Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 116 (2018) 50257-50265

ISSN: 2229-712X

Adaptive Opportunistic Routing for Wireless Ad Hoc Networks

P.N.Varalakshmi.k, Ch.Naga Praneetha, K Harika, B TejGopal and G Anusha
Department of Computer Science Engineering, Tirumala Engineering College.

ARTICLE INFO

Article history:
Received: 20 February 2018;
Received in revised form:
13 March 2018;
Accepted: 24 March 2018;

Keywords
Opportunistic Routing,
Reward Maximization,
Wireless Ad hoc Networks.

ABSTRACT

A distributed adaptive opportunistic routing scheme for multihop wireless ad hoc networks is proposed. The proposed scheme utilizes a reinforcement learning framework to opportunis-tically route the packets even in the absence of reliable knowledge about channel statistics and network model. This scheme is shown to be optimal with respect to an expected average per-packet re-ward criterion. The proposed routing scheme jointly addresses the issues of learning and routing in an opportunistic context, where the network structure is characterized by the transmission success probabilities. In particular, this learning framework leads to a sto-chastic routing scheme that optimally "explores" and "exploits" the opportunities in the network.

I. INTRODUCTION

Opportunistic routing for multihop wireless ad hoc networks has seen recent research interest to overcome deficiencies of conventional routing [1]–[6] as applied in wireless setting. Motivated by classical routing solutions in the Internet, conventional routing in ad hoc networks attempts to find a fixed path along which the packets are forwarded [7]. Such fixed -path schemes fail to take advantage of broadcast nature and opportunities provided by the wireless medium and result in unnecessary packet retransmissions. The opportunistic routing decisions, in contrast, are made in an online manner by choosing the next relay based on the actual transmission outcomes as well as a rank ordering of neighboring nodes. Opportunistic routing mitigates the impact of poor wireless links by exploiting the broadcast nature of wireless transmissions and the path diversity.

The authors in [1] and [6] provided a Markov decision theoretic formulation for opportunistic routing. In particular, it is shown that the optimal routing decision at any epoch is to select the next relay node based on a distance-vector summa-rizing the expected-cost-to-forward from the neighbors to the destination. This "distance" is shown to be computable in a distributed manner and with low complexity using the proba-bilistic description of wireless links. The study in [1] and [6] provided a unifying framework for almost all versions of op-portunistic routing such as SDF [2], Geographic Random For-warding (GeRaF) [3], and ExOR [4], where the variations in [2]–[4] are due to the authors' choices of cost measures to optimize. For instance, an optimal route in the context of ExOR [4] is computed so as to minimize the expected number of trans-missions (ETX), while GeRaF [3] uses the smallest geographical distance from the destination as a criterion for selecting the next-hop.

The opportunistic algorithms proposed in [1]–[6] depend on a precise probabilistic model of wireless connections and local topology of the network. In a practical setting, however, these probabilistic models have to be "learned" and "maintained." In other words, a comprehensive study and evaluation of any opportunistic routing scheme requires an integrated approach to the issue of probability estimation. Authors in [8] provide a sensitivity analysis for the opportunistic routing algorithm given in [6]. However, by and large, the question of learning/estimating channel statistics in conjunction with opportunistic routing re-mains unexplored.

In this paper, we first investigate the problem of opportunistically routing packets in a wireless multihop network when zero or erroneous knowledge of transmission success proba-bilities and network topology is available. Using a reinforcement learning framework, we propose a distributed adaptive opportunistic routing algorithm (d-AdaptOR) that minimizes the expected average per-packet cost for routing a packet from a source node to a destination. This is achieved by both suffi-ciently exploring the network using data packets and exploiting the best routing opportunities.

Our proposed reinforcement learning framework allows for a low-complexity, low-overhead, distributed asynchronous implementation. The significant characteristics of d-AdaptOR are that it is oblivious to the initial knowledge about the network, it is distributed, and it is asynchronous.

The main contribution of this paper is to provide an opportunistic routing algorithm that:1)assumes no knowledge about the channel statistics and network, but 2) uses a reinforcement learning framework in order to enable the nodes to adapt their routing strategies, and 3) optimally exploits the statistical opportunities and receiver diversity. In doing so, we build on the Markov decision formulation in [6] and an important theorem in Q-learning proved in [9]. There are many learning-based routing solutions(both heuristic or analytically driven) for conventional routing in wireless or wired networks [10]–[15]. None of these solutions exploits the receiver diversity gain in the context of opportunistic routing. However, for the sake of completeness, we provide a brief overview of the existing approaches. The authors in [10]–[14] focus on heuristic routing algorithms that adaptively identify the least congested path in a wired network. If the network congestion, hence delay, were to be replaced by time-invariant quantities,1 the heuristics in [10]–[14] would be-come a special case of d-AdaptOR in a network with deterministic channels and with no receiver diversity. In this light, Theorem 1 in Section IV provides analytic guarantees for the heuristics obtained in [10]–[14]. In [15], analytic results for ant routing are obtained in wired networks without opportunism. Ant routing uses ant-like probes to find paths of optimal costs such as expected hop count, expected delay, and packet loss probability.2 This dependence on ant-like probing represents astark difference with our approach where d-AdaptOR relies solely on data packet for exploration.

The rest of the paper is organized as follows. In Section II, we discuss the system model and formulate the problem. Section III formally introduces our proposed adaptive routing algorithm, d-AdaptOR. We then state and prove the optimality theorem for d-AdaptOR in Section IV. In Section V, we present the imple-mentation details and practical issues of d-AdaptOR. We perform simulation study of d-AdaptOR in Section VI. Finally, we conclude the paper and discuss future work in Section VII.

II. SYSTEM MODEL

We consider the problem of routing packets from a source critical and is only assumed for ease of exposition). A packet ered to model the amount of energy used for transmission, the expected time to transmit a given packet, or the hop count when the cost is set to unity. We consider an opportunistic routing set-ting with no duplicate copies of the packets. In other words, at a given time only one node is responsible for routing any given packet. Given a successful packet transmission from node to the set of neighbor nodes , the next (possibly randomized) routing decision includes: 1) retransmission by node ; 2) re-laying the packet by a node $j \in S$; or 3) dropping the packet altogether. If node is selected as a relay, then it transmits the

We define the termination event for packet m to be the event that packet m is either received at the destination or is dropped by a relay before reaching the destination. We denote this ter among the termination events as follows. We assume that upon the termination of a packet at the destination (successful de-livery of a packet to the destination), a fixed and given positive T

i.e., either zero if the packet is dropped prior to reaching the destination node or R if the packet is received at the destination. mission (equal to zero if at time $n_{,m}$ packet m is not transmitted). The routing scheme can be viewed as selecting a (random) sequence of nodes {i nm}for relaying packets m=1,2,3 As such, the expected average per-packet reward associated with time N and the expectation is taken over the events of transmission decisions, successful packet receptions, and packet generation times.

Problem ($\mathbf{P}$ *):* Choose a sequence of relay nodes { i^{nm} } in the absence of knowledge about the network topology such that J_N is maximized as $N \to \infty$.

In Section III, we propose the d-AdaptOR algorithm, which solves Problem $(\mathbf{P})$. The nature of the algorithm allows nodes to make routing decisions in distributed, asynchronous, and adaptive manner.

Remark 1: The problem of opportunistic routing for multiple source–destination pairs, without loss of generality, can be de-composed to the single source–destination problem described

III. DISTRIBUTED ALGORITHM

Before we proceed with the description of d-AdaptOR, we provide the following notations. Let $\mathcal{N}(i)$ denote the set of neighbors of node i including node i itself. Let $\mathfrak{S}^i$ denote the set of potential reception outcomes due to a transmission from node $i \in \Theta$, i.e., $\mathfrak{S}^i = \{S : S \subseteq \mathcal{N}(i), i \in S\}$. We refer to $\mathfrak{S}^i$ as the state space for node i's transmission. Furthermore, let $\mathfrak{S} = \cup_{i \in \Theta} \mathfrak{S}^i$. Let $A(S) = S \cup \{T\}$ denote the space of all allowable actions available to node i upon successful reception at nodes in S. Finally, for each node i, we define a reward function on states $S \in \mathfrak{S}^i$ and potential decisions $a \in A(S)$ as

A. Overview of d-AdaptOR

As discussed before, the routing decision at any given time is made based on the reception outcome and involves retransmission, choosing the next relay, or termination. Our proposed

TABLE I. NOTATIONS USED IN THE DESCRIPTION OF THE ALGORITHM.

Symbol	Definition
S_n^i	Nodes receiving the transmission from node i at time n
a_n^i	Decision taken by node i at time n
$A(S)$	Set of available actions when nodes in S receive a packet
$\mathcal{N}(i)$	Neighbors of node i including node i
$g(S, a)$	Reward obtained by taking decision a when set S of nodes receive a packet
$\nu_n(i, S, a)$	Number of times up to time n, nodes S have received a packet from node i and decision a is taken
$N_n(i, S)$	Number of times up to time n, nodes S have received a packet from node i
$\Lambda_n(i, S, a)$	Score for node i at time n, when nodes S have received the packet and decision a is taken
Λ_{max}^i	Estimated best score for node i

scheme makes such decisions in a distributed manner via the following three-way handshake between node i and its neighbors N(i).

1) At time n, node i transmits a packet.

2) The set of nodes S^i who have successfully received the packet from node i, transmit acknowledgment (ACK) packets to node i. In addition to the node's identity, the acknowledgment packet of node $k \in S^i$ includes a control *message* known as *estimated best score* (EBS) and denoted by Λ^k .

3) Node i announces node $j \in S^i$ as the next transmitter or announces the termination decision T in a forwarding (FO) packet.

B. Detailed Description of d-AdaptOR

The operation of d -AdaptOR can be described in terms of initialization and four stages of transmission, reception and acknowledgment, relay, and adaptive computation as shown in Fig. 1. For simplicity of presentation, we assume a sequential timing for each of the stages. We use $n+$ to denote some

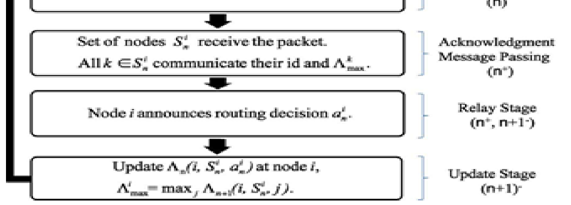

Fig. 1. Flow of the algorithm. The algorithm follows a four-stage procedure: transmission, acknowledgment, relay, and update.

Initialization:

For all $i \in \Theta S \in \mathcal{S} a \in \mathcal{A}(S)$, j

$$\Lambda_0(i, S, \bullet) = \nu_1(i, S, \bullet) = N_1(i, S) = \Lambda^i_{\max} = 0,$$

While

$$\Lambda_{T_{\max}} = -R$$

1) Transmission Stage: n

Transmission stage occurs at time n in which node i transmits if it has a packet.

2) Reception and acknowledgment Stage:

Let S^i_n denote the (random) set of nodes that have received the packet transmitted by node i. In the reception and acknowledgment stage, successful reception of the packet transmitted by node i is acknowledged to it by all the nodes in S^i. We assume that the delay for the acknowledgment stage is small enough (not more than the duration of the time slot) such that node i infers S^i by time $n+$. For all nodes $k \in S^i$, the ACK packet of node k to node i includes the EBS message $\Lambda^k_{\max}$. Upon reception and acknowledgment, the counting random variable N is incremented as follows:

$$N_n(i, S) = \begin{cases} N_{n-1}(i, S) + 1, & \text{if } S = S^i_n \\ N_{n-1}(i, S), & \text{if } S \neq S^i_n \end{cases}$$

- With probability $\epsilon_n(i, S^i_n)$

$$a^i_n \in \mathcal{A}(S^i_n)$$

is selected uniformly with probability $\frac{\epsilon_n(i, S^i_n)}{|\mathcal{A}(S^i_n)|}$

Node i transmits FO, a control packet that information about routing decision a_i at some time strictly between $n+$ and $(n+1)^-$. If $a_i \neq T$, then node a_i pre-pares for forwarding in the next time slot, while nodes $j \in S^i, j \neq a_i$ expunge the packet. If termination action is chosen, i.e., $a_i = T$, all nodes in S^i expunge the packet.

Upon selection of routing action, the counting variable ν is updated

$$\nu_n(i, S, a) = \begin{cases} \nu_{n-1}(i, S, \bullet) + 1, & \text{if } (S, a) = (S^i_n, \bullet^i_n) \\ \nu_{n-1}(i, S, \bullet), & \text{if } (S, a) \neq (S^i_n, \bullet^i_n). \end{cases}$$

4) Adaptive Computation Stage: $(\cdot \cdot)$

- For $S = S^i_n$, $a = a^i_n$

$$\Lambda_{n+1}(i, S, \bullet) = \Lambda_n(i, S, \bullet) + \bullet_{\nu_n(i,S,\bullet)}$$
$$\times \left(-\Lambda_n(i, S, \bullet) + g(S, \bullet) + \Lambda^\bullet_{\max}\right). \quad (2)$$

C. Computational Issues

The computational complexity and control overhead of d-AdaptOR is low.

1) Complexity: To execute stochastic recursion (2), the number of computations required per packet is order of $O(\max_{i \in \Theta} |\mathcal{N}(i)|)$ at each time slot. The space complexity of d-AdaptOR is exponential in the number of neighbors, i.e., $O(\max_{i \in \Theta} 2^{|\mathcal{N}(i)|})$ for each node. The reduction in storage requirement using approximation techniques in [16] is left as future work.

2) Control Overhead: The number of acknowledgments per packet is order of $O(\max_{i \in \Theta} |\mathcal{N}(i)|)$, independent of network size.

3) Exploration Overhead: The adaptation to the optimal performance in the network is guaranteed via a controlled randomized routing strategy that can be viewed as cost of exploration. The cost of exploration is proportional to the total number of packets whose routes deviates from the optimal path. In proof of Theorem 1, we show that this cost increases sublinearly with the number of delivered packets, hence the per-packet exploration cost diminishes as the number of delivered packets grows. to the genie-aided or greedy-based schemes such as ExOR or SR.

IV. ANALYTIC OPTIMALITY OF D-ADAPTOR

We will now state the main result establishing the optimality of the proposed d-AdaptOR algorithm under the assumptions of a time-invariant model of packet reception and reliable control packets. More precisely, we have the following assumptions.

Assumption 1: The probability of successful reception of a packet transmitted by node i at set $S \subseteq \mathcal{N}(i)$ of nodes is $P(S|i)$, independent of time and all other routing decisions.

The probabilities $P(\cdot|\cdot)$ in Assumption 1 characterize a packet reception model that we refer to as *local broadcast model*. Note that for all $S = S'$, successful reception at S

Remark 2: Assumption 1 is in line with the experimentally tested state of the art routing protocols MORE [17] and ExOR [4]. These studies seem to indicate that reasonably simple probabilistic models provide good abstractions of media access control (MAC) and physical (PHY) layers at the routing layer.

Remark 3: In practice, Assumption 2 is hard to satisfy. But as we will see in Section VI, when the rates and power of the control packets are set to maximize the reliability, the impact of violating this assumption can be kept extremely low.

Remark 4: In Section VI, we address the severity as well as the implications of Assumptions 1 and 2. In particular, via a set of QualNet simulations, we will show that d-AdaptOR exhibits many of its desirable properties in a realistic setup despite the relaxation of the analytical assumptions.

Given Assumptions 1 and 2, we are almost ready to present Theorem 1 regarding the optimality of d- AdaptOR among the class of policies that are oblivious to the net-work topology and/or channel statistics. More precisely, let

$$\bigcup_{j \in \mathcal{N}(j)} \{ \bullet \quad \bullet \cdots \quad n \ 1 \quad n \ 1 \quad n \}$$

set of such (P) -admissible policies. Theorem 1 states that d-AdaptOR, denoted by $\phi^* \in \Phi$, is an optimal (P) -admissible policy.

Theorem 1: Suppose $\sum^\infty \alpha = \infty \sum^\infty \alpha^2 < \infty$ and Assumptions 1 and 2 hold. Then, for all $\phi \in \Phi$

$$\lim_{N \to \infty} \mathbf{E}^{\phi^*} \left[\frac{1}{M_N} \sum_{m=1}^{M_N} \left\{ r_m - \sum_{n=\tau_s^m}^{\tau_T^m - 1} c_{i_n, m} \right\} \right]$$
$$\geq \limsup_{N \to \infty} \mathbf{E}^{\phi} \left[\frac{1}{M_N} \sum_{m=1}^{M_N} \left\{ r_m - \sum_{n=\tau_s^m}^{\tau_T^m - 1} c_{i_n, m} \right\} \right]$$

where $\mathbf{E}^{\phi^*}$ and $\mathbf{E}^{\phi}$ are the expectations taken with respect to policies ϕ^* and ϕ, respectively.5

Next, we prove the optimality of d-AdaptOR in two steps. In the first step, we show that Λ^n converges in an almost sure sense. In the second step, we use this convergence result to show that d-AdaptOR is optimal for Problem $(\mathbf{P})$.

A. Convergence of A_n

Let $U : \prod_i \mathbf{R}^{v_i} \rightarrow \prod_i \mathbf{R}^{v_i}$ be an operator on vector such that

$$(U\Lambda)(i\ S\ \bullet) = g(S\ \bullet) + \sum P(S'\ \bullet)\ \max\ \Lambda(\bullet\ S'\ j).$$

Using the convergence of Λ^n, we show that the expected average per-packet reward under d-AdaptOR is equal to the optimal expected average per-packet reward obtained for a genie-aided system where the local broadcast model is known perfectly. In other words, we take cue from known results associated with a closely related Auxiliary Problem ($\mathbf{AP}$). In this Auxiliary Problem ($\mathbf{AP}$), there exists a centralized controller with full knowledge of the local broadcast model $P(\cdot)$ as well as the transmission outcomes across the network [1], [6]. The objective in the Auxiliary Problem ($\mathbf{AP}$) is a single-packet variation of that in Problem ($\mathbf{P}$): the reward

$$\mathbf{E}\left[r_m - \sum_{n=\bullet}^{\tau_T^m\ 1} c_{i_{n,m}}\right]$$

for routing a single packet m from the source to the destination is maximized over a set Π of (AP)-*admissible* policies, where this set Π of (AP)-admissible policies is a superset of (P) admissible policies Φ that also includes all topology-aware and centralized policies. This Auxiliary Problem ($\mathbf{AP}$) has been extensively studied in [1], [6], and [19], where a Markov decision formulation provides the following important result.

Fact 1 [6, Theorem 2.1]: Consider the unique solution $V^*\Theta\cup\{T\}\rightarrow\mathbb{R}^+$ to the following fixed-point equation:

$$V^*(\bullet) = R \tag{6}$$

$$V^*(i) = \max\left(\left\{\ c_i + \sum_{S'} P(S'|i)(\max_{j\in S'} V^*(j))\right\}, \bullet\right) \tag{7}$$

There exists an optimal topology-aware and centralized admis sible policy $\pi^* \in \Pi$ such that

$$\liminf_{N\rightarrow\bullet\bullet} \mathbf{E}^{\bullet^*}\left[\frac{1}{M_N}\sum_{m=1}^{M_N}\left\{r_m - \sum_{n=\tau_s^m}^{\tau_T^m\ 1} c_{i_{n,m}}\right\}\right] \geq V^*(\bullet) - \delta.$$

The proof is given in Appendix- C. Lemmas 2 and 3 imply that ϕ^* [which is (P)-admissible by construction] is an optimal policy under which

$$\lim_{N\rightarrow\bullet\bullet} \mathbf{E}^{\bullet^*}\left[\frac{1}{M_N}\sum_{m=1}^{M_N}\left\{r_m - \sum_{n=\tau_s^m}^{\tau_T^m\ 1} c_{i_{n,m}}\right\}\right]$$

exists and is equal to $V^*(0)$ establishing the proof of Theorem 1.

Corollary 1: When $c_i = 1 i\in\Theta$, the network is connected, and R is greater than the worst-case routing cost,7 d-AdaptOR minimizes

$$D_N = \mathbf{E}^\pi\left[\frac{1}{M_N}\sum_{m=1}^{M_N}\{\tau_T^m\ \ \tau_s^m)\}\right] \tag{9}$$

the expected per-packet delivery time as $N\rightarrow\infty$.

This is because when $c_i = 1R$ is sufficiently large, and the network is connected

$$V^*(\bullet) = R - \inf_{\pi\in\Pi} \mathbf{E}^\pi\left[\frac{1}{M_N}\sum_{m=1}^{M_N}\left\{\sum_{n=\tau_s^m}^{\tau_T^m\ 1} c_{i_{n,m}}\right\}\right]$$
$$= R\ \ \inf_{\pi\in\Pi} D_N,\ \text{as}\ N\rightarrow\bullet\bullet.$$

V. PROTOCOL DESIGN AND IMPLEMENTATION ISSUES

In this section, we describe an 802.11 compatible implementation for d-AdaptOR.

A. 802.11 Compatible Implementation

The implementation of d-AdaptOR, analogous to any opportunistic routing scheme, involves the selection of a relay node among the candidate set of nodes that have received and acknowledged a packet successfully. One of the major challenges in the implementation of an opportunistic routing algorithm in general, and the d -AdaptOR algorithm in particular, is the de-sign of an 802.11 compatible acknowledgment mechanism at the MAC layer. We propose a practical and simple way to implement acknowledgment architecture.

The transmission at any node i is done according to an 802.11 CSMA/CA mechanism. Specially, before any transmission, transmitter i performs channel sensing and starts transmission after the backoff counter is decremented to zero. For each neighbor node $j\in\mathcal{N}i$, the transmitter node i then reserves a virtual time slot of duration $T_{\mathrm{ACK}}+T_{\mathrm{SIFS}}$, where T_{ACK} is the duration of the acknowledgment packet and T_{SIFS} is the duration of Short InterFrame Space (SIFS) [20]. Transmitter i then piggybacks a priority ordering of nodes $\mathcal{N}(i)$ with each data packet transmitted. The priority ordering determines the virtual time slot in which the candidate nodes transmit their acknowledgment. Nodes in the set S^i that have successfully received the packet then transmit acknowledgment packets sequentially in the order determined by the transmitter node. node i transmits a Forwarding control packet (FO). The FO packets contain the identity of the next forwarder, which may be node i again or any node $j\in S^i$. If T^{it} expires and no FO packet is received (FO packet reception is unsuccessful), then the corresponding candidate nodes drop the received data packet. If the transmitter i does not receive any acknowledgment, node i retransmits the packet. The backoff window is doubled after every retransmission. Furthermore, the packet is dropped if the retry limit (set to 7) is reached.

In addition to the acknowledgment scheme, d-AdaptOR requires modifications to the 802.11 MAC frame format. Fig. 2 shows the modified MAC frame formats required by d-AdaptOR. The reserved bits in the type/subtype fields of the frame control field of the 802.11 MAC specification are used to indicate whether the rest of the frame is a d -AdaptOR data frame, a d-AdaptOR ACK, or a, FO.8 The data frame contains

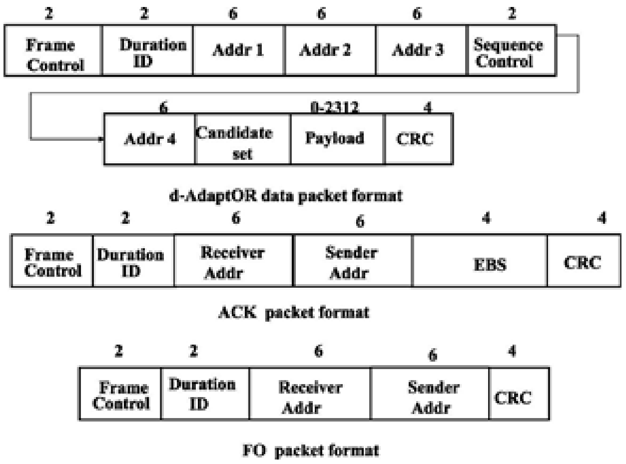

Fig. 2. Frame structure of the data packets, acknowledgment packets, and FO packets.

the candidate set in priority order, the payload, and the 802.11 Frame Check Sequence. The acknowledgment frame includes the data frame sender's address and the feedback EBS Λ_{max}. The FO packet is exactly the same as a standard 802.11 short control frame that uses different subtype value.

d-AdaptOR in a Realistic Setting

1) Loss of ACK and FO Packets: Interference or low signal-to-noise ratio (SNR) can cause loss of ACK and FO packets. Loss of an ACK packet results in an incorrect estimation of nodes that have received the packet, and thus affects the performance of the algorithm. Loss of FO packet negatively impacts the throughput performance of the network. In particular, loss of an FO packet can result in the drop of data packets at all the potential relays, reducing the throughput performance. Hence, in our design, FO packets are transmitted at lower rates to ensure a reliable transmission.

2) Increased Overhead: As it is the case with any opportunistic scheme, d-AdaptOR adds a modest additional overhead to the standard 802.11 due to the added acknowledgment/hand-shake structure. This overhead increases linearly with the number of neighbors. Assuming a 802.11b physical layer operating at 11 Mb/s with an SIFS time of 10 μs, preamble duration of 20 μs, Physical Layer Convergence Protocol (PLCP) header duration of 4 μs, and 512-B frame payloads, Table II compares the overhead in the data packet due to piggybacking and the control overhead due to ACK and FO packets for unicast 802.11, genie-aided opportunistic scheme, and d-AdaptOR. d-AdaptOR requires communication overhead of 4 extra bytes (for EBS) per ACK packet compared to the genie-aided opportunistic scheme, while unicast 802.11 does not require such overhead.

Note that the overhead cost can be reduced by restricting the number of nodes in the candidate list of MAC header to a given number, MAX-NEIGHBOUR. The unique ordering for the nodes in the candidate set is determined by prioritizing the nodes with respect to $\Lambda(i\{i j\} j) j \in \mathcal{N}(i)$ and then

TABLE II. OVERHEAD COMPARISONS.

	Data Frame	Control packets	Total
802.11	397 μs	40 μs (ACK)	437 μs
Genie-aided opportunistic scheme	400 μs	115 μs + 40 μs (ACK+FO)	555 μs
d-AdaptOR	400 μs	124 μs +40 μs (ACK+FO)	564 μs

choosing the MAX-NEIGHBOUR highest priority nodes.9 Such a limitation will sacrifice the diversity gain and, hence, the performance of any opportunistic routing algorithm for lower overhead. In practice, we have seen that limiting the neighbor set to 4 provides most of the diversity gain.

VI. SIMULATIONS

In this section, we provide simulation studies in realistic wire-less settings where the theoretical assumptions of our study do not hold. These simulations not only demonstrate a robust performance gain under d-AdaptOR in a realistic network, but also provide significant insight in the appropriate choice of the design parameters such as damping sequence $\{\alpha_n\}$, delivery re-ward R, etc. We first investigate the performance of d-AdaptOR with respect to the design parameters and network parameters in a grid topology of 16 nodes. We then use a realistic topology of 36 nodes with random placement to demonstrate robustness of d-Adaptor to the violation of the analytic Assumptions 1 and 2.

A. Simulation Setup

In Sections VI-B and VI-C, using the appropriate choice of the design parameters, we compare the performance of d-AdaptOR against suitably chosen candidates. As a bench-mark, when appropriate, we have compared the performance against a genie-aided policy that relies on full network topology information when selecting routes. This is nothing but π^* dis-cussed in Section IV-B. We also compare against Stochastic Routing (SR) [1] (SR is the distributed implementation of policy π^*) and ExOR [4] (an opportunistic routing policy with ETX metric) in which the empirical probabilistic structure of the network is used to implement opportunistic routing algorithms. As a result, their performance will be highly de-pendent on the precision of empirical probability associated with link p_{ij} . To provide a fair comparison, we have considered simple greedy versions of SR and ExOR. These algorithms adapt $\{p_{ij}\}$ to the history of packet reception outcomes and rely on the updates to make routing decisions assuming error-free $\{p_{ij}\}$. We have also compared our performance against a conventional routing SRCR [21] with full knowledge of topology. In this setting, a conventional route is selected with perfect knowledge of link success probability at any given node. This comparison in effect provides a simple benchmark for all learning -based conventional routing policies in the literature such as Q-routing [10] and predictive Q-routing [12] when congestion is taken to be small enough (such that finding least congested paths coincides with finding the path with minimum expected number of transmissions).

Our simulations are performed in QualNet. We consider two sets of topologies in our experimental study.

1) Grid Topology: In Section VI-B, we study a grid topology consisting of 16 indoor nodes such that the nearest neighbors are separated by distance L meters. If unspecified, L is chosen to be 25 m. The source and the destination are chosen at the maximal distance (on diagonal) from each other.

2) Random Topology: In Section VI-C, we study a random topology consisting of 36 indoor nodes placed in an area of 150×150 m^2 . Here, we investigate the performance under a multisource multidestination setting as the number of flows in the network is varied and each flow is specified

via a randomly selected pair of source and destination. The nodes are equipped with 802.11b radios placed in indoor environment transmitting at 11 Mb/s with transmission power 15 dBm. Note that the choice of indoor environment is motivated by the findings in [22], where opportunistic routing is found to provide significant diversity gains. The wireless medium model includes Rician fading with K-factor of 4 and log -normal shadowing with mean 4 dB. The path loss follows the two-ray model in [23] with path exponent of 3. The acknowledgment packets are short packets of length 24 B transmitted at 11 Mb/s, while FO packets are of length 20 B and transmitted at a lower rate of 1 Mb/s to ensure reliability. If unspecified, packets are generated according to a constant bit rate (CBR) source with rate 20 packets/s. The packets are assumed to be of length 512 B equipped with simple cyclic redundancy check (CRC) error detection. The cost of transmission is assumed to be one unit, and the reward R is set to 40.

We have chosen $\alpha_1 = \frac{1}{\sqrt{n}}$ as the exploration parameter of choice.

B. Effects of Design and Network Parameters

Here, we investigate the role and criticality of various design parameters of d- AdaptOR with respect to the expected number of transmission criterion. Let us start with design parameters $\{\alpha_n\}$ and R.

1) Exploration Parameter Sequence $\{\alpha_n\}$:

The convergence rate of stochastic recursion (2) depends strongly on the choice of gence is fast but results in large variance in the estimates of

{ n n l• n }

d-AdaptOR is slower to adapt to the optimal performance it shows a slightly smaller variance. This is because the choice of $\{\alpha_n\}$ controls the rate with which greedy versus (randomly chosen) exploration actions are utilized. The optimization of the choice of $\{\alpha_n\}$ is an interesting topic of study in stochastic approximation [24], [25], far beyond the scope of this work.

2) Per-Packet Delivery Reward R: To ensure an acceptable performance of d-AdaptOR, the value of delivery reward, R, must be chosen sufficiently high. This would ensure the existence of routes under which the value of delivering a packet (as represented in R) is worth (i.e., larger than) the cost of

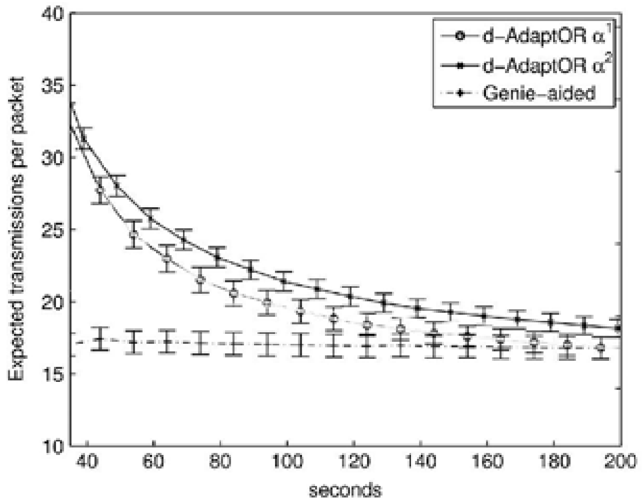

Fig. 3. Comparison for

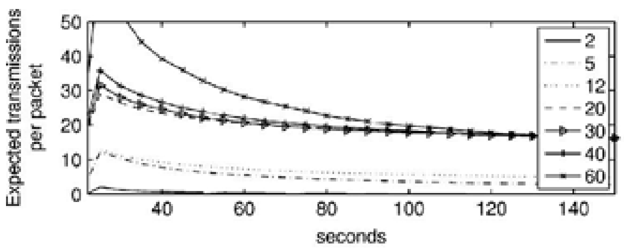

Fig. 4. Expected number of transmissions versus time as is varied.

relaying and routing that packet. A reasonable choice of R is any value larger than the worst-case expected transmission cost. Increasing R beyond such a value does not affect the asymptotic optimality of the algorithm. Next, we study the performance of d-AdaptOR with respect to the convergence rate and delivery ratio.

Fig. 4 plots the expected number of transmissions rate as time progresses for various values of R. As seen in Fig. 4, if R increases beyond a threshold R^0 (in the example provided here, this threshold is 18, but in general it depends on the network diameter), the expected number of transmissions per packet achieve the optimal value of R^0. In contrast, for $R<R^0$, the expected number of transmissions approaches zero as the packets not worth obtaining routing reward are dropped.10 Fig. 4 also shows that the convergence rate of the expected number of transmissions for routing per packet under d-AdaptOR decreases as R increases. The slow convergence for $R>R^0$ for large R is due to the flexibility of exploring longer paths. The slow convergence to zero for $R<R^0$ near R^0 is attributed to the fact that it takes a longer time for d-AdaptOR to realize that the packet is not worth relaying.

Fig. 5 plots the delivery ratio as R is varied. Fig. 5 shows that as R increases beyond a threshold R^0, the delivery ratio remains fixed. However, for sufficiently small R, nearly all the packets are dropped as the cost of transmission of the packet as well as relaying is not worth the obtained delivery reward. Due to very

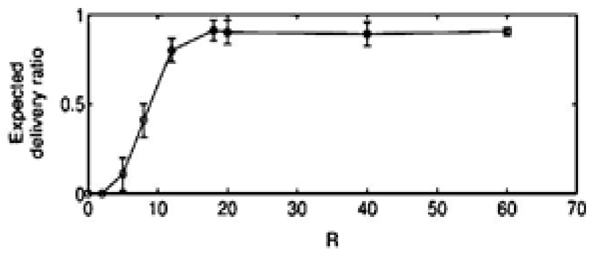

Fig. 5. Delivery ratio as is varied.

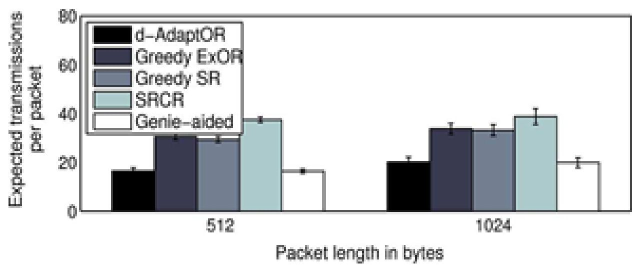

Fig. 6. d-AdaptOR performance as packet length is varied.

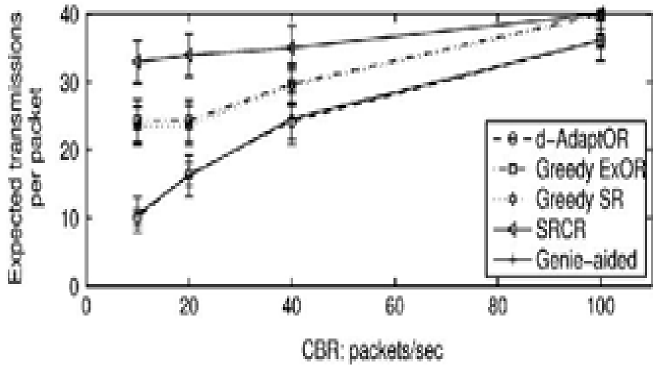

Fig. 7. Performance of d-AdaptOR as CBR traffic is varied.

slow convergence rate around R^0 for $R<R^0$, we observe that a non negligible number of packets is delivered in the duration of experiment.

Next, we investigate the performance of d-AdaptOR with respect to other candidate protocols for the network parameters such as packet length, traffic rate, neighbor distance, and time-varying costs.

3) Packet Length: We have repeated our simulations for 1024-B packets. Fig. 6 plots the performance as the packet length is varied from 512 to 1024 B. Note that due to the de-creasing packet transmission reliabilities, the expected routing cost per packet is increased with the packet size. However, the optimality of d-AdaptOR does not depend on the packet length.

4) Traffic Rate: Fig. 7 plots the mean number of transmissions versus CBR rate for candidate algorithms. Even though the performance gain for d-AdaptOR decreases somewhat with increase in the load, there is always a non negligible advantage over greedy solutions.

5) Average Hop Length L: In an attempt to understand the performance gap between various opportunistic algorithms, specifically the gap between d-AdaptOR versus learning-based conventional routing algorithms [10]–[13] whose performance is bounded by SRCR, one needs to gain insight about the diver-sity gain achieved by opportunistic routing. Fig. 8 compares the expected transmission cost for the three opportunistic routing algorithms (d-AdaptOR, ExOR, and SR) and SRCR as the

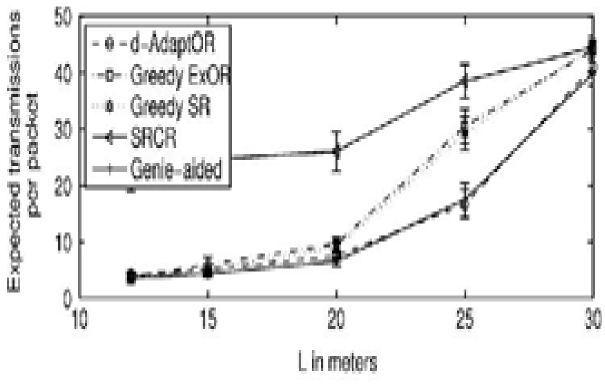

Fig. 8. Small hops provide significant receiver diversity gain.

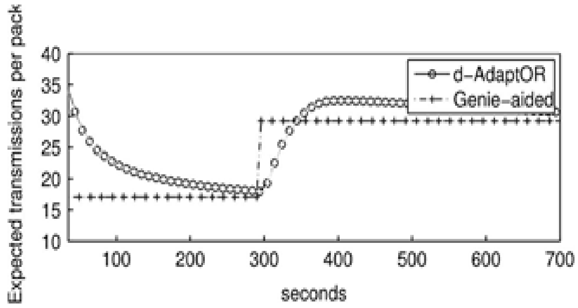

Fig. 9. Time-varying cost: Nodes go into sleep mode at time 300 s.

distance between the neighboring nodes in the grid topology, measured in L meters, is varied from 10 to 30 m. Note that for high values of L, the receiver diversity is low due to retransmission packet losses giving nearly similar performance for candidate protocols, while small L corresponds to a network with large receiver diversity gain. As expected, when L is small, all opportunistic routing schemes provide a significant improvement over conventional routing, but perhaps what is more interesting is the performance gain of learning-based d-AdaptOR over the greedy-based solutions in medium ranges.

6) Time-Varying Cost: In our analytical setup, we assume the transmission costs are fixed. Next, we discuss a simple scenario where the nodes have time-varying transmission costs. Consider a network in which nodes may go into an energy-saving mode when they do not participate in routing (e.g., to recharge their energy sources). Assume that upon entering the energy-saving mode, a node announces a high cost of trans-mission (100 instead of usual transmission cost of 1). Fig. 9 plots the expected average cost of d-AdaptOR when two nodes at the center of the grid move into an energy-saving mode. It shows that d-AdaptOR can track the genie-aided solution after the nodes move into the energy-saving mode.

C. Case Study: Random Network

Here, we study a random network scenario consisting of 36 wireless nodes placed randomly, with the remaining parameters kept the same as the default parameters.

Fig. 10 plots the expected number of transmissions and the expected average per-packet reward for the candidate routing algorithms versus network operation time when a single flow is present in the random topology. We fi rst note that, as expected, SRCR performs poorly compared to the opportunistic schemes as it fails to utilize the receiver diversity gain. This underlines our contribution over all existing learning-based solutions [10]–[13] that ignore receiver diversity. Furthermore,

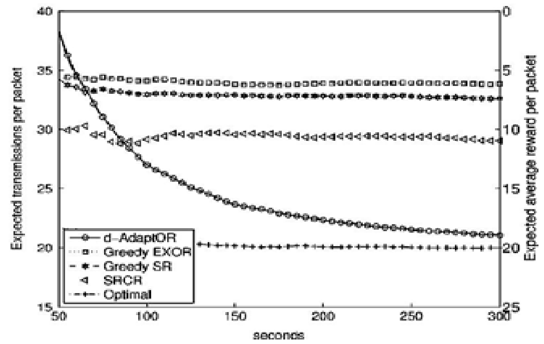

Fig. 10. Expected number of transmissions and average per-packet reward as function of operation time.

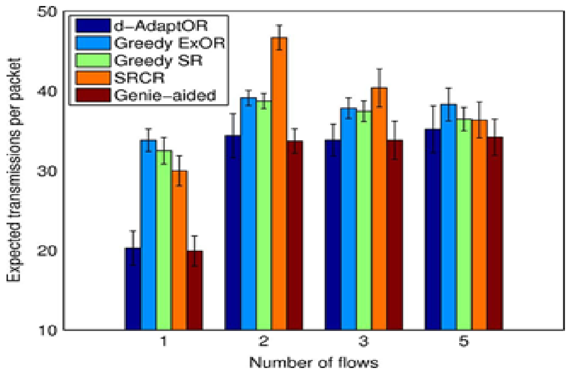

Fig. 11. d-AdaptOR versus distributed SR, ExOR, and SRCR performance for multiple flows.

Fig. 10 shows that the d-AdaptOR algorithm outperforms the greedy opportunistic schemes given sufficient number of packet deliveries. This is because the greedy versions of SR and ExOR fail to explore possible choices of routes and often result in strictly suboptimal routing policies. Fig. 10 also shows that the randomized routing decisions employed by d - AdaptOR work as a double-edged sword. On the one hand, they form a mechanism through which network opportunities are exhaustively explored until the globally optimal decisions are constructed, resulting in an improved long-term performance while these randomized decisions lead to a short-term performance loss. This, in fact, is reminiscent of the well-known exploration/exploitation tradeoff in stochastic control and learning literature.

Next, we study the performance of d-AdaptOR as the number of flows in the network is varied, where each flow is specified via a randomly selected pair of source and destination. Fig. 11 plots the expected number of transmissions and expected average reward for the candidate routing algorithms for the random topology. As seen in Fig. 11, d-AdaptOR maintains an optimal performance. However, Fig. 11 also shows that the gap between d-AdaptOR and the greedy version of SR significantly decreases with an increase in number of flows where the natural pattern of traffic flow renders the (randomized) exploration phase less critical. In other words, while Fig. 11 is consistent with the Remark 1 in Section II regarding the decomposition of multiple flow scenario to multiple single-flow scenarios, it also suggests that a joint design in which the multiplicity of flows provide a natural (and greedy) exploration of the network might be beneficial with regard to the transient/short-term performance measures of interest.

VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed d-AdaptOR, a distributed, adaptive, and opportunistic routing algorithm whose performance is shown to be optimal with zero knowledge regarding network topology and channel statistics. More precisely, under idealized assumptions, d-AdaptOR is shown to achieve the performance of an optimal routing with perfect and centralized knowledge about network topology, where the performance is measured in terms of the expected per-packet reward. Furthermore, we show that d-AdaptOR allows for a practical distributed and asynchronous 802.11 compatible implementation, whose performance was investigated via a detailed set of QualNet simulations under practical and realistic networks. Simulations show that d - AdaptOR consistently outperforms existing adaptive routing algorithms in practical settings.

The long-term average reward criterion investigated in this paper inherently ignores the short-term performance. To cap-ture the performance of various adaptive schemes, however, it is desirable to study the performance of the algorithms over a finite horizon. One popular way to study this is via measuring the incurred "regret" over a finite horizon. Regret is a function of horizon N that quantifies the loss of the performance under a given adaptive algorithm relative to the performance of the topology-aware optimal one. More specifically, our results so far implies that the optimal rate of growth of regret is strictly sublinear in N, but fails to provide a conclusive understanding of the short- term behavior of d-AdaptOR. An important area of future work comprises developing adaptive algorithms that en-sure optimal growth rate of regret.

The design of routing protocols requires a consideration of congestion control along with the throughput performance [26], [27]. Our work, however, does not consider this closely related issue. Incorporating congestion control in op-portunistic routing algorithms to minimize expected delay without the topology and the channel statistics knowledge is an area of future research.

REFERENCES

[1] C. Lott and D. Teneketzis, "Stochastic routing in ad hoc wireless net-works," in Proc. 39th IEEE Conf. Decision Control, 2000, vol. 3, pp. 2302–2307, vol. 3.

[2] P. Larsson, "Selection diversity forwarding in a multihop packet radio network with fading channel and capture," Mobile Comput. Commun. Rev., vol. 2, no. 4, pp. 47–54, Oct. 2001.

[3] M. Zorzi and R. R. Rao, "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: Multihop performance," IEEE Trans. Mo-bile Comput., vol. 2, no. 4, pp. 337–348, Oct.–Dec. 2003.

[4] S. Biswas and R. Morris, "ExOR: Opportunistic multi-hop routing for wireless networks," Comput. Commun. Rev., vol. 35, pp. 33–44, Oct. 2005.

[5] S. Jain and S. R. Das, "Exploiting path diversity in the link layer in wireless ad hoc networks," in Proc. 6th IEEE WoWMoM, Jun. 2005,pp. 22–30.

[6] C. Lott and D. Teneketzis, "Stochastic routing in ad hoc networks," IEEE Trans. Autom. Control, vol. 51, no. 1, pp. 52–72, Jan. 2006.

[7] E. M. Royer and C. K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," IEEE Pers. Commun., vol. 6, no. 2,pp.46–55, Apr. 1999.

[8] T. Javidi and D. Teneketzis, "Sensitivity analysis for optimal routing in wireless ad hoc networks in presence of error in channel quality es-timation," IEEE Trans. Autom. Control, vol. 49, no. 8, pp. 1303–1316, Aug. 2004.

[9] J. N. Tsitsiklis, “Asynchronous stochastic approximation and Q-learning,” in Proc. 32nd IEEE Conf. Decision Control, Dec. 1993, vol. 1, pp. 395–400.
[10] J. Boyan and M. Littman, “Packet routing in dynamically changing networks: A reinforcement learning approach,” in Proc. NIPS, 1994,pp. 671–678.
[11] J. W. Bates, “Packet routing and reinforcement learning: Estimating shortest paths in dynamic graphs,” 1995, unpublished.
[12] S. Choi and D. Yeung, “Predictive Q-routing: A memory-based rein-forcement learning approach to adaptive traffic control,” in Proc. NIPS, 1996, pp. 945–951.
[13] S. Kumar and R. Miikkulainen, “Dual reinforcement Q-routing: An on-line adaptive routing algorithm,” in Proc. Smart Eng. Syst., Neural Netw., Fuzzy Logic, Data Mining, Evol. Program., 2000, pp. 231–238.
[14] S. S. Dhillon and P. Van Mieghem, “Performance analysis of the AntNet algorithm,” Comput. Netw., vol. 51, no. 8, pp. 2104–2125, 2007.
[15] P. Purkayastha and J. S. Baras, “Convergence of Ant routing algorithm via stochastic approximation and optimization,” in Proc. IEEE Conf. Decision Control, 2007, pp. 340–354.
[16] D. P. Bertsekas and J. N. Tsitsiklis, Neuro-Dynamic Programming. Belmont, MA: Athena Scientific, 1996.
[17] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, “Trading structure for randomness in wireless opportunistic routing,” in Proc. ACM SIG-COMM, 2007, pp. 169–180.
[18] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: Wiley, 1994.
[19] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Compu-tation: Numerical Methods. Belmont, MA: Athena Scientific, 1997.
[20] W. Stallings, Wireless Communications and Networks, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2004.
[21] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, “Architecture and eval-uation of an unplanned 802.11b mesh network,” in Proc. ACM Mo-biCom, Cologne, Germany, 2005, pp. 31–42.
[22] M. Kurth, A. Zubow, and J. P. Redlich, “Cooperative opportunistic routing using transmit diversity in wireless mesh networks,” in Proc. IEEE INFOCOM, Apr. 2008, pp. 1310–1318.
[23] J. Doble, Introduction to Radio Propagation for Fixed and Mobile Communications. Boston, MA: Artech House, 1996.
[24] S. Russel and P. Norvig, Artificial Intelligence: A Modern Approach,
2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2003.
[25] R. Parr and S. Russell, “Reinforcement learning with hierarchies of machines,” in Proc. NIPS, 1998, pp. 1043–1049.
[26] P. Gupta and T. Javidi, “Towards throughput and delay optimal routing for wireless ad-hoc networks,” in Proc. Asilomar Conf., Nov. 2007, pp. 249–254.
[27] M. J. Neely, “Optimal backpressure routing for wireless networks with multi-receiver diversity,” in Proc. CISS, Mar. 2006, pp. 18–25.
[28] L. Breiman, Probability. Philadelphia, PA: SIAM, 1992.
[29] S. Resnick, A Probability Path. Boston, MA: Birkhuser, 1998.