Random Walks and Green's Function on Digraphs: A Framework for Estimating Wireless Transmission Costs

Yanhua Li, *Student Member, IEEE*, and Zhi-Li Zhang, *Fellow, IEEE*

***Abstract*—Various applications in wireless networks, such as routing and query processing, can be formulated as random walks on graphs. Many results have been obtained for such applications by utilizing the theory of random walks (or spectral graph theory), which is mostly developed for undirected graphs. However, this formalism neglects the fact that the underlying (wireless) networks in practice contain asymmetric links, which are best characterized by directed graphs (*digraphs*). Therefore, random walk on *digraphs* is a more appropriate model to consider for such networks. In this paper, by generalizing the random walk theory (or spectral graph theory) that has been primarily developed for undirected graphs to digraphs, we show how various transmission costs in wireless networks can be formulated in terms of hitting times and cover times of random walks on digraphs. Using these results, we develop a unified theoretical framework for estimating various transmission costs in wireless networks. Our framework can be applied to random walk query processing strategy and the three routing paradigms—*best path routing*, *opportunistic routing*, and *stateless routing*—to which nearly all existing routing protocols belong. Extensive simulations demonstrate that the proposed digraph-based analytical model can achieve more accurate transmission cost estimation over existing methods.**

***Index Terms*—Digraph, random walk, spectral graph theory, transmission cost, wireless networks.**

I. Introduction

DUE TO the unique characteristics of wireless technologies and the dynamics in the environments (e.g., mobility and interference) they operate in, wireless channels are known to be time-varying, unreliable, and asymmetric [17], [21], [37]–[39]. Furthermore, wireless networks are often designed to support certain applications or missions, and deployed in specific environments. For these reasons, a plethora of wireless mechanisms—especially, routing algorithms and protocols—have been proposed and developed to achieve a range of different objectives such as throughput, latency, energy consumption, network lifetime, and so forth. Evaluating the efficacy of wireless protocols in terms of various transmission cost metrics, and deciding on which one to employ in a specific environment so as to attain certain performance objective, can be a challenging task in practice. The ability to analyze, estimate, and quantify various transmission costs is therefore imperative in the design of wireless networks.

While experimentation and testing in realistic wireless environments are indispensable and provide the most definite and authoritative means to evaluate the efficacy of wireless routing protocols, they are in general very expensive and are typically utilized in the later stage of the network design and evaluation process. Simulation-based evaluation is also important and necessary. However, conducting realistic simulations is hard, and simulation results often hinge on the settings and parameters used. We believe that analytical models and theories also play a critical role in the design of wireless networks, complementing the roles played by real-world experimentation and simulations. By generating performance bounds and theoretical limits, they provide important insights on what is achievable and under what conditions and produce useful metrics for understanding the key design tradeoffs. Such insights and understanding are particularly important in the early stage of wireless network design.

Guided by this belief, in this paper we develop a unified theoretical framework to quantify and estimate various transmission costs of wireless routing protocols. To account for the *stochastic* and *asymmetric* natures of wireless channels, we model a wireless network as a *directed* graph (in short, *digraph*), where each directed edge (link) is associated with a *packet delivery probability*. We consider three wireless routing paradigms, the (traditional) best path routing (e.g., AODV [32], DSR [20], and several energy-aware routing protocols [3], [7]), opportunistic routing (e.g., ExOR [4], MORE [6]), and stateless (stochastic) routing (e.g., as proposed in [9], [10], and [29])—nearly all existing routing protocols fall under one of these paradigms, or use a combination thereof. Under the (simplifying) assumption that packet delivery probabilities are independent, we demonstrate how packet forwarding under each paradigm can be modeled as a Markov chain on a digraph with an appropriately defined transition probability matrix capturing the specifics of the routing algorithm under consideration. In other words, the traversal of a packet being forwarded in a wireless network can be viewed as a *random walk on a digraph*. Consequently, various transmission costs of end-to-end packet delivery (e.g., the expected number of transmissions, end-to-end packet delivery ratio, throughput, latency, energy consumptions) can therefore be formulated using well-known notions such as *hitting times*, *sojourn times* associated with random walks.

The main contributions of this paper are summarized as follows.

Manuscript received October 01, 2010; revised October 30, 2011; accepted March 08, 2012; approved by IEEE/ACM Transactions on Networking Editor S. Subramaniam. Date of publication April 05, 2012; date of current version February 12, 2013. This work was supported in part by the US National Science Foundation under Grants CNS-0626808 and CNS-0905037, the DTRA under Grant HDTRA1-09-1-0050, and the University of Minnesota DTC under a DTI grant. An earlier version of this work appeared in the Proceedings of the IEEE International Conference on Computer Communications (INFOCOM), San Diego, CA, March 15–19, 2010.

The authors are with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: yanhua@cs.umn.edu; zhzhang@cs.umn.edu).

Digital Object Identifier 10.1109/TNET.2012.2191158

- By building on top of our previous work in [24], we utilize the random walk (Markov chain) model to formulate the end-to-end transmission costs for various types of wireless routing strategies.
- The theory of random walk (and the closely related *spectral graph theory*) has been developed primarily for *undirected* graphs (see, e.g., [11] and [28]). We successfully extend the theory of random walks on undirected graphs to *directed graphs (digraphs)*, with a more general definition of normalized (graph) Laplacian matrix $\tilde{\mathcal{L}}_G$. We also show how the hitting times, commute times, sojourn times (or hitting costs), and (partial) cover time can be computed using the Moore–Penrose pseudo-inverse of $\tilde{\mathcal{L}}_G$.
- Using three representative routing protocols (one from each routing paradigm) and a query processing protocol as examples, we systematically illustrate how our proposed theoretical framework based on random walks on digraphs can be used to estimate various transmission costs. Our analysis subsumes earlier results obtained using more *ad hoc* methods. We also perform extensive simulations to show the relative errors in estimation when asymmetric links are artificially symmetrized and undirected graphs are used.

The remainder of this paper is organized as follows. The related work is briefly touched on below. In Section II, we first describe three wireless routing paradigms. We then illustrate how packet forwarding under each of them can be modeled using Markov chains/random walks, and various transmission costs can be formulated using hitting times and hitting costs associated with the random walks. We outline the theory of random walks on undirected and its generalization to directed graphs in Section III. In Sections IV and V, we apply the theory of random walks on digraphs to three representative routing protocols and a query processing protocol. Simulation results are reported in Section VI, and the paper is concluded in Section VII.

Related Work: We will present the three wireless routing paradigms in Section II and discuss some related wireless protocols in that context. Hence we do not discuss them here. Many studies, e.g., [17], [38], and [39], have shown that the wireless channels are unreliable and asymmetric. Reference [38] provides both empirical and analytical results on the asymmetric and unreliable characteristics of wireless links and demonstrates that they can significantly impact the performance of applications. Ganesan *et al.* [17] show that the performance of simple flooding mechanism can be significantly affected due to asymmetric, dynamic long-distance links. Zhou *et al.* [39] report that unreliable and asymmetric wireless links have adverse impact on performance of wireless routing protocols. Digraphs are therefore a more appropriate model to study the wireless networks with *asymmetric links*. Our paper is partly inspired by the work in [9], where results from random walks on *undirected graphs* are used to model and derive a delay estimation formula for stateless routing with heterogenous sojourn time. In contrast, our paper not only supersedes the results in [9], which essentially assumes symmetric wireless links, but develops a general theoretical framework based on random walks on digraphs for estimating various transmission costs under all three wireless routing paradigms. As mentioned earlier, the theory of random walks has been developed primarily for undirected graphs. Relatively fewer attempts have been made to extend it to digraphs. In [12], Chung defines a *symmetrized* Laplacian matrix for directed graph and successfully generalizes the well-known Cheeger Inequality to directed graphs. However, it is unclear whether this generalization can be used to compute hitting times and commute times for random walks on digraphs. Random walks (on digraphs), or more generally irreversible Markov chains, have been a well-studied topic (see, e.g., [1], [19], and [35]) in probability theory. For instance, it is known the hitting times on digraphs can be computed as an infinite sum involving the transition probability or expressed in terms of the fundamental matrix [1]. Our work provides a new *spectral* perspective to study random walks in terms of the digraph Laplacian.

II. Wireless Routing, Transmission Costs, and Random Walks in Directed Graphs

In this section, we briefly describe the three wireless routing paradigms. We then show how packet traversals under each routing paradigm can be modeled using Markov chains, and we use the Markov models to estimate various transmission costs in a wireless network.

A. Wireless Routing and Transition Costs

The existing (unicast) wireless routing schemes can be roughly classified into three categories: the traditional *best path* routing, *opportunistic* routing, and *stateless (stochastic)* routing.

The traditional best path routing protocols (e.g., AODV [32], DSR [20], and their variations/extensions to multipath or energy-aware routing) typically select a single *best* path, sometimes multiple paths, based on certain routing metric. Unlike wired networks, these best paths are selected typically on-demand instead of precomputed. Depending on the objective of the routing protocols, different routing metrics may be designed and used. For example, if the goal is to maximize the packet delivery probability and minimize the number of transmission, the ETX metric [14] may be used, which captures the expected number of transmissions per link, and the best (least-cost) path is the path that minimizes the overall path ETX. If the objective is to minimize the energy consumption and maximize the network lifetime, an energy-aware metric should be used. For example, [8] has proposed a lifetime maximization algorithm for energy-aware routing in wireless sensor networks.

The key idea behind opportunistic routing is to take advantage of the broadcasting nature of wireless communication channels, while at the same time addressing the probabilistic nature of packet reception. Instead of selecting one or multiple *fixed* best paths, opportunistic routing protocols (e.g., [4] and [6]) specify a set of forwarders, often arranged in a prioritized list, referred to as a *forwarder list*. Using the (prespecified) forwarder list, after each packet transmission, the "best" forwarder among those that happen to receive the packet is used to forward the packet towards the destination. Hence, a packet may opportunistically traverse any path from the source, among the set of forwarders, to the destination instead of a fixed path. Through experiments in the MIT Roofnet [34] testbed, ExOR [4]—one of the first practical opportunistic routing protocols—is shown to increase the throughput by a factor of two to four over traditional best path routing schemes. Further improvements to ExOR [6], [22], [23], [26], [33] have also been developed. For instance, in [23] the key problem of how to optimally select the forwarder list is addressed, and an optimal algorithm (MTS) that minimizes the expected total number of transmissions is developed.

By its name, stateless (stochastic) routing does not maintain any routing state (e.g., topology, routing tables) and performs packet forwarding in a purely "random" fashion. In contrast to opportunistic routing, no forwarder list is prespecified in general; any node receiving a packet may decide to forward the packet (some mechanisms to avoid and reduce unnecessary duplicate transmissions are generally employed). Stateless (stochastic) routing is typically designed and best suited for resource-constrained, dynamically varying, and highly unreliable wireless network environments (e.g., sensor or delay/disruption-tolerant networks). For instance, several studies have suggested the stateless (stochastic) forwarding in wireless networks [15], [29] for its simplicity and scalability. A stateless routing protocol is developed in [29] for wireless sensor networks. Due to its stateless feature, stateless routing schemes do not involve control overhead, e.g., exchanging link-state information, thus are easy to implement. However, due to the pure randomness employed in these schemes, their efficacy, e.g., in terms of end-to-end packet delivery and other performance metrics, may suffer.

Thanks to widely disparate wireless network environments and diverse application objectives, no one routing paradigm always overperforms the others in practice. For instance, traditional best path routing may work very well in a static wireless environment with fairly stable and reliable wireless channels, while opportunistic routing may perform better where wireless channels are less reliable with frequently varying conditions. Hence, in the design of practical routing protocols for wireless networks, which routing paradigm (or a hybrid combination thereof) to use will depend critically on the specific wireless environment. The ability to analyze, estimate, and quantify various transmission costs (e.g., the expected number of transmissions, latency, or energy consumption) is therefore imperative in the design of wireless networks. In Section II-B, we illustrate how we can model the packet traversal in a wireless network under each of the wireless routing paradigms using Markov chains. Through these Markov chain models, transactions costs incurred by different routing schemes can then be computed using the notion of *hitting times* and other related quantities (e.g., sojourn times or hitting costs).

B. Modeling Packet Traversal Using Markov Chains

Here, we illustrate how we can model packet forwarding under each of the three routing paradigms using Markov chains. Due to the probabilistic nature of wireless transmissions, when a packet is forwarded from one node, say i, to another node, say j, it only has some probability to "transit" from node i to node j. This suggests that we could model and trace the traversal of a packet when it is forwarded from one node to another in a wireless network as state transitions in a Markov chain. Before we proceed to describe how packet forwarding under each routing paradigm can be modeled using Markov chains, we first present some general notations and basic assumptions.

We model a wireless network as a (weighted) *directed* graph (i.e., a *digraph*) $G = (V, E)$, where V is the set of wireless nodes, and E is the set of *directed* wireless links. Here, each directed link, $\langle i, j\rangle$, represents the relation that node j is within the transmission range of node i. In other words, a packet transmitted by node i may be received by node j with some probability. We denote this probability by a_{ij}. Hence, each link $\langle i, j\rangle$ is associated with a link weight a_{ij}. We will simply refer to a_{ij} as the (link-level) packet delivery probability. More generally, we associate a weight a_{ij} to any (ordered) pair of nodes, $\langle i, j\rangle$. If $\langle i, j\rangle \notin E$ (namely, node j is not within the transmission range of node i), we simply set $a_{ij} = 0$. Hence, for any two distinct nodes $i, j \in V$, $i \neq j$, we have $0 \leq a_{ij} \leq 1$, and $a_{ij} > 0$ if and only if $\langle i, j\rangle \in E$. Due to the asymmetric nature of wireless communications, in general we have $a_{ij} \neq a_{ji}$. In particular, we may have $a_{ij} > 0$, but $a_{ji} = 0$. Furthermore, for any node $i \in V$, we define $a_{ii} = 0$.

Let $n = |V|$ denote the total number of nodes in the wireless topology. Then, the $n \times n$ matrix, $A = [a_{ij}]$, gives us a matrix representation of (one-hop or link-level) packet delivery probabilities of a wireless network. In general, A is *asymmetric*. We call A the adjacency matrix of the (weighted) directed graph $G = (V, E)$. In modeling packet forwarding using Markov chains, we assume that for any $\langle i, j\rangle \in E$, when a packet is forwarded by node i, the probability that the packet is received by node j, i.e., a_{ij}, does not depend on where the packet was before reaching node i. Namely, except for node i, a_{ij} does not depend on who and where the previous forwarders are. In other words, we assume that the Markov property holds. In modeling opportunistic routing and stateless routing, we will also make the simplifying assumption that the (link-level) packet delivery probabilities are independent. More precisely, let $N(i) = \{j : \langle i, j\rangle \in E\}$ be the direct neighbors of node i that are within its transmission range. We assume that for any $j_1, j_2 \in N(i)$, a_{ij_1} and a_{ij_2} are independent. We remark that to model the time-varying dynamics of a wireless network, we can introduce a series of time-dependent graphs $G_t = (V_t, E_t, A_t)$ with time-varying node and edge sets as well as varying link-level packet delivery probabilities a_{ij}'s. For clarity and model simplicity, in this paper we focus only on *one instance* of such a time-varying graph and assume that during this instance except for a few of them, the node/edge sets and a_{ij}'s are largely unchanged. Finally, we assume that the digraph graph $G = (V, E)$ is *strongly* connected, namely, there exists a (directed) path from any node to any other node in G.

Best Path Routing: Consider a specific source–destination pair (s, d). Let $R(s, d) = \{u_0 = s, u_1, \ldots, u_m, u_{m+1} = d\}$ denote the route (i.e., a best path) selected by a best path routing protocol for forwarding packets from s to d. We use $G_R = (V_R, E_R) \subset G$ to denote the subgraph (a path or line subgraph) induced by R, where $V_R = \{u_i, 0 \leq i \leq m+1\}$, and $E_R = \{\langle u_i, u_{i+1}\rangle, 0 \leq i \leq m\}$. We can model the traversal of a packet being forwarded from s to d as a Markov chain with the state space V_R and the transition probability matrix $P_R = [p_{ij}]$ defined as follows:

$$p_{ij} = \begin{cases} a_{i,i+1}, & \text{if } j = i+1, i = 0, \ldots, m \\ 1 - a_{i,i+1}, & \text{if } j = i, i = 0, \ldots, m \\ 1, & \text{if } j = i, i = m+1 \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Using the wireless topology shown in Fig. 1 as an example, let $R_{s,d} = \{s, (u_1 =)v_2, (u_2 =)v_4, d\}$ be the best path (route) for

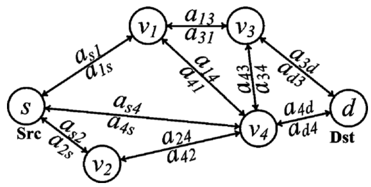


Fig. 1. Example wireless topology.

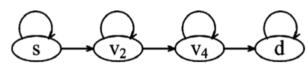


Fig. 2. Markov chain for best-path routing.

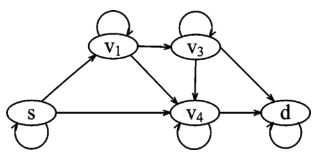


Fig. 3. Markov chain for opportunistic routing ($\mathrm{FL} = \{s, v_1, v_3, v_4, d\}$).

the source–destination pair (s, d). The corresponding Markov chain is schematically depicted in Fig. 2, where the arrows indicate the state transitions. The transition probability matrix P_R captures the fact that when a packet is forwarded by node u_i, $0 \le i \le m$, with probability $p_{i,i+1} = a_{i,i+1}$ the packet may be received by the next hop u_{i+1} (thus it transits or "walks" from node i to node $i+1$ with probability $p_{i,i+1}$), and with probability $p_{ii} = 1 - a_{i,i+1}$ it is not received by node $i+1$ (thus it stays with node i). Hence, packet forwarding under best path routing can be viewed as a *random walk* on the line subgraph G_R with P_R as the transition probability matrix. We note that this is an absorbing Markov chain, with node s as the starting state and d the final absorbing state. As we will see later, using this Markov chain (or random walk on a digraph), we can formulate various transmission costs in terms of quantities associated with the Markov chain (random walk). For instance, the expected number of transmissions is the expected number of steps for a packet to "walk" from the source s to the destination d. Lastly, the above Markov chain model can be also easily generalized to (best-path-based) multipath routing.

Opportunistic Routing: Given a source–destination pair (s, d), let $\mathrm{FL}(s, d) = \{u_0 = s, u_1, \ldots, u_m, u_{m+1} = d\}$ denote the (*prioritized*) forwarder list selected by an opportunistic routing protocol, say, ExOR. We first note that unlike traditional best path routing, the forwarder list $\mathrm{FL}_{s,d}$ used in opportunistic routing represents not a path, but a subgraph $G_{\mathrm{FL}} = (V_{\mathrm{FL}}, E_{\mathrm{FL}})$ connecting the source s to the destination d (see Fig. 3 for an example, where $\mathrm{FL}(s, d) = \{s, (u_1 =) = v_1, (u_2 =)v_3, (u_3 =)v_4, d\}$). Within this subgraph G_{FL}, there are many (directed) paths from s to d; which of them is actually traversed by a packet-during the packet forwarding process depends on which nodes on the forwarder list receive the packet and which nodes forward the packet.

The priority of nodes is used in opportunistic routing to decide which node should forward a packet when several of them on the forwarder list receive the same packet. Here, we use the convention that a node on the right has higher priority than a node to its left; namely, for any $j' < j$, u_j has higher priority than $u_{j'}$. Using these priorities, we can describe the forwarding process of a single packet as follows. Suppose node i, $(0 \le i \le m)$ is the current node to forward the packet. After its transmission, if the destination d receives it, then the forwarding process for this packet ends. Otherwise, suppose node j, $i < j \le m$, receives it. Node j will be the next forwarder if and only if no higher priority node, $k > j$, has received the packet. Hence, to correctly capture the packet forwarding process under an opportunistic routing, *we must track which node is the next forwarder* instead of simply which nodes receive the packet. In other words, we say the packet has successfully "walked" from node i to node j if and only if node j is the highest-priority node that receives the packet. This happens with probability $p_{ij} = a_{ij} \prod_{k>j} (1 - a_{ik})$. The packet will stay with node i if none of the higher-priority nodes have received it. This happens with the probability $p_{ii} = \prod_{k>i}(1 - a_{ik})$. Hence, we have a Markov chain defined on the state space V_{FL} with the following transition probability matrix $P_{\mathrm{FL}} = [p_{ij}]$:

$$p_{ij} = \begin{cases} a_{ij} \prod_{k>j}(1 - a_{ik}), & \text{if } 0 \le i < j \le m+1 \\ \prod_{k>i}(1 - a_{ik}), & \text{if } j = i, 0 \le i \le m \\ 1, & \text{if } j = i, i = m+1 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

It is not too hard to verify that $\sum_j p_{ij} = 1$. Using the topology in Fig. 1 as an example, the corresponding opportunistic routing Markov chain is shown in Fig. 3. Again this is an absorbing Markov chain, with node s as the starting state and d the final absorbing state. Using this Markov chain/random walk, we can again formulate various transmission costs using quantities associated with the chain/walk. As an aside, a key problem in opportunistic routing is to determine the "best" forwarder list FL, or subgraph G_{FL}, for a source and destination pair. This problem is addressed in [23], where an optimal algorithm is developed. In this paper, we will assume that the (optimal) forwarder list is given and used.

Stateless (Stochastic) Routing: As no routing states are maintained or used, given a source–destination pair (s, d), any node in G may be involved in the forwarding process of a packet. Suppose that node i is the current forwarder. After node i's transmission, a subset of its direct neighbors, $N(i)$, may receive the packet. Unlike opportunistic routing where priorities are used to determine which node should be the next forwarder, any of these nodes may become the next forwarder with equal probability. For example, the next forwarder may be selected by using a random backoff mechanism where each node randomly sets a backoff timer value uniformly chosen from $[0, t_0]$, where t_0 is an appropriately chosen contention slot. Hence, to track the packet traversals under stateless routing, we see that the packet stays with node i if and only if none of its neighbors receive the packet. This happens with probability $p_{ii} = \prod_k (1 - a_{ik})$. Otherwise, the packet transits or "walks" from node i to node j, $j \in N(i)$, with probability $p_{ij} = \frac{a_{ij}}{\sum_{k \in N(i)} a_{ik}} (1 - \prod_k (1 - a_{ik}))$. Hence, we have a Markov chain defined on the state space V (the entire node set) with the following transition probability matrix $P_G = [p_{ij}]$:

$$p_{ij} = \begin{cases} \frac{a_{ij}}{\sum_k a_{ik}} (1 - \prod_k (1 - a_{ik})), & \text{if } i \ne j \\ \prod_k (1 - a_{ik}), & \text{if } i = j. \end{cases} \tag{3}$$

It is easy to verify that $\sum_j p_{ij} = 1$. Especially, when the graph is symmetric, then the Markov chain will be reversible. The

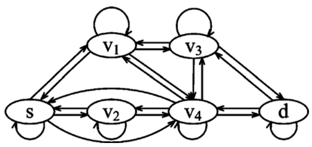


Fig. 4. Markov chain for stateless routing.

traversals of a packet under stateless routing are thus modeled as a random walk on the digraph G with the transition probability matrix P_G. Using the topology in Fig. 1 as an example with (s, d) as the source–destination pair, the resulting Markov chain is shown in Fig. 4.

Modeling the Transmission Costs: Given the Markov chain (or "random walk on a digraph") models of wireless routing, we now briefly discuss how various transmission costs such as the expected number of transmissions, latency, duty-cycle delay, or energy consumption can be modeled using certain standard notions or quantities associated with the Markov chain/random walk.

We first use the expected number of transmissions as an example and show how this cost can be formulated as the *hitting time*. In a Markov chain (or random walk), the hitting time H_{ij} is defined as the (expected) number of transitions (i.e., steps) for a random walker that starts from node (state) i to *first* reach (or hit) node j. The hitting time H_{ij} satisfies the following recursive relation:

$$H_{ij} = \begin{cases} 1 + \sum_{k=1}^{n} p_{ik} H_{kj}, & \text{if } i \neq j \\ 0, & \text{if } i = j. \end{cases} \tag{4}$$

Given the appropriately defined Markov chain for a wireless routing scheme, it is not too hard to see that the (expected) total number of transmissions needed to forward a packet from source s to destination d is exactly H_{sd}. The recursive relation (4) plays a key role in computing the hitting time H_{sd}. The remainder of this paper is devoted to addressing this and other related computation problems.

To account for other transmission costs, we introduce a transition cost matrix $T = [T_{ij}]$ associated with each one-hop transition, $T_{ij} \geq 0, \forall\, i, j$. For example, depending on the context and modeling objective, T_{ii} can be used to represent the per-node processing/transmission latency, duty-cycle delay, or per-node energy consumption, where $T_{ij}, j \in N(i)$ the one-hop forwarding latency, energy consumption, etc. Analogous to the notion of hitting time H_{ij}, we define the *hitting cost* H_{ij}^s (also referred as the *sojourn time* associated with T) as the (expected) total cost (or "delay") incurred by a random walk that starts at node i to first reach node j, where each state at any node k incurs a cost (delay) T_{kk} and each transition from node k to node l incurs a cost (delay) of T_{kl}. As in the case of H_{ij}, H_{ij}^s satisfies the following recursive relation where $s_i = \sum_j p_{ij} T_{ij}$ is the average transmission cost every time a packet visits:

$$H_{ij}^s = \begin{cases} \sum_{k=1}^{n} p_{ik}(T_{ik} + H_{kj}^s) = s_i + \sum_{k=1}^{n} p_{ik} H_{kj}^s, & \text{if } i \neq j \\ 0, & \text{if } i = j. \end{cases} \tag{5}$$

Hence, given the appropriately defined Markov chain for a wireless routing scheme and the transition cost matrix H^s, we can use H_{sd}^s to capture the (expected) total cost of transmission when forwarding a packet from source s to destination d. We note that if $T_{ij} = 1$ for all i, j, i.e., T is the all-1 matrix, then $H_{ij}^s = H_{ij}$.

III. Random Walks on Directed Graphs: Hitting, Commute, and Sojourn Times

In this section, we briefly overview the random walk theory on *undirected graphs* and show how important quantities such as hitting, commute, and sojourn times can be computed. We then outline a generalization of the random walk theory to *directed graphs (digraphs)* and show how the same quantities can be computed.

A. Random Walks on Undirected Graphs

Given an undirected graph $G = (V, E)$ that is finite connected (i.e., any node can reach any other node in G), let A be a *symmetric* weight (or adjacency) matrix appropriately defined on G. For $1 \leq i \leq n = |V|$, define $d_i = \sum_{j=1}^{n} a_{ij}$, the (weighted) degree of node i, and $d = \sum_{i=1}^{n} d_i$, often referred to as the *volume* of G, denoted by $\mathrm{vol}(G)$. Let $D = \mathrm{diag}[d_i]$ be a diagonal matrix of node degrees. Then, $P = D^{-1}A$ is a transition matrix associated with a Markov chain (a random walk) on G, where $p_{ij} = a_{ij}/d_i$. Let $\pi = [\pi_i]_{1 \leq i \leq n}$ be its stationary distribution probability vector. It is well known (see, e.g., [1]) that this Markov chain (random walk) on G is *reversible*, namely

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{6}$$

where for $1 \leq i \leq n$

$$\pi_i = \frac{d_i}{\sum_k d_k} = \frac{d_i}{d}. \tag{7}$$

Given this random walk on an undirected graph, *hitting times* H_{ij} [cf. (4)], *commute times* $C_{ij} = H_{ij} + H_{ji}$, and the *hitting costs* or *(heterogeneous) sojourn times* H_{ij}^s [cf. (5)] can be computed using a number of methods through the well-known connections between the Markov chain/random walk theory, electrical resistance theory [16], and spectral graph theory [11], [13]. Here, we present the results using the spectral graph theory.

In [11], the *normalized Laplacian matrix* for undirected graph G is defined as

$$\mathcal{L} = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = D^{\frac{1}{2}}(I - P)D^{-\frac{1}{2}} \tag{8}$$

where $\mathcal{L}$ is symmetric and positive semi-definite. Let λ_k and μ_k, $1 \leq k \leq n$, be the eigenvalues and the corresponding eigenvectors of $\mathcal{L}$, where λ_k's are arranged in the increasing order where $\lambda_1 = 0 < \lambda_2 \leq \cdots \leq \lambda_n$. Then, the hitting time H_{ij} can be computed as follows (see [28]):

$$H_{ij} = \sum_{k>1} \frac{d}{\lambda_k} \left(\frac{\mu_{kj}^2}{d_j} - \frac{\mu_{ki}\mu_{kj}}{\sqrt{d_i d_j}} \right) \tag{9}$$

and the commute time $C_{ij} = H_{ij} + H_{ji}$ is equal to

$$C_{ij} = \sum_{k>1} \frac{d}{\lambda_k} \left(\frac{\mu_{ki}}{\sqrt{d_i}} - \frac{\mu_{kj}}{\sqrt{d_j}} \right)^2. \tag{10}$$

In [9] and [10], Chau and Basu introduce a *(diagonal) sojourn time* matrix[1] $T = \mathrm{diag}[T_i]$, where T_i represents a *per-node* transition cost or "delay" incurred at node i, and define the following (T-extended) Laplacian matrix $\mathcal{L}^s$:

$$\mathcal{L}^s = T^{-\frac{1}{2}} \mathcal{L} T^{-\frac{1}{2}}. \tag{11}$$

Let σ_k (again arranged in the increasing order) and γ_k, $1 \leq k \leq n$, be the eigenvalues and eigenvectors of $\mathcal{L}^s$. Chau and Basu [9] obtain the following solution for the hitting cost matrix $H^s = [H^s_{ij}]$, extending the above (homogeneous) result for the hitting matrix $H = [H_{ij}]$:

$$H^s_{ij} = \sum_{k>1} \frac{d^s_o}{\sigma_k} \left(\frac{\gamma^2_{kj}}{d_j T_j} - \frac{\gamma_{ki}\gamma_{kj}}{\sqrt{d_i T_i d_j T_j}} \right) \tag{12}$$

where $d^s_o = \sum_k d_k T_k$.

The (normalized) Laplacian matrix $\mathcal{L}$ of an undirected graph G can be viewed as a (discrete) operator on G, analogous to a (continuous) Laplacian operator defined on a (continuous) manifold. In [13], Chung and Yau define the (discrete) normalized *Green's function* (with no boundary) $\mathcal{G}$ as a matrix with its entries, indexed by vertices i and j, that satisfies the following conditions:

$$[\mathcal{G}\mathcal{L}]_{i,j} = I_{i,j} - \frac{\sqrt{d_i d_j}}{d}, \qquad 1 \leq i, j \leq n \tag{13}$$

where $[\cdot]_{i,j}$ indicates the (i,j)th entry of the matrix $[\cdot]$. Expressed in the matrix form, we have

$$\mathcal{G}\mathcal{L} = I - \pi^{\frac{1}{2}} {\pi^{\frac{1}{2}}}^T \tag{14}$$

where $\pi^{\frac{1}{2}} = [\sqrt{\pi}_1, \ldots, \sqrt{\pi}_n]^T = \left[\frac{\sqrt{d_1}}{\sqrt{d}}, \ldots, \frac{\sqrt{d_n}}{\sqrt{d}}\right]^T$ is the (column) eigenvector of $\mathcal{L}$ associated with the eigenvalue $\lambda_1 = 0$.

We claim that the (discrete) normalized Green's function $\mathcal{G}$ corresponding to the (discrete) normalized Laplacian operator $\mathcal{L}$ (with no boundary) is exactly the pseudo-inverse of $\mathcal{L}$, namely, $\mathcal{G} = \mathcal{L}^+$. This follows easily from [13, eq. (16)] and the relation between $\mathcal{L}^+$ and the singular value decomposition of $\mathcal{L}$: Since $\mathcal{L}$ is positive semi-definite, $\mathcal{L} = U \Gamma U^T$ (as defined in Section III-A) and $\mathcal{L}^+ = U^T \Gamma^+ U$, where $\Gamma^+ = \mathrm{diag}[\lambda^+_i]$, and $\lambda^+_1 = 0$, and $\lambda^+_i = 1/\lambda_i$, $i = 2, \ldots, n$.

B. Random Walks on Directed Graphs (Digraphs)

In this section, we develop the random walk theory for digraphs. In particular, we generalize the graph Laplacian defined for undirected graphs, and introduce the *digraph Laplacian* matrix. We prove that the Moore–Penrose pseudo-inverse of this digraph Laplacian is exactly equal to (a normalized version of) the fundamental matrix of the Markov chain governing random walks on digraphs and show that it is also the Green's function of the digraph Laplacian. Using these connections, we illustrate how hitting and commute times of random walks on digraphs can be directly computed using the singular values and vectors of the digraph Laplacian. We also show that when the underlying graph is undirected, our results reduce to the well-known results for undirected graphs. Hence, our theory includes undirected graphs as a special case.

1) Random Walks on Directed Graphs and Fundamental Matrix: Let $G = (V, E)$ be a (weighted) digraph defined on the node set $V = \{1, 2, \ldots, n\}$ and A be a nonnegative, but generally *asymmetric* weight matrix such that $a_{ij} > 0$ *if and only if* the directed edge (or arc) $\langle i, j \rangle \in E$. Denote $D = \mathrm{diag}[d_i]$ as a diagonal matrix of the node out-degrees, and define $P = D^{-1}A$. Then, $P = [p_{ij}]$ is the transition probability matrix of the Markov chain associated with random walks on G, where at each node i, a random walk has the probability $p_{ij} = a_{ij}/d_i$ to transit from node i to node j, if $\langle i, j \rangle \in E$. Unlike undirected graphs, the Markov chain associated with random walks on directed graphs is generally *nonreversible*, and (6) and (7) for undirected graphs do not hold. We assume that G is *strongly connected and aperiodic*. i.e., there is a (directed) path from any node i to any other node j, and the greatest common divisor of the lengths of all closed directed walks for any node i in G is equal to 1. The assumption that G is strongly connected and aperiodic yields the following properties [1]: 1) the Markov chain is irreducible with no transient states; 2) P has a simple eigenvalue equal to 1 with right (column) eigenvector $e = [1, 1, \ldots, 1]^T$ and left (row) eigenvector π, the stationary probability vector; and 3) $\pi_i > 0$, $1 \leq i \leq n$.

For random walks on directed graphs, quantities such as hitting times and commute times can be defined exactly as in the case of undirected graphs. However, since the (normalized) Laplacian matrix $\mathcal{L}$ is (*so far!*) defined *only* for undirected graphs, we cannot use the relations (9) and (10) to compute hitting times and commute times for random graphs on directed graphs. On the other hand, using results from the standard Markov chain theory, we can express the hitting times and commute times in terms of the *fundamental matrix*. In [1], Aldous and Fill define the fundamental matrix $Z = [z_{ij}]$ for an *irreducible* Markov chain with the transition probability matrix P

$$z_{ij} = \sum_{t=0}^{\infty} \left(p^{(t)}_{ij} - \pi_j \right), \ 1 \leq i, j \leq n \tag{15}$$

where $p^{(t)}_{ij}$ is the (i,j)th entry in the t-step transition probability matrix $P^t = \underbrace{P \cdots P}_{t}$.

Let $\Pi = \mathrm{diag}[\pi_i]$ be the diagonal matrix containing the stationary probabilities π_i's on the diagonal, and $J = [J_{ij}]$ the all-one matrix, i.e., $J_{ij} = 1$, $1 \leq i, j \leq n$. We can express Z alternatively as the sum of an infinite matrix series

$$Z = \sum_{t=0}^{\infty} \left(P^t - J\Pi \right) = \sum_{t=0}^{\infty} \left(P^t - \mathbf{1}\pi^T \right) \tag{16}$$

where $\mathbf{1} = [1, \ldots, 1]^T$ is the all-one column vector. Hence $J = \mathbf{1}\mathbf{1}^T$, and $\mathbf{1}^T \Pi = \pi^T$.

While the physical meaning of the fundamental matrix Z may not be obvious from its definition (15) [or (16)], it plays a crucial role in computing various quantities related to random walks or, more generally, various *stopping time* properties of Markov chains [1]. For instance, the hitting times and commute times of

[1] In contrast, in Section II-B, we have introduced a general transition cost matrix $T = [T_{ij}]$, $T_{ij} \geq 0$, with a transition cost associated with each node T_{ii} and each link T_{ij}, where T is not necessarily diagonal nor symmetric.

random walks on a directed graph can be expressed in terms of Z as follows (see [1]):

$$H_{ij} = \frac{z_{jj} - z_{ij}}{\pi_j} \tag{17}$$

and

$$C_{ij} = \frac{z_{jj} - z_{ij}}{\pi_j} + \frac{z_{ii} - z_{ji}}{\pi_i}. \tag{18}$$

In (15) and (16), the fundamental matrix Z is defined as an infinite sum. We show that Z in fact satisfies a simple relation (19) and hence can be computed directly using the standard matrix inverse.

Theorem 1: Let P be the transition probability matrix for an irreducible Markov chain. Then, its corresponding fundamental matrix Z as defined in (15) satisfies the following relation[2]:

$$Z + J\Pi = (I - P + J\Pi)^{-1}. \tag{19}$$

Proof: Note that $J\Pi = \mathbf{1}\pi^T$. From $\pi^T P = \pi^T$ and $P\mathbf{1} = \mathbf{1}$, we have $J\Pi P = J\Pi$ and $PJ\Pi = J\Pi$. Using these two relations, it is easy to prove the following equation by induction.

$$P^m - J\Pi = (P - J\Pi)^m, \qquad \text{for any integer } m > 0. \tag{20}$$

Plugging this into (16) yields Theorem 1. ■

As undirected graphs are a special case of directed graphs, (17) and (18) provide an alternative way to compute hitting times and commute times for random walks on fully connected undirected graphs. In this paper, we will show that (9) and (10) are in fact equivalent to (17) and (18).

2) Normalized Digraph Laplacian and Green's Function for Digraphs: We now generalize the existing spectral graph theory defined for undirected graphs to directed graphs by introducing an appropriately generalized Laplacian matrix for (strongly connected and aperiodic) diagraphs. Let $G = (V, E)$ be a strongly connected and aperiodic (weighted) digraph defined on the vertex set $V = \{1, 2, \ldots, n\}$, where in general the weight (or adjacency) matrix A is asymmetric.

For a strongly connected and aperiodic digraph G, let $\Pi^{\frac{1}{2}} = \mathrm{diag}[\sqrt{\pi_i}]$. We introduce the *generalized normalized Laplacian* matrix for a strongly connected and aperiodic digraph $G = (V, E)$ (and an associated Markov chain with the transition matrix P) as follows.

Definition 1 (Generalized Normalized Laplacian $\tilde{\mathcal{L}}$):

$$\tilde{\mathcal{L}} = \Pi^{\frac{1}{2}}(I - P)\Pi^{-\frac{1}{2}} \tag{21}$$

or in the scalar form

$$\tilde{\mathcal{L}}_{ij} = \begin{cases} 1 - p_{ii}, & \text{if } i = j \\ -\pi_i^{\frac{1}{2}} p_{ij} \pi_j^{-\frac{1}{2}}, & \text{if } \langle i, j \rangle \in E, i \neq j \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

We see that the (normalized) digraph Laplacian[3] for (strongly connected and aperiodic) digraphs is redefined using the diagonal stationary distribution matrix Π instead of the diagonal degree matrix D. In the case of a (fully connected) undirected graph, as $\Pi^{\frac{1}{2}} = D^{\frac{1}{2}}/\sqrt{d}$, we have $\tilde{\mathcal{L}} = D^{\frac{1}{2}}(I - P)D^{-\frac{1}{2}} = \mathcal{L}$. Hence, this definition of (generalized) normalized Laplacian for digraphs reduces to the usual definition of normalized Laplacian for undirected graphs.

Treating this (normalized) digraph Laplacian matrix $\tilde{\mathcal{L}}$ as an (*asymmetric*) operator on a digraph G, we now define the (discrete) *Green's function* $\tilde{\mathcal{G}}$ (without boundary conditions) for digraphs in exactly the same manner as for undirected graphs [1]. Namely, $\tilde{\mathcal{G}}$ is a matrix with its entries, indexed by vertices i and j, that satisfies the following conditions:

$$[\tilde{\mathcal{G}}\tilde{\mathcal{L}}]_{i,j} = I_{i,j} - \sqrt{\pi_i \pi_j}, \qquad 1 \leq i, j \leq n \tag{23}$$

and expressed in the matrix form

$$\tilde{\mathcal{G}}\tilde{\mathcal{L}} = I - \pi^{\frac{1}{2}}\pi^{\frac{1}{2}^T}. \tag{24}$$

In the following, we will show that $\tilde{\mathcal{G}}$ is precisely $\tilde{\mathcal{L}}^+$, the pseudo-inverse of the Laplacian operator $\tilde{\mathcal{L}}$ on the digraph G. Furthermore, we will relate $\tilde{\mathcal{L}}^+$ directly to the fundamental matrix Z of the Markov chain associated with random walks on the digraph G. Before we establish this main result of the section, we first introduce a few more notations and then prove the following useful lemma.

Lemma 1: Define $\tilde{\mathcal{Z}} = \Pi^{\frac{1}{2}} Z \Pi^{-\frac{1}{2}}$ (the normalized fundamental matrix), and $\tilde{J} = \Pi^{\frac{1}{2}} J \Pi^{\frac{1}{2}} = \pi^{\frac{1}{2}}\pi^{\frac{1}{2}^T}$. The following relations regarding $\tilde{\mathcal{Z}}$ and $\tilde{J}$ hold: 1) $\tilde{J} = \tilde{J}^2$; 2) $\tilde{J}\tilde{\mathcal{L}} = \tilde{\mathcal{L}}\tilde{J} = \tilde{\mathcal{L}}\pi^{\frac{1}{2}} = \pi^{\frac{1}{2}^T}\tilde{\mathcal{L}} = 0$; and 3) $\tilde{J}\tilde{\mathcal{Z}} = \tilde{\mathcal{Z}}\tilde{J} = \tilde{\mathcal{Z}}\pi^{\frac{1}{2}} = \pi^{\frac{1}{2}^T}\tilde{\mathcal{Z}} = 0$.

Proof Sketch: These relations can be established using the facts that $J = \mathbf{1}\mathbf{1}^T$, $\mathbf{1}^T\Pi = \pi^T$, $\pi^T J = \mathbf{1}^T$, $\Pi J = \pi\mathbf{1}^T$, $J\Pi J = J$, $\pi^T(I - P) = 0$, $(I - P)\mathbf{1} = 0$, $\pi^T Z = 0$, and $Z\mathbf{1} = 0$. The last four equalities imply that the matrices $I - P$ and Z have the same left and right eigenvectors, π^T and $\mathbf{1}$, corresponding to the eigenvalue 0. ■

We are now in a position to prove the following main theorem of this section, which states the Green's function for the (normalized) digraph Laplacian is exactly its Moore–Penrose pseudo-inverse, and it is equal to the normalized fundamental matrix. Namely, $\tilde{\mathcal{G}} = \tilde{\mathcal{L}}^+ = \tilde{\mathcal{Z}}$. For completeness, we also include the key definitions in the statement of the theorem.

Theorem 2 (Laplacian Matrix and Green's Function for Digraphs): Given a strongly connected and aperiodic digraph $G = (V, E)$ where $V = \{1, \ldots, n\}$, let A be a (generally asymmetric) nonnegative weight/adjancency matrix of G such that $a_{ij} > 0$ if and only if $\langle i, j \rangle \in E$, and define $D = \mathrm{diag}[d_i]$ as the diagonal (out-)degree matrix, i.e., $d_i = \sum_j a_{ij}$. Then, $P = D^{-1}A$ is the transition probability matrix for the (irreducible and generally nonreversible) Markov chain associated with random walks on the digraph G. Let $\pi = [\pi_1, \ldots, \pi_n]^T$ be the stationary probability distribution (represented as a column vector) for the Markov chain P, and $\Pi = \mathrm{diag}[\pi_i]$ be the diagonal stationary probability matrix. We define the (normalized) digraph Laplacian matrix $\tilde{\mathcal{L}}$ of G as in (21), i.e., $\tilde{\mathcal{L}} = \Pi^{\frac{1}{2}}(I - P)\Pi^{-\frac{1}{2}}$. Define $\tilde{\mathcal{Z}} = \Pi^{\frac{1}{2}} Z \Pi^{-\frac{1}{2}}$, where Z is the fundamental matrix of the Markov chain P as defined in (15).

Then, $\tilde{\mathcal{Z}} = \tilde{\mathcal{L}}^+$, is the pseudo-inverse of the Laplacian matrix $\tilde{\mathcal{L}}$. Furthermore, $\tilde{\mathcal{Z}}$ is the (discrete) Green's function for $\tilde{\mathcal{L}}$. Namely

$$\tilde{\mathcal{Z}}\tilde{\mathcal{L}} = I - \Pi^{\frac{1}{2}} J \Pi^{\frac{1}{2}} = I - \pi^{\frac{1}{2}}\pi^{\frac{1}{2}^T} \tag{25}$$

[2]In [18], the matrix $Y := (I - P + J\Pi)^{-1}$ is directly defined as the fundamental matrix of an irreducible Markov chain P instead of Z.

[3]An *unnormalized* digraph Laplacian matrix, $\mathcal{L}$, can also be defined as $\mathcal{L} = \Pi(I - P)$ [5].

where J is the all-1 matrix and $\pi^{\frac{1}{2}} = [\pi_1^{\frac{1}{2}}, \ldots, \pi_n^{\frac{1}{2}}]^T$ (a column vector).

Proof Sketch: From (19) in Theorem 1

$$Z + J\Pi = (I - P + J\Pi)^{-1}$$

we have

$$\Pi^{\frac{1}{2}}(Z + J\Pi)\Pi^{-\frac{1}{2}} = \left(\Pi^{\frac{1}{2}}(I - P + J\Pi)\Pi^{-\frac{1}{2}}\right)^{-1}.$$

Hence

$$\tilde{\mathcal{Z}} + \tilde{J} = (\tilde{\mathcal{L}} + \tilde{J})^{-1}. \tag{26}$$

Multiplying (26) from the right by $\tilde{\mathcal{L}} + \tilde{J}$, and using Lemma 1, it is easy to see that

$$\tilde{\mathcal{Z}}\tilde{\mathcal{L}} = I - \tilde{J} \tag{27}$$

which establishes that $\tilde{\mathcal{Z}}$ is the Green's function of the digraph Laplacian $\tilde{\mathcal{L}}$.

Multiplying (26) from the left by $\tilde{\mathcal{L}} + \tilde{J}$, and using Lemma 1, we can likewise prove that

$$\tilde{\mathcal{L}}\tilde{\mathcal{Z}} = I - \tilde{J}. \tag{28}$$

Hence, $\tilde{\mathcal{Z}}\tilde{\mathcal{L}} = \tilde{\mathcal{L}}\tilde{\mathcal{Z}} = I - \tilde{J}$, which is a real symmetric matrix. Hence, $(\tilde{\mathcal{L}}\tilde{\mathcal{Z}})^T = \tilde{\mathcal{L}}\tilde{\mathcal{Z}}$ and $(\tilde{\mathcal{Z}}\tilde{\mathcal{L}})^T = \tilde{\mathcal{Z}}\tilde{\mathcal{L}}$. Furthermore, as $\tilde{J}\tilde{\mathcal{Z}} = 0$, (27) yields $\tilde{\mathcal{Z}}\tilde{\mathcal{L}}\tilde{\mathcal{Z}} = \tilde{\mathcal{Z}}$. Similarly, as $\tilde{\mathcal{L}}\tilde{J} = \Pi^{\frac{1}{2}}(I - P)J\Pi^{\frac{1}{2}} = 0$, (27) yields $\tilde{\mathcal{L}}\tilde{\mathcal{Z}}\tilde{\mathcal{L}} = \tilde{\mathcal{L}}$. These establish that $\tilde{\mathcal{Z}}$ is also the Moore–Penrose pseudo-inverse of $\tilde{\mathcal{L}}$. Therefore, $\tilde{\mathcal{G}} = \tilde{\mathcal{Z}} = \tilde{\mathcal{L}}^+$. ■

3) Computing Hitting and Commute Times for Digraphs Using Digraph Laplacian: Using the relationship between the (normalized) digraph Laplacian $\tilde{\mathcal{L}}$, its pseudo-inverse $\tilde{\mathcal{L}}^+$, and the (normalized) fundamental matrix $\tilde{\mathcal{Z}}$, we can now express the hitting times and commute times of random walks on digraphs in terms of $\tilde{\mathcal{L}}^+$, or alternatively in terms of the *singular values* and *singular vectors* of the digraph Laplacian matrix $\tilde{\mathcal{L}}$, as stated in Theorem 3.

Theorem 3 (Hitting and Commute Times for Random Walks on Digraphs): Given the generalized normalized Laplacian matrix $\tilde{\mathcal{L}}$ defined in (21) and (22), and let $\tilde{\mathcal{L}}^+$ be its pseudo-inverse. Then, we have the hitting time H_{ij} as

$$H_{ij} = \frac{\tilde{\mathcal{L}}^+_{jj}}{\pi_j} - \frac{\tilde{\mathcal{L}}^+_{ij}}{\sqrt{\pi_i \pi_j}} \tag{29}$$

or in the matrix form

$$H = J \cdot \mathrm{diag}\left(\Pi^{-\frac{1}{2}}\tilde{\mathcal{L}}^+\Pi^{-\frac{1}{2}}\right) - \Pi^{-\frac{1}{2}}\tilde{\mathcal{L}}^+\Pi^{-\frac{1}{2}}. \tag{30}$$

The commute times, $C_{ij} = H_{ij} + H_{ji}$, can be computed as follows:

$$C_{ij} = \frac{\tilde{\mathcal{L}}^+_{jj}}{\pi_j} + \frac{\tilde{\mathcal{L}}^+_{ii}}{\pi_i} - \frac{\tilde{\mathcal{L}}^+_{ij}}{\sqrt{\pi_i \pi_j}} - \frac{\tilde{\mathcal{L}}^+_{ji}}{\sqrt{\pi_i \pi_j}}. \tag{31}$$

Proof Sketch: From $Z = \Pi^{-\frac{1}{2}}\tilde{\mathcal{Z}}\Pi^{\frac{1}{2}} = \Pi^{-\frac{1}{2}}\tilde{\mathcal{L}}^+\Pi^{\frac{1}{2}}$, and using (17) and (18), we can compute the hitting times and commute times for random walks on digraphs directly in terms of the entries of $\tilde{\mathcal{L}}^+$ as stated in Theorem 3

$$H_{ij} = \frac{\tilde{\mathcal{L}_{jj}}^+}{\pi_j} - \frac{\tilde{\mathcal{L}_{ij}}^+}{\sqrt{\pi_i \pi_j}} \tag{32}$$

and

$$C_{ij} = H_{ij} + H_{ji} = \frac{\tilde{\mathcal{L}_{jj}}^+}{\pi_j} + \frac{\tilde{\mathcal{L}_{ii}}^+}{\pi_i} - \frac{\tilde{\mathcal{L}_{ij}}^+}{\sqrt{\pi_i \pi_j}} - \frac{\tilde{\mathcal{L}_{ji}}^+}{\sqrt{\pi_i \pi_j}} \tag{33}$$

where $\tilde{\mathcal{L}_{ij}}^+$ is the (i,j)th entry of $\tilde{\mathcal{L}}^+$, and π_i is the stationary probability of vertex i. ■

We note that if the underlying graph G is undirected, and the transition probability matrix P of the Markov chain is $P = D^{-1}A$ (as defined in Section III-A, where A is *symmetric*), then $\tilde{\mathcal{L}} = \mathcal{L}$. Furthermore, one can show that (30) and (31) are equivalent to (9) and (10). Hence, our theory of random walks on digraphs subsumes the existing theory of random walks on undirected graphs as a special case. Moreover, computing the Moore–Penrose pseudo-inverse is equivalent to solving a singular value decomposition, thus the computational time complexity is $O(n^3)$ [36].

We now extend the above results for hitting and commuting times to hitting and commute costs. Given a (asymmetric) transition cost matrix $T = [T_{ij}]$, $T_{ij} \geq 0$, with T_{ij} as the per-link transition cost, and T_{ii} as the per-node transmission cost, define $S = \mathrm{diag}[s_i]$ a diagonal matrix with $s_i = \sum_{j=1}^n T_{ij}p_{ij}$ as the average transmission cost every time a packet visits. We define the following *normalized cost Laplacian* matrix, $\tilde{\mathcal{L}}^s$ as

Definition 2 (Normalized Cost Laplacian $\tilde{\mathcal{L}}^s$):

$$\begin{aligned} \tilde{\mathcal{L}}^s &= S^{-\frac{1}{2}}\tilde{\mathcal{L}}S^{-\frac{1}{2}} \\ &= S^{-\frac{1}{2}}\Pi^{\frac{1}{2}}(I - P)\Pi^{-\frac{1}{2}}S^{-\frac{1}{2}} \end{aligned} \tag{34}$$

where the $\tilde{\mathcal{L}}$ is the generalized normalized Laplacian matrix.

The corresponding scalar form is

$$\tilde{\mathcal{L}}^s_{ij} = \begin{cases} s_i^{-1}(1 - p_{ii}), & \text{if } i = j \\ -s_i^{-\frac{1}{2}}\pi_i^{\frac{1}{2}}p_{ij}\pi_j^{-\frac{1}{2}}s_j^{-\frac{1}{2}}, & \text{if } \langle i,j\rangle \in E, i \neq j \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

The hitting costs H^s_{ij} as defined by the relation in (5) and the commute costs C^s_{ij} can be computed using the (Moore–Penrose) pseudo-inverse of Laplacian $\tilde{\mathcal{L}}^s$ as stated in Theorem 4, and the proof is delegated to the Appendix.

Theorem 4 (Hitting and Commute Costs for Random Walks on Digraphs): Let $\tilde{\mathcal{L}^s}^+$ denote the (Moore–Penrose) pseudo-inverse of the cost Laplacian $\tilde{\mathcal{L}}^s$, and $d^s = \sum_i \pi_i s_i$. Then, the hitting costs and commute costs can be computed as follows:

$$H^s_{ij} = d^s\left(\frac{\tilde{\mathcal{L}^s}^+_{jj}}{\pi_j s_j} - \frac{\tilde{\mathcal{L}^s}^+_{ij}}{\sqrt{\pi_i s_i \pi_j s_j}}\right) \tag{36}$$

and

$$C^s_{ij} = d^s\left(\frac{\tilde{\mathcal{L}^s}^+_{jj}}{\pi_j s_j} + \frac{\tilde{\mathcal{L}^s}^+_{ii}}{\pi_i s_i} - \frac{\tilde{\mathcal{L}^s}^+_{ij}}{\sqrt{\pi_i s_i \pi_j s_j}} - \frac{\tilde{\mathcal{L}^s}^+_{ji}}{\sqrt{\pi_i s_i \pi_j s_j}}\right). \tag{37}$$

Moreover, we note that in our recent paper [25], we further analyzed and obtained an upper bound on the mixing time of random walks on digraphs in terms of the eigenvalues and singular values of the symmetric and skew-symmetric components of the digraph Laplacian.

IV. Estimating Transmission Cost for Different Routing Strategies

In this section, we apply the theory of random walks on digraphs to various routing strategies and see how it can be used to estimate various transmission costs. We consider three specific examples: the *keep-connect* routing, a class of energy-aware best-path routing algorithms proposed in [31]; the *opportunistic routing* protocol as defined in [4] and analyzed in [23]; and the *stateless routing* protocol introduced and analyzed in [9].

A. Keep-Connect Routing for Network Lifetime Maximization

Many energy-aware routing protocols have been developed for energy-constrained wireless networks, such as wireless sensor networks. These protocols take the energy cost as a key metric in selecting routes and attempt to forward packets along a path that minimizes the energy consumption or maximizes the overall network lifetime. The *keep-connect* routing (thereafter referred to as KC) is a class of energy-aware routing algorithms proposed in [31] that take into account both the energy cost and the "importance" of nodes in the overall network connectivity so as to maximize the network lifetime. Here, the network lifetime is defined as the time until the network becomes disconnected. In [31], the authors apply the spectral graph theory—in particular, use the Fiedler value (the second smallest eigenvalue of a graph Laplacian)—to the design and analysis of the KC routing algorithms. We use KC as a simple example to illustrate how to estimate the wireless transmission costs (hitting costs).

In KC, the "importance" of a node i in terms of its connectivity in a graph $G=(V,E)$ is defined as follows: $W(i)=1/\lambda_2(L(G_{-i}))$, where G_{-i} is a graph resulting from G with node i and its adjacent edges removed, $L(G_{-i})$ is the graph Laplacian of G_{-i}, and $\lambda_2(L(G_{-i}))$ is the second smallest eigenvalue (the Fiedler value) of $L(G_{-i})$. For each link $\langle i,j\rangle\in E$, let e_{ij} denote the one-hop transmission energy cost from i to j. We introduce the following transition cost matrix $T=[T_{ij}]$, where $T_{ij}=e_{ij}W(i)$, if $\langle i,j\rangle\in E$, and $T_{ij}=0$ otherwise. Given a path or route $R=\{u_0=s,u_1,\ldots,u_m,u_{m+1}=d\}$ for a source–destination pair (s,d). Then, using the Markov chain transition matrix P_R defined on the line (sub)graph G_R (with $V_R=\{u_0,u_1,\ldots,u_{m+1}\}$) as given in (1) of Section II-B, we can easily solve the hitting cost (5) and compute the hitting cost matrix H_R^s for G_R as follows:

$$H^s_{i,m+1}=\begin{cases}\sum_{k=i}^{m}e_{k,k+1}W(k)\frac{1}{a_{k,k+1}}, & 0\le i<m+1\\ 0, & \text{otherwise.}\end{cases} \tag{38}$$

In particular, $H^s_{sd}=H^s_{0,m+1}$ is the expected total cost associated with the route R. Hence, to find the best path in KC that minimizes the energy consumption and maximizes the network lifetime is equivalent to finding the best route R that minimizes H^s_{sd}, for any given source–destination pair (s,d).

B. Opportunistic Routing

Following the description in Section II-B, given a (prioritized) forwarder list $\mathrm{FL}(s,d)=\{u_0=s,u_1,\ldots,u_m,u_{m+1}=d\}$ for a source–destination pair (s,d), let P_{FL} be the transition matrix for the Markov chain on the subgraph G_{FL} (with $V_{\mathrm{FL}}=\{u_0,u_1,\ldots,u_{m+1}\}$), as given in (2). Hence, the total expected number of transmissions using this forwarder list FL can be computed using the hitting time matrix H associated with P_{FL} defined on G_{FL}. In particular, $H_{sd}=H_{0,m+1}$ is the total expected number of transmissions from source s to destination d. We can apply the theory of random walks on digraphs to compute H_{sd}. In this case, we can explore the special structure of P_{FL} and find a *closed-form* recursive formula for H directly. From (2), we note that P_{FL} is an upper triangular matrix. Hence, the hitting times $H_{i,m+1}$ satisfy the following recursive relations:

$$\begin{cases}H_{0,m+1}=1+\sum_{k\ge 0}^{m+1}p_{0k}H_{k,m+1}\\ \cdots\\ H_{i,m+1}=1+\sum_{k\ge i}^{m+1}p_{ik}H_{k,m+1}\\ \cdots\\ H_{m,m+1}=1+p_{m,m}H_{m,m+1}+p_{m,m+1}H_{m+1,m+1}\\ H_{m+1,m+1}=0.\end{cases} \tag{39}$$

We observe that these (upper triangular) linear recursive equations can be easily solved, starting from the bottom to the top. Plugging the transition probabilities p_{ij}'s in, using (2), we can write the above recursive equation as follows:

$$H_{i,m+1}=\begin{cases}\frac{1+\sum_{k>i}^{m+1}a_{ik}\prod_{j>k}(1-a_{ij})H_{k,m+1}}{1-\prod_{k>i}(1-a_{ik})}, & 0\le i<m+1\\ 0, & i=m+1.\end{cases} \tag{40}$$

The above equation is exactly the same formula obtained in [23] through an event-based direct probability analysis approach. The random walk method proposed here is far simpler. We can further extend the above analysis of the expected number of transmissions to other transmission costs. Let $S=[s_i]$ be a diagonal average transition cost matrix in subgraph G_{FL}, as defined in Section III-B. We can derive the following recursive formula for computing the expected total transmission cost matrix defined on G_{FL}:

$$H^s_{i,m+1}=\begin{cases}\frac{s_i+\sum_{k>i}^{m+1}a_{ik}\prod_{j>k}(1-a_{ij})H^s_{k,m+1}}{1-\prod_{k>i}(1-a_{ik})}, & 0\le i\le m\\ 0, & i=m+1.\end{cases} \tag{41}$$

Hence, for a given $\mathrm{FL}(s,d)=\{u_0=s,u_1,\ldots,u_m,u_{m+1}=d\}$, the total expected transmission cost of FL is $H^s_{sd}=H^s_{0,m+1}$. In [23], using the recursive relations (40) and (41), an optimal algorithm (MTS stands for *minimum transmission selection*), and generalized MTS algorithms have been developed to minimize respectively the total expected number of transmission, total expected energy consumption, total end-to-end transmission latency, and so forth.

C. Stateless Routing

In [9], Chau and Basu analyze the stateless routing algorithm using the random walk theory for *undirected graphs* and apply (12) to derive an end-to-end delay estimation formula with heterogeneous sojourn times. Since the undirected graph model, they assume that wireless links are symmetric. Given the theory of random walks on digraphs we have developed, we can now solve the same problem using (36), where the transition matrix P for the Markov chain/random walk is given in (3). A step-by-step procedure for constructing the Markov chain and then computing the hitting time matrix H is given in Algorithm 1.

Algorithm 1: Delay Estimation Algorithm for Stateless routing, given the adjacency matrix A.

1: Form the transition probability matrix P using (3).
2: Compute the stationary probability matrix Π.
3: Compute the generalized normalized Laplacian matrix $\tilde{\mathcal{L}}$ (See DEFINITION 1).
4: Compute the pseudo-inverse of $\tilde{\mathcal{L}}$.
5: Compute the hitting time matrix H using (29).

Moreover, by introducing an arbitrary transmission matrix $T=[T_{ij}], T_{ij} \geq 0$ to represent per-node and/or per-hop transmission cost, we can apply Algorithm 1 (with $\tilde{\mathcal{L}}$ and H replaced by $\tilde{\mathcal{L}}^s$ and H^s) to compute various transmission costs associated with the stateless routing, such as latency, energy consumption, and so forth.

V. Estimating (Partial) Cover Time for Query Processing

In this section, we consider another important practical problem, in particular the (broadcast) query processing problem in wireless sensor networks. We illustrate how quantities such as hitting times and cover times of random walks on digraphs can be used to provide theoretical estimates for (partial) broadcast covering time in wireless networks. This application is based on the study in [2].

In [2], the authors proposed and analyzed random-walk-based query processing techniques and demonstrated that such techniques possess many desirable properties such as robustness, simplicity, load balancing, and scalability in dynamic wireless environments. We briefly describe the basic procedure behind such techniques as follows. At the beginning, the base station constructs a query message with description of the query information. The message is transmitted in a "random walk manner" in the network. Once the query message reaches a node v, it updates its information with the local data stored at node v. Then, it is randomly forwarded to one of v's neighbors. When the answer is satisfying or a sufficient number of steps have been taken or a sufficient number of nodes have been visited, the query message is sent back to the base station. For instance, if the minimum temperature information is requested in the network, every time the query message meets a lower temperature, it updates the answer. After visiting enough nodes, the query message is forwarded back to the base station.

The analysis of the techniques relies on a key quantity associated with random walks, namely, the *cover time*. Let $G(V,E)$ be a digraph on which the random walk is performed, and let $n=|V|$. The cover time $c(G)$ is the (worst) expected number of steps taken by a random walk to visit every node in G from any node. The following theorem provides, in many cases, tighter bounds on the cover time $c(G)$ in terms of the maximum (minimum) hitting time $H_{\max}$ ($H_{\min}$) over all ordered pairs of nodes.

Theorem 5 (Matthews' Theorem [30]): For any Graph $G=(V,E)$ and $n=|V|$

$$H_{\min}\mathcal{H}_k \leq c(G) \leq H_{\max}\mathcal{H}_k \tag{42}$$

where $\mathcal{H}_k=\sum_{i=1}^{n}\frac{1}{i}$ is the kth harmonic number.

Extending the notation of cover time, the authors [2] introduce the notion of *partial cover time* (PCT) to better model the operations and performance of random-walk-based query processing techniques in wireless sensor networks. For instance, in many sensor network applications, it is not necessary to consult every node in the network to gather and process information. The PCT is defined as the expected time required to cover only a constant fraction ($0 \leq c < 1$) of the network (digraph). Based on the well-known Matthews' bounds (Theorem 5), the authors establish the following bound on the PCT in terms of the maximum hitting time.

Theorem 6 (Upper Bound on PCT): For any Graph $G=(V,E)$, let $n=|V|$ and $0 \leq c \leq \frac{n-1}{n}$, we have

$$\mathrm{PCT} \leq 2H_{\max}\log_2\left(\frac{1}{1-c}\right) = \mathcal{O}(H_{\max}). \tag{43}$$

This result shows that by relaxing full coverage to a partial coverage of the network, the partial cover time can be up to a factor of $\mathcal{O}(\log n)$ more efficient.

Given the bounds in Theorems 5 and 6, our random walk theory on digraphs provides a method to compute the bounds on the cover time and partial cover time using the digraph Laplacian (and its Moore–Penrose pseudo-inverse)

$$\min_{1\leq i,j\leq n}\left(\frac{\tilde{\mathcal{L}}^+_{jj}}{\pi_j}-\frac{\tilde{\mathcal{L}}^+_{ij}}{\sqrt{\pi_i\pi_j}}\right)\mathcal{H}_k \leq c(G) \leq \max_{1\leq i,j\leq n}\left(\frac{\tilde{\mathcal{L}}^+_{jj}}{\pi_j}-\frac{\tilde{\mathcal{L}}^+_{ij}}{\sqrt{\pi_i\pi_j}}\right)\mathcal{H}_k \tag{44}$$

and

$$\mathrm{PCT} \leq 2\max_{1\leq i,j\leq n}\left(\frac{\tilde{\mathcal{L}}^+_{jj}}{\pi_j}-\frac{\tilde{\mathcal{L}}^+_{ij}}{\sqrt{\pi_i\pi_j}}\right)\log_2\left(\frac{1}{1-c}\right). \tag{45}$$

VI. Performance Evaluation and Comparison

In this section, we use the stateless routing [9], and the query processing protocol [2] as examples and conduct simulations to compare the transmission cost and the (partial) broadcast covering time estimation results obtained using the *random walks on digraphs* model versus using the *random walks on undirected graphs* model, where the asymmetric packet delivery probabilities of a link, a_{ij} and a_{ji}, are symmetrized using their average, $\bar{a}_{ij}=\bar{a}_{ji}=(a_{ij}+a_{ji})/2$. The simulation results demonstrate that compared to the random walks on (symmetrized) *undirected* graph model, our random walks on digraph model improves the accuracies of stateless routing transmission cost estimation and partial broadcast covering time estimation by up to 70% and 90%, respectively. Similar results are obtained for other routing schemes, which are not presented here due to space limitation.

Simulation Setting: Before we present the detailed simulation results, we first briefly describe the simulation settings, especially how parameters are chosen and varied. The wireless topologies are generated with n nodes distributed in a (square) area of size Ar measured in squared meters (m^2). For a given n, the n nodes are randomly placed in the area with a bivariate uniform distribution. The (overall) density of the topology is therefore measured by n/Ar (i.e., nodes per m^2). Each node has the same transmission radius of R. The *connectivity digraph* $G = (V, E, A)$ is constructed as follows: 1) an edge from node i to node j and an edge from node j to node i present if and only if the Euclidean distance from i to j is smaller than the transmission radius, i.e., $d(i, j) < R$; and 2) *asymmetric* edge weights a_{ij}'s are assigned to edges in G using a uniform distribution $U(0, 1)$. The edge weights represent the link qualities (delivery probabilities) among wireless nodes (with 1 representing 100% delivery probability, or 0 no connectivity). Via symmetrization $\bar{A} = (A + A^T)/2$, we obtain an undirected graph $\bar{G} = (V, \bar{E}, \bar{A})$. Depending on the simulation setting, we vary one or more of the simulation parameters Ar, n, and R as well as the variability and asymmetry in link qualities to study how these factors affect the errors introduced when using the random walks on undirected graph model versus digraph model in estimating transmission costs and partial broadcast covering times. For each simulation setting, we generate 1000 topologies with different seeds. The average of these 1000 simulations is used for comparison. The simulations are conducted using MATLAB on a standard PC server.

A. Estimating Transmission Costs in Stateless Routing

In this part, we discuss how different topology parameters, such as network density, degree of asymmetry, and link quality variation, affect the transmission cost estimation in stateless routing.

Let H^d_{ij} represent the hitting time from node i to node j computed using the digraph model, and H^u_{ij} the hitting time computed using the undirected graph model. We use the *relative error* in (46) to measure the inaccuracy in hitting time computation introduced by the symmetrization of link qualities (i.e., when using the undirected graph model)

$$\mathrm{ERR}_{ij} = |H^d_{ij} - H^u_{ij}|/H^d_{ij}, \tag{46}$$

1) Effect of Network Density on ERR: To evaluate how the network density affects the transmission cost estimation, we generate topologies with size $\mathrm{Ar} = 300 \times 300\ \mathrm{m}^2$, where each node has the same transmission radius $R = 100$ m. We change the total number of wireless nodes as $n = 15, 25, 40$, which in turn varies the network density as $n/\mathrm{Ar} = 15, 25, 40$ nodes per 300 ×300 m^2. Fig. 5 shows the results obtained for the three network density settings. Here, we sort the node pairs in each topology by their ERR values, so that we can plot the ERR's in a monotonically decreasing order. We see that the relative errors are overall lower in denser networks. To better illustrate the effect of network density, we group the node pairs based on the ranges of their ERR's—$[0, 30\%)$, $[30\%, 60\%)$ and $[60\%, \infty)$—and compute the percentage of node pairs that fall within each range, and the results are shown in Fig. 6. We see

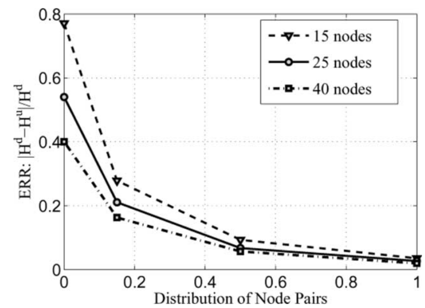


Fig. 5. Distribution of ERR.

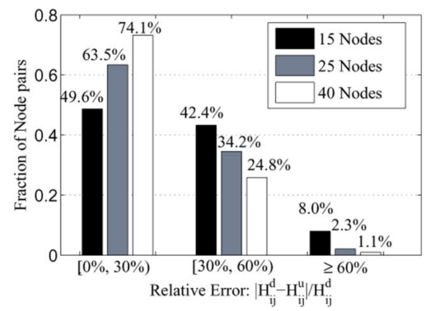


Fig. 6. ERR distribution for different network densities.

that using the undirected graph model, from 25% up to 50% of all node pairs have an average relative error at least 30%, and a few percentage have an average relative error of more than 60%. Both figures indicate that when the network density increases, the percentage of node pairs having large relative errors decreases. This is in fact not surprising: As the network is dense (e.g., 40 nodes in a 300× 300-m^2 area, each having transmission radius of 100 m), path diversity is high. In other words, the number of random (and shorter) paths between two nodes is typically high, thus reducing the hitting time between them. The asymmetric links thus likely have less impact on the overall results. On the other hand, when the network is relatively sparse, the asymmetric links have much higher impact, and therefore how to perform routing effectively becomes more critical.

2) Effect of Degree of Asymmetry and Link Quality Variation: In the second set of evaluations, we generate wireless topology with size $\mathrm{Ar} = 300 \times 300\ \mathrm{m}^2$ and $n = 25$ nodes and choose node transmission radius as $R = 100$ m. Then, we vary the degree of asymmetry (percentage of asymmetric links) and the (asymmetric) link quality variation (the extent to which a_{ij} and a_{ji} differ). For the former, we utilize the asymmetric distribution probability $S\% \in [0, 1]$ to control the distribution of asymmetric links in the topology. For the latter, we use a parameter ϵ defined below to control the link variation: For each node pair chosen to have a pair of asymmetric links (determined by a given $S\%$), we first randomly generate $a_{ij} \sim U(0, 1)$, and then randomly generate $a_{ji} (\neq a_{ij})$ with uniform distribution $a_{ji} \sim U[(1 - \epsilon)a_{ij}, \min((1 + \epsilon)a_{ij}, 1)]$. In the evaluations, we select $S\% = 0$, 25%, 50%, 75%, 100%, and $\epsilon = 0$, 25%, 50%, 75%, 100%.

Fig. 7 shows the effect of the degree of link asymmetry, $S\%$ on the worst-case ERR performance. We show the results with $\epsilon = 25\%$, 75%, and 100%, while varying the degree of link asymmetry, $S\%$. We see that as the degree of asymmetry $S\%$

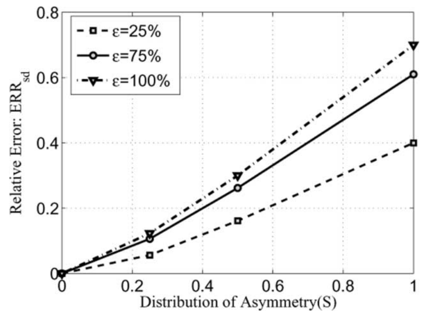


Fig. 7. Effect of degree of asymmetry.

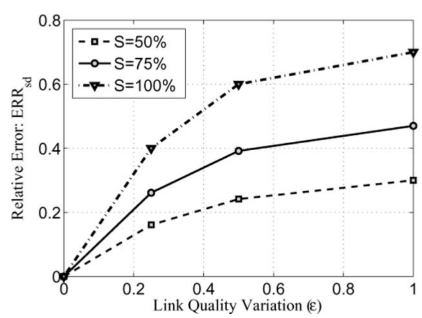


Fig. 8. Effect of (asymmetric) link quality variation.

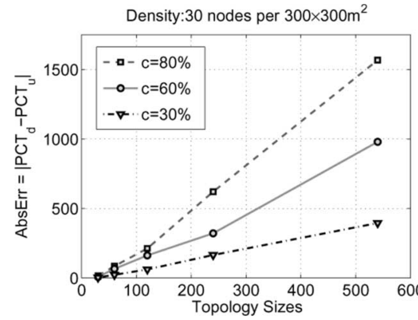


Fig. 9. Error distribution for different topology sizes in estimating the bounds of partial cover time.

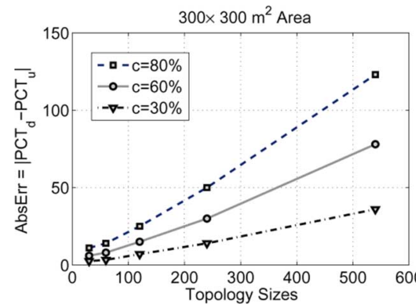


Fig. 10. Error distribution for different densities in estimating the bounds of partial cover time.

increases, the (worst-case) ERR increases rapidly, up to more than 0.6 when $S\% = 100\%$ and $\epsilon = 75\%$ or 100%.

Fig. 8 shows the effect of the (asymmetric) link-quality variation on the worst-case ERR performance. We show the results with $S\% = 50\%$, 75%, and 100%, while varying ϵ from 0 to 1. Similarly, we see that when we allow a larger extent that the qualities of asymmetric link pair, a_{ij} and a_{ji}, can differ the higher the (worst-case) ERR is. All in all, we conclude that when the degree of asymmetry is high, and the link qualities of asymmetric link pairs can differ to a larger extent, it is important to take into account the asymmetric link qualities in routing decision making and estimation of various transmission costs. This is especially the case in a relatively sparse network.

B. Estimating (Partial) Broadcast Cover Time

In this section, we use the query processing application discussed in Section V as an example, comparing the bounds on partial cover times estimated using the *random walks on digraphs* model versus using the *random walks on undirected graphs* model.

Let $\mathrm{PCT}^d(G)$ represent the upper bound of the partial cover time for a given topology G computed using the digraph model, and $\mathrm{PCT}^u(G)$ be the upper bound of the partial cover time computed using the undirected graph model. To compare and gauge the inaccuracy introduced by the symmetrization of link qualities (i.e., when using the undirected graph model) in the PCT bound estimation, we compute the *absolute error* and *relative error* formally defined as follows:

$$\mathrm{AbsErr} = |\mathrm{PCT}^d(G) - \mathrm{PCT}^u(G)|$$
$$\mathrm{RelErr} = \frac{|\mathrm{PCT}^d(G) - \mathrm{PCT}^u(G)|}{\mathrm{PCT}^d(G)}.$$

1) Effect of Network Density and Size on ERR: To evaluate how network density affects the partial cover time estimation, we generate the topology with size $\mathrm{Ar} = 300 \times 300\ \mathrm{m}^2$ and transmission radius $R = 200$. We vary the network density n/Ar by increasing the total number of nodes as $n = 30, 60, 120, 240, 540$.

Fig. 9 shows, given a fixed topology density (30 nodes per $300 \times 300\ \mathrm{m}^2$), how network size affects the absolute error of the PCT upper bound estimation. We see that the absolute error increases as the topology scales up. Let c represent the topology cover ratio of query algorithm, we see that larger c leads to higher absolute error of PCT. When topology size grows up to 540 with 80% cover ratio, the PCT obtained using the undirected graph-based random walk theory can be more than 1500 steps different from result got with the digraph based random walk theory. We could see that symmetrizing the directed links to undirected links in fact introduces high inaccuracy for estimating the PCT upper bound.

To evaluate how network size affects the partial cover time estimation, we choose the node transmission radius of wireless nodes as $R = 200$ m and generate topologies by changing the total number of nodes as $n = 30, 60, 120, 240, 540$ and scaling the topology size up accordingly as $\{300\times300, 424\times424, 600\times600, 848 \times 848, 1273 \times 1273\}\ \mathrm{m}^2$, which guarantees that all topologies generated have the same network density as 30 nodes per $300 \times 300\ \mathrm{m}^2$.

Fig. 10 shows the similar results as in Fig. 9 that absolute error increases as the density goes up. However, we notice that the absolute error of PCT in fact grows much slower than in Fig. 9 as the topology size increases. This happens because increasing density not only introduces more transient states in the random walks, but also establishes more possible paths

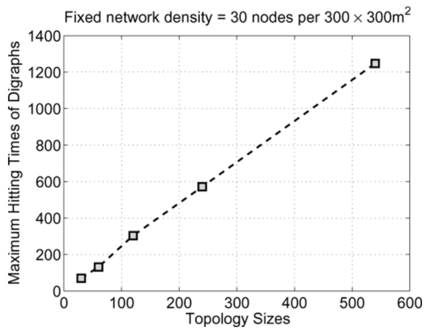


Fig. 11. Effect of topology size on the absolute value of maximum hitting time on digraphs.

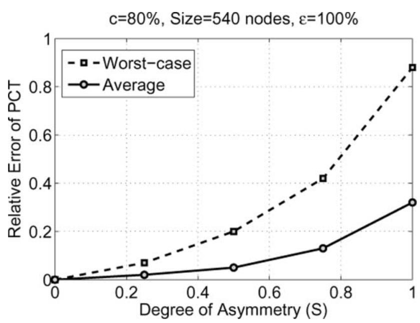


Fig. 12. Effect of degree of asymmetry in estimating the bounds of partial cover time.

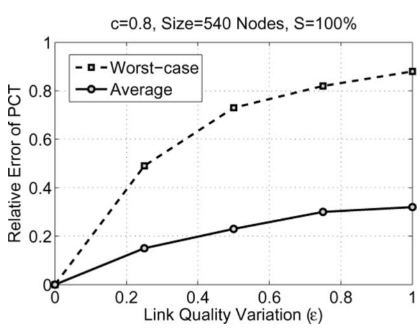


Fig. 13. Effect of (asymmetric) link quality variation in estimating the bounds of partial cover time.

among the topology, which must increase graph connectivity, thus counteract part of the increase of the entire network PCT. Moreover, Fig. 11 shows that as the network size (i.e., number of wireless nodes) increases (with $S\% = 80\%$ asymmetric links and the fixed network density as 30 nodes per 300×300 m^2), the maximum of random walk hitting time on the digraph increases almost linearly.

2) Effect of Degree of Asymmetry and Link Quality Variation: We generate topologies with $n = 540$ nodes in $\mathrm{Ar} = 1800 \times 1800$ m^2 square area, where the node radius is $R = 200$. Fig. 12 shows that when the link quality variation $\epsilon = 1$ and the degree of asymmetry increases as $S = 0$, 25%, 50%, 75%, 100%, the average relative errors of estimated PCT bounds increase up to 33%. For the worst case, the relative errors of estimated PCT bounds could achieve as high as 88%. Fig. 13 shows that while the degree of asymmetry $S = 100\%$, and the link quality variation increases as $\epsilon = 0$, 25%, 50%, 75%, 100%, the average relative errors of estimated PCT bounds increase accordingly.

Hence, the above simulation studies show that it is important to take into account the asymmetric link qualities in wireless networks when assessing and evaluating various wireless protocols and algorithms.

VII. Conclusion

In this paper, we have developed a unified theoretical framework for estimating various transmission costs of packet forwarding and query processing in wireless networks. We illustrated how packet forwarding under each of three routing paradigms—best routing, opportunistic routing, and stateless routing—can be modeled as *random walks on digraphs*. By generalizing the theory of random walks that has primarily been developed for undirected graphs to digraphs, we showed how various transmission costs can be formulated in terms of hitting times and cover times of random walks on digraphs. As representative examples, we applied the theory to three specific routing protocols and a query processing protocol. Extensive simulations demonstrate that the proposed digraph-based analytical model can achieve more accurate transmission cost estimation over existing methods.

Appendix

Proof Sketch for Theorem 4: First, let $\phi = [\phi_i]$ be a column vector whose ith entry is given by $\phi_i = \pi_i^{\frac{1}{2}} s_i^{\frac{1}{2}}$, and let $\Phi = \mathrm{diag}[\phi_i]$ be a diagonal matrix whose ith diagonal entry is ϕ_i. Clearly, $\Phi = \mathrm{diag}[\phi_i] = S^{\frac{1}{2}}\Pi^{\frac{1}{2}} = \Pi^{\frac{1}{2}}S^{\frac{1}{2}}$. Similar to Lemma 1 and Theorem 2, one can verify that the (normalized) cost Laplacian matrix $\tilde{\mathcal{L}}^s$ satisfies the following properties:

$$\tilde{\mathcal{L}^s}^{+}\tilde{\mathcal{L}}^s = \tilde{\mathcal{L}^s}^{+}\tilde{\mathcal{L}}^s = I - \frac{1}{s}\phi\phi^T \tag{47}$$

$$\tilde{\mathcal{L}}^s\phi = \phi\tilde{\mathcal{L}}^s = \tilde{\mathcal{L}^s}^{+}\phi = \phi\tilde{\mathcal{L}^s}^{+} = 0. \tag{48}$$

Equation (47) states that $\tilde{\mathcal{L}^s}^{+}$ is the Green's function for the cost Laplacian operator $\tilde{\mathcal{L}}^s$. Equation (48) implies that ϕ is the left and right eigenvector of $\tilde{\mathcal{L}}^s$ and $\tilde{\mathcal{L}^s}^{+}$ corresponding to the eigenvalue 0.

Given the hitting cost matrix $H^s = [H^s_{ij}]$, it is not too hard to rewrite (5) in the matrix form as follows:

$$(I - P)H^s = SJ - s\Pi^{-1}. \tag{49}$$

Multiplying (49) by $S^{-\frac{1}{2}}\Pi^{\frac{1}{2}}$ from the left, and by $\Pi^{\frac{1}{2}}S^{\frac{1}{2}}$ from the right, we have

$$\begin{aligned} &S^{-\frac{1}{2}}\Pi^{\frac{1}{2}}(I - P)\Pi^{-\frac{1}{2}}\Pi^{\frac{1}{2}}H^s\Pi^{\frac{1}{2}}S^{\frac{1}{2}} \\ &\qquad = S^{-\frac{1}{2}}\Pi^{\frac{1}{2}}(SJ - s\Pi^{-1})\Pi^{\frac{1}{2}}S^{\frac{1}{2}}. \end{aligned}$$

As $\Pi^{\frac{1}{2}}S = S\Pi^{\frac{1}{2}}$, plugging the definitions for Φ and $\tilde{\mathcal{L}}$ yields

$$\tilde{\mathcal{L}}^s\Phi H^s\Phi = \Phi J\Phi - sI. \tag{50}$$

We can now solve the matrix equation (50) using the Green's function, $\tilde{\mathcal{L}^s}^{+}$. Multiplying $\tilde{\mathcal{L}^s}^{+}$ from the left and applying (47) and (48), we have

$$\left(I - \frac{1}{s}\Phi J\Phi\right)\Phi H^s\Phi = -s\tilde{\mathcal{L}^s}^{+}. \tag{51}$$

Since the diagonal entries of H^s are all zero, we can separate the diagonal entries and nondiagonal entries of (51) and solve them separately to obtain Theorem 4. ■

References

[1] D. Aldous and J. A. Fill, “Reversible Markov chains and random walks on graphs,” 1999 [Online]. Available: http://www.stat.berkeley.edu/~aldous/RWG/book.html

[2] C. Avin and C. Brito, “Efficient and robust query processing in dynamic environments using random walk techniques,” in *Proc. 3rd IPSN*, New York, NY, 2004, pp. 277–286.

[3] S. Banerjee and A. Misra, “Minimum energy paths for reliable communication in multi-hop wireless networks,” in *Proc. ACM MobiHoc*, 2002, pp. 146–156.

[4] S. Biswas and R. Morris, “ExOR: Opportunistic routing in multi-hop wireless networks,” in *Proc. ACM SIGCOMM*, 2005, pp. 133–144.

[5] D. Boley, G. Ranjan, and Z.-L. Zhang, “Commute times for a directed graph using an asymmetric Laplacian,” *Linear Algebra Appl.*, vol. 435, no. 2, pp. 224–242, 2011.

[6] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, “Trading structure for randomness in wireless opportunistic routing,” in *Proc. ACM SIGCOMM*, 2007, pp. 169–180.

[7] J.-H. Chang and L. Tassiulas, “Energy conserving routing in wireless ad-hoc networks,” in *Proc. IEEE INFOCOM*, 2000, pp. 22–31.

[8] J.-H. Chang and L. Tassiulas, “Maximum lifetime routing in wireless sensor networks,” *IEEE/ACM Trans. Netw.*, vol. 12, no. 4, pp. 609–619, Aug. 2004.

[9] C.-K. Chau and P. Basu, “Exact analysis of latency of stateless opportunistic forwarding,” in *Proc. IEEE INFOCOM*, 2009, pp. 828–836.

[10] C.-K. Chau and P. Basu, “Analysis of latency of stateless opportunistic forwarding in intermittently connected networks,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 1111–1124, Aug. 2011.

[11] F. R. K. Chung, “Spectral graph theory,” American Mathematical Society, 1997 [Online]. Available: http://www.math.ucsd.edu/~fan/research/revised.html, 2nd ed.

[12] F. R. K. Chung, “Laplacians and the Cheeger inequality for directed graphs,” *Ann. Combinat.*, vol. 9, pp. 1–19, Sep. 2005.

[13] F. R. K. Chung and S. T. Yau, “Discrete Green’s functions,” *J. Combinat. Theory*, ser. A, pp. 191–214, Jul. 2000.

[14] D. De Couto, D. Aguayo, J. Bicket, and R. Morris, “High-throughput path metric for multi-hop wireless routing,” in *Proc. ACM MobiCom*, 2003, pp. 134–146.

[15] S. Dolev, E. Schiller, and J. Welch, “Random walk for self-stabilizing group communication in ad-hoc networks,” in *Proc. 21st IEEE Symp. Rel. Distrib. Syst.*, 2002, pp. 70–79.

[16] P. G. Doyle and L. J. Snell, “Random walks and electric networks,” Mathematical Association of America, Jan. 1984.

[17] D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker, “Complex behavior at scale: An experimental study of low-power wireless sensor networks,” Department of Computer Science, University of California, Los Angeles, CA, Tech. Rep. UCLA/CSD-TR, 02-0013, 2002.

[18] C. M. Grinstead and J. L. Snell, “Introduction to probability,” American Mathematical Society, 2006 [Online]. Available: http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html, 2nd edition

[19] B. Hughes, *Random Walks and Random Environments*. Oxford, U.K.: Clarendon, 1995, vol. 2.

[20] D. B. Johnson and D. A. Maltz, “Dynamic source routing in ad hoc wireless networks,” in *Mobile Computing*. Norwell, MA: Kluwer, 1996.

[21] D. Kotz, C. Newport, and C. Elliott, “The mistaken axioms of wireless-network research,” Department of Computer Science, Dartmouth College, Hanover, NH, Tech. Rep., 2003.

[22] D. Koutsonikolas, C. Wang, and Y. Hu, “Efficient network-coding-based opportunistic routing through cumulative coded acknowledgments,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1368–1381, Oct. 2011.

[23] Y. Li, W. Chen, and Z.-L. Zhang, “Optimal forwarder list selection in opportunistic routing,” in *Proc. IEEE MASS*, 2009, pp. 670–675.

[24] Y. Li and Z.-L. Zhang, “Random walks on digraphs: A theoretical framework for estimating transmission costs in wireless routing,” in *Proc. IEEE INFOCOM*, San Diego, CA, 2010, pp. 1–9.

[25] Y. Li and Z.-L. Zhang, “Random walks on digraphs, the generalized digraph Laplacian and the degree of asymmetry,” in *Proc. 7th LNCS WAW*, Stanford, CA, 2010, pp. 74–85.

[26] Y. Lin, B. Li, and B. Liang, “CodeOR: Opportunistic routing in wireless mesh networks with segmented network coding,” in *Proc. IEEE ICNP*, 2008, pp. 13–22.

[27] H. Liu, Z.-L. Zhang, J. Srivastava, and V. Firoiu, “Pwave: A multi-source multi-sink anycast routing framework for wireless sensor networks,” in *Proceedings IFIP-TC6 NETWORKING*, 2007, pp. 179–190.

[28] L. Lovàsz, “Random walks on graphs: A survey,” *Combinatorics, Paul Erdös Is Eighty*, vol. 2, pp. 1–46, 1993.

[29] I. Mabrouki, X. Lagrange, and G. Froc, “Random walk based routing protocol for wireless sensor networks,” in *Proc. ValueTools*, 2007, pp. 1–10.

[30] P. Matthews, “Covering problems for Brownian motion on spheres,” *Ann. Probab.*, vol. 16, no. 1, pp. 189–199, 1988.

[31] C. Pandana and K. Liu, “Robust connectivity-aware energy-efficient routing for wireless sensor networks,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3904–3916, Oct. 2008.

[32] C. E. Perkins and E. M. Royer, “Ad hoc on-demand distance vector routing,” in *Proc. 2nd IEEE Workshop Mobile Comput. Syst. Appl.*, 1999, pp. 90–100.

[33] B. Radunović, C. Gkantsidis, P. Key, and P. Rodriguez, “Toward practical opportunistic routing with intra-session network coding for mesh networks,” *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 420–433, Apr. 2010.

[34] MIT Roofnet “Roofnet,”, 2005 [Online]. Available: http://pdos.csail.mit.edu/roofnet

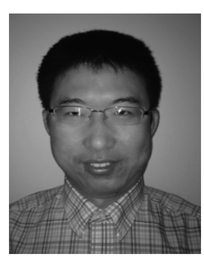
[35] F. Spitzer, *Principles of Random Walk*. New York: Springer-Verlag, 2001, vol. 34.

[36] L. Trefethen and D. Bau, *Numerical Linear Algebra*. Philadelphia, PA: SIAM, 1997.

[37] X. Yang, G. Feng, and D. Kheong, “Call admission control for multi-service wireless networks with bandwidth asymmetry between uplink and downlink,” *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 360–368, Jan. 2006.

[38] M. Zamalloa and B. Krishnamachari, “An analysis of unreliability and asymmetry in low-power wireless links,” *Trans. Sensor Netw.*, vol. 3, no. 2, 2007, Article no. 7.

[39] G. Zhou, T. He, S. Krishnamurthy, and J. Stankovic, “Models and solutions for radio irregularity in wireless sensor networks,” *Trans. Sensor Netw.*, vol. 2, no. 2, pp. 221–262, 2006.



Yanhua Li (S’09) received the B.S. and M.S. degrees in electrical engineering from Sichuan University, Chengdu, China, in 2003 and 2006, respectively, and the Ph.D. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2009, and is currently pursuing the Ph.D. degree in computer science at the University of Minnesota, Twin Cities, Minneapolis.

His research interests are wireless networking and social network analysis in many contexts, including communication protocols design and high-performance networking scheduling strategies, social graph analysis, and measurements.

Zhi-Li Zhang (M’97–SM’11–F’12) received the B.S. degree from Nanjing University, Jiangsu, China, in 1986, and the M.S. and Ph.D. degrees from the University of Massachusetts, Amherst, in 1992 and 1997, respectively, all in computer science.

In 1997, he joined the Computer Science and Engineering faculty with the University of Minnesota, Minneapolis, where he is currently a Professor. From 1987 to 1990, he conducted research with the Computer Science Department, Aarhus University, Aarhus, Denmark, under a fellowship from the Chinese National Committee for Education. He has held visiting positions with Sprint Advanced Technology Labs, Burlingame, CA; IBM T. J. Watson Research Center, Yorktown Heights, NY; Fujitsu Labs of America, Sunnyvale, CA; Microsoft Research China, Beijing, China; and INRIA, Sophia-Antipolis, France.