Joint Optimal Data Rate and Power Allocation in Lossy Mobile Ad Hoc Networks with Delay-Constrained Traffics

Songtao Guo, *Member, IEEE*, Changyin Dang, *Senior Member, IEEE*, and Yuanyuan Yang, *Fellow, IEEE*

Abstract—In this paper, we consider lossy mobile ad hoc networks where the data rate of a given flow becomes lower and lower along its routing path. One of the main challenges in lossy mobile ad hoc networks is how to achieve the conflicting goal of increased network utility and reduced power consumption, while without following the instantaneous state of a fading channel. To address this problem, we propose a cross-layer rate-effective network utility maximization (RENUM) framework by taking into account the lossy nature of wireless links and the constraints of rate outage probability and average delay. In the proposed framework, the utility is associated with the effective rate received at the destination node of each flow instead of the injection rate at the source of the flow. We then present a distributed joint transmission rate, link power and average delay control algorithm, in which explicit broadcast message passing is required for power allocation algorithm. Motivated by the desire of power control devoid of message passing, we give a near-optimal power-allocation scheme that makes use of autonomous SINR measurements at each link and enjoys a fast convergence rate. The proposed algorithm is shown through numerical simulations to outperform other network utility maximization algorithms without rate outage probability/average delay constraints, leading to a higher effective rate, lower power consumption and delay. Furthermore, we conduct extensive network-wide simulations in NS-2 simulator to evaluate the performance of the algorithm in terms of throughput, delay, packet delivery ratio and fairness.

Index Terms—Mobile computing, mobile ad hoc networks, cross-layer optimization, network utility maximization, congestion control, power control, outage probability

1 Introduction

Mobile nodes in ad hoc networks are usually powered by battery with energy constrained [1], [2], [3]. Hence, wise allocation of power is critical in wireless mobile ad hoc networks (MANETs) for both prolonging battery life of the mobile devices and increasing utilization of the scarce wireless spectrum. The aim of power control is to intelligently adjust the transmission power to the least that is required to send the data packet to meet the required signal-to-interference-plus-noise ratio (SINR) threshold and achieve the quality of service (QoS) goals in wireless channels.

Existing power-control schemes, either centralized, such as [4], [5], [6], or distributed, such as [7], [8], [9], [10]), always assume quasi-stationarity of the fading wireless channels and are based on the observed SINR at the receiver or the knowledge of the gains of all the links. Thus, the implicit assumption made is that the power-control updates are made every time the fading state of the channel changes. However, in wireless communication channels, which exhibit fast fading where the fades can change within milliseconds, this might not always be practical. Very frequent power-control updates can also consume a lot of energy. Therefore, it is desirable that the power-control does not need to be updated when the channel meanders from a fading state to another. This is one of our motivations of this work.

Another fundamental task performed frequently by ad hoc networks is to regulate the allowed source rates to maximize the total utility of the users subject to resource constraints. One common approach to achieving this objective is network utility maximization (NUM) through the cross-layer design of joint power and rate control. The cross-layer design takes users' service requirements and interference-limited channel into account to jointly optimize network performance. In fact, the link capacities determined by the power allocation at the physical layer affect the arrival rates at the transport layer, and vice versa. With this coupling, resource optimization within layers is not competent.

Therefore, motivated by the benefits, a cross-lay design of NUM has extensively been studied [11], [12], [13], [14], [15], [16]. In [11], Chiang analyzed such a joint congestion and power control problem by using Lagrangian primal-dual approach. In [12], the authors considered joint routing and rate allocation that maximizes the lifetime of an ad hoc network supporting variable-rate transmissions (link adaptation). In [13], a joint opportunistic power scheduling and end-to-end rate control (JOPRC) scheme was presented for

- *S. Guo is with the College of Electronic and Information Engineering, Southwest University, Chongqing, 400715, P.R. China. E-mail: stguocqu@gmail.com.*
- *Y. Yang is with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook NY 11794, USA. E-mail: yuanyuan.yang@stonybrook.edu.*
- *C. Dang is with the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong. E-mail: mecdang@cityu.edu.hk.*

Manuscript received 9 Sept. 2012; revised 17 Aug. 2013; accepted 11 Dec. 2013. Date of publication 15 Jan. 2014; date of current version 11 Feb. 2015. Recommended for acceptance by M.Y. Eltoweissy.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TC.2013.2296043

wireless ad hoc networks by modeling the time-varying wireless channel as a stochastic process. In [16], the authors modeled an efficient spatial-TDMA based adaptive power and rate cross-layer scheduling problem as a mixed integer linear program (MILP). A cross-layer design of power- and rate-allocation was studied in [17] for MANETs with lossy links. All of these works similarly assume the channel to be quasi-stationary and track the instantaneous channel state; meanwhile, the data rate of a given flow is treated unchanged from hop to hop along its route. In wireless networks, however, wireless links are typically lossy and the packet loss rate is often high due to channel impairment such as fading and interference. As a result, the data rate of a flow decreases along its route because of the lossy nature of wireless links. Clearly, these previous studies that follow fast fades directly are not feasible or desirable in this situation due to mobility of the nodes and/or environment. This fact also motivates our work in this paper.

More recently, there have been many studies on the problem of power control with constraints on outage probability in wireless networks with lossy links, in which outage probability is referred to as the probability that the smallest maximum achievable rate among all of users is smaller than a specified transmission rate. In [18], Kandukuri and Boyd proposed a new method of optimal power control in interference-limited fading wireless channels with outage-probability specifications, where interior point methods are employed to find the solution. In [19], the authors addressed a joint power control and multiuser detection problem with outage-probability constraints in a Rayleigh fast-fading environment. Furthermore, in [20], they proposed a generalized framework for solving the problem under modest assumption on the underlying channel fading. Unlike previous work, which dealt with Rayleigh fast-fading model, each user is allowed to have a different fading distribution. In [21], the outage probability of multiple antenna multicast channels was investigated and its upper and lower-bounds were derived by using extreme value theory. A link outage probability based resource allocation scheme for multi-radio multi-channel wireless mesh networks was proposed in [22]. However, the objective of these works is to minimize the total power of all users while meeting their outage targets or minimize the outage probability, whereas they do not consider the maximization problem of network utility.

An emerging challenging issue to be addressed for dynamic power and rate allocation in next generation MANETs is how to meet with user's heterogeneous transmission QoS requirements. Among others, the demand for wireless high-speed connectivity for both delay-tolerant "packet" data and delay-sensitive "circuit" data is expected to rise significantly in the next decade. Accordingly, how to optimally allocate both power and rate to support delay-constrained data traffic while sufficiently considering the lossy nature of links becomes an important research issue. Therefore, this paper aims to provide concrete answers to the aforementioned problems.

In this paper, we first propose a novel rate-effective network utility maximization (RENUM) framework, which takes account of the lossy nature of links in the objective function and the constraints of rate outage and average delay. In particular, the minimization of the overall power and delay of all links is used as the optimization objective as well, which is the distinct difference between effective network utility maximization (ENUM) with outage constraints in [23] and our work. In our framework, the transmission rate at the source node of each flow is called the *injection rate*, and the data rate correctly received at the destination node is called the *effective rate*. Since the effective rate is typically lower than the injection rate, it is natural to examine the utility corresponding to the effective rate. Then the original non-convex RENUM is converted into a convex one by employing some logarithmic transformation. The convex RENUM problem is further decomposed into three separate maximization sub-problems of rate control, power and delay allocation. Three corresponding distributed sub-algorithms are developed, in which some forms of broadcast message passing is required for power-allocation. In practice, it may be desirable to avoid such overhead and thus we include a near-optimal scheme that makes use of autonomous SINR measurements at each link for power-allocation. To the best of our knowledge, this paper is the first work to address the limits of dynamic power and rate allocation over the lossy links under various average delay and rate-outage probability constraints.

The contributions and findings of the paper can be briefly summarized below.

- The problem of the frequent rate and power updates induced by fast-fading is addressed, which is imperative especially in a MANET with limited energy, by taking account of the average SINR and rate-outage probability, where flow rate changes per hop.
- A rate-effective network utility maximization framework is proposed by considering the lossy nature of wireless links and the rate outage constraints with a small tolerable value, which is necessary because the loss of wireless links and packets occurs frequently due to channel impairment and mobility of nodes in MANETs.
- The minimization of the overall power and average delay of all links is used as the optimization objective while the effect of the loss and the rate outage on optimal joint power, rate and delay control is studied to support delay-constrained data traffic.
- The distributed cross-layer power and rate control algorithms with/without explicit broadcast message passing are proposed by solving the non-convex RENUM problem. Furthermore, our extensive numerical and simulation results demonstrate that the RENUM outperforms other NUM schemes without rate outage probability/average delay constraints in terms of effective rates, link delay and power consumption.

The reminder of this paper is organized as follows. In Section 2, we introduce the system model. Section 3 outlines the RENUM framework. In Section 4, we propose a distributed joint rate, power and average delay control algorithm. In Section 5, we further give the near-optimal rate and power control algorithm making use of noise measurements instead. Illustrative numerical and simulation results are provided in Section 6. Finally, Section 7 concludes the paper.

2 System Model

2.1 Network Model

In this paper, we consider a MANET that consists of a set $L = \{1, 2, \ldots, K_L\}$ of logical links, shared by a set $S = \{1, 2, \ldots, K_S\}$ of sources, where K_L and K_S denote the number of links and sources, respectively. Each source $s \in S$ emits one flow to its destination node using a set of links $L(s) \subseteq L$ for its route. Due to the unreliability of the network, routing may be adapted and optimized with the improved routes communicated to the transport layer. In practice, our formulated optimization model in Section 3 for the RENUM problem is suitable for the scenario of varying routes. This is because our model is based on hop-by-hop rate control and the proposed RENUM algorithm in Section 4 is a distributed hop-by-hop algorithm by considering the lossy nature of each link. Each source s injects data into the network at an average rate of $x_s \in [m_s, M_s]$ where m_s and M_s denote the minimum and maximum rates, respectively. For each link l, the set $S(l) = \{s \,|\, l \in L(s)\}$ indicates the set of sources that use link l. Each source s attains a utility function $U_s(x_s)$, which is a function of the effective rate at the destination of source s. A utility function can reflect the satisfaction of a source with resource allocation. Different shapes of utility functions can lead to different types of fairness among users [24]. Take the following utility function, parameterized by $\alpha > 0$, as an example

$$U_s(x_s) = \begin{cases} (1-\alpha)^{-1}(x_s)^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \log x_s, & \text{otherwise.} \end{cases}$$

If $\alpha = 0$, system throughput maximization is achieved; if $\alpha = 1$, proportional fairness among users is attained; and if $\alpha = 2$, then harmonic mean fairness is achieved. In previous work, the utility function is usually assumed to be strictly concave, strictly increasing, and twice differentiable [17]. Concavity of $U_s(\cdot)$ means that the marginal utility (i.e., the increase in utility obtained for a fixed increase in the resource) decreases as the amount of resource increases. It is reasonable to assume that $U_s(\cdot)$ is increasing since a source always prefers to have more resource, when the amount of all other resources stays the same. In general, all the nodes in the network are assigned the same utility function which aims at providing a uniform measurement of network performance.

The wireless channel is a shared medium and interference-limited. In the physical layer, code-division-multiple-access (CDMA) is used so that each link may simultaneously communicate within the same spectrum allocation, at the expense of multiple-access interference. We have an underlying assumption that nodes can transmit and receive simultaneously. It is indeed possible since simultaneous transceivers have been implemented in [25]. The main notations and their corresponding definitions are summarized in Table 1.

TABLE 1
List of Notations

Notation	Definition
L	Set of logical links
S	Set of sources
K_L	Number of links
K_S	Number of sources
x_s	Data rate of source s
m_s	Minimum rate at which source s transmits data
M_s	Maximum rate at which source s transmits data
$U_s(\cdot)$	Utility function of source s
γ_l	SINR for link l
$\overline{\gamma}_l$	Average SINR for link l
p_l	Transmission power in link l
$p_l^{\min}$	Minimum transmission power
$p_l^{\max}$	Maximum transmission power
P	Vector of transmission powers
c_l	Instantaneous capacity of link l
$P_r(R_l)$	Outage probability at link l with transmission rate R_l
G_{lk}	Path gain with slow-varying
F_{lk}	Path gain with Rayleigh fast-fading
R_l^{avg}	Average rate correctly received over link l
y_s	Average rate correctly received at destination node
$\zeta_l^{\max}$	Maximum value of outage probability on link l
$\phi_l(\boldsymbol{P})$	Probability that the data is correctly received
R_l^{th}	Threshold of data rate R_l
τ_l	Sum of transmission delay and queuing delay on link l
$E(\tau_l)$	Average packet delay on link l
ϑ_l	Specific delay-tolerant amount
H_s	Number of hops for flow s

2.2 Fast-Fading and Average SINR

We let p_l denote the power used by the transmitter of link l, which is subject to a simple power limit $p_l^{\min} \leq p_l \leq p_l^{\max}$, where $p_l^{\min}$ and $p_l^{\max}$ denote the minimum and maximum transmission powers respectively, and γ_l be the signal to interference and noise ratio of link l. The signal power at the receiver of link l is given by $G_{ll}F_{ll}p_l$ and the total interference power at the receiver of link l is given by $\sum_{k \neq l} G_{lk}F_{lk}p_k$. Thus, the SINR for link l is

$$\gamma_l(\boldsymbol{P}) = \frac{G_{ll}F_{ll}p_l}{\sum_{k \neq l} G_{lk}F_{lk}p_k + \sigma^2}, \tag{1}$$

where σ^2 is the thermal noise power at each receiver; $\boldsymbol{P} = (p_1, p_2, \ldots, p_{K_L})^T$ denotes a vector of transmission powers; G_{lk} is the effective positive slow-varying path gain between the transmitter of link k and the receiver of link l. In the analysis below, we assume that it is constant; F_{lk} models Rayleigh fast-fading, which is assumed to be independent exponentially distributed random variables with unit mean [26]. In other words, the received signal power at receiver l from transmitter k is an exponentially distributed random variable with mean $E(G_{lk}F_{lk}p_k) = G_{lk}p_k$.

Furthermore, in Rayleigh fast-fading environment, the instantaneous capacity of each link l is defined on the Shannon capacity as

$$c_l(\boldsymbol{P}) = W \log(1 + \varsigma \cdot \gamma_l(\boldsymbol{P})), \tag{2}$$

where ς is "SINR-gap" that reflects a particular modulation and coding scheme. Without loss of generality, we assume that $\varsigma = 1$; W is the base-band bandwidth, which is normalized by a fixed packet size.

For fast-fading channel, very frequent power updates will induce the fast change of the instantaneous capacity $c_l(\boldsymbol{P})$ according to (2). The high fading rate means that iteration update-rate would need to be increased to keep track of the dynamic fading state, which makes the instantaneous state vary too quickly to track properly. Indeed, we

explicitly take into account the average SINR and the notion of a rate-outage probability, whose definition can be found in Section 2.4 to allow the network to experience a limited amount of fading-induced congestion so as to avoid the fast-update problem.

Therefore, instead of using instantaneous SINR, we will use statistical variation of SINR, i.e., average SINR, in our scheme.

$$\overline{\gamma}_l(\boldsymbol{P}) = \frac{E(G_{ll}F_{ll}p_l)}{E\left(\sum_{k\neq l} G_{lk}F_{lk}p_k\right) + \sigma^2} = \frac{G_{ll}p_l}{\sum_{k\neq l} G_{lk}p_k + \sigma^2}. \quad (3)$$

2.3 Capacity and Data Rate with Outage

It is clear from (2) that the capacity of link depends on channel state information (CSI) such as channel gain between transmitter and receiver. In this paper, similar to [23], we assume that there is receiver CSI and no transmitter CSI at each link, i.e., the receiver knows the value of CSI, and both the transmitter and the receiver know the distribution of CSI, since transmitter CSI results in extra overhead and considerable additional complexity. In this case there are two capacity definitions: Shannon capacity and capacity with outage. It was shown in [19] that the capacity of a fading channel with receiver CSI (i.e., capacity with outage) is less than the Shannon capacity of an additive white Gaussian noise (AWGN) channel with the same average SINR. This indicates that fading and poor channel states can reduce Shannon capacity when only the receiver has CSI, because the transmission strategy must incorporate the effect of these poor states.

The capacity with outage for fading channel with receiver CSI is defined as the maximum rate that can be transmitted over a channel with an outage probability that the transmission cannot be decoded with negligible error probability. The data is correctly received if $c(\boldsymbol{P})$ is greater than or equal to the ingress date rate R. Otherwise, the data received over the transmission burst cannot be decoded correctly with probability approaching 1, and the receiver declares outage. Here, the outage probability is thus given by $P_r(R_l) = P_r(R_l > c_l(\boldsymbol{P}))$ at link l with transmission rate R_l.

The outage probability characterizes the probability of data loss or, equivalently, the probability of deep fading. At a link l with transmission rate R_l, the probability that data is received correctly is $1 - P_r(R_l)$. The rate-effective correctly received over link l is thus

$$R_{l,eff} = R_l \cdot (1 - P_r(R_l)). \quad (4)$$

For a flow s traversing multiple hops, the effective data rate correctly received at the destination node y_s satisfies

$$y_s \leq x_s \prod_{l\in L(s)} (1 - P_r(R_l)), \quad (5)$$

where x_s is the source rate of flow s, and $L(s)$ is the set of link along the route. It can be seen that the data rate of a flow decreases every hop along its route and the data rate correctly received at the destination node decreases as the link or hop count along the route of a flow increases.

2.4 Rate-Outage Probability

As the fading rate increases, more timely feedback of the end-to-end congestion state, which must be feedback to the source to adapt its transmission data rate explicitly through message passing, becomes increasingly difficult. To alleviate this problem, the rate-outage probability should be constrained to a small tolerable value instead of zero-level tolerance on fading-induced congestion.

Similar to [17], the rate-outage probability is defined as the probability of experiencing fading-induced congestion, i.e., the probability that the ingress rate R_l to link l exceeds the instantaneous channel capacity c_l. Therefore, the rate outage constraint can be expressed as

$$P_r(R_l > c_l(\boldsymbol{P})) \leq \zeta_l^{\max}, \forall l \in L, \quad (6)$$

where $\zeta_l^{\max} \in (0, 1)$ is the maximum value of the rate outage probability on each link l. Accordingly, the ζ-outage capacity on link l is defined as the highest rate that the outage probability $P_r(R_l)$ is less than $\zeta_l^{\max}$. It is noted that there is a one-to-one mapping between the outage probability and the corresponding outage capacity. For instance, the ζ-outage capacity on each link l over Rayleigh fading is given by [27]

$$c_l(\boldsymbol{P}, \zeta) = W \log\left[1 + \overline{K}_l \overline{\gamma}_l(\boldsymbol{P})\right], \quad (7)$$

where $\overline{K}_l = -\log(1 - \zeta_l^{\max})$.

For the Rayleigh fading model, we can write the rate-outage probability in closed form by using (3) as [17]

$$\begin{aligned} P_r(R_l > c_l) &= P_r\left(W \log\left[1 + \overline{K}_l \overline{\gamma}_l(\boldsymbol{P})\right] < R_l\right) \\ &= 1 - \phi_l(\boldsymbol{P}), \end{aligned} \quad (8)$$

where

$$\phi_l(\boldsymbol{P}) = \exp\left(-\frac{R_l^{th}}{\overline{\gamma}_l(\boldsymbol{P})}\right) \quad (9)$$

is regarded as the probability that the data is correctly received, and $R_l^{th} = \exp(R_l/W) - 1$ is defined as the data rate threshold.

2.5 Average Delay

We now consider the delay property to be included in constraints and objective function. The average delay a packet experiences when traversing a network is an important design consideration in some real-time applications. The delay includes both queuing and transmission delays. The packet traffic entering the multihop network at the transmitter of link is assumed to be a Poisson distribution and to have an exponentially distributed length with a mean of K bits. In this study, we model each link as an M/M/1 queue, which is a good approximation for the behavior of individual links in densely connected networks satisfying the above assumptions [28]. Under the feed-forward routing and Poisson input assumptions, by applying Burke's theorem [28], the average packet delay on link l is thus given by

$$E(\tau_l) = \frac{K}{c_l(\boldsymbol{P}, \zeta) - \sum_{s\in S(l)} x_s},$$

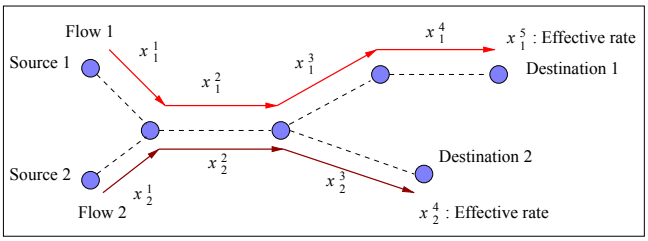

Fig. 1. An example network with lossy links.

where τ_l is the sum of the transmission delay and queuing delay on link l. For real-time applications, generally the average delay $E(\tau_l)$ is required to be no more than a specific delay-tolerant amount $\vartheta_l \geq 0$, i.e., $E(\tau_l) \leq \vartheta_l$. Equivalently, $\sum_{s\in S(l)} x_s \leq c_l(\boldsymbol{P}, \zeta) - \frac{K}{\vartheta_l}$.

3 RENUM with Rate Outage Constraint

In this section, a novel rate-effective network utility maximization framework is proposed by taking into account the lossy nature of links and the rate outage constraints with a small tolerate value. At the ith link of flow s, the data rate is denoted by $x_s^{l(s,i)}$ where $l(s,i)$ denotes the unique link number of the ith link of flow s. For clearer presentation, notation x_s^i will substitute for $x_s^{l(s,i)}$. At the destination of flow s, the correctly received data rate can be represented as $x_s^{H_s+1}$, where $H_s = |L(s)|$ denotes the number of hops for flow s. It can be considered as the data rate over a link which connects the receiver and itself [23]. For example, in Fig. 1, the effective rate of flow 1 is given by x_1^5. Thus similar to [23], the utility function of flow s is associated with the effective rate $x_s^{H_s+1}$.

There are two main objectives to be achieved in this problem formulation, i.e., to maximize the overall rate-effective network utility across all flows and minimize the overall power and average delay of all links, which is necessary to reduce interference and prolong the life of nodes powered by batteries. The first objective can be expressed mathematically by

$$\max_{x} \sum_{s\in S} U_s\left(x_s^{H_s+1}\right)$$

and the second objective can be characterized by

$$\min_{\vartheta, p\geq 0} w_1 \sum_{l\in L} p_l + w_2 \sum_{l\in L} \vartheta_l.$$

Based on the fact that $\min f(\cdot)$ is equivalent to $\max - f(\cdot)$, by combining the above two objectives and considering the constraints in Section 2, the RENUM problem can be formulated as

$$\max_{x,\vartheta,p\geq 0} \sum_{s\in S} U_s\left(x_s^{H_s+1}\right) - w_1 \sum_{l\in L} p_l - w_2 \sum_{l\in L} \vartheta_l \qquad (10)$$

Subject to

$$x_s^{i+1} \leq x_s^i(1 - P_r(l(s,i))), i = 1, 2, \ldots, H_s, \forall s, \qquad (11)$$

$$P_r\left(\sum_{s\in S(l)} x_s^l > c_l(\boldsymbol{P})\right) \leq \zeta_l^{\max}, \forall l, \qquad (12)$$

$$\sum_{s\in S(l)} x_s \leq c_l(\boldsymbol{P},\zeta) - \frac{K}{\vartheta_l}, \forall l, \qquad (13)$$

$$m_s \leq x_s \leq M_s, \forall s, \qquad (14)$$

$$p_l^{\min} \leq p_l \leq p_l^{\max}, \forall l, \qquad (15)$$

where $P_r(l(s,i))$ denotes the outage probability of the ith link of flow s, and w_1 and w_2 are fixed weights to trade these two conflicting objectives. The constraints can be explained as follows:

- *Flow conservation constraint.* Constraint (11) shows that for each flow, the effective data rate received at the receiver is no more than the injection rate at the transmitter for route links by considering the lossy link property.
- *Rate-outage probability constraint.* Constraint (12) describes that at each link, there exists an upper bound (a small tolerable value) on the rate-outage probability.
- *Average delay constraint.* Constraint (13) ensures that the average delay on link l is guaranteed not to exceed a specific delay-tolerant amount ϑ_l.
- *Rate fairness constraint.* Constraint (14) demonstrates that the transmitting rate of each source is bounded by the maximum and minimum transmission rates, M_s and m_s, in order to keep the fairness and improve the efficiency of wireless links.
- *Energy constraint.* Constraint (15) describes that the transmission power at each transmitter on link l is constrained by the maximum and minimum transmission powers, $p_l^{\max}$ and $p_l^{\min}$, in order to save the finite energy of each transmitter and satisfy the diverse quality of service requirement, respectively.

Note that the outage parameter $\zeta_l^{\max}$ reflects the trade-off between rate and outage. With the increase of $\zeta_l^{\max}$, higher data rate can be transmitted, but higher data loss may occur.

Remark 1. Compared with other NUM problems, the RENUM framework has two unique features: (i) the network utility is based on the achieved rates at the destinations instead of the injected rates at the sources; (ii) the effect of rate-outage probability and average delay on the optimal rate and power allocation is considered accurately.

For wireless links experiencing Rayleigh fading, the inequality constraint (11) becomes

$$x_s^{i+1} = x_s^i \phi_{l(s,i)}(\boldsymbol{P}), i = 1, 2, \ldots, H_s, \forall s. \qquad (16)$$

In general, the RENUM problem (10) is a non-convex problem due to the utility function associated with the effective rate $x_s^{H_s+1}$ and the existence of constraint (11). Therefore, the problem needs to be converted into a convex one under some regularity conditions by employing some auxiliary variables and appropriate transformation. Define $\tilde{x}_s^l = \log(x_s^l)$, $\tilde{p}_l = \log(p_l)$, $\tilde{\vartheta}_l = \log(\vartheta_l)$, $\tilde{\Lambda}_s = [\log(m_s),$

$\log(M_s)]$, $\tilde{I}_l = \left[\log(p_l^{\min}), \log(p_l^{\max})\right]$, then the optimization problem (10) can be transformed as follows:

$$\max_{\tilde{x}_s, \tilde{\vartheta}_l, \tilde{p}_l} \sum_{s\in S} U_s\left(\exp\left(\tilde{x}_s^{H_s+1}\right)\right) - w_1 \sum_{l\in L} \exp(\tilde{p}_l) - w_2 \sum_{l\in L} \exp(\tilde{\vartheta}_l) \tag{17}$$

Subject to

$$\tilde{x}_s^{i+1} \le \tilde{x}_s^i - \psi^{l(s,i)}(\tilde{\boldsymbol{P}}), i = 1, \ldots, H_s, \forall s, \tag{18}$$

$$\psi^l(\tilde{\boldsymbol{P}}) \le \overline{K}_l, \forall l, \tag{19}$$

$$\sum_{s\in S(l)} \exp\left(\tilde{x}_s^l\right) \le c_l(\exp(\tilde{\boldsymbol{P}}), \zeta) - \frac{K}{\exp(\tilde{\vartheta}_l)}, \forall l, \tag{20}$$

where $\tilde{x}_s \in \tilde{\Lambda}$, $\tilde{p}_l \in \tilde{I}_l$ and $\psi^{l(s,i)}(\tilde{\boldsymbol{P}}) = -\log(\phi_{l(s,i)}(\tilde{\boldsymbol{P}}))$.

Next, we will explore the convexity of the RENUM problem (17)-(20) in preparation for the development of an algorithm attaining a global solution. At first, we will show the convexity of the constraint set in (18)-(20).

Theorem 1. *The constraint set comprised by (18)-(20) is convex.*

Proof. The source rates x_s are transformed logarithmically to ensure that the left-hand side of the constraints is convex as sum-exp is convex. Since $\psi^{l(s,i)}(\tilde{\boldsymbol{P}})$ is twice differential with respect to $\tilde{\boldsymbol{P}}$ and its Hessian matrix is positive semidefinite, the right-hand side of constraint (18) is concave and the left-hand side of constraint (19) is also convex. It is easy to verify that the Hessian matrix of the right-hand side of constraint (20) is negative semidefinite in $\tilde{\vartheta}_l$ and $\tilde{p}_l$, which means that it is concave. We therefore conclude that the constraint set in (18)-(20) is convex. □

The next issue to be addressed is to show the convexity of the objective of the transformed problem (17).

Theorem 2. *The objective of the transformed problem (17) is concave over $\tilde{x}_s^{H_s+1}$, $\tilde{p}_l$ and $\tilde{\vartheta}_l$.*

Proof. Since the objective is separable in each variable, it is clearly concave in $\tilde{p}_l$ and $\tilde{\vartheta}_l$ (sum of negative exponentials). The concavity in $\tilde{x}_s^{H_s+1}$ follows from the concavity $U_s(\exp(\tilde{x}_s^{H_s+1}))$ of in $\tilde{x}_s^{H_s+1}$. Since the utility function $U_s(x_s)$ is twice-differentiable, increasing and strictly-concave in x_s, it is also (log, x_s)–concave. We further have that the objective function in (17) is concave about $\tilde{x}_s^{H_s+1}$. Problem (17) is thus a convex problem with a unique global optimal point. □

It is not difficult to see that the Slater condition [29] is satisfied, i.e., there exist feasible solutions of $\tilde{x}_s^{H_s+1}$, $\tilde{p}_l$ and $\tilde{\vartheta}_l$ such that the constraints hold with strict inequality, and strong duality holds, which implies that the optimal values of the primal and dual problems are equal. Thus, the optimization problem (17) can be efficiently tackled by using modern convex programming schemes.

4 Distributed Algorithm for RENUM Problem

In this section, we use Lagrange-duality method for solving RENUM problem to develop a distributed algorithm. The first step for the Lagrange-duality method is to introduce Lagrangian Multipliers associated with some constraints of the original problem. For RENUM problem that has multiple constraints, there are also various ways to introduce Lagrangian Multipliers that might result in different dual problems. For the following proposed solution, Lagrangian Multipliers are chosen aiming to facilitate implementing it in the real time. We refers to the following development as *Algorithm A*.

As a first step, the Lagrangian multipliers λ_s^i, μ_l, υ_l, $i = 1, 2, \ldots, H_s, s \in S, l \in L$ are introduced for users and links with respect to constraints (18), (19) and (20), respectively. Here, λ_s^i, μ_l and υ_l can be interpreted as flow conservation price for the ith lossy link of flow s, i.e., congestion price, the price for keeping the rate-outage probability at each link l within a small tolerable value, and the price for the average delay no more than a specific delay-tolerant amount, i.e., queueing delay price, respectively. The objective function of the Lagrangian dual problem of (17) can be then expressed as

$$\begin{aligned} D(\lambda, \mu, \upsilon) &= \max_{\{\tilde{x}\in\tilde{\Lambda}, \tilde{\vartheta}, \tilde{p}\in\tilde{I}\}} L(\tilde{x}, \tilde{\vartheta}, \tilde{p}, \lambda, \mu, \upsilon) \\ &= \max_{\{\tilde{x}\in\tilde{\Lambda}, \tilde{\vartheta}, \tilde{p}\in\tilde{I}\}} L_{\tilde{x}}(\tilde{x}, \lambda, \mu, \upsilon) + L_{\tilde{p}}(\tilde{p}, \lambda, \mu, \upsilon) \\ &\quad + L_{\tilde{\vartheta}}(\tilde{\vartheta}, \upsilon), \end{aligned} \tag{21}$$

where $L_{\tilde{x}}(\tilde{x}, \lambda, \mu, \upsilon)$ is the Lagrangian function with respect to source rate $\tilde{x}$ and Lagrangian multipliers λ, μ and υ; $L_{\tilde{p}}(\tilde{p}, \lambda, \mu, \upsilon)$ represents the Lagrangian function with respect to transmission power $\tilde{p}$ and Lagrangian multipliers λ, μ and υ; $L_{\tilde{\vartheta}}(\tilde{\vartheta}, \upsilon)$ denotes the Lagrangian function with respect to transmission delay $\tilde{\vartheta}$ and Lagrangian multiplier υ. The expressions of these Lagrangian functions can be obtained by multiplying the Lagrangian multipliers by the corresponding constraints (18), (19), (20) and combining the similar terms.

By the linearity of the differentiation operator, the objective (21) can be decomposed into three separate maximization subproblems due to the separable nature of the function

$$P1: D_{\tilde{x}}(\lambda, \mu, \upsilon) = \max_{\tilde{x}\in\tilde{\Lambda}} L_{\tilde{x}}(\tilde{x}, \lambda, \mu, \upsilon),$$

$$P2: D_{\tilde{p}}(\lambda, \mu, \upsilon) = \max_{\tilde{p}\in\tilde{I}} L_{\tilde{p}}(\tilde{p}, \lambda, \mu, \upsilon),$$

$$P3: D_{\tilde{\vartheta}}(\lambda, \mu, \upsilon) = \max_{\tilde{\vartheta}} L_{\tilde{\vartheta}}(\tilde{\vartheta}, \upsilon),$$

$s.t.\ \tilde{\vartheta} \in \mathbb{R}$, where $\mathbb{R}$ denotes the real number.

Subproblem $P1$ is the rate control subproblem in terms of rate variable $\tilde{x}$, $P2$ represents the power control subproblem to achieve the minimum power consumption, and $P3$ denotes the average delay control subproblem to give an optimal delay-tolerant amount $\tilde{\vartheta}$. Note that indices s, i and l are dropped in $P1$, $P2$ and $P3$ since they are applicable for all s, i and l. The maximization involved in the calculation of the dual function has been decomposed into three simple

maximization subproblems, which implies that the optimal resource allocation problems can be separated and solved in a distributed manner.

Accordingly, the dual problem of the original (primal) problem is given by

$$\min_{\lambda,\mu,\upsilon} D(\lambda,\mu,\upsilon). \tag{22}$$

Because the primal problem is convex and also satisfies the Slater's condition [29], the duality gap between the optimal value of the primal problem and that of the dual problem becomes zero. This suggests that we can design primal-dual algorithms to find the optimal powers, rates, delay-tolerant amount and prices. More specifically, they can be updated iteratively in such a way that the Lagrangian function $L(\tilde{x},\tilde{\vartheta},\tilde{p},\lambda,\mu,\upsilon)$ is maximized with respect to $\tilde{x}$, $\tilde{p}$ and $\tilde{\vartheta}$ for the primal, and the dual function $D(\lambda,\mu,\upsilon)$ is minimized with respect to λ, μ, and υ for the dual.

Next, we will take an iterative approach to solving the dual problem (22): at iteration t, the maximization problems $P1$, $P2$ and $P3$ are solved for fixed λ, μ, and υ, then a subgradient method is employed to update the dual variables. Subsequent iterations $t+1$, $t+2$, ... would repeat this process until convergence to the globally optimum resource allocation.

4.1 Source Rate Subalgorithm

The source rates are updated by solving the maximization problem $P1$. Since problem $P1$ is strictly concave for fixed λ, μ, and υ, we can use a subgradient projection with a sufficiently small fixed step-size to find a maximizer. For convenience, we define $\lambda_s^0 = 0$, $\forall s$. Taking partial derivative of $L_{\tilde{x}}(\tilde{x},\lambda,\mu,\upsilon)$ with respect to $\tilde{x}_s^i$, we have the subgradient on the data rate $\tilde{x}_s^i$ as follows:

for $i = 1, 2, \ldots, H_s$,

$$\nabla D_1\big(\tilde{x}_s^i\big) = \lambda_s^i - \lambda_s^{i-1} - \upsilon_{l(s,i)}\exp\big(\tilde{x}_s^i\big), \tag{23}$$

and for $i = H_s + 1$,

$$\nabla D_1(\tilde{x}_s^i) = U_s'\big(\exp\big(\tilde{x}_s^{H_s+1}\big)\big)\exp\big(\tilde{x}_s^{H_s+1}\big) - \lambda_s^{H_s}, \tag{24}$$

where $U_s'(\cdot)$ is the first derivative of the utility function.

Due to the strict concavity of the Lagrangian function $L_{\tilde{x}}(\tilde{x},\lambda,\mu,\upsilon)$ by Theorem 1 and Theorem 2, the dual function $D(\lambda,\mu,\upsilon)$ is differential everywhere by [30], Prop. 6.1.1]. Taking partial derivative of with respect to λ_s^i, the subgradient on the price λ_s^i is given by

$$\nabla D\big(\lambda_s^i\big) = \tilde{x}_s^i - \tilde{x}_s^{i-1} - \psi^{l(s,i)}(\tilde{\boldsymbol{P}}). \tag{25}$$

Based on (24) and (25) and by utilizing the complementary slackness condition, we can obtain the destination of flow s, the flow conservation price and the effective rate as follows:

$$\lambda_s^{H_s} = U_s'\big(\exp\big(\tilde{x}_s^{H_s-1}\big)\big)\exp\big(\tilde{x}_s^{H_s-1}\big), \tag{26}$$

$$\tilde{x}_s^{H_s+1} = \tilde{x}_s^{H_s} - \psi^{l(s,H_s)}(\tilde{\boldsymbol{P}}). \tag{27}$$

Remark 2. $\psi^{l(s,i)}(\tilde{\boldsymbol{P}})$ is referred to as the data loss rate at the ith link of flow s. Since $\psi^{l(s,i)}(\tilde{\boldsymbol{P}})$ is a decreasing function on $\tilde{\boldsymbol{P}}$, which means that the larger the power is, the smaller the loss rate. We can observe from (26) and (27) that at the destination of flow s, the effective rate becomes larger as the power increases. Since $\lambda_s^{H_s}$ is a non-increasing function of the effective rate $\tilde{x}_s^{H_s+1}$, it can be viewed as the effective flow weight, which guarantees the fairness among destination nodes.

Next, we use the subgradient projection method to find the optimal solution for $\tilde{x}_s^i$ and λ_s^i. We have the following theorem.

Theorem 3. *At the ith link of flow s, the flow rate $\tilde{x}_s^i$ can be updated as follows:*

$$\tilde{x}_s^i(t+1) = \big[\tilde{x}_s^i(t) + \varepsilon\nabla D_1(\tilde{x}_s^i)\big]_{\tilde{\Lambda}_s}, i = 1, 2, \ldots, H_s, \tag{28}$$

where ε is a sufficiently small fixed step-size.

It is noted that the rate update depends on the link congestion and the QoS price induced by tolerant delay and the difference of available buffer spaces between adjacent nodes $\lambda_s^i - \lambda_s^{i-1}$.

At the receiver of the ith link of flow s, by using the subgradient projection method to solve the dual problem (22), the price λ_s^i can be updated as follows:

$$\lambda_s^i(t+1) = \big[\lambda_s^i(t) - \varepsilon\nabla D\big(\lambda_s^i\big)\big]^+, i = 1, 2, \ldots, H_s. \tag{29}$$

Equation (29) reveals that λ_s^i is intimately related to the available buffer space at the receiver of the link, which determines in turn the difference between the incoming rate $\tilde{x}_s^i - \psi^{l(s,i)}(\tilde{\boldsymbol{P}})$ and the outgoing rate $\tilde{x}_s^{i+1}$. For instance, the more the available buffer space the receiver has, the smaller the difference between the incoming rate and the outgoing one, which results in an increasing price λ_s^i. In fact, this is a back-pressure algorithm where the flow with greater queue length difference is granted higher priority. If each node is assumed to have the same total buffer size, then $\lambda_s^i - \lambda_s^{i-1}$ is exactly the queue length between the transmitter and receiver at the ith link of flow s. The detailed rate control subalgorithm for subproblem $P1$ is summarized in Table 2. It is clear that the complexity of the subalgorithm is $O(H_s^2)$.

4.2 Link Power and Delay Subalgorithm

Since the twice order derivatives of $L_{\tilde{p}}(\tilde{p},\lambda,\mu,\upsilon)$ and $L_{\tilde{\vartheta}}(\tilde{\vartheta},\upsilon)$ with regard to $\tilde{p}$ and $\tilde{\vartheta}$ respectively are negative semidefinite, i.e., they are concave, there exist optimal solutions of $\tilde{p}$ and $\tilde{\vartheta}$. Taking partial derivative of $L_{\tilde{p}}(\tilde{p},\lambda,\mu,\upsilon)$ with respect to $\tilde{p}_l$, and coming back to the solution space $\boldsymbol{P}$ instead of $\tilde{\boldsymbol{P}}$, it is easy to verify that the subgradient on the transmission power p_l is

$$\nabla D_2(p_l) = \Omega_l(\boldsymbol{P})G_{ll} + \sum_{n\neq l} G_{nl}M_n(\boldsymbol{P}) - w_1, \tag{30}$$

where

$$\Omega_l(\boldsymbol{P}) = \frac{\upsilon_l W\overline{K}_l}{1+\overline{K}_l\overline{\gamma}_l(\boldsymbol{P})} - (\lambda_s^l + \mu_l)\frac{R_l^{th}}{(\overline{\gamma}_l(\boldsymbol{P}))^2}, \tag{31}$$

TABLE 2
Distributed Rate Control Subalgorithm for Flow s

Initialize $\lambda_s^0 = 0$ and $\lambda_s^i(0)$ for $i = 1, 2, \ldots, H_s$;
For $i = 1 : H_s$
  Repeat
    Receive $\upsilon_{l(s,i)}$ from the Lagrangian update subalgorithm;
    Measure and compute $\psi^{l(s,i)}$ by (9);
    Compute the subgradient $\nabla D_1(\widetilde{x}_s^i)$ by (23);
    Compute $\widetilde{x}_s^i(t)$ by (28);
    Compute the subgradient $\nabla D(\lambda_s^i)$ by (25);
    Update Lagrangian multiplier $\lambda_s^i(t+1)$ by (29);
  Until $\{\lambda_s^i(t)\}$ converges to λ_s^{i*}
End for
If $i = H_s + 1$ then
  Compute $\lambda_s^{H_s}$ by (26);
  Compute $\widetilde{x}_s^{H_s+1}$ by (27);
End if

TABLE 3
Link Power and Delay Subalgorithm for Flow s

For each link $l \in L(s)$;
  Initialize $\zeta_l^{\max}$, $\mu_l(0)$ and $\upsilon_l(0)$;
  Repeat
    Receive λ_s^l from the rate control subalgorithm in Table 2:
    Measure average SINR $\overline{\gamma}_l$;
    Measure R_l and compute R_l^{th};
    Compute $\Omega_l(\boldsymbol{t})$ by (31);
    Compute message $M_n(t)$ by (32) for all $n \neq l$
    Broadcast the message to its neighbors;
    Compute power $p_l(t)$ by (33);
    Compute delay $\vartheta_l(t)$ by (34);
    Update Lagrangian multiplier $\mu_l(t+1)$ by (35);
    Update Lagrangian multiplier $\upsilon_l(t+1)$ by (36);
  Until $\{\mu_l(t)\}$ and $\{\upsilon_l(t)\}$ converges to μ_l^* and υ_l^*, respectively
End for

$$M_n(\boldsymbol{P}) = \frac{\Omega_n(\boldsymbol{P})(\overline{\gamma}_n(\boldsymbol{P}))^2}{G_{nn}p_n}. \tag{32}$$

Therefore, we have the following theorem.

Theorem 4. *The transmission power is updated for each link $l \in L(s)$ with message passing by*

$$\begin{aligned} p_l(t+1) &= [p_l(t) + \varepsilon\nabla D_2(p_l)]_{p_l^{\min}}^{p_l^{\max}} \\ &= \left[p_l(t) + \varepsilon G_{ll}\Omega_l(t) - \varepsilon w_1 + \varepsilon\sum_{n\neq l} G_{nl}M_n(t)\right]_{p_l^{\min}}^{p_l^{\max}}, \end{aligned} \tag{33}$$

where each $M_n(t)$ is referred to as the messages comprising local information at each link n, which is a scaled measurement of the receiver SINR. A distributed power-update algorithm can be implemented through message passing: each receiver on link n broadcasts its message $M_n(t)$ and estimates $M_n(t)$ through training sequence.

Similarly, the link average delay is obtained by solving $L_{\tilde{\vartheta}}(\tilde{\vartheta}, \upsilon)$ and updated by

$$\vartheta_l(t+1) = (\upsilon_l(t)K/w_2)^{1/2}. \tag{34}$$

The link power and delay subalgorithm for subproblems $P2$ and $P3$ is summarized in Table 3. It is clear that the complexity of the subalgorithm is $O(K_L^2)$.

4.3 Lagrangian Multiplier Update Subalgorithm

The updates in (33) and (34) are shown after transformation back to the original co-ordinate space. Furthermore, by using the subgradient projection, prices μ_l and υ_l are updated by

$$\mu_l(t+1) = \left[\mu_l(t) - \varepsilon\left(\overline{K}_l - \psi^l(\tilde{\boldsymbol{P}})\right)\right]^+, \tag{35}$$

and

$$\upsilon_l(t+1) = [\upsilon_l(t) - \varepsilon\nabla D(\upsilon_l)]^+, \tag{36}$$

respectively, where

$$\nabla D(\upsilon_l) = -\sum_{s\in S(l)} \exp(\tilde{x}_s^l) + c_l(\exp(\tilde{\boldsymbol{P}}),\zeta) - \frac{K}{\exp(\tilde{\vartheta}_l)}.$$

It is worth noting that in principle, mobile ad hoc networks are independent of any established infrastructure or centralized mechanism. Each node operates in a distributed peer-to-peer mode, acting as an independent router and an end host. It can move independently in any direction, thus change its links to other devices frequently. In such an infrastructure-less network, there does not exist any central coordination or control. Thus, centralized solutions are impractical. The proposed RENUM algorithm has to be implemented in a distributed manner in the absence of central coordination, through the cooperation of nodes. To elaborate, on one hand, each node must give the feedback such as message M_n, based on the local information such as SINR $\overline{\gamma}_l$ and prices λ_s^i, μ_l and υ_l of other links in the interference set, rather than the individual link, and broadcast the message to its neighbor nodes. On the other hand, the distributed nature of the RENUM algorithm can cope with the absence of the central coordination in ad hoc networks: source s can solve Equation (28) separately without need to coordinate with other sources because λ_s^i summarizes all the congestion information source s needs to know; Equations (33) and (34) can be implemented by an individual link l using only the accumulated information such as message M_n from its interference set and its own information such as ingress rate R_l, average SINR $\overline{\gamma}_l$ and prices λ_s^l, μ_l and υ_l of link l.

The complexity of the algorithm can be derived as follows. Assuming that the ellipsoid method is used to iteratively update dual variables, the required number of iterations for convergence is $O(m^2)$, where m denotes the size of the problem. Let K_L and K_S denote the number of links and sources, respectively. The ellipsoid method will need $O(K_S^2K_L^2)$ iterations to obtain $\{\lambda_s^{i*}\}$ and $\{\tilde{x}_s^{i*}\}$ and $O(K_L^2)$ iterations to obtain $\{\mu_l^*\}$, $\{\upsilon_l^*\}$, $\{p_l^*\}$ and $\{\vartheta_l^*\}$. Hence, the total complexity of the algorithm is $O(K_S^2K_L^4)$.

4.4 Relationship between RENUM Algorithm and TCP

In this section, we discuss the relationship between the proposed RENUM algorithm and transmission control protocol (TCP).

Our proposed source rate control subalgorithm is closely related to the congestion control in TCP. It is well known that a TCP sender maintains two windows, the congestion window and the TCP maximum window, for congestion

control. The congestion window is the flow control imposed by the sender, while the TCP maximum window, i.e., the window advertised by the receiver, is the flow control imposed by the receiver. A source/sender s keeps a variable called congestion window size W_s that determines the number of packets it can transmit in a round-trip time (RTT) τ_i over link i, the time from sending a packet to receiving its acknowledgment from the destination. This implies that the source rate x_s^i in RENUM is approximately equal to the ratio of window size to RTT, i.e., $x_s^i = W_s/\tau_i$, in packets/second. Compared to the wired networks, for multihop MANETs, the traditional TCP cannot distinguish effectively between congestion loss induced by buffer overflow and packet loss caused by mobility and shared channel contention among wireless nodes. This is because that buffer overflow-induced packet losses are rare and packet drops due to wireless link contention dominate.

To make the congestion control mechanism of TCP adapt to multihop MANETs, we employ a cross-layer method, which adaptively adjusts the TCP maximum window size according to the number of RTS (request to send) retry counts of the IEEE 802.11 DCF (distributed coordination function) protocol of the MAC layer at the TCP sender, to control the number of TCP packets in the network and thus reduce the channel contention. In our method, not only the TCP receiver but also the TCP sender participate in adapting the maximum window size. More specifically, the receiver still determines the maximum window size according to its buffer space, and the sender also adjusts the maximum window size according to the average RTS retry counts of its MAC layer.

Equation (28) in Theorem 3 reflects the TCP congestion window update while Equation (29) describes the link congestion price update. The source rate control subalgorithm shows that if the data loss rate $\psi^{l(s,i)}(\tilde{\boldsymbol{P}})$ at the ith link of flow s becomes higher, which implies that congestion occurs at the ith link of flow s, then congestion price λ_s^i will increase. In this case, source s decreases its window size W_s and transmission rate. Otherwise, if the demand for the weighted incoming flow rate $\upsilon_{l(s,i)} \exp(\tilde{x}_s^i)$ at the ith link of flow s is less than the available buffer spaces between adjacent nodes of flow s or the allowed maximum queue length between the transmitter and receiver at the ith link of flow s, $\lambda_s^i - \lambda_s^{i-1}$, then source s increases its window size and transmission rate. Herein, $\lambda_s^i - \lambda_s^{i-1}$ can be regarded as the TCP receiver's advertised window, i.e., the TCP maximum window. Thus, the proper reduction of the advertised window by intermediate transmitters/receivers can control the number of packets sent from a TCP source.

To achieve the adaptive adjustment for the TCP maximum window size $\lambda_s^i - \lambda_s^{i-1}$ according to the number of RTS retry counts, in the RTS/CTS (request to send/clear to send) scheme of the IEEE 802.11 MAC protocol, each node has a variable, $ssrc$ (station short retry count), to record the retry count of the current RTS frame. Each node determines its current channel access according to the value of $ssrc$. To make the node know its channel state, we introduce a new variable, rrc, to record the retry count of the last RTS frame. The bigger the rrc is, the more RTS collisions the node experiences and thus the worse the channel state is. According to the value of rrc from the MAC layer, the TCP sender can know its local channel contention state and estimate the channel contention state of the forwarding nodes. Therefore, the TCP sender can reduce shared channel collisions in the network by dynamically adjusting the TCP maximum window size $\lambda_s^i - \lambda_s^{i-1}$ based on its local RTS retry counts. Such a method improves the TCP performance by controlling the packets injected by TCP senders and thus avoids the congestion caused by packet losses due to low link quality, which is owing to that we tackle TCP degradation in the wireless domain through the cooperation between the transport layer and the MAC layer.

In addition, in RENUM algorithm, some information is required for the subalgorithm updates such as the local information of SINR $\overline{\gamma}_l(\boldsymbol{P})$, prices $\upsilon_l, \lambda_s^l, \mu_l$, ingress rate R_l and message M_n from the neighbor set. The local information can be computed by each node itself. Message M_n can be included in a distributed protocol by making use of explicit message passing. To achieve low communication overhead, we employ a type of piggybacking similar to the method in [31], i.e., the information of M_n is piggybacked onto acknowledgment (ACK) packets. A field is reserved in an ACK packet header that is sent from the receiver back to the transmitter. Along the path, each intermediate link accumulates the information of message M_n into this field. When the ACK reaches the transmitter, this reserved field forms the required information for those updates. This way, the network can also provide explicit feedback information to the TCP sources to dynamically adjust their window size when congestion is detected.

To reach an inference and update for data rate and transmission power, each node maintains a local broadcast message pool. The local broadcast message pool includes n entities (where n is the number of downstream links of the node). Each entity includes the following fields: (1) ID of sender and ID of receiver, which are used to uniquely identify the link, (2) sequence no., which is used to judge the freshness of the entity, (3) message type, which is set as "1" to identify that the information is received from other nodes, (4) fields to record message M_n. At the beginning, the sender of the link generates an entity of the broadcast message packet for a link. In this entity, the ID of sender, ID of receiver and message M_n are filled in the corresponding field. When a data packet is received successfully, the ACK packet piggybacking a broadcast message is transmitted. The broadcast message packet contains all the entities in the local broadcast message pool that have been broadcast. When a broadcast message packet is received, each entity is handled as follows: According to ID of sender and ID of receiver, if an entity did not exist earlier, this entity is added in the local broadcast message pool. If an entity already exists, and the newer entity has a larger sequence number, the newer entity is also added in the pool, and the older entity is discarded. If the newer entity has a smaller sequence number, this newer entity is discarded.

4.5 Packet Routing and Flow Contention

In this section, we extend our rate control subalgorithm to handle the route outages by jointly considering the packet

routing at the network layer and the flow contention at the MAC layer.

The routing of data packets is determined by congestion price λ_s^i and flow distribution over the links they traverse. Intuitively, a node prefers to forward the flow to a neighbor with lower congestion price or on the route with a lower flow rate. Such a scheme guarantees that each flow travels through links in a decreasing order of congestion prices. This indicates that the packet routing is in accordance with the rate control since we can observe from (25) and (28) that the increase of data loss rate $\psi^{l(s,i)}(\tilde{P})$ would lead to the increase of congestion price λ_s^i, and in turn the decrease of flow rate $\tilde{x}_s^i$. Accordingly, the routing of data packets will inevitably make the adjustment based on the variations of congestion price and flow rate. Therefore, the routing implementation requires a supplementary mechanism which coordinates the necessary message exchange. Node i needs to collect local measures such as average SINR $\overline{\gamma}_l$ and message M_n as well as reports of congestion prices λ_s^j from its neighbor j to which it forwards flow s. Moreover, node i is responsible for calculating its own congestion price by (28), and then providing λ_s^i to its neighbors from which it receives traffic of flow s.

Remark 3. It is known that the ad hoc on-demand distance vector (AODV) routing protocol uses the messages of route requests (RREQs), route replies (RREPs) and route errors (RERRs) to discover and maintain routes. As long as the endpoints of a communication connection have valid routes to each other, AODV does not play any role. Therefore, there exists a flow between the source and the destination for the AODV routing protocol as long as the communication links connecting them are valid. Clearly, our proposed RENUM algorithm is effective for the case that AODV keeps the discovered routes. As for the case when AODV needs to discover a new route, the RENUM algorithm will employ the messages of RREQs, RREPs and RERRs to show that a new route is found or a link break has occurred, and accordingly update the set of links $L(s)$ connecting source s and its destination. During discovery, the broken link hinders the data transmission, which in turn causes the increase of congestion price λ_s^i. As a result, source s will have to choose the set of links with lower congestion prices to transmit data. This adaptive routing shows the proposed hop-by-one source rate subalgorithm can also work well for the case of discovering a new route.

In the meanwhile, both the interflow contention and the intraflow contention at the MAC layer could cause the route outage of data transmission. The interflow contention is referred to as the contention that each multiple flow encounters from other flows which pass by the neighborhood. The intraflow contention is the MAC layer contention for the shared channel among the nodes of the same flow, which are in each other's interference range. These two types of flow contentions could result in severe collisions and congestion. A good solution to the flow and congestion control problem in ad hoc networks must consider the MAC layer characteristics and respond quickly to the congestion. The essence of addressing the intraflow contention problem is to prevent the first few nodes on the path from injecting more packets than that next nodes can forward. Therefore, in this paper, one way to prevent the intraflow contention is to assign high channel access priority to each node when it only receives a packet, and those packets which have traversed more hops. That is to say, the source node tries to hold the current packets until the preceding packet is transmitted out of its interference range. This mechanism only considers the interference in a single flow. If the next hop of the current receiver is busy or interfered by other transmission, the receiver cannot seize the channel even with the highest priority.

Hence, we introduce the backward-pressure scheduling mechanism to deal with the interflow contention. Its basic idea is to keep nodes from transmitting to their already congested downstream nodes, hence, yield the channel access to the congested nodes to clear congestion as well as to avoid severe medium contention. The mechanism includes a transmission blocking procedure and a transmission resuming procedure. It requires that each node monitors the number of individual flow in the shared outgoing queue. Let n_s denote the number of packets of flow s. If n_s reaches a *backward-pressure threshold*, the transmission of flow s from its upstream node will be blocked. When the node successfully forwards some packets to its downstream node so that n_s is less than the backward-pressure threshold, it initiates the transmission resuming procedure to allow the upstream node to transmit packets of flow s. The transmission blocking procedure takes advantage of the RTS/CTS exchange in the IEEE 802.11 MAC protocol to restrict the transmission from the upstream nodes. The transmission resuming procedure adopts the receiver-initiated transmission and uses the three-way handshake CTS/DATA/ACK instead of the normal four-way handshake RTS/CTS/DATA/ACK, because the downstream node already knows the restricted upstream node has packets designated to it.

5 Disposing of Power Allocation Overhead

In pursuit of the globally optimum solution to the RENUM problem (10)-(13), explicit message passing between links is required for power allocation (33). While broadcast is a viable way to realize such inter-link communication, the ideal scheme is to use some type of indirect measurement to achieve a similar goal.

In this section, we provide an algorithm with the following property: each link's power-allocation is based on the locally measured interference level caused by other links. No explicit message passing is required. These messages are embedded in a header field within ACK packets that traverse the source's reverse-path and incur little overhead. The algorithm is referred to as *Algorithm B*.

5.1 Link Power Subalgorithm

The following development concentrates on the fact that at the optimum solution, the ingress rate to a link is matched to its capacity (all constraints are active). Considering a particular link l, we can rearrange the constraint in terms of the link transmission power, resulting in the following power update

$$p_l(t+1) = \frac{p_l(t)}{\overline{\gamma}_l(\boldsymbol{P}(t))}\varphi_l(x_s, \vartheta_l), \forall l, \tag{37}$$

where $\varphi_l(x_s, \vartheta_l) = \exp(\frac{1}{W}(\sum_{s\in S(l)} x_s + \frac{K}{\vartheta_l})) - 1$.

It is observed from (37) that $\overline{\gamma}_l$ is a scaled noise measurement, while the term $\varphi_l(x_s, \vartheta_l)$ makes use of an estimate of the ingress rate and the link average delay. When both the rate and the link delay are fixed, we can interpret $\varphi_l(x_s, \vartheta_l)$ as a SINR-threshold, and convergence of the update follows from [32]. Clearly, the measurement of noise and SINR can be carried out by the interference level caused by local other links. It is also found that the update rate is a decreasing function of the average delay ϑ_l. In other words, the longer the average delay is, the lower the rate of power update becomes. It is also consistent with the power update in practical wireless networks.

5.2 Source Rate and Delay Subalgorithm

We now obtain a rate allocation under the assumption that link powers are fixed. Plugging (37) into (10) results in the following optimization:

$$\max_{x,\vartheta,p\geq 0} \sum_{s\in S} U_s\big(x_s^{H_s+1}\big) - w_2 \sum_{l\in L} \vartheta_l - w_1 \sum_{l\in L} \frac{p_l}{\overline{\gamma}_l(\boldsymbol{P})}\varphi_l(x_s, \vartheta_l) \tag{38}$$

$$s.t. \quad x_s^{i+1} \leq x_s^i \boldsymbol{\phi}_{l(s,i)}, i = 1, 2, \ldots, H_s, \forall s, \tag{39}$$

$$P_r\left(\sum_{s\in S(l)} x_s^l > c_l(\boldsymbol{P})\right) \leq \zeta_l^{\max}, \forall l. \tag{40}$$

Remark 4. The above method does not guarantee the optimality of the joint rate and power allocation of the original problem (10) since the optimization with the fixed link powers $\boldsymbol{P}$ replaces function $\boldsymbol{P}^*(x)$, which denotes the simultaneous minimum power solution to (37) for a specified $\boldsymbol{x}$. However, we do not have a closed-form expression for this function.

Similar to (17), by employing some auxiliary variables and logarithmic transformation, the optimization problem (38) can be transformed to the unconstrained optimization

$$\begin{aligned} DD(\lambda, \mu, \upsilon) = &\max_{\{\tilde{x}\in\tilde{\Lambda}, \tilde{\vartheta}, \tilde{p}\in\tilde{I}\}} L_1(\tilde{x}, \lambda, \mu, \upsilon) + L_2(\tilde{p}, \lambda, \mu, \upsilon) \\ &+ L_3(\tilde{\vartheta}, \upsilon), \end{aligned} \tag{41}$$

where $L_1(\tilde{x}, \lambda, \mu, \upsilon)$ is the Lagrangian function with respect to source rate $\tilde{x}$ and Lagrangian multipliers λ, μ, υ; $L_2(\tilde{p}, \lambda, \mu, \upsilon)$ represents the Lagrangian function with respect to transmission power $\tilde{p}$ and Lagrangian multipliers λ, μ, υ; $L_3(\tilde{\vartheta}, \upsilon)$ denotes the Lagrangian function with respect to transmission delay $\tilde{\vartheta}$ and Lagrangian multipliers υ.

For (41) with $\tilde{\boldsymbol{P}}$ fixed, we observe that the objective is strictly concave in $\tilde{x}$. We have the subgradient on the data rate $\tilde{x}_s^i$ as follows for $i = 1, 2, \ldots, H_s$,

$$\nabla DD(\tilde{x}_s^i) = \lambda_s^i - \lambda_s^{i-1} - \frac{w_1 \exp\big(\tilde{x}_s^i\big) M_s}{W}, \tag{42}$$

where $M_s = \frac{\tilde{p}_l}{\overline{\gamma}_l(\tilde{\boldsymbol{P}})}\exp(\frac{1}{W}(\sum_{s\in S(l)} \exp(\tilde{x}_s) + \frac{K}{\exp(\tilde{\vartheta}_l)}))$ and for $i = H_s + 1$, the subgradient is the same as that in (24).

Theorem 5. *At the ith link of flow s, the flow rate $\tilde{x}_s^i$ without explicit message passing can be given by*

$$\tilde{x}_s^i(t+1) = \big[\tilde{x}_s^i(t) + \varepsilon \nabla DD(\tilde{x}_s^i)\big]_{\tilde{\Lambda}_s}, i = 1, 2, \ldots, H_s, \tag{43}$$

where ε is a sufficiently small fixed step-size.

Similarly, the objective in (41) is also strictly concave in $\tilde{\vartheta}_l$ for $\tilde{\boldsymbol{P}}$ fixed. We can therefore make use of the delay update

$$\tilde{\vartheta}_l(t+1) = \big[\tilde{\vartheta}_l(t) + \varepsilon \nabla DD(\tilde{\vartheta}_l(t))\big]^+, \forall l, \tag{44}$$

where ε is a sufficiently small fixed step-size and

$$\nabla DD(\tilde{\vartheta}_l(t)) = \frac{w_1 K M_s}{W \exp(\tilde{\vartheta}_l(t))}. \tag{45}$$

The update of λ_s^i and μ_l is the same as that in (29) and (35), respectively.

M_s are messages that are accumulated from each link along the route of a source s. Each of these terms involves a noise measurement that is scaled by the total ingress rate and the delay a per-link, a locally measurable quantity. As in Section 4, these terms may be progressively accumulated into an ACK packet header as it traverses from the receiver back to the source.

This source rate update is guaranteed to converge if (43) and (44) can be iterated until convergence for a single source, keeping other sources fixed, and before proceeding to the next source. In practice, we have observed such convergence even when these messages are used for continuous asynchronous updates.

6 Performance Evaluation

In this section, we evaluate the performance of the proposed RENUM algorithm by numerical results and network-wide simulation.

6.1 Numerical Results

In this section, we implement the proposed RENUM algorithm in MATLAB and provide numerical examples to illustrate the advantages of RENUM over other network utility maximization (NUM) algorithms without link outage/average delay constraints in terms of convergence, effective rates, link delay and power consumption.

Consider a MANET with k links, where the links are numbered from 1 to k. The link outage constraint is set to be $\zeta_l^{\max} = 0.1,\ l \in 1, 2, \ldots, k$. An example network with four sources having fixed routes is shown in Fig. 2. Flows 2 and 3 traverse all the links from link 2 to link 4 and from link 1 to link 5, respectively, and flow 1 is three hops to its destination over links 3 and 4. The receiver of link 1 is interfered by the transmitters of flows 1 and 4. A transmission bandwidth is 1 MHz, giving a $W = 125$ KHz baseband on each link. We let $m_s = 0.1$ kbps, $M_s = 0.5$ Mbps, $K = 1$ Mbit, $G_{lk} = d_{lk}^4$, $p_l^{\min} = 1$ mw and $p_l^{\max} = 100$ mw. We average over 50 trials to obtain a small relative error (within 10 percent of the average value).

For clearer illustration, we use α-fairness utility function $U_s(x_s) = (1-\alpha)^{-1} x_s^{1-\alpha}$ with $\alpha = 5$. We compare the performance of RENUM with JOPRC in [13], both of which

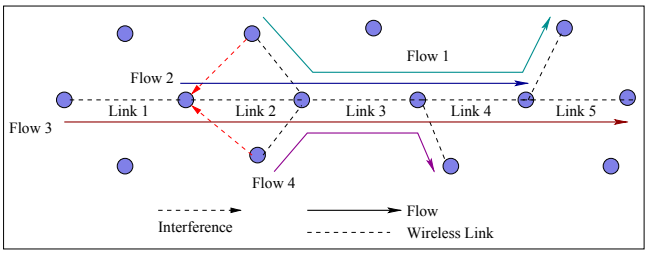

Fig. 2. MANET topology.

propose joint power scheduling and rate control schemes in wireless ad hoc networks. However, the latter employs opportunistic power scheduling and a stochastic subgradient algorithm. Figs. 3a and 3b show the comparison of the injection rates and the effective rates on flows 1 and 3, respectively. Clearly, the effective rates received at destinations are smaller than the injection one at transmitters. As shown in Fig. 3b, both RENUM and JOPRC can converge fast, whereas RENUM yields higher effective rates for flows 1 and 3 than JOPRC does on their destination nodes. It is seen from Fig. 3a that the injection rates for both flows 1 and 3 in RENUM is larger than those in JOPRC while the total power consumption of the two for all flows are almost equivalent, as shown in Fig. 3c. This is because that RENUM explicitly takes into account the data loss in its objective function, whereas JOPRC is concerned with the injection rates only. The data loss can induce the increase of source rate to maximize the network utility.

Fig. 4a compares the convergence of Algorithms A and B proposed in this paper and their power consumption. It is clear that Algorithm B converges much faster than Algorithm A, whereas compared with Algorithm A, Algorithm B is not optimal and it consumes more power to achieve the convergence. It is justified since no explicit message is passed to the transmitters; moreover, it needs more power to measure interference level caused by other links. We now consider the channels with fast fading. Only algorithm B is considered due to its attractive convergence rate and fully-distributed power-updates. At each system period t, each link fading gain G_{lk} is computed with Clarke's model [33, Ch. 5] at a carrier frequency of 2.4 GHz and we use $v = 10$ km/hr to calculate the maximum Doppler frequency for the worst-case situation of node moving in opposite directions. Fig. 4b shows the effect of node mobility velocity on the effective rate of flow 3 by Algorithm B. It is clear that as the node mobility velocity increases, the effective rate of flow 3 decreases rapidly. The RENUM with Algorithm B can converge fast in low mobility velocity, whereas the effective rate of flow 3 oscillates severely in high mobility velocity, in this case, it is controlled by a rate-outage target 20 percent.

Figs. 5a and 5b show the effect of maximum rate outage probability on efficient rate and flow delay when RENUM converges to optimum. We can observe from Fig. 5a that for a given maximum rate outage probability, the flow effective rate on destination decreases as the number of hops increases. Especially, due to the competition for wireless links among multiple flows, the decrease becomes sharper. We can also find that as the maximum rate outage probability increases, there exits a bound on the effective rate increase. Once crossing this bound, the effective rate begins to decrease. As shown in Fig. 5b, its effect on flow delay is opposite to that. This is reasonable since at the beginning, the larger rate outage probability means more tolerance on fading-induced congestion and more packets forwarded, which further induces the increase of the effective rate and the decrease of flow delay. However, once the tolerance is beyond a bound, more packets injected into the wireless links will make the congestion more serious. Furthermore, a lot of packets are dropped and flow delay increases as well.

We also compare the performance of RENUM with ENUM proposed in [23]. Fig. 6 depicts the effective rate, power consumption and delay obtained by RENUM and ENUM for flow 3. We can find that both can achieve the similar effective rate, whereas flow 3 by RENUM consumes less transmission power and has smaller average delay in comparison to ENUM. It is justified since ENUM only takes into account the effect of the lossy feature of wireless links on rate control to maximize the network utility, however, does not consider the effect of the lossy feature on transmission power and link delay. Especially, the objective of ENUM is not to minimize the total power consumption and flow delay.

6.2 Network-Wide Performance

To evaluate the network-wide performance of the proposed RENUM algorithm, we also implement it in NS-2 simulator [34].

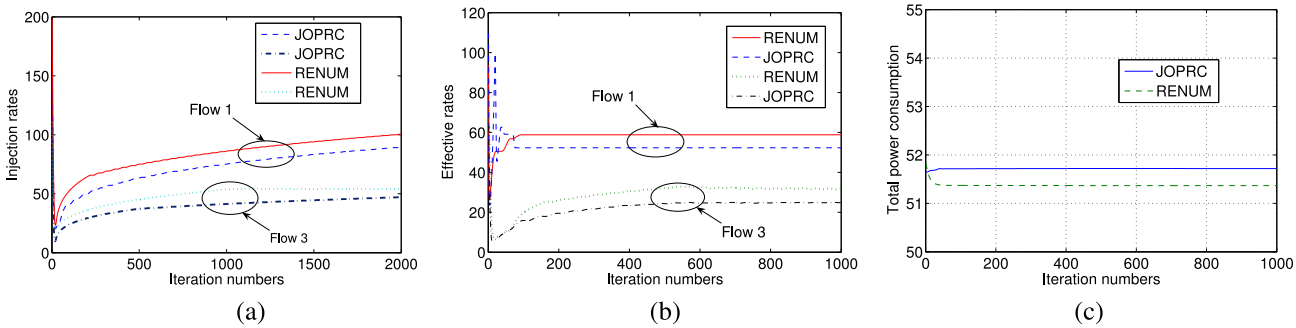

Fig. 3. Performance comparison between RENUM and JOPRC. (a) Injection rates of flows 1 and 3 on their own sources. (b) Effective rates of flows 1 and 3 on their own destinations. (c) Total power consumption comparison between ERNUM and JOPRC on all links.

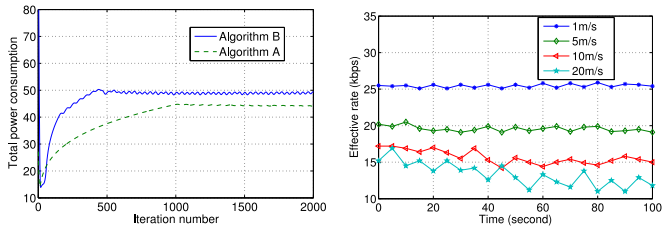

Fig. 4. (a) Comparison of convergence and power consumption between Algorithms A and B. (b) Effect of the node mobility on the effective rate by Algorithm B.

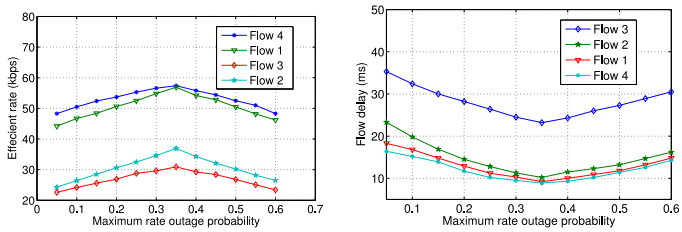

Fig. 5. (a) Effect of maximum rate outage probability on efficient rate. (b) Effect of maximum rate outage probability on flow delay.

To elaborate, the source rate control subalgorithm is implemented in the transport layer while the power allocation subalgorithm is in the physical layer. Unless otherwise stated, the default parameter settings are as follow: the two-ray ground reflection model is used as the radio propagation model, IEEE 802.11 DCF (distributed coordination function) as the MAC protocol, Drop Tail as the interface queue type, Omniantenna as the antenna model and AODV (ad hoc on-demand distance vector routing) as the routing protocol. An interface queue at the MAC layer could hold 50 packets before they are sent out to the physical link. Link breakage is detected from MAC layer feedbacks. A routing buffer at the network layer could store up to 64 data packets. This buffer keeps data packets waiting for a route, such as packets for which route discovery has started but no reply arrived yet.

In the simulation, 60 nodes are randomly placed in a 1,000 m $\times$ 1,000 m square area. The source of each TCP flow randomly selects one node as the destination, which is at least m hops away from the source. In our studies, we choose $m = 1, 3, 5$. There are a total of 20 flows with the same CBR/UDP traffic in the network. In these topologies, each node is 200 m away from its closest neighbors. The transmission range and the interference range are 250 and 550 m, respectively. The data transmission rate is 2 Mbps. The packets generated are of size 1,000 bytes in the simulation. The mobility model used for topology generation is the random waypoint model [35]. We conduct simulation using five different moving speeds, 0, 1, 5, 10 and 20 m/s. The speed of 0 m/s means all nodes are static. The pause time of each node is equal to 0, and therefore each node moves continuously in the network if its maximum speed is greater than 0. All results are averaged over 30 random simulations with 80 s simulated time each.

In the simulation, the following several important performance metrics are evaluated.

- *Aggregate end-to-end throughput*. The amount of data delivered to the destinations per second.
- *End-to-end delay*. The accumulative delay in data packet delivery due to buffering of packets, new route discoveries, queuing delay, MAC-layer retransmission, and transmission and propagation. delays.
- *Packet delivery ratio*. Percentage of data packets received at the destinations out of the total data packets generated by CBR traffic sources.
- *Fairness index*. The commonly used fairness index for all flows $x_i (1 \leq i \leq n)$, i.e.,

$$f = \frac{(\sum_{i=1}^{n} x_i)^2}{(n \cdot \sum_{i=1}^{n} x_i^2)}, \tag{46}$$

where x_i denotes the end-to-end throughput of the ith flow.

We first verify the performance of RENUM in terms of the aggregate end-to-end throughput and end-to-end delay.

Fig. 7a depicts the aggregate end-to-end throughput of RENUM under different number of hops. We observe that, when the number of hops for each flow increases, the aggregated end-to-end throughput of RENUM decreases. This is reasonable because multihop flows with longer paths have to pass more lossy links, thus lose more data packets for the same injection traffic, which results in that the effect rates received at destinations decrease. In addition, we also find that a larger number of hops can induce unstable aggregate

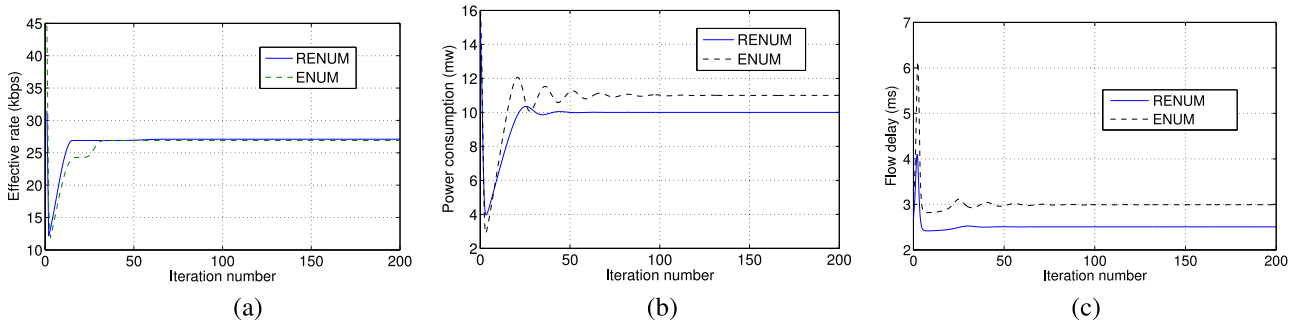

Fig. 6. Performance comparison between RENUM and ENUM. (a) Comparison of effective rate for flow 3 between the two. (b) Comparison of power consumption for flow 3 between the two. (c) Comparison of delay for flow 3 between the two.

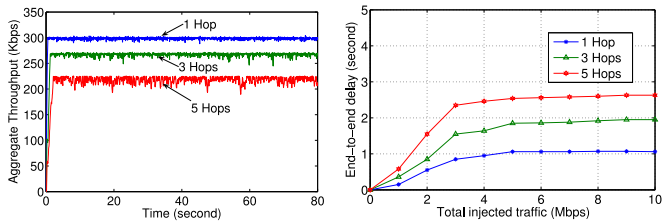

Fig. 7. Performance evaluation of RENUM in the random topology (speed of node mobility = 1 m/s). (a) Aggregate end-to-end throughput of RENUM under different number of hops. (b) End-to-end delay of RENUM under different total injected traffic.

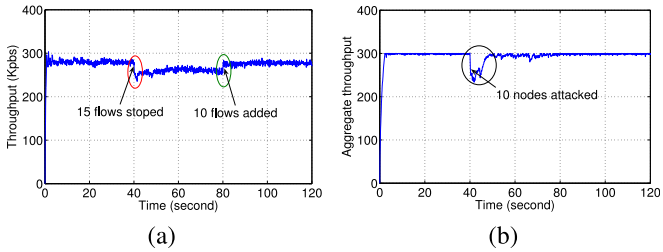

Fig. 9. Impact of sudden flow changes and cyberattacks on throughput in the random topology (speed of node mobility = 1 m/s). (a) Effect of sudden flow changes on throughput (number of hops = 3). (b) Evolution of aggregate throughput when 10 nodes are attacked at 40th second (number of hops = 1).

end-to-end throughput. The underlying reason is that the flows with longer paths will encounter not only the interflow contention, but also the intraflow contention, which could result in collisions and congestion, and significantly limit the performance of ad hoc networks.

Fig. 7b shows end-to-end delay of RENUM with different total injected traffic when the speed of node mobility is 1 m/s. We can observe that the end-to-end delay is low when the network is in light traffic load, then rapidly increases with traffic load and finally trends to be stable in spite of high traffic load. This is because as the traffic load becomes heavy, more flows take part in the contention for shared channels, and the network begins to get congested, which results in the longer round-trip time (RTT) of TCP packets. In the meanwhile, we can see that a large number of hops leads to long end-to-end delay. The reason is that a large number of hops can prolong the time of ACK feedback, data retransmission and new route discovery, and lead to a lot of accumulated packets in the outgoing queue at each node, which can greatly increase the queueing delay.

We now explore the impact of node mobility on packet delivery ratio, throughput and fairness.

Fig. 8 specifies data packet delivery ratio with different node moving speeds. The purpose of considering mobility is only to illustrate that our RENUM framework can work well in mobile scenarios with an on-demand routing scheme. We have the following observations. First, the data packet delivery ratio of nodes with lower moving speeds is larger than that with higher moving speeds. Second, the larger number of hops, the smaller packet delivery ratio. The reason behind these observations is that if nodes move rapidly in the network, the mobility would cause packet losses and make TCP connection unreliable. Moreover, when the path length of TCP flows

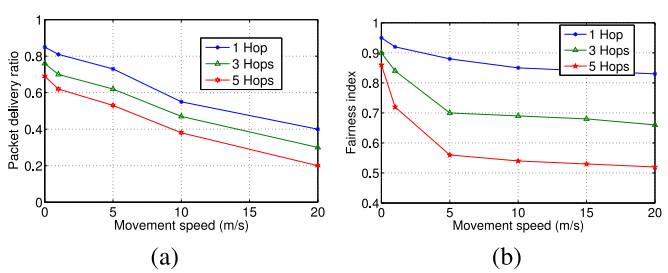

Fig. 8. Impact of node mobility on performance in the random topology. (a) Data packet delivery ratio with different node moving speeds. (b) Fairness index with different node moving speeds.

increases, the probability that TCP packets collide with their ACKs would become larger, and the TCP sender has fewer possibilities to accurately evaluate the channel state of forwarding nodes according to its local information.

Fig. 8b indicates fairness index with different speeds of node mobility. We can see that RENUM algorithm achieves comparatively stable fairness among all flows and allocates bandwidth fairly even for the nodes with higher moving speed. Moreover, we also observe that the number of hops becomes the main factor that influences the fairness of resource allocation among flows, i.e., the longer the length of TCP flows is, the worse the fairness among the flows is. This is because that in the RENUM framework, we consider rate fairness constraint, which makes it possible that the flows with poor channel conditions have an opportunity to transmit data, and employ the rate-effective based congestion control mechanism with an explicit congestion notification including the link congestion prices and the channel states.

We now study the effect of variability and unexpected events such as sudden flow changes and cyberattacks on throughput in the random topology.

Fig. 9a shows the effect of sudden flow changes on throughput in the random topology. In this simulation, the number of TCP flows is 50 at time 0.0 s. At the 40th second, 15 flows are terminated. At the 80th second, additional 10 flows are added. The whole simulation lasts 120 s. We can observe that RENUM algorithm is sensitive to these sudden flow changes and rapidly becomes stable after these disturbances. As the flows depart suddenly, the throughput becomes small, and the change is opposite for sudden arrivals. This experiment verifies that RENUM algorithm is much robust for sudden flow changes. The reason behind these observations is that we explicitly take into account the average SINR and take advantage of rate-outage probability to allow the network to experience a limited amount of congestion.

Fig. 9b depicts the evolution of aggregate throughput when 10 nodes are attacked at 40th second in the random topology. This simulation aims to verify whether RENUM algorithm can resist and tolerate the cyberattacks, and can hold on the normal performance when the network is attacked. In this simulation, we assume that at the 40th second, the locations of 10 nodes are disclosed and the hacker captures a part of these nodes so that in the network there forms a "black hole" in which nodes can neither transmit nor receive, while the attacker controls the rest of these nodes to wirelessly interfere with their neighbors by using a

huge amount of power or sending a huge amount of rubbish messages to their neighbors so that in the network traffic jamming occurs and the neighbor nodes are not able to transmit and receive normal data packets. We can observe from Fig. 9b that at the 40th second, the aggregate throughput decreases sharply and then oscillates slightly at a low level; After a short time, the throughput gradually increases, then recovers up to the level before the attack and finally becomes stable. The underlying reason is that both connections and communications between nodes in a MANET do not rely on a preexisting infrastructure, and each node is an equally privileged participant in routing by forwarding data for other nodes. Even though some nodes are disabled due to cyberattacks, the sources could rapidly adjust their forwarding nodes and search for new routes to the destinations by cooperation among nodes so as to avoid the degradation of network performance. On the other hand, the fact that data transmissions are decentralized among nodes and the forwarding nodes are hidden among plenty of network components reduces the probability that private information is eavesdropped.

7 Conclusions

This paper has revisited a cross-layer design problem for MANETs involving rate- and power-allocation. By considering the lossy feature of wireless links and a small degree of toleration for fading-induced congestion, we formulate a RENUM problem with the constraints of flow conservation, rate-outage probability and average delay, whose objective is to maximize the effective rate related network utility while minimizing the total power consumption and average delay of all links. The non-convex RENUM problem is decomposed into three subproblems of the rate control at the transport layer and the power and average delay control at the physical layer. Moreover, we propose three corresponding distributed subalgorithms with explicit message passing. In order to alleviate the explicit message passing, we also develop three distributed subalgorithms having near-optimal performance with a convergence rate order of magnitude faster than previous schemes. Numerical examples illustrate that the proposed RENUM outperforms other NUM such as JOPRC and ENUM in terms of convergence, effective rates, link delay and power consumption. Our NS-2-based network-wide simulation has also confirmed the effectiveness of RENUM algorithm and demonstrated that RENUM algorithm can achieve efficient and fair resource allocation even for nodes with high moving speed in ad hoc networks.

As a future work, we will implement and evaluate the proposed RENUM algorithm in a real network testbed/system instead of simulations, and further study how to integrate the proposed RENUM framework with TCP-SACK so as to improve the network performance since TCP-SACK allows ACKs to carry information about which packet they are acknowledging.

Acknowledgments

The work in this paper was supported in part by the grants from the Research Grants Council of Hong Kong, China (No. CityU 112910), the National Natural Science Foundation of China (61170248, 61373179, 61373178), New Century Excellent Talents in University (No. NCET-10-0877), Natural Science Foundation of Chongqing (CSTC, cstcjjA40003), Science and Technology Leading Talent Promotion Project of Chongqing (cstc2013kjrc-ljrccj40001), Fundamental Research Funds for the Central Universities (XDJK2013A018), and US National Science Foundation (NSF) under grant number ECCS-1307576 and US Army Research Office under grant number W911NF-09-1-0154.

References

[1] C. Ma and Y. Yang, "A Battery-Aware Scheme for Routing in Wireless Ad Hoc Networks," *IEEE Trans. Vehicular Technol.*, vol. 60, no. 8, pp. 3919-3932, 2011.

[2] C. Ma and Y. Yang, "Battery-Aware Routing for Streaming Data Transmissions in Wireless Sensor Networks," *Mobile Netw. Appl.*, vol. 11, no. 5, pp. 757-767, 2006.

[3] X. Xiang, X. Wang and Y. Yang, "Stateless Multicasting in Mobile Ad Hoc Networks," *IEEE Trans. Comput.*, vol. 59, no. 8, pp. 1076-1090, 2010.

[4] J. Lee, R. Mazumdar, and N. Shroff, "Opportunistic Power Scheduling for Dynamic Multi-Server Wireless Systems," *IEEE Trans. Wireless Comm.*, vol. 5, no. 6, June 2006.

[5] L. Zhang, Y.-C. Liang, and Y. Xin, "Joint Beamforming and Power Allocation for Multiple Access Channels in Cognitive Radio networks," *IEEE J. Selected Areas in Comm.*, vol. 26, no. 1, Jan. 2008.

[6] J.T. Wang, "SINR Feedback-Based Integrated Base-Station Assignment, Diversity, and Power Control for Wireless Networks," *IEEE Trans. Vehicular Technol.*, vol. 59, no. 1, pp. 473-484, Jan. 2010.

[7] J. Gomez and A.T. Campbell, "Variable-Range Transmission Power Control in Wireless Ad Hoc Networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 1, pp. 87-99, Jan. 2007.

[8] B. Alawieh, C.M. Assi, and W. Ajib, "Distributed Correlative Power Control Schemes for Mobile Ad Hoc Networks Using Directional Antennas," *IEEE Trans. Vehicular Technol.*, vol. 57, no. 3, pp. 1733-1744, May 2008.

[9] Q. Qu, L.B. Milstein, and D.R. Vaman, "Cross-Layer Distributed Joint Power Control and Scheduling for Delay-Constrained Applications over CDMA-Based Wireless Ad-Hoc Networks," *IEEE Trans. Comm.*, vol. 58, no. 2, pp. 669-680, Feb. 2010.

[10] D. Wang, X. Wang, and X. Cai, "Optimal Power Control for Multi-User Relay Networks over Fading Channels," *IEEE Trans. Wireless Comm.*, vol. 10, no. 1, pp. 199-207, Jan. 2011.

[11] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control," *IEEE J. Selected Areas in Comm.*, vol. 23, no. 1, pp. 104-116, Jan. 2005.

[12] R. Jantti and S. Kim, "Joint Data Rate and Power Allocation for Lifetime Maximization in Interference Limited Ad Hoc Networks," *IEEE Trans. Wireless Comm.*, vol. 5, no. 5, pp. 1086-1094, May 2006.

[13] J.-W. Lee, R.R. Mazumdar, and N.B. Shroff, "Joint Opportunistic Power Scheduling and End-to-End Rate Control for Wireless Ad Hoc Networks," *IEEE Trans. Vehicular Technol.*, vol. 56, no. 2, pp. 801-809, Mar. 2007.

[14] P. Soldati, B. Johansson, and M. Johansson, "Distributed Cross-Layer Coordination of Congestion Control and Resource Allocation in S-TDMA Wireless Networks," *Wireless Netw.*, vol. 14, no. 6, pp. 949-965, Dec. 2008.

[15] S. Guo, C. Dang, and X. Liao, "Distributed Algorithms For Resource Allocation of Physical and Transport Layers in Wireless Cognitive Ad Hoc Networks," *Wireless Netw.*, vol. 17, no. 2, pp. 33-356, Feb. 2011.

[16] K. Hedayati, I. Rubin, and A. Behzad, "Integrated Power Controlled Rate Adaptation and Medium Access Control in Wireless Mesh Networks," *IEEE Trans. Wireless Comm.*, vol. 9, no. 7, pp. 2362-2370, July 2010.

[17] J. Papandriopoulos, S. Dey, and J. Evans, "Optimal and Distributed Protocols for Cross-Layer Design of Physical and Transport Layers in MANETs," *IEEE-ACM Trans. Netw.*, vol. 16, no. 6, pp. 1392-1405, Dec. 2008.

[18] S. Kandukuri and S. Boyd, "Optimal Power Control in Interference-Limited Fading Wireless Channels with Outage-Probability Specifications," *IEEE Trans. Wireless Comm.*, vol. 1, no. 1, pp. 46-55, Jan. 2002.
[19] J. Papandriopoulos, J. Evans, and S. Dey, "Optimal Power Control for Rayleigh-Faded Multiuser Systems with Outage Constraints," *IEEE Trans. Wireless Comm.*, vol. 4, no. 6, pp. 2705-2715, Nov. 2005.
[20] J. Papandriopoulos, J. Evans, and S. Dey, "Outage-Based Optimal Power control for Generalized Multiuser Fading Channels," *IEEE Trans. Comm.*, vol. 54, no. 4, pp. 693-703, Apr. 2006.
[21] S.Y. Park and D.J. Love, "Outage Performance of Multi-Antenna Multicasting for Wireless Networks," *IEEE Trans. Wireless Comm.*, vol. 8, no. 4, pp. 1996-2005, Apr 2009.
[22] C. Shao, X. Hua, and A. Huang, "Outage Probability Based Resource Allocation in Wireless Mesh Networks," *Proc. IEEE Global Telecomm. Conf.*, pp. 1-5, 2010.
[23] Q. Gao, J. Zhang, and S.V. Hanly, "Cross-Layer Rate Control in Wireless Networks with Lossy Links: Leaky-Pipe Flow, Effective Network Utility Maximization and Hop-by-Hop Algorithms," *IEEE Trans. Wireless Comm.*, vol. 8, no. 6, pp. 3068-3076, June 2009.
[24] J.-W. Lee, M. Chiang, and A.R. Calderbank, "Jointly Optimal Congestion and Contention Control Based on Network Utility Maximization," *IEEE Comm. Letters*, vol. 10, no. 3, pp. 216-218, 2006.
[25] S. Chen, M. Beach, and J. McGeehan, "Division-Free Duplex for Wireless Applications," *Electronics Letters*, vol. 34, no. 2, pp. 147-148, Jan 1998.
[26] G. Stuber, *Principles of Mobile Communication*. Kluwer, 1997.
[27] A. Goldsmith, *Wireless Communication*. Cambridge Univ. Press, 2005.
[28] D.P. Bertsekas and R.G. Gallager, *Data Networks*. Prentice Hall, 1987.
[29] S. Boyd and L. Vandenbergh, *Convex Optimization*. Cambridge Univ. Press, 2004.
[30] D.P. Bertsekas, *Nonlinear Programming*. second ed. ed., Athena Scientific, 1999.
[31] W. Xu, Y. Wang, J. Chen, G. Baciu, and Y. Sun, "Dual Decomposition Method for Optimal and Fair Congestion Control in Ad Hoc Networks: Algorithm, Implementation and Evaluation," *J. Parallel and Distributed Computing*, vol. 68, no. 7, pp. 997-1007, 2008.
[32] R. Yates, "A Framework for Uplink Power-Control in Cellular Radio Systems," *IEEE J. Selected Areas in Comm.*, vol. 13, no. 7, pp. 1341-1347, Sept 1995.
[33] T.S. Rappaport, *Wireless Communication: Principles and Practice*. second ed. ed., Prentice-Hall, 2002.
[34] NS-2, Network Simulator, http://www.isi.edu/nsnam/ns/, 2012.
[35] D. Johnson and D. Maltz, *Ad Hoc Networking*. Addison-Wesley, 2001.

Songtao Guo received the BS, MS, and PhD degrees in computer software and theory from Chongqing University, Chongqing, China, in 1999, 2003, and 2008, respectively. He was a professor from 2011 to 2012 at Chongqing University. He is currently a full professor at Southwest University, China. He was a senior research associate at the City University of Hong Kong from 2010 to 2011, and a visiting scholar at Stony Brook University, New York, from May 2011 to May 2012. His research interests include wireless sensor networks, wireless ad hoc networks and parallel and distributed computing. He has published more than 30 scientific papers in leading refereed journals and conferences. He has received many research grants as a principal investigator from the National Science Foundation of China and Chongqing and the Postdoctoral Science Foundation of China. He is a member of the IEEE.

Changyin Dang received the PhD degree (cum laude) in operations research/economics from Tilburg University, The Netherlands, in 1991, the MSc degree in applied mathematics from Xidian University, China, in 1986, and the BSc degree in computational mathematics from Shanxi University, China, in 1983. His research interests include operations research/optimization and related fields. Because of his significant contributions, he has been named an excellent researcher by Tilburg University, The Netherlands, in 1990 and invited to give talks at universities such as Cornell University, Stanford University, University of California, University of Michigan, University of Minnesota, etc., and to work as a research fellow at Yale University in 1994. He has published more than 100 papers in journals including *SIAM Journal on Discrete Mathematics*, *SIAM Journal on Optimization*, *Artificial Intelligence*, *IEEE Transactions on Automatic Control*, *IEEE Transactions on Evolutionary Computation*, etc. He is a senior member of the IEEE.

Yuanyuan Yang received the BEng and MS degrees in computer science and engineering from Tsinghua University, Beijing, China, and the MSE and PhD degrees in computer science from Johns Hopkins University, Baltimore, Maryland. She is currently a professor of computer engineering and computer science at Stony Brook University, New York, and the director of the Communications and Devices Division at the New York State Center of Excellence in Wireless and Information Technology (CEWIT). Her research interests include wireless networks, data center networks, optical networks and high-speed networks. She has published 280 papers in major journals and refereed conference proceedings and holds seven US patents in these areas. She is currently an associate editor-in-chief for the *IEEE Transactions on Computers* and an associate editor for the *Journal of Parallel and Distributed Computing*. She has served as an associate editor for the *IEEE Transactions on Computers* and *IEEE Transactions on Parallel and Distributed Systems*. She has served as a general chair, program chair, or vice chair for several major conferences and a program committee member for numerous conferences. She is a fellow of the IEEE.