Opportunistic Routing under Unknown Stochastic Models

Pouya Tehrani†, Qing Zhao†, Tara Javidi‡
† University of California, Davis, CA 95616, {potehrani,qzhao}@ucdavis.edu
‡University of California San Diego, La Jolla, CA 92093 USA, tjavidi@ucsd.edu

***Abstract*—We consider opportunistic routing in wireless ad-hoc networks under an unknown probabilistic local broadcast model. The objective is to design online learning algorithms that govern the sequential selection of relaying nodes based on the realizations of the probabilistic wireless links. The performance measure of interest is regret, defined as the expected additional cost accumulated over time when compared with the optimal centralized opportunistic routing policy under a known model of the wireless links. We propose both centralized and distributed online learning algorithms that achieve the optimal logarithmic regret order.**

***Index Terms*—Opportunistic routing, multi-armed bandit, regret, distributed algorithms, cognitive radio.**

I. Introduction

Classic routing protocols aim to find a fixed path (under a specific link metric). It fails to take advantages of the broadcast nature and the opportunities provided by the wireless medium, thus resulting 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. This mitigates the impact of poor wireless links by exploiting the broadcast nature of wireless transmissions and the path diversity.

The authors in [1] 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 an index, i.e. a distance-vector summarizing the expected-cost-to-forward from the neighbors to the destination. This "index" is shown to be computable in a distributed manner and with low complexity using the probabilistic description of wireless links. This index induces a rank ordering of all the nodes in the network, and at each time, the optimal action is to let the node with the highest rank among all nodes that have received the packet to relay. The study in [1] provided a unifying framework for existing opportunistic routing schemes such as [2], Geographic Routing and Forwarding (GeRaF) [3] and EXOR [4].

The opportunistic algorithms proposed in [1]–[4] depend on a precise probabilistic model of the wireless links in the network. In practice, however, it is not realistic to assume that the full knowledge of the wireless medium is available *a priori*. Authors in [5] considered this problem under the unknown probabilistic model and proposed an adaptive opportunistic routing (AdaptOR) scheme which is shown to minimize the expected long-run average routing cost per packet. In particular, over a horizon sufficiently long, AdaptOR is shown to achieve the same per packet expected average routing cost of the opportunistic routing algorithms proposed in [1], despite zero or erroneous knowledge about the underlying probabilistic model and/or local topology of the network.

The metric of long-run average cost per packet, however, does not reveal the convergence rate of the performance under an unknown model to that of a known model. A finer performance measure is the so-called regret, defined as the total expected additional cost accumulated over a horizon of T packets when compared to the optimal routing algorithm under a known model as given in [1]. All learning algorithms with a regret growing at a sublinear order with T achieve the same optimal long-run average performance. The difference in their total expected routing cost, however, can be arbitrarily large as the horizon length T increases. Regret thus not only indicates whether the optimal average performance under a known model is achieved, but also measures the convergence rate of the average performance of the learning algorithm, or the effectiveness of learning. The minimization of the regret growth rate is of great interest, and is the focus of this paper.

A centralized version of the above problem can be directly cast as a classic multi-armed bandit (MAB) problem. In the classic MAB [6], there are N independent arms. At each time, a player needs to decide which arm to play. An arm, when played, incurs i.i.d. random cost drawn from an unknown distribution. The performance of a sequential arm selection policy is measured by regret, defined as the total additional cost over a time horizon of length T when compared to an omniscient player who knows the cost model and always plays the best arm. It has been shown by Lai and Robbins that the minimum regret growth rate is logarithmic with time [6]. Since arms are assumed independent under the classic model, observations from one arm do not provide information about other arms. The optimal regret thus grows linearly with the number of arms.

It is not difficult to see that the optimal opportunistic routing policy developed in [1] can be considered as a classic multi-armed bandit (MAB) problem by treating each rank ordering of the nodes as an arm. Consequently, any MAB policy can be applied to achieve centralized learning of the optimal node

[0]The work of P. Tehrani and Q. Zhao was supported by the Army Research Office under Grant W911NF-12-1-0271 and by the Army Research Lab under Grant W911NF-09-D-0006. The work of T. Javidi was partially supported by the industrial sponsors of UCSD Center for Wireless Communication (CWC), L3 Communications Inc, and NSF Grants AST-1247995.

978-1-4673-3146-3/13/$31.00 ©2013IEEE

ordering with the optimal logarithmic regret growth rate. Such an approach, however, leads to poor regret order with respect to the size of the network measured by the number N of nodes. Since there are $N!$ possible rank ordering of nodes, regret thus grows at the order of $N! \sim O(N^N)$ with the network size, which is higher than exponential. When the wireless links in the network are independent, the total number of unknowns, however, is only in the order of N^2 (i.e, the number of links). This sharp contrast demonstrates the inefficiency of the direct approach of treating each rank ordering of nodes as an arm. This direct mapping to MAB also does not allow distributed implementation. In particular, all classic MAB policies rely on the number of times that each arm has been played to balance the tradeoff between exploration and exploitation. In a distributed setting where each node only interacts and observes its neighbors, an individual node does not have the global information on how many times a specific rank ordering of all the nodes in the network has been tried.

In this paper, we propose both centralized and distributed learning algorithms for opportunistic routing under an unknown wireless broadcast model. Rather than focusing on each global rank ordering of the network as in the direct mapping approach discussed above, the proposed learning algorithms focus on local learning at each node to allow better regret scaling with the network size as well as distributed implementation. The proposed learning algorithms achieves a polynomial regret order with respect to the number of unknowns while maintaining the optimal logarithmic regret order with time.

The same problem was also considered in [7], where it was shown that the optimal logarithmic regret order can be achieved by letting each node send probing packets to its neighbors for learning the wireless links. This approach is fundamentally different from the learning algorithms developed in this paper that allow transmissions of only information-bareing packets. Other related work includes [8], [9], which considers centralized learning of the shortest path routing (i.e., the conventional fixed-path routing approach), thus differs from this paper in the problem scope as well as the specific proposed learning schemes.

II. Problem Statement

We consider the problem of routing packets from a source node to a destination node in a wireless ad-hoc network of N nodes denoted by $\Omega = \{1, 2, \ldots, N\}$ with node 1 being the source node and N the destination node. Time is slotted and indexed by $n \geq 0$ (this assumption is not technically critical and is only assumed for ease of exposition). A packet indexed by $t \geq 1$ is generated at the source node 0 at time σ_t according to an arbitrary distribution with rate $\lambda > 0$.

We adopt a general probabilistic local broadcast model. Specifically, the wireless links at node i are characterized by a probability distribution $\mathbf{P}_i \triangleq \{P(S|i)\}_{S \subseteq \mathcal{N}(i)}$, where $\mathcal{N}(i)$ denotes the neighbor set of i including i itself and $P(S|i)$ the probability that all nodes in S and only nodes in S receive the transmission from i. Note that for all $S \neq S'$, successful reception at S and S' are mutually exclusive and $\sum_{S \subseteq \mathcal{N}(i)} P(S|i) = 1$. Furthermore, node i is always a recipient of its own transmission, thus $P(S|i) = 0$ if $i \notin S$. It is easy to see that this general model allows dependencies across the wireless links rooted at i.

We assume a fixed transmission cost $c_i > 0$ is incurred upon a transmission from node i. Transmission cost c_i can be considered 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 setting 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 i to the set of neighbor nodes S, the next (possibly randomized) routing decision includes 1) retransmission by node i, 2) relaying the packet by a node $j \in S$, or 3) dropping the packet altogether. If node j is selected as a relay, then it transmits the packet at the next slot, while other nodes $k \neq j, k \in S$, expunge that packet.

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 define termination time τ_t to be the stopping time when packet t is terminated. We discriminate amongst the termination events as follows: we assume that upon the termination of a packet at the destination (successful delivery of a packet to the destination), a fixed and given positive delivery reward R is obtained, while no reward is obtained if the packet is terminated before it reaches the destination. Let r_t denote this random reward obtained at the termination time τ_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.

Let $i_{n,t}$ denote the index of the node which transmits packet t at time n. The routing scheme can be viewed as selecting a (random) sequence of nodes $\{i_{n,t}\}$ for relaying packets $t = 1, 2, \ldots$. As such, the total expected cost (minus reward) associated with routing T packets along a sequence of $\{i_{n,t}\}$ up to the termination time of packet T is:

$$J_T = \mathbb{E}\left[\sum_{t=1}^{T}\left\{\sum_{n=\sigma_t}^{\tau_t - 1} c_{i_{n,t}} - r_t\right\}\right], \tag{1}$$

where the expectation is taken over the events of transmission decisions, successful packet receptions, and packet generation times. The regret U_T is thus obtained by considering the performance achieved when the underlying probabilistic local broadcast model of the network is known perfectly. In other words,

$$U_T = J_T - TV^*(1),$$

where $V^*(1)$ denotes the optimal cost (minus reward) of delivering a packet from source node 1 to the destination under the perfect knowledge of network topology and the underlying local broadcast model.

Problem (P) Choose a sequence of relay nodes $\{i_{n,t}\}$ in the absence of knowledge about the network topology to minimize the order of the regret U_T with respect to T.

In the next sections we propose centralized and distributed algorithms which solve Problem **(P)**. The nature of the algorithms allow nodes to make routing decisions in distributed, asynchronous, and adaptive manner while optimally balancing the exploration and exploitation costs.

III. Centralized opportunistic Routing

A. The Optimal Centralized Policy under A Known Model

When the probabilistic model $\mathbf{P}_i$ is known for all $i \in \Omega$, it is shown in [1] that the optimal routing policy is of an index type: each node is associated with an index that summarizes the expected-cost-to-forward from this node to the destination, and at each time, the node with the smallest index among all nodes that have received the packet is chosen to relay the packet (or to retire). It is shown in [1] that the index $V^*(i)$ of node i is the unique solution $V^* : \Omega \to \mathbb{R}^+$ to the following fixed-point equation:

$$\begin{aligned} V^*(N) &= -R \\ V^*(i) &= \min\{-r_i, \{c_i + \sum_S P(S|i)(\min_{j\in S} V^*(j))\}\}. \end{aligned}$$

It is easy to see that the index V^* leads to a ranking of the nodes in the network regarding their priority in serving as relays. Noticing that the computation of $V^*(i)$ only requires the indexes of those nodes with higher rank (i.e., smaller indexes) than node i, the authors of [1] proposed a Dijkstra-type algorithm for solving the above fixed-point equation.

B. Centralized Opportunistic Routing with Learning

In this section, we propose a centralized learning algorithm for opportunistic routing under an unknown probabilistic link model $\{\mathbf{P}_i\}_{i\in\Omega}$. Referred to as ORL (Opportunistic Routing with Learning), this algorithm partitions the sequence of packets generated at the source into two types: the exploration packets and the exploitation packets. The exploration packets are routed through the currently least traversed node in the network to ensure sufficient learning of all links in the network. The exploitation packets are routed through a sequence of opportunistic relaying nodes calculated based on the estimated link success probabilities $\{\hat{\mathbf{P}}_i\}_{i\in\Omega}$. Specifically, let $\mathcal{E}(t)$ denote the index set of the exploration packets up to (and possibly including) the tth packet. Let $n_i(t)$ denote the number of times that node i has served as a relaying node for the packets in $\mathcal{E}(t)$. Define

$$l(t-1) \stackrel{\Delta}{=} \min_{i\in\Omega-\{N\}} n_i(t-1) \tag{2}$$

as the least traversed node before the transmission of packet t. Consider packet t. If $t \notin \mathcal{E}(t)$, then packet t is routed opportunistically based on the current estimated priority indexes of all nodes (see below). If $t \in \mathcal{E}(t)$, packet t is routed to node $l(t-1)$ through a route with the least hop count. Node $l(t-1)$ then relays the packet to the destination (if the destination has not received the packet) by treating the packet as a regular exploitation packet. After the transmission of each exploration packet, the central controller updates $n_i(t-1)$. For each node i that served as a relay for the transmission of packet t, the central controller also updates the estimate of the local broadcast model of node i as follows:

$$\hat{P}(S|i) = \frac{\sum_{t\in\mathcal{E}(t)} 1_{S|i}(t)}{n(i)}, \tag{3}$$

where $1_{S|i}(t)$ denotes the event that all nodes and only nodes in S received the transmission of packet t from node i. Based on $\hat{P}(S|i)$, the central controller then computes the priority indexes (as given in [1]) of all nodes to be used for routing the next exploitation packet.

An important design parameter in ORL is the number of exploration packets in a sequence of T packets generated by the source. The cardinality of $\mathcal{E}(T)$ (denoted by $|\mathcal{E}(T)|$) balances the tradeoff between exploration and exploitation. It is not difficult to see that the regret order is lower bounded by $|\mathcal{E}(T)|$. Nevertheless, the sequence of exploration packets needs to be chosen sufficiently dense to ensure effective learning of $\{\mathbf{P}_i\}_{i\in\Omega}$ and subsequently, the opportunistic routing indexes. The key issue here is to find the minimum cardinality of the exploration packets that ensures the additional cost caused by incorrectly identified node routing priorities during the transmissions of the exploitation packets having an order no larger than $|\mathcal{E}(T)|$. As shown in the theorem below, $|\mathcal{E}(T)|$ can be set to a logarithmic order with T, leading to the optimal logarithmic regret order of the learning algorithm ORL.

Theorem 1: Let B be an upper bound on the node degree in the network (i.e., $|\mathcal{N}(i)| \le B, \forall i \in \Omega$). Define $\alpha = R\sum_{i=1}^{N}\frac{1}{c_i}$ and choose $\Delta \in (0, \min_{i\in\Omega}\{\min_{j\in\mathcal{N}(i)}|V^*(i) - V^*(j)|\})$. Then ORL has the following regret performance:

(1) *General local broadcast model with link dependencies*
Set $G = \frac{N\alpha^2 4^B}{2\Delta^2}$. For each packet $t > 1$, if $|\mathcal{E}(t-1)| < G\log t$, then include t in $\mathcal{E}(t)$. Under this sequence of exploration packets, policy ORL achieves regret $O(NG\log T)$ which is logarithmic with the number of packets and polynomial with the number of unknowns.

(2) *Independent wireless links*
Set $G = \frac{N\alpha^2 B^2}{2\Delta^2}$. For each packet $t > 1$, if $|\mathcal{E}(t-1)| < G\log t$, then include t in $\mathcal{E}(t)$. Under this sequence of exploration packets, policy ORL achieves regret $O(NG\log T)$ which is logarithmic with the number of packets and polynomial with the network size.

Proof: Omitted due to space limit. ■

IV. Distributed Opportunistic Routing

A. Distributed Opportunistic Routing under A Known Model

The optimal index policy described in section III-A can be implemented in a distributed manner once the optimal value of the indexes are computed and the node priorities are determined. This is because after each transmission at a node, say i, the next relay must be one of node i's neighbors that just received the packet. The selection of the next relay can thus be done through local communications among neighbors. A distributed implementation of the opportunistic routing policy thus only requires a distributed computation of the routing

indexes $V^*(i)$, which as shown in [10], can be done through recursive local exchanges and updates among neighbors.

B. Distributed Opportunistic Routing with Learning

We now consider a distributed implementation of the learning algorithm proposed in Sec. III-B. The basic structure of ORL allows distributed implementation: the classification of exploration and exploitation packets can be easily carried out at the source by adding a header to each exploration packet; the estimates of the link probabilities $\mathbf{P}_i$ are computed at each node i using local observations obtained during the transmissions of exploration packets; the opportunistic routing index of each node is then obtained based on the estimated $\mathbf{P}_i$ using the distributed algorithm given in [10]. The only difficulty is in finding the least traversed node in a distributed manner for routing each exploration packet. This can be solved by a distributed algorithm as detailed in Fig. 1. Specifically, through local exchanges of n_i (the number of exploration packets that each node i has relayed), the least traversed node and a path to reach this node from the source are identified for the transmission of the next exploration packet.

Note that we do not require a node knows whether an exploration packet has been delivered in order to start the distributed algorithms for finding the least traversed node and computing the routing index. The local information exchange for these algorithms will be initiated when a node sees a change in its current local information (e.g., an increase in n_i after relaying an exploration packet). Furthermore, the exploitation packets are routed based on the current routing index at each node without assuming the convergence in the distributed calculation of the indexes. As a consequence, the algorithm is fully distributed; each individual node does not need to maintain a global count on the number of exploration packets that have been delivered (except the source) or to know whether its current local information reflects convergence.

The theorem below gives the regret performance of the distributed learning algorithm.

Theorem 2: In a horizon of T packets if

$$|\mathcal{E}(t)| \geq G \log t, \tag{5}$$

where $G = \frac{N\alpha^2 B^2}{2\Delta^2}$, the distributed implementation of ORL (referred to as DORL) achieves regret $O(\log T)$ which is logarithmic with the number of packets.

Proof: Omitted due to space limit. ■

Finding the Least Traversed Node

- *Initialization:*

  Each node i ($i \in \Omega - \{N\}$) sets $n_j = m_j = 0$ for each $j \in \mathcal{N}(i)$.

- *Local Updates and Exchanges at Node i:*

  When any of the following events occurs:
  - n_i increases by one;
  - node i receives a new value of n_j from a neighbor j that is bigger than the current stored value of n_j;
  - node i receives a new value of m_j from a neighbor j that is bigger than the current stored value of m_j.

  then
  - update m_i and the exploration neighbor o_i:

$$\begin{aligned} m_i &\triangleq \min_{j\in\mathcal{N}(i)} n_j \\ o_i &\triangleq \arg\min_{j\in\mathcal{N}(i)} n_j \end{aligned} \tag{4}$$

  - send the new values (if changed) of n_i or m_i or both to neighbors.

  When the following event occurs:
  - node i receives a new value of m_j from a neighbor j that is smaller than the current stored values of both m_j and m_i.

  then
  - update m_i with the new value of m_j; set $o_i = j$.
  - send the new value of m_i to neighbors.

Fig. 1. A distributed algorithm for finding the least traversed node.

V. CONCLUSION

In this paper an opportunistic routing problem in an ad hoc wireless network is considered. A local broadcast model is assumed for transmission. Dynamic centralized and distributed learning algorithms are proposed for this problem under unknown local broadcast models.

REFERENCES

[1] C. Lott and D. Teneketzis, "Stochastic Routing in Ad-Hoc Networks", *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 52-72, January 2006.

[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. 4754, Oct. 2001.

[3] M. Zori and R. R. Rao, "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance", *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, 2003.

[4] S. Biswas and R. Morris, "ExOR: Opportunistic Multihop Routing for Wireless Networks", *ACM SIGCOMM Computer Communication Review*, vol. 35, pp. 3344, October 2005.

[5] A. Bhorkar, M. Naghshvar, T. Javidi and B. Rao, "An Adaptive Opportunistic Routing Scheme for Wireless Ad-hoc Networks", *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, January 2012.

[6] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules", in *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 422, 1985.

[7] A. Bhorkar and T. Javidi, " No Regret Routing for ad-hoc wireless networks", *in Proceedings of Asilomar Conference on Signals, Systems, and Computers*, November 2010.

[8] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards and Individual Observations", *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, 2012.

[9] K. Liu and Q. Zhao, "Adaptive Shortest-Path Routing under Unknown and Stochastically Varying Link States", *in Proc. of the 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May, 2012.

[10] C. Lott, "Stochastic Routing in Ad Hoc Networks", Rep. TR-362, Univ. Michigan, Ann Arbor, MI. Available:www.eecs.umich.edu/systems/TechReportsList.html, May 2005.