as the coefficient of $y_{\rm rs}$, which is 1. This maximum weight matching problem can be effectively solved in polynomial time using the Hungarian method, and the solution is optimal.

In our case, the computational complexity of using Hungarian method to solve problem **P0** is $\mathcal{O}((\|\mathcal{R}\| + \|\mathcal{S}_j\|)(\|\mathcal{R}\| \times \|\mathcal{S}_j\|))$, where $(\|\mathcal{R}\| + \|\mathcal{S}_j\|)$ is the total number of vertices

TABLE I
SIMULATION PARAMETERS

Parameter	*Value*
$\|\mathcal{S}_{i\cup j}\|$	10
$\|\mathcal{R}\|$	10
$\|\mathcal{L}_{\mathrm{sr}}\|, \forall s, r$	random $\in [3, 7]$
$\|\mathcal{L}_{\mathrm{rd}(s)}\|, \forall s, r$	random $\in [3, 7]$
$\|\mathcal{L}_{\mathrm{sd}(s)}\|, \forall s$	random $\in [0, 3]$
$\max_{\{l\in(\mathcal{L}_{\mathrm{sr}}\cup\mathcal{L}_{\mathrm{rd}(s)}\cup\mathcal{L}_{\mathrm{sd}(s)}),\forall s,r\}} q_l$	0.9
$p_l, \forall l \in (\mathcal{L}_{\mathrm{sr}} \cup \mathcal{L}_{\mathrm{rd}(s)} \cup \mathcal{L}_{\mathrm{sd}(s)}), \forall s, r$	random $\in [0.3, 0.7]$
Chanel capacity	1 Gbps
Time slot duration	1 s

and $(\|\mathcal{R}\| \times \|\mathcal{S}_j\|)$ is the total number of possible edges in the bipartite graph representing problem **P0** [30].

The computational complexity of the heuristic algorithm, when there are no enough relays, as shown in Algorithm 1 consists of the following three parts:

1) Procedures in Lines 5 – 7, with computational complexity $\mathcal{O}(\|\mathcal{S}_i\|)$.
2) Procedures in Line 8, with computational complexity $\mathcal{O}(1)$, since problem **P0** has been solved.
3) Procedures in Lines 9 – 10, according to (23) and (40), the computational complexity of determining $l^+_{\mathrm{sr}}, \forall r \in \mathcal{R}(s), s \in \mathcal{S}_j$ is $\mathcal{O}(\sum_{s\in\mathcal{S}_j}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{\mathrm{sr}}\|)$, and the computational complexity of determinging $l^+_{\mathrm{rd}(s)}, \forall r \in \mathcal{R}(s), s \in \mathcal{S}_j$ is $\mathcal{O}(\sum_{s\in\mathcal{S}_j}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{\mathrm{rd}(s)}\|)$.

In summary, the computation complexity of the Heuristic Algorithm when $Y < \|\mathcal{S}_j\|$ is $\mathcal{O}((\|\mathcal{R}\| + \|\mathcal{S}_j\|)(\|\mathcal{R}\| \times \|\mathcal{S}_j\|)) + \mathcal{O}(\|\mathcal{S}_i\|) + \mathcal{O}(\sum_{s\in\mathcal{S}_j}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{\mathrm{sr}}\|) + \mathcal{O}(\sum_{s\in\mathcal{S}_j}\sum_{r\in\mathcal{R}(s)}\|\mathcal{L}_{\mathrm{rd}(s)}\|)$.

IV. SIMULATION STUDY

A. Simulation Configuration

In this section we validate the performance of the proposed Decomposition Principle by Matlab simulations. Unless otherwise specified, the values of simulation parameters are as given in Table I. Each simulation result presented in the figures is obtained by repeating the simulation 50 times with different random seeds, while 95% confidence intervals are computed and plotted as error bars in the figures.

We compare the performance of the proposed algorithm in Theorem 4 (termed *Proposed*) with two existing schemes designed for mmWave networks. The first one (termed *Benchmark 1*) is proposed in [18], where a source tries to maximize its throughput by choosing the optimal Access Points (APs), and the source-AP channels are modeled as Markov chains. A heuristic algorithm is used to solve the formulated NP-hard problem in [18]. The second one (termed *Benchmark 2*) is proposed in [17], where relay paths are determined for multiple SD pairs with a heuristic to maximize the total throughput under static channel conditions.

The performance metrics to evaluate the proposed algorithm are delay, MEDT among all SD pairs, and network throughput.

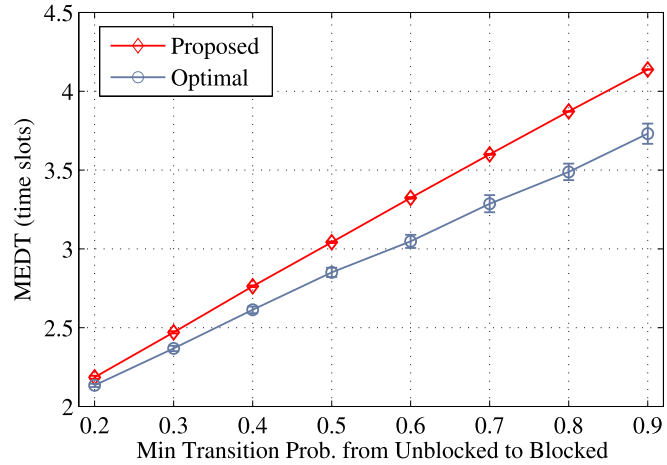


Fig. 2. The OFV of the proposed decomposition principle and that of the optimal solution versus $\min_l\{q_l\}$, while $P_G = 0.8$.

The delay of a packet is the time it spends at the source queue plus the packet delivery time from source to destination. Traffic is generated with a Bernoulli process [3]. At each time slot, the source generates a number of packets with a predetermined probability, denoted as P_G, and the total volume of bits of the packets generated at each time slot does not exceed the channel capacity.

B. Simulation Results and Discussions

The performance of the proposed algorithm is demonstrated in Fig. 2 by comparing the OFV of problem **SP2'** with that of problem **P1** (i.e., the *Optimal*) under increasing channel transition probability q_l. Problem **P1** is an NIP whose solution takes a very long time to obtain using exhaustive search even for a moderately-sized network. Therefore we simulate a relative small network with 2 SD pairs, 2 relays, 2 links in hop 1, 2 links in hop 2, and 1 link in the direct path, for each SD pair and relay, to obtain the optimal solution within a reasonable time. From Fig. 2, we fine the difference between the OFV of problem **SP2'** and that of problem **P1** is strictly within one time slot over the entire range of q_l. The gap is actually much smaller than one time slot. Furthermore, the gap increases slightly as q_l grows, since a sub-optimal schedule may result in a relatively worse performance when channel conditions are bad, which means a greater MEDT.

In Fig. 3, we compare the simulation execution times of solving problem **P1** and solving problem **SP2'**. The number of SD pairs in the network is increased from 2 to 10, and the simulation is executed on a PC with an Intel(R) Core(TM) i5-3470 CPU @ 3.20 GHz. The version of Matlab is R2013a. We can see that the time of solving problem **SP2'** is strictly smaller than that of solving problem **P1**, and the gap becomes greater as the network size is increased. The increment of the time of solving problem **SP2'** has a relatively smoother trend compared with that of solving problem **P1**. As we can see from the computational complexity analysis of problems **P1** and **SP2'**, the number of decision variables in problem **P1** increases much faster than that

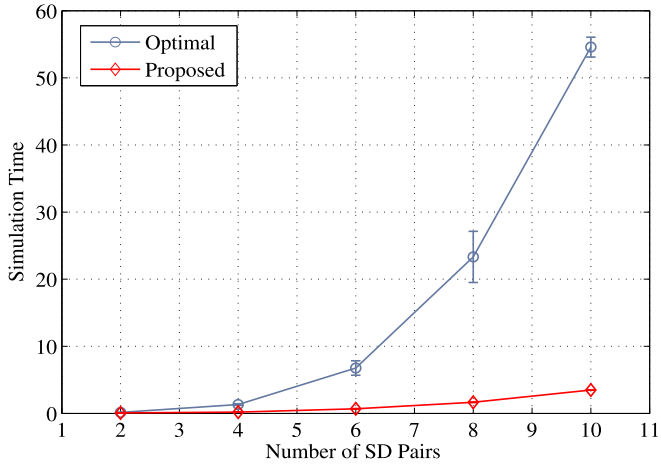


Fig. 3. Simulation time of solving problem **P1** and solving problem **SP2'** under various numbers of SD pairs in the network.

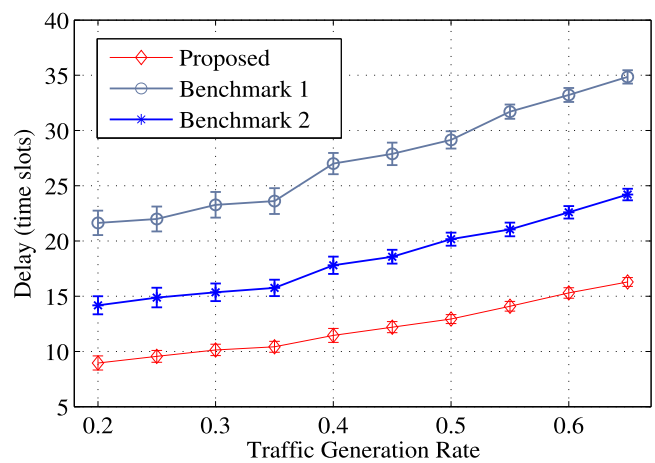


Fig. 4. Delay versus minimum traffic generation probability P_G, while $\min_{\{l\}}\{eq_l\} = 0.2$.

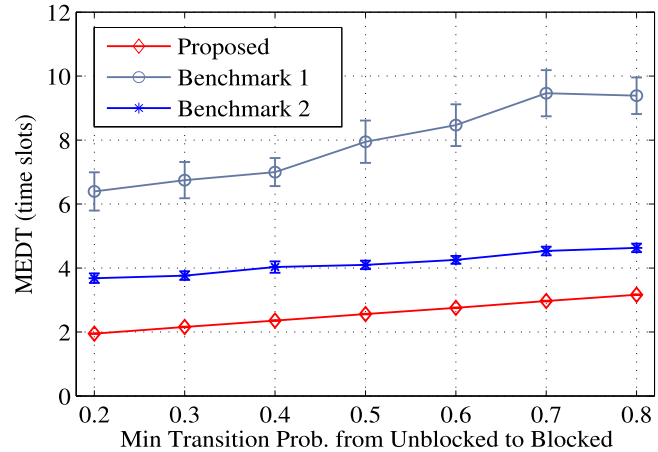


Fig. 5. MEDT versus $\min_{\{l\}}\{q_l\}$, while $P_G = 0.8$.

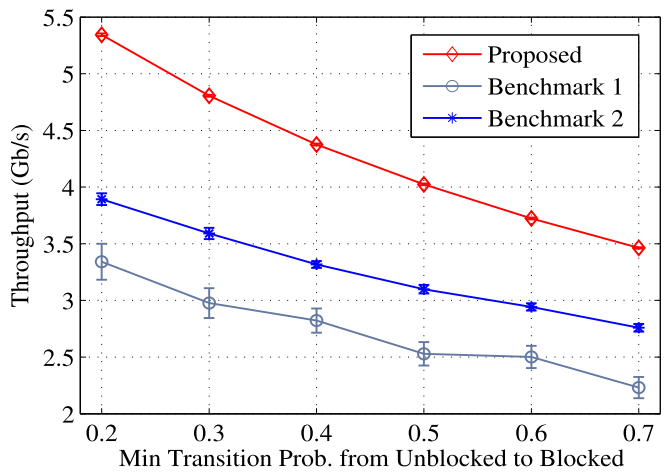


Fig. 6. Throughput versus $\min_{\{l\}}\{q_l\}$, while $P_G = 0.8$.

in in problem **SP2'**. Since problem **P1** is an NIP, it is expected that the time of solving problem **P1** will have a sharper increase.

We next compare the delay performance of the proposed scheme with that of the two benchmark schemes in Fig. 4 under various traffic generation rate P_G. As P_G is increased, the average delays of all the three schemes increase due to the increased traffic load, while the average delay of our proposed algorithm is always considerably lower than that of the two benchmark schemes. Benchmark 1 does not consider coordinating concurrent transmissions of SD pairs. Therefore different SD pairs may select the same relay and thus collision happens, resulting in an increased delivery time. Benchmark 2 does not consider channel dynamics and thus its schedules may be sub-optimal. This comparison also demonstrates that traffic collision has a serious negative effect on delay performance.

Fig. 5 shows the MEDT among all SD pairs versus channel state transition probability q_l. The proposed scheme achieves the lowest MEDT among the three. The confidence interval of Benchmark 1 is greater than that of the other two schemes, indicating Benchmark 1 is less stable in terms of the number of trails until the first successful transmission is achieved. Benchmark 2 lacks in adaptation to the channel dynamics, which certainly has an effect on the instantaneous scheduling decision for the current time slot.

The throughput performance achieved by the three schemes is presented in Fig. 6. The network throughput is defined as the total number of bits delivered for all the SD pairs per time slot, i.e., per second. As channel condition degrades, the number of links available for transmission is decreased at each time slot. So the number of bits that can be delivered at each time slot is reduced. For Benchmark 1, due to collisions, the number of bits successfully delivered per time slot is less than that of the proposed algorithm. For Benchmark 2, due to lack of consideration of channel dynamics, although it tends to maximize the total expected throughput of all SD pairs, it still makes sub-optimal scheduling decisions under dynamic channel conditions, thus achieving a lower throughput.

In Fig. 7, we compare the MEDT of the Heuristic link and relay selection algorithm described in Algorithm 1 (termed *Heuristic*) with that of the two benchmarks under the condition that there is an insufficient number of relays to serve all the SD pairs. Here we set the number of relays equals to 6,

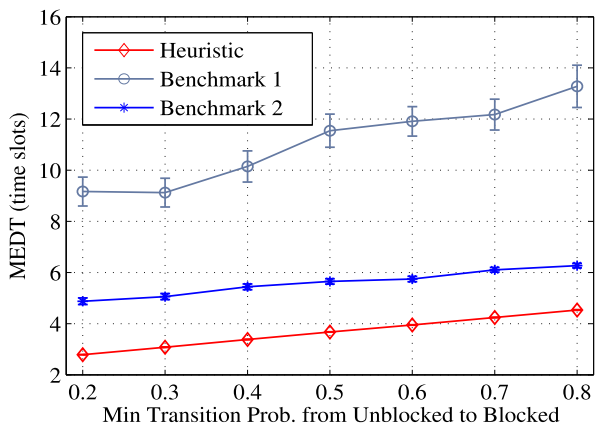


Fig. 7. MEDT versus $\min_{\{l\}}\{q_l\}$ for the proposed Heuristic algorithm, while $P_G = 0.8$ and the number of relays is 6.

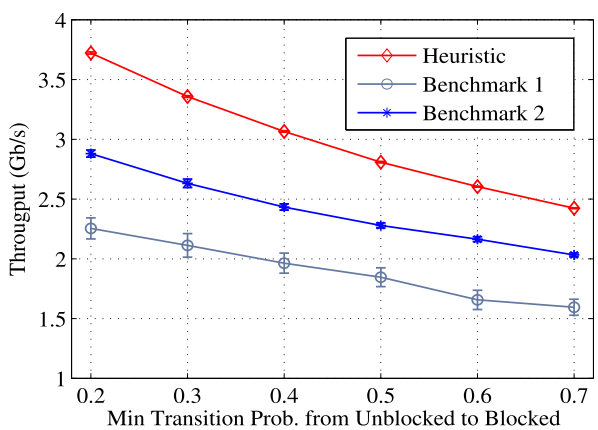


Fig. 8. Throughput vs $\min_{\{l\}}\{q_l\}$ for the proposed Heuristic algorithm, while $P_G = 0.8$ and the number of relays is 6.

while other parameters are as given in Table I. Comparing to the results in Fig. 5, the MEDT of Heuristic is slightly higher, due to the insufficient number of realys to serve the 10 SD pairs. For example, when $\min_{\{l\}}\{q_l\} = 0.2$, the MEDT is increased from 2 to 2.7881. However, it can be observed that Heuristic still outperforms both benchmark schemes with considerable gains. This result makes sense since letting type $\mathcal{S}_i$ SD pairs to communicate using direct links, instead of using a relay, will save more relaying opportunity to Type $\mathcal{S}_j$ SD pairs, so these SD pairs may need less time to successfully delivery their packets. A similar reasoning can be applied to the comparison of throughput performance of Heuristic with that of the two benchmarks, as shown in Fig. 8.

Finally, we compare the fairness performance of the three schemes, in terms of average delay of the SD pairs. Fig. 9 shows the fairness performance comparison between the proposed scheme and the benchmark schemes. We adopt Jain's fairness index as in [3]:

$$f(e_1, e_2, \ldots, e_N) = \frac{(e_1 + e_2 + \ldots + e_N)^2}{N(e_1^2 + e_2^2 + \ldots + e_N^2)}, \qquad (50)$$

where e_n is the average delay of SD pair n, $n = 1, 2, \ldots, N$. The fairness index ranges from 0 (worst) to 1 (best). We can

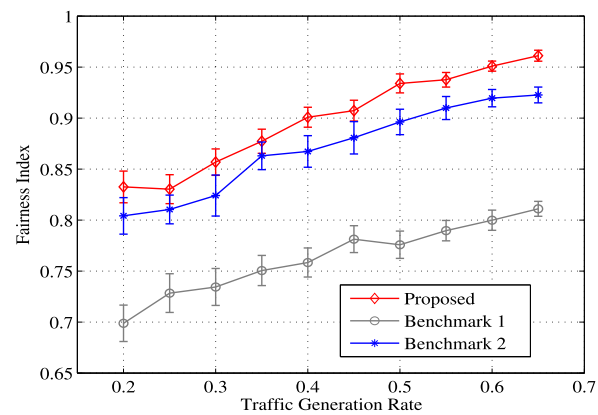


Fig. 9. Fairness performance versus minimum traffic generation probability, for Proposed Scheme, while $\min_{\{l\}}\{q_l\} = 0.2$ and $P_G \leq 0.8$.

see that our proposed algorithm consistently achieves a higher fairness index than the other two schemes do, due to the *minimax* approach adopted in the problem formulation.

V. Related Work

There have been some interesting work on link scheduling in 60 GHz networks. The authors in [19] propose a Partially Observable Markov Decision Process (POMDP) framework to model the link status in 60 GHz networks, and a greedy scheduling strategy that aims to maximize the instant throughput at each time slot. However, this strategy is only applicable for single-hop centralized networks, and the multiple potential links between a node pair is not explored in this paper. A similar problem is studied in [18] with a special scenario of a single-transmitter. To improve network throughput, the authors in [15] propose a fast relay selection algorithm to reduce the overhead of relay selection time, so that there will be more time for data transmission. The basic idea is to determine the sectors where the best relay may be located, and then find the best relays in the selected sector. However, the authors do not consider coordinating concurrent transmissions of multiple transmitters. It is possible that different transmitters may select the same relay and thus collision happens.

In [11], the authors consider the fact that different relays may have different path losses, and thus having different outage probabilities. A relay selection scheme is proposed to minimize the outage probability for a single transmitter. In both indoor and outdoor environments, the obstacles may change over time (e.g., pedestrians move) and thus the blockage of a 60 GHz link is actually not static. Such dynamic channel condition is not considered in [11]. A network throughput maximization problem for a dual-hop network is studied in [10], where different relays may provide different capacities for a SD pair. Relays assignment for multiple SD pairs is optimized to maximize the network throughput. The path loss and blockage model considered in this work are also time-invariant, and thus the proposed algorithm may not be suitable for 60 GHz networks with dynamic link conditions.

To overcome the problem of link breakage and degradation in point-to-point 60 GHz networks, the authors of [31] propose to use *repeaters* to provide alternate paths when the direct path between transmitter and receiver degrades. It is assumed that the nodes and repeaters can beamform in any direction and thus by tuning the transmitting and receiving antenna to the repeater a new link between the transmitter and receiver can be established. However, it is worth noting that in complex environments, the placement and selection of repeaters is a non-trivial problem. In [32], to reinforce transmission efficiency and also reduce power consumption of 60 GHz devices, the authors propose a fast beam-switching scheme, which employs an efficient beamforming training algorithm based on the direct numerical search. Only a small portion of beam-pairs will be sequentially tested while most other beam-pairs will never be probed, so that the search complexity can be significantly reduced.

In [3], the authors propose a heuristic scheduling scheme for given traffic demands under static channel conditions, aiming to minimize the time needed to clear all the traffic demands. The pseudo-wired 60 GHz channel model is adopted in this work. The authors in [16] study the relationship between the collision probability of two concurrent transmissions on two links and the link distances. It is found that the collision probability is an increasing function of link distance. Based on this finding, the authors propose a hop selection metric based on link distance, to reduce the collision probability of concurrent transmissions. By replacing a single long hop with multiple short hops, the proposed scheme can improve the number of concurrent transmission flows while constraining the harmful interference below an acceptable level. However, the algorithm is heuristic and lacks consideration of multiple coexisting links. The time slot allocation problem in multi-hop 60 GHz networks is investigated in [17], where the direct path shares time slots with the relay path. Different time slot allocation schemes may result in different system throughput, and the effective system throughput is optimized with time slot allocation. A sub-optimal solution is proposed to solve the formulated NP-hard problem. Besides time slot allocation, channel allocation is also significant to the improvement of network throughput. The authors of [33] investigate the problem of channel allocation in 60 GHz indoor WLANs in order to maximize throughput, and two SDMA (Spatial Division Multiple Access) algorithms are proposed, for the single-channel case and the multiple-channel case respectively, to exploit the peculiar propagation properties so that data rates to end users can be improved.

There are also some papers on designing MAC protocols for 60 GHz wireless networks and performance analysis of MAC protocols. In [34], the authors claim that conventional directional CSMA/CA protocols do not work well at 60 GHz networks due to the impaired carrier sensing at the transmitters. To overcome this difficulty, the authors propose a novel protocol which adopts virtual carrier sensing instead of physical carrier sensing, and relies on a central coordinator to distribute network allocation vector (NAV) information. The authors of [35] present an analytical model for computing the saturation throughput of a Medium- Transparent MAC protocol in 60 GHz radio-over-fiber networks. Both of the contention at the optical and the wireless layer are considered. The authors derive the saturation throughput performance of the Medium Transparent MAC protocol under various scenarios. To provide a more comprehensive performance analysis of the Medium Transparent MAC protocol for 60 GHz radio-over-fiber networks, the authors of [36] analysis the delay fairness performance of the Medium Transparent MAC protocol, and it is shown that delay equalization can be achieved even for highly varying user population patterns among the different antenna units when certain wavelength availability conditions are satisfied.

VI. Conclusion

We developed a Decomposition Principle for the problem of link and relay selection in centralized dual-hop 60 GHz networks. The objective was to minimize the MEDT, and the main idea was to decompose the original problem into sub-problems for link selection and relay selection. When there are a sufficient amount of relays, we proved that the two sub-problems together can provide a sub-optimal solution to the original problem with an optimality gap bounded by one time slot, which also has a greatly reduced complexity. We also developed a heuristic scheme to handle the case when there are no enough relays to serve all the SD pairs. Through simulations, we showed that both proposed schemes outperformed two 60 GHz network scheduling schemes with considerable gains.

References

[1] Z. He and S. Mao, "A decomposition principle for link and relay selection in dual-hop 60 GHz networks," in *Proc. IEEE INFOCOM*, San Francisco, CA, USA, Apr. 2016, pp. 1–9.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5 g cellular: It will work!" *IEEE Access J.*, vol. 1, pp. 335–349, May 2013.

[3] I. K. Son, S. Mao, Y. Li, M. Chen, M. X. Gong, and T. S. Rappaport, "Frame-based medium access control for 5G wireless networks," *Mobile Netw. Appl.*, vol. 20, no. 6, pp. 763–772, Dec. 2015.

[4] *Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs), IEEE Unapproved Draft STD P802.15.3c/DIO*, Jun. 2009.

[5] H. S Hokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[6] R. C. Daniels and R. W. Heath, "60 GHz wireless communications: Emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol. 2, no. 3, pp. 41–50, Sep. 2007.

[7] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.

[8] R. Mudumbai, S. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2871–2875.

[9] S. Singh, R. Mudumbai, and U. Madhow, "Distributed coordination with deaf neighbors: Efficient medium access for 60 GHz mesh networks," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[10] H. Chu and P. Xu, "Relay selection with feedback beamforming information for NLoS 60 GHz mmWave WLANs/WPANs," in *Proc. IEEE Int. Conf. Commun.*, Sydney, NSW, Australia, Jun. 2014, pp. 5514–5519.

[11] Z. Liu, X. Tao, W. U. Rehman, Z. Xu, and X. Xu, "Opportunistic relay selection and outage performance analysis for 60 GHz wireless system," in *Proc. IEEE IEEE Global Telecommun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 328–332.

[12] S. Singh, F. Ziliotto, U. Madhow, E. Belding, and M. Rodwell, "Blockage and directivity in 60 GHz wireless personal area networks: From cross-layer model to multihop MAC design," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1400–1413, Oct. 2009.

[13] C. Yiu and S. Singh, "Empirical capacity of mmWave WLANs," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1479–1487, Oct. 2009.

[14] X. An *et al.*, "Beam switching support to resolve link-blockage problem in 60 GHz WPANs," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Toyko, Japan, Sep. 2009, pp. 390–394.

[15] K. Song, R. Cai, and D. Liu, "A fast relay selection algorithm over 60 GHz mm-wave systems," in *Proc. 15th IEEE Int. Conf. Commun. Technol.*, Guilin, China, Nov. 2013, pp. 676–680.

[16] J. Qiao, L. X. Cai, X. Shen, and J. Mark, "Enabling multi-hop concurrent transmissions in 60 GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3824–3833, Nov. 2011.

[17] Z. Lan *et al.*, "Relay with deflection routing for effective throughput improvement in Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1453–1465, Oct. 2009.

[18] X. Zhang *et al.*, "Improving network throughput in 60 GHz WLANs via multi-AP diversity," in *Proc. IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, Jun. 2012, pp. 4803–4807.

[19] Z. He, S. Mao, and T. S. Rappaport, "Minimum time length link scheduling under blockage and interference in 60 GHz networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, New Orleans, LA, USA, Mar. 2015, pp. 837–842.

[20] Z. He, S. Mao, S. Kompella, and A. Swami, "Minimum time length scheduling under blockage and interference in multi-hop mmWave networks," in *Proc. IEEE IEEE Global Telecommun. Conf.*, San Diego, CA, USA, Dec. 2015, pp. 1–7.

[21] Z. He, S. Mao, and T.S. Rappaport, "On link scheduling under blockage and interference in 60 GHz ad hoc networks," *IEEE Access J.*, vol. 3, pp. 1437–1449, Sep. 2015.

[22] T. Bai, R. Vaze, and R. W. Heath Jr., "Using random shape theory to model blockage in random cellular networks," in *Proc. Int. Conf. Signal Process. Commun.*, Bangalore, India, Jul. 2012, pp. 1–5.

[23] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 922–928, May 2004.

[24] P. Hyunhee, Y. Kim, T. Song, and S. Pack, "Multiband directional neighbor discovery in self-organized mmWave ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1143–1155, Mar. 2015.

[25] L. Li, K. Josiam, and R. Taori, "Feasibility study on full-duplex wireless millimeter-wave systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 2769–2773.

[26] Federal Communications Commission, "Revision of part 15 of the commissions rules regarding operation in the 57–64 GHz band," Federal Commun. Commission, Washington, DC, USA, Tech. Rep. FCC 13-112, ET Docket No. 07-113, RM-11104, Aug. 2013.

[27] Airlinx, "60 GHZ wireless links for high-bandwidth data application," [Online]. Available: http://www.airlinx.com/files/AIRLINX, Accessed on: Jun. 2017.

[28] G. Zheng, H. Cunqing, R. Zheng, and Q. Wang, "A robust relay placement framework for 60 GHz mmWave wireless personal area networks," in *Proc. IEEE Global Telecommun. Conf.*, Atlanta, GA, USA, Dec. 2013, pp. 4816–4822.

[29] W. K. Ching, X. Huang, M. K. Ng, and T. K. Siu, *Markov Chains: Models, Algorithms and Applications*, Berlin, Germany: Springer, 2006.

[30] M. X. Goemans, "Lecture notes on bipartite matching," Massachusetts Inst. Technol., Cambridge, MA, USA. [Online]. Available: http://math.mit.edu/ goemans/18433S09/matching-notes.pdf, Accessed on: Jun. 2017.

[31] C. Yiu and S. Singh, "Link selection for point-to-point 60 GHz networks," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May. 2010, pp. 1–6.

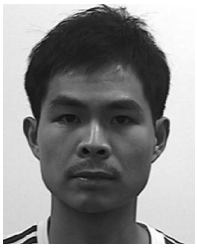
[32] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, "On the efficient beamforming training for 60 GHz wireless personal area networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 504–515, Dec. 2013.

[33] C. Yiu and S. Singh, "SDMA for 60 GHz gigabit wireless networks," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–6.

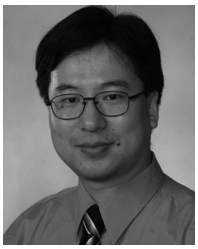
[34] M. X. Gong, R. Stacey, D. Akhmetov, and S. Mao, "A directional CSMA/CA protocol for mmWave wireless PANs," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Sydney, NSW, Australia, Apr. 2010, pp. 1–6.

[35] G. Kalfas, N. Pleros, K. Tsagkaris, L. Alonso, and C. Verikoukis, "Performance analysis of a medium-transparent MAC protocol for 60 GHz radio-over-fiber networks," in *Proc. IEEE IEEE Global Telecommun. Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–5.

[36] P. Maniotis, G. Kalfas, L. Alonso, C. Verikoukis, and N. Pleros, "Throughput and delay fairness through an agile medium-transparent MAC protocol for 60 GHz fiber-wireless LAN networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Shanghai, China, Apr. 2012, pp. 3769–3773.



Zhifeng He received the Ph.D. degree in electrical and computer engineering from Auburn University, Auburn, AL 36849, USA, in December 2016, the M.S. degree in microelectronics and solid state electronics from Beijing University of Posts and Telecommunications, Beijing, China, in 2012, and the B.S. degree in electronics information science and technology from Shandong University of Technology, Zibo, China in 2009. His research interests include cognitive radio, mmWave communications and networking, multimedia communications and optimization. He received the IEEE GLOBECOM 2015 Best Paper Award.



Shiwen Mao (S'99–M'04–SM'09) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, USA (now NYU Tandon School of Engineering). He is currently the Samuel Ginn Distinguished Professor in the Department of Electrical and Computer Engineering, and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, applications, and multimedia communications. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society for 2014–2018. He is on the editorial board of the IEEE TRANSACTIONS ON MULTIMEDIA, IEEE INTERNET OF THINGS JOURNAL, IEEE MULTIMEDIA, *ACM GetMobile*, etc. He is the Chair of the IEEE ComSoc Multimedia Communications Technical Committee. He received the 2015 IEEE ComSoC TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and the NSF CAREER Award in 2010. He received the IEEE SECON 2017 Best Demo Award, the IEEE GLOBECOM 2016 Best Paper Award, the IEEE GLOBECOM 2015 Best Paper Award, the IEEE WCNC 2015 Best Paper Award, the IEEE ICC 2013 Best Paper Award, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems.

Sastry Kompella (S'04–M'07–SM'12) received the B.E. degree in electronics and communication engineering from Andhra University, Vishakhapatnam, India, the M.S. degree in electrical engineering from Texas Tech University, Lubbock, TX, USA, and the Ph.D. degree in electrical and computer engineering from the Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 1996, 1998, and 2006, respectively. He is currently the Head of the wireless network research section under the Information Technology Division, U.S. Naval Research Laboratory, Washington, DC, USA. His research interests include wireless networks, from mobile ad hoc to underwater acoustic networks, with specific focus toward the development of cognitive and cooperative network optimization techniques for efficient wireless link scheduling.

Ananthram Swami (S'79–M'79–SM'96–F'08) received the B. Tech. degree from Indian Institute of Technology Bombay, Mumbai, India, the M.S. degree from Rice University, Houston, TX, USA, and the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA, all in electrical engineering. He is currently with the U.S. Army Research Laboratory (ARL) as the Army's Senior Research Scientist for Network Science. He has held positions with Unocal Corporation, the University of Southern California (USC), CS-3 and Malgudi Systems. He was a Statistical Consultant to the California Lottery, developed a MATLAB-based toolbox for non-Gaussian signal processing, and has held visiting faculty positions at INP, Toulouse. His research interests include the broad area of network science: The study of interactions and co-evolution, prediction and control of interdependent networks, with applications in composite tactical networks. He is currently an ARL Fellow.