Multipath Routing and Max-Min Fair QoS Provisioning under Interference Constraints in Wireless Multihop Networks

Preetha Thulasiraman, *Member*, *IEEE*, Jiming Chen, *Member*, *IEEE*, and Xuemin (Sherman) Shen, *Fellow*, *IEEE*

Abstract—In this paper, we investigate the problem of flow routing and fair bandwidth allocation under interference constraints for multihop wireless networks. We first develop a novel isotonic routing metric, RI^3M, considering the influence of interflow and intraflow interference. The isotonicity of the routing metric is proved using virtual network decomposition. Second, in order to ensure QoS, an interference-aware max-min fair bandwidth allocation algorithm, LMX:M^3F, is proposed where multiple paths (determined by using the routing metric) coexist for each user to the base station. In order to solve the algorithm, we develop an optimization formulation that is modeled as a multicommodity flow problem where the lexicographically largest bandwidth allocation vector is found among all optimal allocation vectors while considering constraints of interference on the flows. We compare our RI^3M routing metric and LMX:M^3F bandwidth allocation algorithm with various interference-based routing metrics and interference-aware bandwidth allocation algorithms established in the literature. We show that RI^3M and LMX:M^3F succeed in improving network performance in terms of delay, packet loss ratio, and bandwidth usage.

Index Terms—Interference, multicommodity flow, fairness, routing, quality of service.

1 Introduction

The use of relays to improve the performance of broadband wireless access (BWA) networks has been the subject of intense research activities in recent years [1]. With the use of multihop relaying, increasing number of users, and limited spectrum, wireless multihop BWA networks are limited by two main resources: bandwidth and network capacity. While bandwidth refers to the achievable data rate, capacity refers to the data transport capacity available for each link in the network. Achieving high throughput and fair allocation of resources among competing users (or flows) in wireless networks is one of the most important problems in data communications. However, these two objectives may conflict with each other [2]. Max-min fairness (MMF) is considered to be an efficient approach that balances these two conflicting objectives by preventing starvation of any flow, and at the same time, increases the bandwidth of a flow as much as possible. The concept of fairness in wireless networks is a quality of service (QoS) policy and can be applied to various design issues such as scheduling and routing [3], [4].

- *P. Thulasiraman is with the Department of Electrical and Computer Engineering, Naval Postgraduate School, 833 Dyer Road, Monterey, CA, 93943. E-mail: pthulas1@nps.edu.*
- *J. Chen is with the Institute of Industrial Process Control, Department of Control, Zhejiang University, Zheda Road 38#, Hangzhou 310027, P.R. China. E-mail: jmchen@ieee.org.*
- *X. (Sherman) Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L-3G1, Canada. E-mail: xshen@bbcr.uwaterloo.ca.*

Manuscript received 2 Mar. 2010; revised 1 June 2010; accepted 11 June 2010; published online 6 Aug. 2010.
Recommended for acceptance by V. Misic.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2010-03-0137.
Digital Object Identifier no. 10.1109/TPDS.2010.145.

1.1 Interference and Fairness in Routing

Efficient routing between pairs of nodes in communication networks is a basic problem of network optimization; discovering available relaying paths (routes) between a source and destination node is a critical prerequisite for the success of multihop wireless networks. In this context, the classic MMF problem was originally defined for wired networks in order to allocate bandwidth to a set of given routes [5]. Research on MMF routing in the wired environment can be split into two categories: nonsplittable and splittable (multipath). In the nonsplittable case [5], [6], an MMF distribution of resources (bandwidth) to connections is done for fixed single path routing. In the splittable (multipath) MMF routing case, the traffic demands are allowed to be split among multiple flows (paths) [7], [8], [9]. Multipath routing has long been recognized as an effective strategy to achieve load balancing and increase reliability. It has been shown in [9] that multipath (splittable) demand routing is a linear relaxation of the nonsplittable case, thus rendering the problem computationally tractable. To improve the transmission reliability and increase the probability of network survivability, the multiple paths can be selected to be link disjoint. In this case, the multipath routing approach is referred to as disjoint multipath routing.

An important feature of multipath routing is the ability to provide QoS in terms of fair bandwidth allocation. Fairness-based routing protocols that use the max-min model have been recently proposed [10], [11], [12], [13], which focus on the lexicographic (node ordering) optimization of routing for fair bandwidth allocation. These solutions can lead to high throughput with guaranteed max-min bandwidth allocation values. However, they are

formulated in ideal scenarios. Specifically, the inherent influence of wireless interference has been neglected.

In the wireless environment, allocation of bandwidth to paths sharing a set of links is complicated by the interference that is generated by simultaneous transmissions. Interference can be divided into two categories: *interflow* and *intraflow*. Interflow interference is generated when two links belonging to different flows are active on the same channel at the same time. Intraflow interference is when two links belonging to the same flow are active on the same channel at the same time. The effects of interference using the MMF approach have been quantified using graph-theoretic approaches (i.e., conflict/contention graph) which ultimately exploit the protocol interference model (i.e., transmissions interfere only within a specific range) [14], [15]. The authors in [14], [15] have provided a theoretical foundation for fairness in wireless networks. However, the reliance on such graph-based models induces binary conflicts which means any two links either interfere with each other or they are active simultaneously, regardless of the other ongoing transmissions, which is not true in practice [16]. Thus, a more practical interference model such as the physical model (also known as Signal to Interference Noise Ratio (SINR) model) can provide a less restrictive and realistic quantification of interference. Although the SINR model has been used for achieving channel assignment and scheduling [17], [18], little research on SINR-based fair routing and bandwidth allocation exists.

1.2 Interference-Based Routing Metrics

Providing fault tolerance and QoS provisioning in the presence of interference are major issues that must be studied jointly in wireless networks in order to gauge a realistic sense of network performance. Developing routing metrics has long been the central focus of network-layer protocol design. To compute paths using an interference-aware routing metric is essentially equivalent to computing minimum weight (shortest) paths where the link weight is generated by the routing metric. In order to efficiently compute minimum weight paths using algorithms such as Dijkstra's shortest path or Bellman-Ford, the routing metric must be *isotonic*. The isotonic property essentially means that a routing metric should ensure that the order of the weights of two paths is preserved if they are appended by a common third path. In addition, isotonicity ensures loop-free routing. If a routing metric is not isotonic, only algorithms with exponential complexity can calculate minimum weight paths, which is not tractable for networks of even moderate size [19].

The two most prominent metrics are Expected Transmission Count (ETX) [20] and Expected Transmission Time (ETT) [21]. ETX is defined as the expected number of MAC-layer transmissions needed to successfully deliver a packet through a wireless link. ETT improves upon ETX by considering the differences in transmission rates. Although both metrics are isotonic, neither considers interference. The earliest metric to consider interference is Weighted Cumulative ETT (WCETT) [21]. This metric essentially captures intraflow interference by reducing the number of nodes on a path of a flow that transmit on the same channel; it gives low weight to paths that have more diversified channel assignments. However, WCETT does not capture interflow interference and is not isotonic which prevents the use of an efficient loop-free routing algorithm to compute minimum weight paths. The Metric for Interference and Channel switching (MIC) [19] improves WCETT by capturing interflow interference and overcomes the nonisotonicity problem. However, MIC does not measure interference dynamically, meaning that changes to interference level over time due to signal strength and traffic load may not be captured accurately. The Interference AWARE (iAWARE) routing metric [22] computes paths with lower interflow and intraflow interference than MIC and WCETT. It uses Signal-to-Noise and Signal-to-Interference-Noise ratios, SNR and SINR, respectively, to continuously monitor neighboring interference variations. Yet, iAWARE is not isotonic. Recently, improvements to the ETX and ETT metrics such as Interferer Neighbor Count (INX) were proposed in [23]. Similar to MIC, INX takes into account interference through the number of links that can interfere on a link l. This metric performs better only in low traffic load conditions, and therefore, load balancing is not completely resolved.

According to the main requirements of interference, load awareness, and isotonicity, existing routing metrics address only some specific requirements. For this reason, in this paper, a new routing metric is proposed in order to simultaneously address all of these aspects.

1.3 Contributions and Organization

In this paper, we study the issues of routing and fair bandwidth allocation in the presence of interference in wireless networks.[1] Our contributions are twofold and can be summarized as follows: First, we design an isotonic routing metric which is cognizant of interference and provides reliable multipath routing. The routing metric is used to quantify the interference on the network links such that least interfering paths can be obtained. The Routing with interflow and Intraflow Interference Metric (RI^3M) captures both interflow and intraflow interference while balancing link load. We illustrate the isotonicity of the RI^3M routing metric through virtual network decomposition. We then use RI^3M to find disjoint paths from each user to the base station. Second, we develop an MMF optimization formulation that finds the fairest (lexicographically largest) bandwidth allocation vector for the demands. The MMF optimization formulation explicitly considers the constraints of wireless interference on the individual flows. We refer to this algorithm as the Lexicographic MMF Multipath Flow (LMX:M^3F) algorithm.

The remainder of this paper is organized as follows: Section 2 discusses the system model and relevant assumptions. In Section 3, the RI^3M routing metric is discussed along with proof of isotonicity using virtual decomposition, while in Section 4, the LMX:M^3F optimization formulation is developed. The performance evaluation through simulations is given in Section 5. We conclude the paper in Section 6.

1. Parts of this work were presented at the IEEE ICC 2010 [24] and IEEE WCNC 2010 [25] conferences.

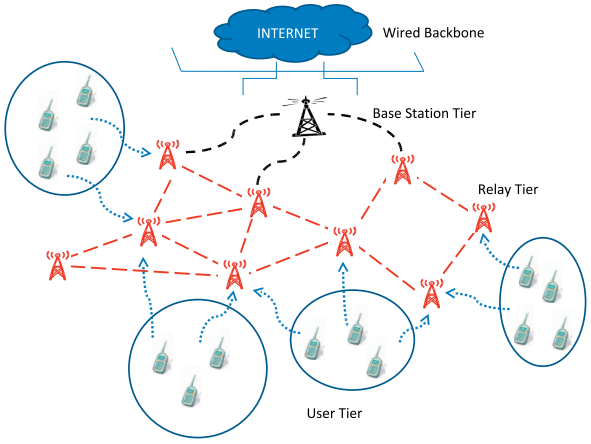

Fig. 1. Network architecture.

2 System Model

Our network topology is based on the multihop cellular network (MCN) model used in emerging BWA networks [26]. As shown in Fig. 1, the network architecture has three tiers of wireless devices: 1) the set of user nodes which are the lowest tier has limited functionality (i.e., do not communicate with one another and have no routing capability); 2) the set of relay nodes that route packets between the user and BS (and also communicate with one another) is the second tier; and 3) the base station is the highest tier and is connected to the wired infrastructure. In order to avoid single points of failure (i.e., failure of a relay node which will disrupt traffic flow), the relays are connected in a mesh manner so that multiple paths are available between the user and BS thereby increasing service availability and fault tolerance.

This topology setup ensures that the network is at least two-link connected (i.e., each node has at least two-link connections to other nodes). Through Menger's Theorem [27], it has been shown that for two distinct nodes x and y, the minimum number of edges whose removal disconnects x and y is equal to the maximum number of pairwise link disjoint paths from x to y. Thus, in our case, two-connectivity is a necessary and sufficient condition to find a solution for two disjoint paths for each user node to the base station. Two-connectivity in wireless networks has been studied in [28] and [29].[2]

We assume that each relay node is equipped with omnidirectional transceivers and relays are used purely for packet forwarding (i.e., relays do not inject traffic into the network). We assume that each user and relay node have fixed transmission power P_{max}, where the P_{max} value is different for the user and relay nodes. Node power level determines the transmission range and interference range which, in turn, determines the SINR of a specific transmission from that node. By keeping the power level of each node the same, we can simplify the interference calculation (since the SINR is dependent on the node's transmission power) and keep the topology of the network unchanged. We also assume that channels have been assigned to the links in the network using a generic link coloring approach. In addition, each node knows the geographic location of all the other nodes in the cell via location discovery schemes [30]. This information is necessary for the receivers to feedback SINR measurements to their respective transmitters.

2. It must be noted that maintaining two-connectivity is a necessary condition for finding two disjoint paths from each user to the base station. Guaranteeing two-connectivity is feasible in a static wireless environment as considered in this paper. However, in the presence of mobility, two-connectivity of the network cannot be ensured due to time varying changes in the topology. Thus, this constraint and the solutions obtained in this paper are pertinent for static wireless networks.

2.1 Interference Model

We represent the network architecture by a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (relays, users, and BS) and $\mathcal{E}$ is the set of edges. In the literature, there are two prominent interference models: *protocol model and physical model*. The protocol model states that two simultaneous transmissions will interfere only within a certain predefined interference range. The physical interference model is less restrictive than the protocol model. It states that a communication between nodes u and v is successful if the SINR at v (the receiver) is above a certain threshold. The SINR for transmission between u and v is given as follows:

$$SINR_{uv} = \frac{P_v(u)}{N + \sum_{w \in V'} P_v(w)} \geq \beta, \tag{1}$$

where $P_v(u)$ is the received power at node v due to node u, N is the noise power, V' is the subset of nodes in the network that are transmitting simultaneously, and β is the SINR threshold.

In this paper, we consider both the protocol and physical interference models, similar to the approach given in [31]. To be specific, we use the following variation of the protocol model, used to accurately mimic the behavior of CSMA/CA-relay-based cellular networks [26]. Let R_T^{max} (r_T^{max}) and R_I^{max} (r_I^{max}) represent the maximum transmission and interference ranges of each relay (user) node, respectively. All relay nodes use the same maximum transmission range (R_T^{max}) as do all the user nodes (r_T^{max}). Each wireless node i (either relay or user node) has a transmission range which is a circle in a 2D plane, centered at i with radius R_T^{max} (r_T^{max}). The transmission range represents the maximum distance up to which a packet can be received, while the interference range represents the maximum distance up to which simultaneous transmissions interfere. In the literature, the interference range is usually chosen to be twice as large as the transmission range which is not necessarily a practical assumption [16]. The actual values of the transmission and interference ranges depend on the transmission power used by the nodes. To provide realistic limits for R_T^{max} (r_T^{max}) and R_I^{max} (r_I^{max}), we use a method called a "reality check." The reality check method, introduced in [32], essentially sets a realistic interference range in which links are assumed to interfere. For the protocol model, R_T^{max} (r_T^{max}) and R_I^{max} (r_I^{max}) are the only two parameters used. Since the underlying physical-layer mechanism is the same, the parameter R_T^{max} (r_T^{max}) should be consistent with the β parameter in the physical model, as shown in (1). Two nodes with distance R_T^{max} (r_T^{max}) should be able to communicate with each other under the maximum transmission power P_{max} and the SINR should be β. As a result, according to [32], R_T^{max} (r_T^{max}) is $\frac{P_{max}}{\beta}$.

Note that the maximum interference range $R_T^{max}(r_T^{max})$ is a parameter introduced by the protocol model and there is no corresponding parameter in the physical model. The only requirement on R_I^{max} (r_I^{max}) is R_I^{max} $(r_I^{max}) > R_T^{max}$ (r_T^{max}), i.e., a lower bound for R_I^{max} (r_I^{max}) is R_T^{max} (r_T^{max}). Thus, if we set the interference range to be slightly higher than the transmission range, $R_T^{max}(r_T^{max}) = \frac{P_{max}}{\beta}$, then the solution is more realistic.

Since link-layer availability is required for CSMA/CA, an ACK packet is generated by each receiver for every data packet it receives. Due to carrier sensing and RTS/CTS/ACK exchanges, a transmission along link $e = (u, v)$ (in either direction) blocks all simultaneous transmissions within the interference ranges of u and v. In the physical interference model, successful reception of a packet sent by node u to node v depends on the SINR at v. To be coherent with the link-layer availability, we extend the physical interference model as follows: We assume that a packet sent by node u is correctly received by node v if and only if the packet is successfully received by v, and the ACK sent by node v is correctly received by node u. Furthermore, for a transmission from node x to node y that is concurrent with the packet on (u, v), we account for the interference both from node $x's$ data packet and node $y's$ ACK. Although only one of x and y transmits at a time, their data and ACK packets could both overlap with either the data packet or the ACK along (u, v). Thus, we choose the maximum of the interferences from x and y when calculating the total interferences at u and v. Note that which of the two (x or y) contributes the maximum interference could be different at u and v. Thus, a packet sent along link (u, v) (in either direction) is correctly received if and only if:

$$SINR_{uv} = \frac{P_v(u)}{N + \sum_{(x,y)\in E'} max(P_v(x), P_v(y))} \geq \beta, \quad (2)$$

and

$$SINR_{vu} = \frac{P_u(v)}{N + \sum_{(x,y)\in E'} max(P_u(x), P_u(y))} \geq \beta, \quad (3)$$

where E' contains all links that have simultaneous transmissions concurrent with the one on (u, v).

It must be noted that optimization techniques to find an efficient algorithm that determines the collision domain and backoff times for each node based on the interference range have been studied [33]. The authors propose closed-form expressions for the mean backoff time in terms of path flow variables, making it possible to optimize the network based on multipath routing. However, their approach is analytically complex. In addition, since the focus of this paper is to incorporate the physical-layer interference into the protocol model, determining the *optimal* collision domain and wait periods is not relevant.

2.2 Isotonicity

As mentioned earlier, isotonicity reflects the ability of a routing metric to compute minimum weight and loop-free paths. Assume that for any path a, its weight is defined by a routing metric, which is a function of a, denoted as $W(a)$. Denoting the concatenation of two paths a and b by $a \oplus b$, isotonicity can be defined as follows:

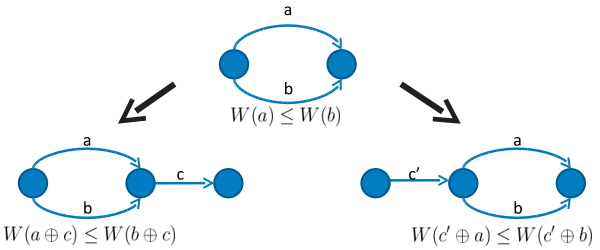

Fig. 2. Example of the isotonic property.

Definition 2.1 (Isotonicity). *A routing metric $W(\cdot)$ is isotonic if $W(a) \leq W(b)$ implies that both $W(a \oplus c) \leq W(b \oplus c)$ and $W(c' \oplus a) \leq W(c' \oplus b)$, for all a, b, c, c'.*

Fig. 2 illustrates the isotonicity property. In [19], it was shown that isotonicity is a sufficient and necessary condition for both the Bellman-Ford and Dijkstra's algorithms to find minimum weight paths that are loop-free. Therefore, if a routing metric can be proven to be isotonic, any variation of a shortest path algorithm can be used to route packets in a wireless network.

3 Routing with Interflow and Intraflow Interference Metric (RI^3M)

3.1 Problem Formulation

The RI^3M interference routing metric takes into consideration the following three factors: interflow interference, intraflow interference, and traffic load. Interflow interference generally results in bandwidth starvation for some nodes since a flow contends for bandwidth along its own path and its neighboring area. To prevent such starvation, the routing metric must balance the traffic load along the path of the flow and reduce the interflow interference imposed in the neighboring area. RI^3M consists of two components. The first component, IL, deals with interflow interference and load awareness. The second component, channel switching cost CSC, captures intraflow interference. We now formalize our routing metric. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be an undirected, two-connected network, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of links. Let p be a path from a user node to the BS. We define RI^3M as follows:

$$\sum_{\forall(i,j)\in p} IL_{ij} + \sum_{\forall i\in p} CSC_i, \quad (4)$$

where node i represents a node on path p and link (i, j) represents a link on the path p.

3.1.1 IL_{ij} Component

The IL_{ij} component is intended to depict information about the interflow interference and traffic load simultaneously. It consists of two separate subcomponents. To capture the interflow interference, we use the concept of the interference ratio (IR) [22], which is based on the physical interference model. The IR depicts the interference based on the ratio between SNR and SINR. The IR captures interference by monitoring the signal strength values. When there is no interference (i.e., no interfering neighbors or no traffic generated by interfering neighbors), the SINR of link

(i, j) is independent of the interflow interference and the quality of the link is determined by the intraflow interference component. Equation (5) shows the IR ratio:

$$IR_{ij} = \frac{SINR_{ij} + SINR_{ji}}{SNR_{ij}}, \tag{5}$$

where SNR is given by $\frac{P_j(i)}{N}$ and the SINR in the numerator is the sum of the SINR values given in (2) and (3).

To estimate the traffic load on a wireless relay node, a typical approach is to measure the traffic volume going through the corresponding node in terms of byte rate or packet rate. Unfortunately, this approach is unable to give an accurate estimate of the usage of the radio channel at which the node operates because the total capacity of the network is not fixed and depends on many factors, such as the physical transmission rate of each relay node, frame size, number of retransmissions, interference, etc. Simply counting the bytes or even packets going through a relay node fails to take into account these factors. In light of these limitations, the authors of [34] adopt an alternative approach to estimate the traffic load, which is based on the percentage of channel time of the relay node that is consumed for frame transmission.

To measure the traffic load, we use the concept of Channel Busy Time (CBT). A radio channel's time consists of a series of interleaved busy periods and idle periods. A busy period is a time period in which one node attempts to transmit frames, while other nodes hold off their transmission. An idle period is a time period in which every node considers the radio medium available for access. Using the CBT, it is possible to estimate the traffic load (channel utilization) on each link. The CBT calculation is the percentage of time that a channel is busy (transmitting). In order to compute this time, we first define the different states that a node can be assigned:

- Success: This state refers to the case where a node has successfully received the acknowledgment of the packet it has sent.
- Backoff: Even though a node has some data to transmit and the medium is free, there is a random waiting period (during which the wireless medium has to remain idle) before it starts sending its data.
- Wait: If there are ongoing transmissions within the interference range of the node which causes the SINR threshold to drop below β, it has to wait until the ongoing communications are completed before starting its own.
- Collision: In this state, a node which has sent a packet never receives an acknowledgment for this packet.

Let $T_{success}$, $T_{backoff}$, T_{wait}, and $T_{collision}$ be the time spent, respectively, in the states Success, Backoff, Wait, and Collision. The idle time (i.e., time where there is no data to keep the channel busy), T_{idle}, considers backoff times, collision times, and the waiting times. Thus, the percentage of time the channel spends idle is defined as

$$T_{idle} = \frac{T_{backoff} + T_{collision} + T_{wait}}{T_{backoff} + T_{collision} + T_{wait} + T_{success}}. \tag{6}$$

Let us denote the denominator of (6) as the total time T_{total}. Then, the CBT for a link (i, j) is defined as

$$CBT_{ij} = \frac{T_{total} - T_{idle}}{T_{total}}. \tag{7}$$

The CBT is used as a smoothing function, weighted over IR_{ij}. Using the IR_{ij} and CBT_{ij} subcomponents, IL_{ij} is defined as follows:

$$IL_{ij} = (1 - IR_{ij}) * CBT_{ij}, \tag{8}$$

where $0 \leq IR \leq 1$ and $0 \leq CBT \leq 1$.

3.1.2 CSC Component

To reduce the intraflow interference, the RI^3M routing metric uses the CSC component. CSC, originally defined in [19], designates paths with consecutive links using the same channel with higher weight than paths that alternate their channel assignments. This allows paths with more diversified channel assignments to be favored in the routing process. Intraflow interference can occur between successive nodes on a path, however, depending on the interference range, it can also occur between nodes further away along the path. In this case, it is necessary to consider the channel assignments at more hops in order to choose an effective path that reduces intraflow interference. To eliminate the intraflow interference between node i and its previous hop, $prev(i)$, node i must transmit to the next hop, $next(i)$, using a different channel from the one it uses to receive from $prev(i)$. CSC denotes $CH(i)$ as the channel that node i transmits on to $next(i)$. The CSC of node i for intraflow interference reduction of successive nodes is given as

$$CSC_i = \begin{cases} w_1, & \text{if} \quad CH(prev(i)) \neq CH(i), \\ w_2, & \text{if} \quad CH(prev(i)) = CH(i), \end{cases} \tag{9}$$

where $w_2 > w_1 \geq 0$ to ensure that a higher cost is imposed for those nodes that transmit on the same channel consecutively. In order to capture intraflow interference between two nodes that are 2 hops away, node i interferes with both nodes $prev(i)$ and $prev^2(i)$, where $prev^2(i)$ is the node that is the 2 hop precedent of i. According to [19], the multihop extension of the CSC equation of (9) is

$$CSC_i = \begin{cases} w_2, & \text{if} \quad CH(prev^2(i)) \neq CH(i) = CH(prev(i)), \\ w_3, & \text{if} \quad CH(prev^2(i)) = CH(i) \neq CH(prev(i)), \\ w_2 + w_3, & \text{if} \quad CH(prev^2(i)) = CH(i) = CH(prev(i)), \\ w_1, & \text{otherwise}, \end{cases} \tag{10}$$

where w_3 captures the intraflow interference between nodes $prev^2(i)$ and i and w_2 captures the intraflow interference between nodes $prev(i)$ and i. The weight w_3 must be strictly less than the weight w_2 because since the further away that two nodes are, the less interference exists between them. We consider intraflow interference up to the limit of a node's interference range which is typically within a 3 hop range.

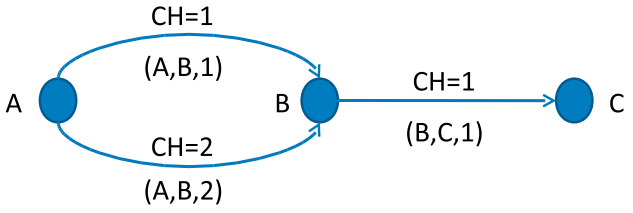

Fig. 3. Example to show nonisotonicity of RI^3M routing metric.

3.2 Virtual Network Decomposition to Illustrate Isotonicity

The RI^3M routing metric is not isotonic if used directly. We can see this in the example network given in Fig. 3. In the example, a link is represented by three parameters: starting node of the link, ending node of the link, and the channel the link transmits on. If we assume that link $(A, B, 1)$ has a smaller RI^3M value than link $(A, B, 2)$, the weights of paths $(A, B, 1)$ and $(A, B, 2)$ satisfy: $RI^3M(A, B, 1) < RI^3M(A, B, 2)$. However, adding path $(B, C, 1)$ to path $(A, B, 1)$ introduces a higher cost than adding $(B, C, 1)$ to $(A, B, 2)$ because of the reuse of channel 1 on path $(A, B, 1) \oplus (B, C, 1)$. Thus, $RI^3M((A, B, 1) \oplus (B, C, 1)) > RI^3M((A, B, 2) \oplus (B, C, 1))$, which does not satisfy the definition of isotonicity as given in Section 2.2.

To make RI^3M into an isotonic routing metric, we use a decomposition technique that creates a virtual network from the real network and decomposes RI^3M into isotonic link weight assignments on the virtual network. First introduced in [19] to prove the isotonicity of the MIC routing metric, the decomposition of RI^3M is based on the fact that the nonisotonic behavior of RI^3M is caused by the different increments of path weights due to the addition of a link on a path. Whether a cost increment will be different by adding a link is only related to the channel assignment of the previous link on the path. Since the possible assignments of channels for the precedent link are limited, we introduce several virtual nodes to represent these possible channel assignments. Namely, for every channel c that a node $X's$ radios are configured to, two virtual nodes $X_i(c)$ and $X_e(c)$ are introduced. $X_i(c)$ indicates that node $prev(X)$ transmits to X on channel c. $X_e(c)$ indicates that node X transmits to its next hop $next(X)$ on channel c. The subscript i stands for ingress and the subscript e stands for egress. In addition, two additional virtual nodes are introduced, $X-$ and $X+$, which represent the start and end nodes of a flow (i.e., $X-$ is used as the virtual destination node for flows destined to node X and $X+$ is used as the virtual source node for flows starting at node X). Hence, $X+$ has a link weight with 0 pointing to each egress node and $X-$ has a link weight 0 with each ingress virtual node of X.

Links from the ingress virtual nodes to the egress virtual nodes at node X are added and the weights of these links are assigned to capture different CSC costs. Link $(X_i(c), Xe(c))$ represents that node X does not change channels while forwarding packets, and hence, weight w_2 is assigned to this link. Similarly, weight w_1 is assigned to link $(X_i(c), X_e(c1))$, where $c \neq c1$, to represent the low cost of changing channels while forwarding packets. Links between the virtual nodes belonging to different real nodes are used to capture the IL weight. Fig. 4 shows the virtual decomposition of Fig. 3.

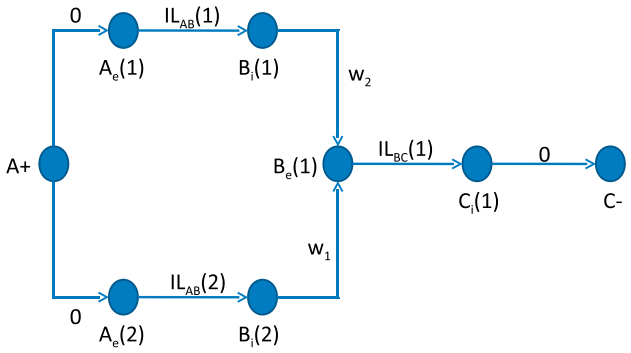

Fig. 4. Decomposition of the network in Fig. 3 into a virtual network to show that RI^3M is isotonic.

By building the virtual network from a real network, RI^3M is essentially decomposed in the real network into weight assignments to the links between virtual nodes. This is because the RI^3M weight of a real path in a real network can be reconstructed by aggregating all of the weights of the virtual links on the corresponding virtual path. The IL part of RI^3M is reflected in the weight of the links between virtual nodes in different real nodes. The CSC costs are captured by routing through different virtual links inside real nodes. Table 1 illustrates the real network mapping into the virtual network.

3.3 Multipath Routing Using RI^3M

Now that RI^3M has been shown to be isotonic using a virtual network decomposition, it can be used with any shortest path algorithm to find least interfering (minimum weight) paths. The problem of finding two link disjoint paths (primary and backup) of minimum total weight across a network has been dealt with efficiently by Suurballe's algorithm [35]. The algorithm developed by Suurballe has become the reference algorithm for finding link disjoint paths in wireless networks. Suurballe's algorithm always finds two link disjoint paths from a source node to the destination, as long as the paths exist in the network, assuring the total weight of both paths is the minimum among all pairs of paths in the network. We run Suurballe's algorithm on the virtual network $\mathcal{G}_v(\mathcal{V}_v, \mathcal{E}_v)$, where $\mathcal{V}_v$ and $\mathcal{E}_v$ are the nodes and links of the virtual network, respectively. The link weights are determined by the values of the RI^3M routing metric. Due to space constraints, the steps of Suurballe's routing algorithm are omitted in this paper. For further details of Suurballe's algorithm, refer to [35].

TABLE 1
Real Network Mapping to the Virtual Network

Real Path	Virtual Path	RI³M
$(A, B, 1) \oplus (B, C, 1)$	$A_e(1) \rightarrow B_i(1) \rightarrow B_e(1) \rightarrow C_i(1)$	$IL_{AB}(1) + IL_{BC}(1) + w_2$
$(A, B, 2) \oplus (B, C, 1)$	$A_e(2) \rightarrow B_i(2) \rightarrow B_e(1) \rightarrow C_i(1)$	$IL_{AB}(2) + IL_{BC}(1) + w_1$

4 Lexicographic MMF Multipath Flow (LMX:M^3F) Routing Algorithm with Interference Constraints

4.1 Problem Formulation and Definitions

In this paper, we model the MMF bandwidth allocation problem as a multicommodity flow (MCF) problem. The MCF problem is a network flow problem where multiple commodities (demands) flow through the network. We consider the case that each demand has two candidate paths (where the paths are determined by using RI^3M). Thus, the flows realizing each demand volume are split among the allowable paths. In the remainder of this paper, we will denote vectors with bold letters and an arrow overhead. We will denote optimal vectors as regular vectors except with an additional star (*).

Definition 4.1 (Multicommodity Flow). *Given $\mathcal{D}$ demands, let $\delta_{edp}x_{dp} \geq 0$ be the flow allocated to path p of commodity (demand) d, $d \in \mathcal{D}$ on link $e \in \mathcal{E}$, where δ_{edp} is a binary variable that denotes whether link e belongs to path p or not. Also, consider a vector $\vec{\mathbf{X}}_d = (x_{dp} : \forall p, d \in \mathcal{D})$ as a single commodity flow of commodity d. A multicommodity flow is the union of flows for each commodity. Specifically, $\vec{\mathbf{X}} = (\vec{\mathbf{X}}_d : d \in \mathcal{D})$ is a feasible multicommodity flow if $\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}_d} \delta_{edp}x_{dp} \leq C_e$.*

The capacity of link $e \in \mathcal{E}$ is denoted C_e and is mathematically expressed as

$$C_e = log_2(1 + SINR_e) \geq \beta, \tag{11}$$

where $SINR_e$ is given in (1).

In this paper, our objective is to attain the MMF bandwidth allocation vector under interference constraints where the allocation vector is lexicographically the largest possible.

Definition 4.2. *An n-vector $\vec{\mathbf{x}} = (x_1, x_2, \ldots, x_n)$ sorted in nondecreasing order $(x_1 \leq x_2 \leq \cdots \leq x_n)$ is **lexicographically greater** than another n-vector $y = (y_1, y_2, \ldots, y_n)$ sorted in nondecreasing order $(y_1 \leq y_2 \leq \cdots \leq y_n)$ if an index k, $0 \leq k \leq n$ exists, such that $x_i = y_i$ for $i = 1, 2, \ldots, k$ and $x_k > y_k$.*

In the following section, we will discuss how our lexicographic bandwidth allocation algorithm is formulated using the interference-aware routing metric developed in Section 3.

4.2 LMX:M^3F Algorithm

Given the network $\mathcal{G}$, paths for routing the traffic flow are found by using the routing metric given in Section 3 and running Suurballe's multipath routing algorithm. Given these paths, we provide the formulation of the lexicographically largest allocation vector using MMF considering interference constraints and the subsequent methodology used to solve it. The LMX:M^3F formulation is given in (12)-(15) (referred to as Problem A in the remainder of the paper) and follows a multicommodity flow approach.

LMX:M^3F: Problem A

Objective. Find total bandwidth allocation vector $\vec{\mathbf{X}}$ such that it is lexicographically maximal among all total bandwidth allocation vectors.

$$\text{lexicographically maximize } \vec{\mathbf{X}} \tag{12}$$

subject to

$$\sum_{p\in\mathcal{P}_d} x_{dp} = X_d, \quad \forall d \in \mathcal{D}, \tag{13}$$

$$\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}_d} \delta_{edp}x_{dp} \leq C_e, \quad \forall e \in \mathcal{E}, \tag{14}$$

$$x_{dp} \geq 0, \tag{15}$$

where $\mathcal{P}_d$ are the paths for demand d, x_{dp} is the flow (bandwidth) allocated to path p of demand d, and X_d is the total flow (bandwidth) allocated to demand d, $\vec{\mathbf{X}} = (X_1, X_2, \ldots, X_D)$.

In order to find the MMF allocation vector for the corresponding paths, we define the *demand satisfaction vector* $\vec{\mathbf{t}}$. Let $\gamma_d \geq 0$ be the flow value of x_{dp}, and $\zeta^+(v)$ and $\zeta^-(v)$ be the outgoing and incoming links to node v, respectively. The law of flow conservation states that

$$\sum_{e\in\zeta^+(v)} x_{dp} - \sum_{e\in\zeta^-(v)} x_{dp} = \begin{cases} \gamma_d, & \text{if } v = BS, \\ -\gamma_d, & \text{if } v = source, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

A feasible multicommodity flow $\vec{\mathbf{X}}$, with $\gamma_d \geq h_d$, $d \in \mathcal{D}$, defines an admissible flow (bandwidth), where h_d is the amount of demand to be routed. Assume that $\vec{\mathbf{X}}$ is feasible and also consider a vector $\vec{\mathbf{t}} = (t_d \geq 0 : d \in \mathcal{D})$ such that $\gamma_d = t_d h_d$ in (16). If $t_d \geq 1$ for all $d \in \mathcal{D}$, then the flow is admissible (i.e., it fulfills the demand requirement h_d, $d \in \mathcal{D}$). Thus, $\vec{\mathbf{t}}$ is denoted as the demand satisfaction vector for routing vector $\vec{\mathbf{X}}$. Specifically, the physical meaning of the value t is the amount that is added to saturate/satisfy x_{dp}. We solve for t using the optimization formulation given in (17)-(21) (referred as Problem B in the remainder of the paper).

Problem B.

$$\text{maximize } t \tag{17}$$

subject to

$$X_d = \sum_{p\in\mathcal{P}_d} x_{dp}, \quad \forall d \in \mathcal{D}, \tag{18}$$

$$t - X_d \leq 0, \quad \forall d \in \mathcal{D}, \tag{19}$$

$$\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}_d} \delta_{edp}x_{dp} \leq C_e, \quad \forall e \in \mathcal{E}, \tag{20}$$

$$x_{dp} \geq 0. \tag{21}$$

The objective function in (17) and the constraint in (19) are equivalent to the ultimate objective to be achieved, given in (22):

$$\max \min X_d : d \in \mathcal{D}. \tag{22}$$

Problem A can be solved by computing consecutively the value of the demand satisfaction vector of Problem B. Primarily, the idea is that first, the lowest value among the components of $\vec{\mathbf{t}}$ has to be maximized before the second lowest value is maximized. In order to ensure that the demands are satisfied, we have to check which total demand allocations X_d can be further increased. A demand d whose satisfaction value t_d cannot be further increased is called blocking [36]. To check the satisfaction of a demand, the following linear program (LP) ((23)-(27)), referred to as Problem C, is solved for each demand d

LMX:M^3F Algorithm

Step1: Solve Problem B. Let $(t^*, \vec{\mathbf{x}}^*, \vec{\mathbf{X}}^*)$ be the optimal solution of Problem A. Initialize: $k := 0$ (number of iterations), $Z_0 := \emptyset$ (set of demands that are blocking/saturated) $Z_1 = \{1, 2, ..., \mathcal{D}\}$, and $t_d := t^*$ for each $d \in Z_1$.

Step2: $k := k + 1$. Consider each demand, $d \in Z_1$, one by one to check whether the total allocated bandwidth X_d^* can be increased more than t^* without decreasing the already found maximal allocations t'_d for all other demands, d'. To check the demands, solve Problem C. If there are no blocking demands in Z_1, go to **Step3**. Otherwise for blocking demand d, add d to set Z_0 and delete it from set Z_1, $Z_0 := Z_0 \cup \{d\}$, $Z_1 := Z_1 \setminus \{d\}$. If $Z_1 = \emptyset$, STOP.
Then $\vec{\mathbf{X}}^* = (X_1^*, X_2^*, ..., X_D^*) = (t_1, t_2, ..., t_d)$ is the solution of Problem A.

Step3: To improve the current best bandwidth allocation, solve the following LP (**Problem D**).

maximize t
subject to
$X_d = \sum_{p \in \mathcal{P}_d} x_{dp}, \forall d \in Z_1$
$t - X_d \leq 0, \forall d \in Z_0$
$\sum_{d \in \mathcal{D}} \sum_{p \in \mathcal{P}_d} \delta_{edp} x_{dp} \leq C_e, \forall e \in \mathcal{E}$
$x_{dp} \geq 0$

Let $(t^*, \vec{\mathbf{x}}^*, \vec{\mathbf{X}}^*)$ be the optimal solution of Problem D. Put $t_d := t^*$ for each $d \in Z_1$. Go to **Step2**.

Fig. 5. Algorithm for LMX:M^3F.

Problem C.

$$\text{maximize } X_d \tag{23}$$

subject to

$$X_{d'} = \sum_{p \in \mathcal{P}_{d'}} x_{d'p}, \quad \forall d' \in \mathcal{D}, \tag{24}$$

$$t_{d'} - X_{d'} \leq 0, \quad \forall d' \in \mathcal{D}, \tag{25}$$

$$\sum_{d' \in \mathcal{D}} \sum_{p \in \mathcal{P}_{d'}} \delta_{ed'p} x_{d'p} \leq C_e, \quad \forall e \in \mathcal{E}, \tag{26}$$

$$x_{d'p} \geq 0, \tag{27}$$

where $t_{d'}$ are constants. To put Problem C in perspective, let t^* be the optimal solution of the LP. A demand is nonblocking (can be further increased) if the optimal X_d value, X_d^*, is strictly greater than t^* (i.e., $X_d^* > t^*$).

The components of Problem B and Problem C are used in conjunction to solve the original LMX:M^3F (Problem A) problem. The algorithm for solving LMX:M^3F is given in Fig. 5.

5 Performance Evaluation

5.1 Simulation Model and Performance Metrics

We consider a two-connected cellular network $\mathcal{G}$ in a $900 \times 900\ \mathrm{m}^2$ region where all nodes are stationary. Each user generates traffic and the flows are routed to and from the base station. We use NS-2 to simulate the networks and use CPLEX to solve the optimization formulation for LMX:M^3F. The base station is located in the center of the network. Locations for the set of relay nodes that form the mesh network and the users are randomly generated. We assume that the BS and relays have an infinite buffer, thus eliminating complications due to buffer overflow. The simulation parameters used are as follows: System Bandwidth $(\mathrm{W}) = 1$ MHz, AWGN $\mathrm{Noise} = -90$ bBW/Hz; Transmission power: Relay (35 dBm), User (24 dBm) (note that the power levels of the nodes are such that it is sufficient to allow nodes to connect to at least two of its neighbors, ensuring two-connectivity); PHY Specification: 802.11; Number of channels per radio: 12; and Antenna: Omnidirectional. To evaluate the performance of RI^3M, we study the following performance metrics: 1) end-to-end delay (amount of time it takes to deliver packets from the client node to the BS); 2) flow throughput; and 3) packet loss ratio. We simulate 20 runs for each set of data and show the average results. To evaluate the performance of LMX:M^3F, we adopt the following performance metrics: 1) bandwidth blocking ratio (BBR): BBR represents the percentage of the amount of blocked traffic over the amount of bandwidth requirements of all traffic requests (connection requests) during the entire simulation period; 2) total bandwidth usage: this measurement helps us examine whether our LMX:M^3F algorithm can save more network resources (use less) than other established MMF routing algorithms that incorporate interference; and 3) link load: measurement

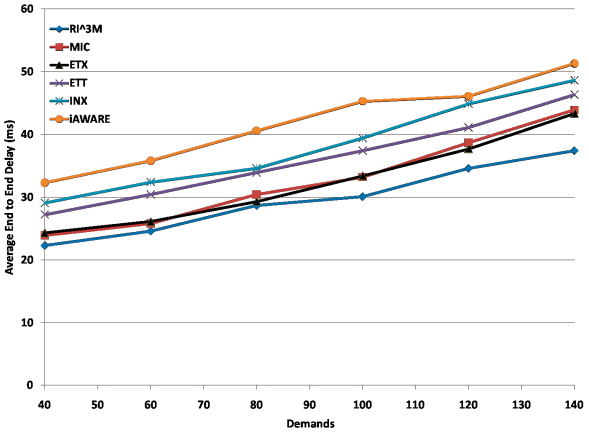

Fig. 6. Average end-to-end delay values for RI^3M compared to prominent routing metrics in the literature.

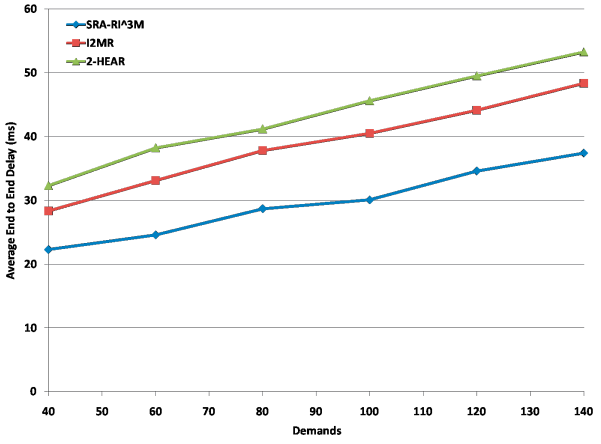

Fig. 7. Comparison of average end-to-end delay for Suurballe's disjoint multipath routing algorithm using RI^3M (SRA-RI^3M) and two established disjoint multipath routing algorithms: I2MR and 2-HEAR.

that indicates the traffic load on each link due to different routing approaches. Note that the performance evaluation of LMX:M^3F is based upon the paths determined from using the RI^3M routing metric.

As benchmarks for evaluating the effectiveness of our proposed metric we compare with five other routing metrics in the literature, specifically, ETX [20], ETT [21], MIC [19], iAWARE [22], and INX [23]. Each metric is used with Suurballe's disjoint multipath routing algorithm. We also compare our proposed approach with two disjoint multipath routing algorithms. First, the algorithm developed in [37] develops a routing metric where a node calculates the SINR to its neighboring links based on a 2-Hop interference estimation algorithm (2-HEAR). Second, the algorithm developed in [38] provides an interference minimized multipath routing (I2MR) algorithm that increases throughput by discovering zone disjoint paths using the concept of path correlation. As benchmarks for evaluating the effectiveness of our bandwidth allocation algorithm, we compare LMX:M^3F to two MMF bandwidth allocation algorithms that consider interference when allocating bandwidth. First, the algorithm developed in [15] is an interference-based routing and bandwidth algorithm, known as MICB. The protocol model is used to create an auxiliary graph such that the maximum interference level within the network does not exceed a maximum value. Second, the algorithm described in [14] quantifies interference through the creation of contention graphs where interfering flows are captured in multihop wireless networks. We modify the implementations of these algorithms so that multiple paths are considered.

5.2 Simulation Results and Discussion

We first evaluate RI^3M in terms of end-to-end delay. We use the end-to-end user demand delivery delay as a metric to evaluate the impact of the interference quantification method of RI^3M in comparison to the existing routing metrics and the two established disjoint multipath routing algorithms. To measure the end-to-end delay, the transmitting rate of the user and relay nodes are set to 4.5 Mbps. All routing flows are CBR flows with 512 byte packets. To model the packet dropping error, for a given SINR value, we use the packet error ratio (PER) [39], which is readily available in NS-2.

5.2.1 Performance Evaluation of RI^3M

We first compare RI^3M with the existing routing metrics. We simulate networks with 100 nodes (1BS, 6 relays, and 93 user nodes). Fig. 6 shows the average end-to-end delay values of RI^3M versus the other routing metrics, measured against varying demands (traffic load). We see that the proposed RI^3M achieves the lowest delay in comparison to the other metrics, particularly as demands increase. It can be said that RI^3M quantifies interference more accurately because it considers the influence of interflow and intraflow interference which thereby allows us to avoid paths with high interference, and reduce the time taken to deliver a packet. INX performs most closely to our algorithm since it quantifies interference through the number of links that interfere with another link l. The remaining metrics behave somewhat similarly because most of them are derived from one another (as discussed in Section 1.2). Therefore, despite small implementation differences, there is no overarching performance improvement among the remaining metrics (i.e., ETT, ETX, MIC, and iAWARE). As can be seen from Fig. 6, the delay values for all the metrics (including RI^3M) increase as demands increase, which intuitively is true.

In Fig. 7, the average end-to-end delay values for RI^3M with Suurballe's algorithm, referred to as SRA-RI^3M in the simulation graphs, is compared to the two aforementioned disjoint multipath routing algorithms. They are referred to as 2-HEAR and I2MR in the simulation graphs. The SRA-RI^3M achieves the lowest end-to-end delay compared to the other algorithms. We can justify the better performance of our results as follows: In both 2-HEAR and I2MR, the paths are formed using incomplete interference information. In 2-HEAR, the SINR calculated by each node only includes those nodes within a 2-hop range which means that even if interference beyond this range occurs, it is not captured in the routing metric (interflow and intraflow interference not fully accounted for). If the interference level is high beyond the 2-hop range, then any paths built may not be successful as interference may cause a drop in packets and a retransmission is required. This obviously

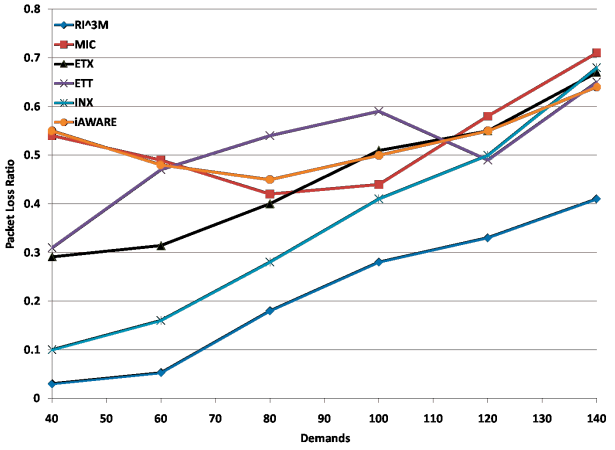

Fig. 8. Comparison of packet loss ratio when using RI^3M versus prominent routing metrics in the literature.

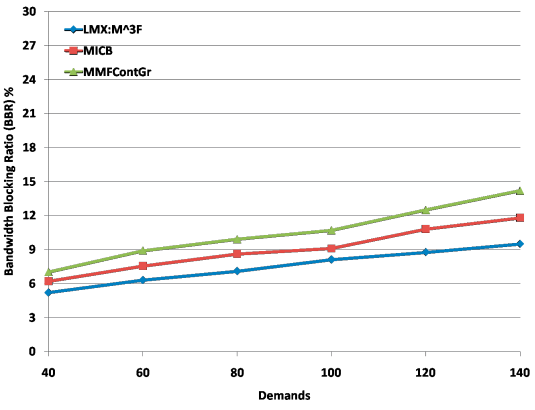

Fig. 10. BBR comparison for networks with 46 nodes (6 relays and 40 users).

incurs delay. A similar argument can be used with the I2MR algorithm. In our case, RI^3M quantifies the interference from both within flows and in the neighboring area.

Next, we show the average packet loss incurred from the various routing metrics and the average flow throughput when each metric is used. Fig. 8 shows the packet loss ratio and Fig. 9 shows the average flow throughput. It can be seen that MIC and iAWARE have the lowest throughput and highest packet loss ratio at low traffic demands in comparison to the other metrics. ETX and INX have better throughput and loss ratios with low loads, but their performance decreases with high traffic demands. In Fig. 8, the ETT metric exhibits unstable behavior primarily because it overestimates link quality by inaccurately probing the channel. Moreover, ETT does not depend on the traffic load. Although MIC and iAWARE partially rely on ETT, these metrics employ normalization functions to smoothen ETT values, and therefore, become more stable. This indicates the unpredictability of the results for the three metrics: ETT, MIC, and iAWARE. The remaining metrics behave intuitively, i.e., greater packet loss as demands increase. The ETT, MIC, and iAWARE routing metrics behave in a similar unpredictable manner for the throughput results shown in Fig. 9 for the same reason given above. Overall, RI^3M is able to achieve higher throughput and lower loss ratio than the remaining metrics over the varying traffic demands shown.

5.2.2 Performance Evaluation of LMX:M^3F

For the LMX:M^3F algorithm, we first evaluate it in terms of BBR. We compare it with MICB [15] and MMFContGr [14], respectively, as shown in the simulation graphs. We run all three algorithms on networks with different densities. Figs. 10 and 11 show the BBR results from the simulated networks with 46 (6 relays and 40 users) and 24 (4 relays and 20 users) nodes, respectively (each network has one base station). It can be seen that our LMX:M^3F algorithm performs the best in most cases. The blocking ratio increases no matter which algorithm is used because of heavier traffic load. The average blocking ratio difference between our solution and that of MICB and MMFContGr is 16 and 13 percent, respectively, for the network of size 46 nodes. Similarly, the average difference between our algorithm and MICB and MMFContGr for the network of size 24 nodes is 18 and 32 percent, respectively. Essentially, the BBR indicates if a connection request for traffic is blocked. If

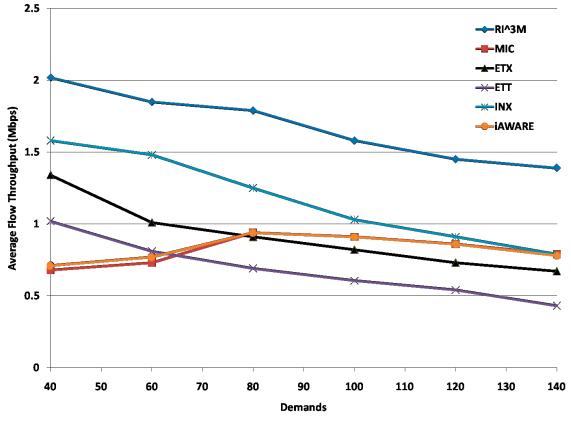

Fig. 9. Average flow throughput generated by RI^3M versus prominent routing metrics in the literature.

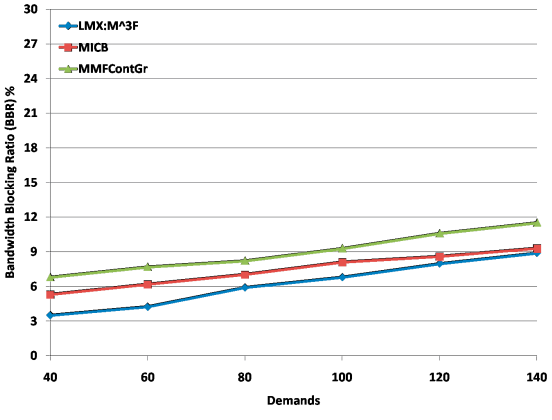

Fig. 11. BBR comparison for networks with 24 nodes (4 relays and 20 users).

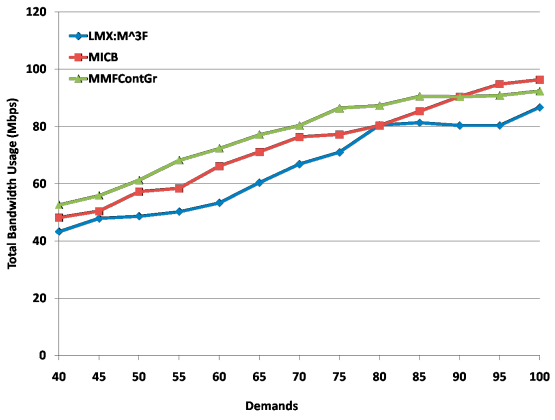

Fig. 12. Comparison of total bandwidth usage for networks with 46 nodes (6 relays and 40 users).

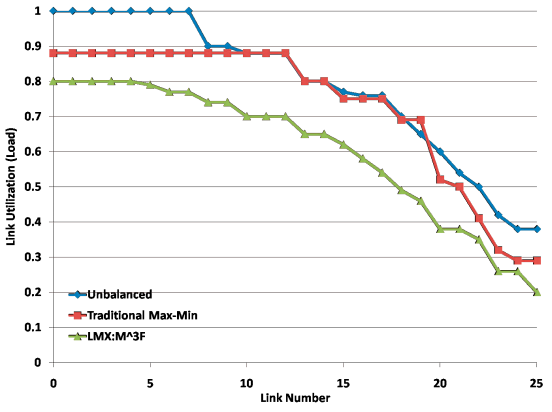

Fig. 14. Link loads on various links for network with 10 nodes.

traffic is blocked, it means that there is less bandwidth on a link than there should be to accommodate the offered traffic. For the best performance, the BBR should be kept as low as possible. Given the BBR results in Figs. 10 and 11, the BBR of LMX:M^3F is lower than that of the MICB and MMFContGr algorithms. Therefore, we can claim that the network performance improves under our proposed algorithm.

Next, we show the real-time network resource usage for all three algorithms. Figs. 12 and 13 show the results of the bandwidth usage for the three algorithms. As expected, LMX:M^3F uses the least amount of bandwidth for varying demands. In the case of 46 nodes, on average, the bandwidth usage of LMX:M^3F compared to MICB and MMFContGr is 11 and 14 percent less, respectively. The bandwidth usage of the LMX:M^3F for the case of 24 nodes is, on average, 2 and 6 percent less than for the other two algorithms. The bandwidth usage shown in Fig. 12 shows that the LMX:M^3F algorithm clearly uses less bandwidth than the other two approaches. However, there is less clarity in the case of Fig. 13 (24 nodes) because the density of the network is less. Therefore, there is not a great deal of difference between the performances of the individual algorithms even though we are simulating against the same number of varying demands. The conclusion is that our approach is more effective in network resource usage in higher density networks. Given that BWA networks are generally used in dense urban settings, our approach fits the application. However, the LMX:M^3F algorithm is time-consuming to solve for very large networks with thousands of demands because each demand must be checked for bandwidth satisfaction (see Problem C). Thus, our algorithm is limited to a certain extent because of scalability.

Lastly, we look at the impact that our algorithm has on the load balancing of the network across various links. We compare the LMX:M^3F algorithm with that of an unbalanced routing scheme (no fairness incorporated) and a traditional max-min fair routing approach, which minimizes the load of only the maximally loaded link in the network (does not look for the lexicographically highest). We simulate networks with 10 (two relays and eight users) and 15 (2 relays and 13 users) nodes (each network has one base station). Figs. 14 and 15 show the link load on various links for networks with 10 nodes and 15 nodes, respectively. The link number represents each individually numbered link in the network. We see that the unbalanced routing scheme has some links with 100 percent utilization. When the traditional max-min routing approach is used, the link

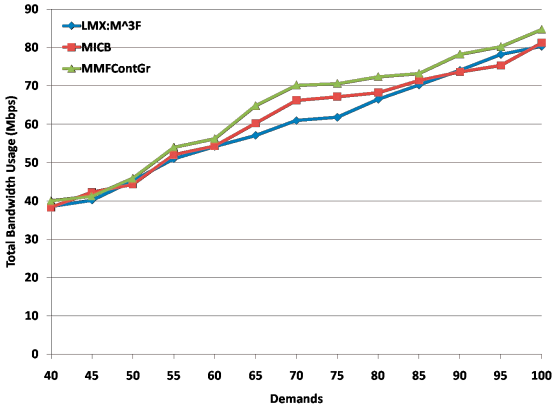

Fig. 13. Comparison of total bandwidth usage for networks with 24 nodes (4 relays and 20 users).

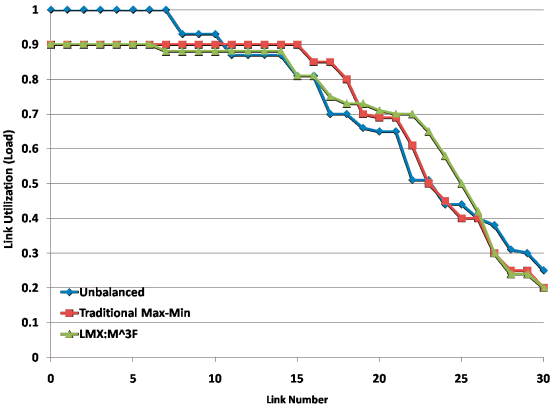

Fig. 15. Link loads on various links for network with 15 nodes.

load utilization is better, but there are still some links that are nearly 90 percent loaded. Our lexicographic bandwidth allocation algorithm performs an optimization of all the links and presents a better load balance of the traffic load, as can be in the results. We observe that the LMX:M^3F algorithm generally results in approximately 75 percent of the links having the same load. We also see that the maximum load of any link is less than 1. This allows for spare capacity to exist on the link so that a proportionate increase in demands can be tolerated.

6 Conclusion

In this paper, we have proposed a novel routing metric, RI^3M, by considering both interflow and intraflow interference to enhance the selection of good quality paths. Using virtual network decomposition, we have shown that RI^3M is an isotonic routing metric that outperforms the most prominent and relevant routing metrics used in the literature in terms of end-to-end delay, packet loss, and throughput. In addition, we have developed an MMF bandwidth allocation algorithm for multipath flow routing in multihop wireless networks. To ensure QoS, our LMX:M^3F optimization formulation has been shown to provide better utilization of bandwidth resources in comparison to well-respected MMF algorithms established in the literature particularly in terms of blocking ratio and link load. In our future work, we will incorporate the effect of mobility on MMF and routing protocols and simplify the implementation of RI^3M to make it more seamless to use with other routing protocols.

References

[1] D. Soldani and S. Dixit, "Wireless Relays for Broadband Access," *IEEE Comm. Magazine,* vol. 46, no. 3, pp. 58-66, Mar. 2008.

[2] H.T. Cheng and W. Zhuang, "An Optimization Framework for Balancing Throughput and Fairness in Wireless Networks with qos Support," *IEEE Trans. Wireless Comm.,* vol. 7, no. 2, pp. 584-593, Feb. 2008.

[3] L. Tassiulas and S. Sarkar, "Maxmin Fair Scheduling in Wireless Ad Hoc Networks," *IEEE J. Selected Areas in Comm.,* vol. 23, no. 1, pp. 163-173, Jan. 2005.

[4] Y. Wang, W. Wang, X.-Y. Li, and W.-Z. Song, "Interference-Aware Joint Routing and tdma Link Scheduling for Static Wireless Networks," *IEEE Trans. Parallel and Distributed Systems,* vol. 19, no. 12, pp. 1709-1725, Dec. 2008.

[5] D. Bertsekas and R. Gallagher, *Data Networks.* Prentice-Hall, 1992.

[6] J. Kleinberg, Y. Rabani, and E. Tardos, "Fairness in Routing and Load Balancing," *J. Computer and System Sciences,* vol. 63, no. 1, pp. 568-578, Aug. 2001.

[7] M. Allalouf and Y. Shavitt, "Centralized and Distributed Algorithms for Routing and Weighted Max-Min Fair Bandwidth Allocation," *IEEE/ACM Trans. Networking,* vol. 16, no. 5, pp. 1015-1024, Oct. 2008.

[8] D. Nace, L. Doan, O. Klopfenstein, and A. Bashllari, "Max-Min Fairness in Multi-Commodity Flows," *Computers and Operations Research,* vol. 35, no. 2, pp. 557-573, Feb. 2008.

[9] D. Nace, N. Doan, E. Gourdin, and B. Liau, "Computing Optimal Max-Min Fair Resource Allocation for Elastic Flows," *IEEE/ACM Trans. Networking,* vol. 14, no. 6, pp. 1272-1281, Dec. 2006.

[10] X. Huang, S. Feng, and H. Zhuang, "Cross-Layer Fair Resource Allocation for Multi-Radio Multi-Channel Wireless Mesh Networks," *Proc. Int'l Conf. Wireless Comm., Networking and Mobile Computing (WiCom),* pp. 2639-2643, Sept. 2009.

[11] V.S. Mansouri, A.H. Mohsenian-Rad, and V.W.S. Wong, "Lexicographically Optimal Routing for Wireless Sensor Networks with Multiple Sinks," *IEEE Trans. Vehicular Technology,* vol. 58, no. 3, pp. 1490-1500, Mar. 2009.

[12] J. Tang, R. Hincapie, G. Xue, W. Zhang, and R. Bustamante, "Fair Bandwidth Allocation in Wireless Mesh Networks with Cognitive Radios," *IEEE Trans. Vehicular Technology,* vol. 59, no. 3, pp. 1487-1496, Mar. 2010.

[13] J. Chou and B. Lin, "Optimal Multi-Path Routing and Bandwidth Allocation under Utility Max-Min Fairness," *Proc. IEEE Int'l Workshop Quality of Service (IWQoS),* pp. 1-9, July 2009.

[14] P. Wang, H. Jiang, W. Zhuang, and H. Poor, "Redefinition of Max-Min Fairness in Multi-Hop Wireless Networks," *IEEE Trans. Wireless Comm.,* vol. 7, no. 12, pp. 4786-4791, Dec. 2008.

[15] J. Tang, G. Xue, and C. Chandler, "Interference Aware Routing and Bandwidth Allocation for Qos Provisioning in Multihop Wireless Networks," *Wireless Comm. and Mobile Computing,* vol. 5, no. 8, pp. 933-943, Dec. 2005.

[16] A. Iyer, C. Rosenberg, and A. Karnik, "What Is the Right Model for Wireless Channel Interference," *IEEE Trans. Wireless Comm.,* vol. 8, no. 5, pp. 2662-2671, May 2009.

[17] P. Thulasiraman and X. Shen, "Interference Aware Subcarrier Assignment for Throughput Maximization in OFDMA Wireless Relay Mesh Networks," *Proc. IEEE Int'l Conf. Comm. (ICC),* pp. 1-6, June 2009.

[18] P. Thulasiraman and X. Shen, "Decoupled Optimization of Interference Aware Routing and Scheduling for Throughput Maximization in Wireless Relay Mesh Networks," *Proc. IEEE Workshop Wireless Mesh Networks (WiMesh) (Colocated with IEEE Conf. Sensor, Mesh and Ad Hoc Comm. and Networks (SECON)),* pp. 1-6, June 2009.

[19] Y. Yang, J. Wang, and R. Kravets, "Designing Routing Metrics for Mesh Networks," *Proc. IEEE Workshop Wireless Mesh Networks (WiMesh),* Sept. 2005.

[20] D.D. Couto, D. Aguayo, J. Bicket, and R. Morris, "A High-Throughput Path Metric for Multi-Hop Wireless Routing," *Proc. ACM MobiCom,* pp. 134-146, Sept. 2003.

[21] R. Draves, J. Padhye, and B. Zill, "Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks," *Proc. ACM MobiCom,* pp. 114-128, Sept. 2004.

[22] A. Subramanian, M. Buddhikot, and S. Miller, "Interference Aware Routing in Multi-Radio Wireless Mesh Networks," *Proc. IEEE Workshop Wireless Mesh Networks (WiMesh),* pp. 55-63, Sept. 2006.

[23] R. Langar, N. Bouabdallah, and R. Boutaba, "Mobility-Aware Clustering Algorithms with Interference Constraints," *Computer Networks,* vol. 53, no. 1, pp. 25-44, Jan. 2009.

[24] P. Thulasiraman, J. Chen, and X. Shen, "Max-Min Fair Multipath Routing with Physical Interference Constraints for Multihop Wireless Networks," *Proc. IEEE Int'l Conf. Comm. (ICC '10),* May 2010.

[25] P. Thulasiraman and X. Shen, "Disjoint Multipath Routing and qos Provisioning under Physical Interference Constraints," *Proc. IEEE Wireless Comm. and Networking Conf. (WCNC '10),* Apr. 2010.

[26] W. Park and S. Bahk, "Resource Management Policies for Fixed Relays in Cellular Networks," *Elsevier Computer Comm.,* vol. 32, no. 4, pp. 703-711, Mar. 2009.

[27] A. Bagchi, A. Chaudhary, and P. Kolman, "Short Length Menger's Theorem and Reliable Optical Routing," *Theoretical Computer Science,* vol. 339, no. 2, pp. 315-332, June 2005.

[28] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint Multipath Routing in Dual Homing Networks Using Colored Trees," *Proc. IEEE Global Telecomm. Conf. (GLOBECOM),* pp. 1-5, Nov./Dec. 2006.

[29] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint Multipath Routing to Two Distinct Drains in a Multi-Drain Sensor Network," *Proc. IEEE INFOCOM,* pp. 643-651, May 2007.

[30] M. Mauve, J. Widmer, and H. Hartenstein, "A Survey on Position Based Routing in Mobile Ad Hoc Networks," *IEEE Networks,* vol. 15, no. 6, pp. 30-39, Nov. 2001.

[31] G. Brar, D. Blough, and P. Santi, "Computationally Efficient Scheduling with the Physical Interference Model for Throughput Improvement in Wireless Mesh Networks," *Proc. ACM MobiCom,* pp. 2-13, Sept. 2006.

[32] Y. Shi, Y. Hou, J. Liu, and S. Kompella, "How to Correctly Use the Protocol Interference Model for Multi-Hop Wireless Networks," *Proc. ACM MobiHoc,* pp. 239-248, May 2009.

[33] J. Zhou and K. Mitchell, "A Scalable Delay Based Analytical Framework for csma/ca Wireless Mesh Networks," *Computer Networks,* vol. 54, no. 2, pp. 304-318, Feb. 2010.

[34] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas, "Cross-Layer Framework for Association Control in Wireless Mesh Networks," *IEEE Trans. Mobile Computing*, vol. 8, no. 1, pp. 65-80, Jan. 2009.
[35] R. Bhandari, *Survivable Networks: Algorithms for Diverse Routing*. Kluwer Academic Publishers, 1999.
[36] D. Nace and M. Pioro, "Max-Min Fairness and Its Applications to Routing and Load Balancing in Communication Networks: A Tutorial," *IEEE Comm. Surveys and Tutorials*, vol. 10, no. 4, pp. 5-17, Jan. 2008.
[37] R. Kortebi, Y. Gourhant, and N. Agoulmine, "On the Use of Sinr for Interference-Aware Routing in Wireless Multi-Hop Networks," *Proc. ACM Int'l Workshop Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, pp. 395-399, Oct. 2007.
[38] J.-Y. Teo, Y. Ha, and C.-K. Tham, "Interference Minimized Multipath Routing with Congestion Control in Wireless Sensor Network with High Rate Streaming," *IEEE Trans. Mobile Computing*, vol. 7, no. 9, pp. 1124-1137, Sept. 2008.
[39] M. Takai, J. Martin, and R. Bragodia, "Effects of Wireless Physical Layer Modeling in Mobile Ad Hoc Networks," *Proc. ACM MobiHoc*, pp. 87-94, Oct. 2001.

Preetha Thulasiraman received the PhD degree from the Electrical and Computer Engineering Department at the University of Waterloo, Waterloo, Ontario, Canada in 2010, the MSc degree in computer engineering from the University of Arizona in 2006, and the BSc degree in electrical engineering from the University of Illinois, Urbana-Champaign in 2004. She has been an assistant professor in the Electrical and Computer Engineering Department at the Naval Postgraduate School in Monterey, CA, since January 2011. Her research interests include network and MAC-layer design of resource allocation algorithms, wireless routing and fault tolerance, wireless mesh and relay networks, combinatorial optimization, and general applications of graph theory. She is a member of the IEEE and the IEEE Computer Society.

Jiming Chen (M'08) received the PhD degree in control science and engineering from Zhejiang University in 2005. He was a visiting scholar at INRIA, NUS. He is an associate professor in the Institute of Industrial Process Control, and the coordinator of the Networked Sensing and Control Group in the State Key Laboratory of Industrial Control Technology at Zhejiang University, China. Currently, he is also a visiting researcher with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests are estimation and control over sensor networks, sensor and actuator networks, target tracking in sensor networks, and optimization in mobile sensor networks. He currently serves as an associate editor for the *International Journal of Communication System* (Wiley), the *Ad Hoc and Sensor Wireless Networks* (Old City Publishing), etc. He also serves as a guest editor for the *IEEE Transactions on Automatic Control*, *Wireless Communication and Mobile Computing* (Wiley), etc. He serves as the general symposia cochair of ACM IWCMC 2009 and ACM IWCMC 2010, WiCON 2010 MAC track cochair, Chinacom 2010 Publicity cochair, and TPC for IEEE ICDCS 2010, IEEE MASS 2010, IEEE INFOCOM 2011, etc. He is a member of the IEEE.

Xuemin (Sherman) Shen received the BSc degree in electrical engineering from Dalian Maritime University, China, in 1982, and the MSc and PhD degrees in electrical engineering from Rutgers University, New Jersey, in 1987 and 1990, respectively. He is a professor and university research chair in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications networks, wireless network security, wireless body area networks, and vehicular ad hoc and sensor networks. He is a coauthor of three books, and has published more than 400 papers and book chapters in wireless communications and networks, control, and filtering. He served as the tutorial chair for the IEEE ICC '08, the technical program committee chair for the IEEE Globecom '07, the general cochair for the Chinacom '07 and QShine '06, and the founding chair for the IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as a founding area editor for the *IEEE Transactions on Wireless Communications*; the editor in chief for *Peer-to-Peer Networking and Application*; and an associate editor for the *IEEE Transactions on Vehicular Technology*, *KICS/IEEE Journal of Communications and Networks, Computer Networks*, *ACM/Wireless Networks*, *Wireless Communications and Mobile Computing* (Wiley), etc. He has also served as a guest editor for the *IEEE JSAC*, *IEEE Wireless Communications*, *IEEE Communications Magazine*, *ACM Mobile Networks and Applications*, etc. He received the Excellent Graduate Supervision Award in 2006, the Outstanding Performance Award in 2004 and 2008 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a registered professional engineer in Ontario, Canada, and a distinguished lecturer of the IEEE Communications Society. He is a fellow of the IEEE.