The Impact of Blacklisting on Data Gathering Trees in Wireless Sensor Networks

Rahul Urgaonkar and Bhaskar Krishnamachari
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089
Email: {urgaonka, bkrishna}@usc.edu

Abstract— **In realistic wireless sensor networks, lossy unreliable links and high density pose considerable challenges to efficient routing and data gathering. Blacklisting techniques provide a mechanism to address these challenges by removing weak links and reducing networking complexity, however they run the risk of reducing path optimality and introducing network disconnections. We undertake an original analysis of blacklisting techniques on two tree structures of particular importance for routing in wireless sensor networks: the Directed Minimum Spanning Tree (DMST), which is best suited for data gathering with aggregation, and the Shortest Path Tree (SPT), which is best suited for data gathering without aggregation. Using a realistic model of link loss statistics, we show interesting tradeoffs between the connectivity and optimality of these structures for different blacklisting techniques. For the specific model considered, we show that up to 70% of the links can be blacklisted with negligible impact on the optimality and connectivity of the SPT with even higher degree of blacklisting possible for the DMST.**

I. Introduction

Tree-based routing techniques are a natural choice for a large class of wireless sensor network applications that require periodic data gathering. These networks are envisioned to be densely deployed in an ad hoc fashion without much control over the placement of the individual sensors. The high density of these devices would mean that, on an average, a node can communicate with a large number of other nodes. This gives rise to the problem of neighborhood table management which is to determine a small and finite subset of neighbors, irrespective of network size, whose information is sufficient for the purpose of data gathering (see for example [1]).

One approach to neighborhood table management in wireless sensor networks in the presence of unreliable links is to *blacklist* a subset of bad links and use only the remaining better links. In the context of cost-based routing, at the node level, this translates into each node discarding some of its high-cost outgoing links and not using them for routing. The key intuition behind this approach is that a low-cost path between any pair of nodes is unlikely to have a high-cost link as a component. Thus, low-cost paths (based on a cost metric) constructed *after* blacklisting are likely to be same as or close to the paths constructed without any blacklisting.

In this paper, we study the impact of blacklisting on data gathering trees in wireless sensor networks. Specifically, we consider two standard edge-weighted tree structures, namely the Directed Minimum Spanning Tree (DMST) and the Shortest Path Tree (SPT) on the network graph. Minimum Spanning Tree (MST) and Shortest Path Tree (SPT) are two basic Tree structures that have been extensively used for Tree-based routing in traditional networks. Note that, for directed graphs, the tree structure corresponding to MST over undirected graph is called Directed Minimum Spanning Tree (DMST). We illustrate the DMST and SPT rooted at node R for a directed graph using an example in Fig 1, 2, 3.

In wireless sensor networks, where data aggregation is a highly promising technique for reducing the cost of data gathering, these structures become even more important. The DMST is the optimal edge-weighted tree structure when perfect aggregation is possible. This is because under perfect aggregation, the optimal tree structure is the one with minimum *total* cost, and DMST is such a structure by definition. Similarly, the SPT is the optimal structure (in terms of cost) when no aggregation is performed. Again, this is because under no aggregation, the optimal structure minimizes the cost from *each* node to the sink.

The main advantages offered by a blacklisting-based approach are:

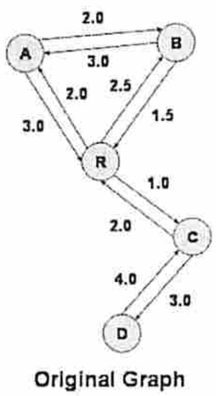

Fig. 1. Example showing a directed edge-weighted graph with root at node R.

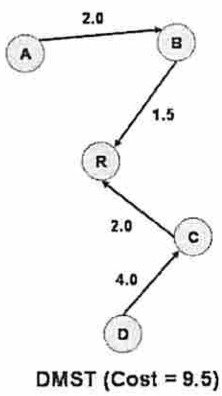

Fig. 2. Directed Minimum Spanning Tree (DMST) rooted at R for the graph in Fig 1.

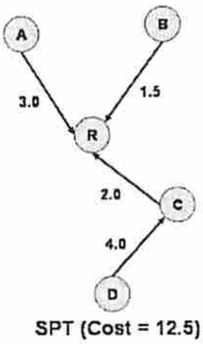

Fig. 3. Shortest Path Tree (SPT) rooted at R for the graph in Fig 1.

1) Reduced cost of building standard tree structures: The complexity of all the standard algorithms for constructing these tree structures (e.g. Dijkstra's algorithm for SPT or Edmonds' algorithm for DMST rooted at a sink) is proportional to number of edges in the network graph ([11], [12]). Blacklisting reduces this complexity as it essentially removes some edges from the graph.
2) Reduced storage requirements and easier neighborhood table management: Typical sensor network applications would require dense deployments of sensor nodes. In such scenarios, the average number of neighbors that a node can communicate with can be very high. This would lead to a bigger neighborhood table which would increase storage requirements. Blacklisting a subset of the outgoing links would reduce the table size and make the table management easier.
3) Reduced cost of flooding: Flooding is one of the basic communication primitives in a broadcast medium like wireless. However, it can be an expensive operation because of repeated broadcasts by nodes. Blacklisting a subset of the links would mean that those links are not used for broadcast and thus, the cost of flooding would be reduced.

Note that a blacklisting-based approach also has the following disadvantages:

1) Disconnected Network Graph: Blacklisting links could lead to scenarios where it is not possible to construct *any* path between a pair a nodes.
2) Suboptimal Tree structures: It is also possible that the tree structures constructed after blacklisting are worse than the optimal trees that are obtained with no blacklisting.

With these ideas, the problem that we consider in this paper can be now be formulated. We model the wireless sensor network as an edge-weighted directed graph where the sensor nodes are represented by vertices and the communication links between them are represented by edges of the graph. These links are assigned a weight that represents the expected number of transmissions required to successfully send a packet over a link. We use an empirically obtained model (see section III) for the link characteristics which is used to obtain these weights.

Given this model, we ask the following questions:

1) How much can we blacklist without disconnecting the graph?
2) What are the tradeoffs between the degree of blacklisting and optimality of the tree structures

constructed after blacklisting?

3) What policies/strategies can be used for blacklisting?

We investigate these issues in the remaining sections which are organized as follows. To gain insights into the connectivity-optimality tradeoffs, we first consider similar scenarios in Geometric Random Graphs (GRG) in section II. To generalize to realistic scenarios, we use an empirical model for link level dynamics which is discussed in section III. We then propose four blacklisting schemes and compare them using extensive simulations followed by a discussion on the results in section IV. The related work is presented in section V before concluding in section VI.

II. Ideas from Geometric Random Graphs

It can be seen that when a blacklisting policy progressively removes high-cost links, the graph would remain connected as long as the worst edge of the DMST or SPT has not been removed. Moreover, the DMST and SPT for the graph obtained after removing the links would be same as those for the original graph. Thus, for such a blacklisting policy, the first question can be answered as follows: We can blacklist all the links that are worse than the worst link of the DMST/SPT. And we are guaranteed that the optimality of these tree structures would be preserved.

The properties of the worst edge of the MST of a Geometric Random Graphs (GRG) have been extensively studied. Note that in a GRG, the cost associated with an edge is its Euclidean length.

It has been shown in [6], [7] that the longest edge of the MST in a GRG is always the critical radius required for connectivity. The connectivity problem in geometric random graphs has been addressed independently in [9] and [10]. Gupta and Kumar [9] show that if n nodes are placed uniformly and independently in a disc D of unit area in $\Re^2$, and each node transmits at a power level so as to cover an area of $\pi R^2 = (log(n) + c(n))/n$, then the network is connected with probability asymptotically tending to one if and only if $c(n) \rightarrow \infty$. Penrose [10] has shown that the longest edge M_n of the minimum spanning tree of n points randomly and uniformly distributed in a unit square S satisfies the $lim_{n \rightarrow \infty} Pr(n\pi M_n^2 - log(n) \leq \alpha) = e^{-e^{-\alpha}}$ for any real number α.

Based on the above, we can conclude the following:

1) In a GRG, we can blacklist all the links longer than the longest edge of the minimal spanning tree without getting disconnected.
2) Moreover, the minimal spanning tree constructed after blacklisting is guaranteed to be same as the optimal MST as long as the graph is connected.
3) Since the longest edge of the MST is also the critical radius for connectivity, this value quantifies the blacklisting limit.

We note that MST on a undirected graph can be shown to be unique if the edge-weights are all different. On the other hand, DMST on a directed graph would be different for different choices of the root node. However, DMST for a *fixed* root node would still satisfy the above properties.

The Shortest Path Tree is another tree structure of interest in cost-based routing. This structure is different from an *undirected* MST in the following fundamental ways:

1) The *undirected* MST is defined for the entire graph and it can be shown to be unique if the edge-weights are all different. On the other hand, SPT is defined specific to a source/sink node. Thus, it is expected to be different for different source/sink nodes.
2) MST minimizes the *total* cost of the tree structure whereas SPT minimizes the cost of path from *each node* to the source/sink node.

These properties make the characterization of the longest edge of the SPT analytically difficult. Specifically, adding edges to the graph would not preserve the SPT. Thus, blacklisting of links from the graph can lead to suboptimal shortest path trees. Note that the connectivity condition is still the same, i.e., we can blacklist up to the critical connectivity radius and still remain connected. In this limit, the MST becomes the SPT for the blacklisted graph.

The properties of the MST and SPT for a randomly chosen node in a GRG are illustrated in Fig. 4. Here, the transmission radius R associated with the GRG is gradually reduced till the point at which the graph becomes disconnected. The end points of the curves represent the disconnected graph, where the cost of MST and SPT becomes infinite. As discussed above, it can be seen that while the cost of MST always remains optimum, the cost of SPT shows a transition from optimum to sub-optimum.

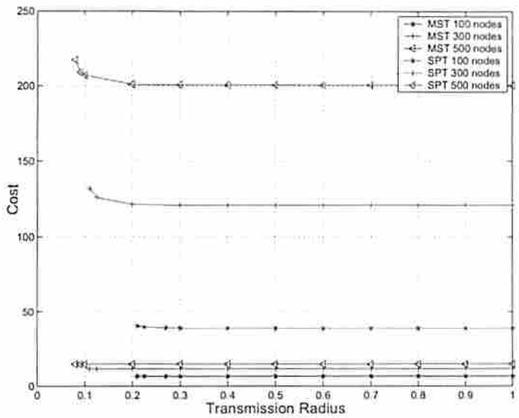

Fig. 4. Cost of Minimum Spanning Tree (MST) and Shortest Path Tree (SPT) rooted at a randomly chosen node for a GRG on square of unit area

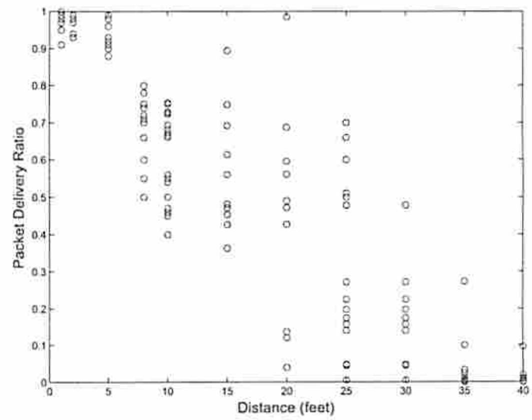

Fig. 5. A scatter plot showing packet delivery ratios with distance

III. A Realistic Model for Link Layer

The GRG model assumes that nodes within each other's transmission range can communicate with each other perfectly. On the other hand, nodes with distance exceeding the radius cannot communicate at all. Although useful for analysis, this is a very simplistic view of the environment. Recent studies (see Section V) have shown that the real wireless medium can deviate highly from this ideal scenario. Thus, cost-based routing where distance is taken as the cost metric is not a good approach. The authors of [1], [5] propose another metric (they call it MT and ETX respectively) which takes into account link layer dynamics. This metric denotes the *expected* number of transmissions (including retransmissions) required to successfully deliver a packet over a link. This metric is thus a good indicator of the energy-cost of packet transmission associated with the link. We therefore use this metric in our simulations in Section IV.

To capture the effects of lossy channels in realistic sensor networks, we used an empirical model obtained by curve fitting on the data collected by Woo *et al.* [1]. In this model, the bit error rate on a link is assumed to be a Gaussian random variable whose mean μ is the following function of the distance d (measured in feet)

$$\mu(d) = \begin{cases} 2.0\times10^{-4.0} + 8.0\times10^{-4.0} & d < 11.0 \\ \frac{d^{2.17}}{5.0\times10^{4.0}} & d < 40.0 \\ 0.0 & d \geq 40.0 \end{cases} \tag{1}$$

and the standard deviation σ is related to the distance d as

$$\sigma(d) = \begin{cases} \frac{e^{-\frac{(d-\alpha)^{2.0}}{2.0\times\beta^{2.0}}}}{7.0\times\beta} & d < 40.0 \\ 0.0 & d \geq 40.0 \end{cases} \tag{2}$$

Here $\alpha = 23.0$ and $\beta = 7.0$ for the specific environment considered in [1].

Using the above model, bit error rates for all the links in both the directions can be obtained. Note that this model allows us to take into account the asymmetric nature of links, which is a commonly observed characteristic of link layer in real sensor networks ([4], [1], [5]). Having obtained the bit error rate BER, the packet delivery ratio on a link can be calculated as $(1.0 - BER)^{size}$ where $size$ is the packet size in bits. Fig. 5 shows a scatter plot of the packet delivery ratios with increasing distance using this model.

IV. Simulations and Results

In this section, we describe the simulations carried out to examine the connectivity-optimality tradeoffs associated with various blacklisting techniques on the two tree structures of interest to us, i.e., DMST and SPT. Specifically, we consider four blacklisting schemes that differ in the way they find the high-cost edges and compare their performance in terms of two metrics: cost of the tree structure generated after blacklisting and connectivity of the tree structure.

A. Methodology

In the simulations, in each run, n nodes were uniformly and randomly distributed in a square of dimension 100 feet X 100 feet. The bit error rates for the links were obtained using the model described in section III. Using these bit error rates, the forward delivery ratio P_f and backward delivery ratio P_b for each link were obtained. The expected number of transmissions (including re-transmissions) required for successful packet delivery on a link is given by $\frac{1.0}{P_f \times P_b}$ ([1], [5]). This metric was assigned as the cost of each link. In the simulations, we assume the packet size to be 240 bits. Further, we assume that acknowledgement packet ACK is sent on the reverse path and its size is taken to be 40 bits.

With this setup, the experiment consisted of constructing the two tree structures of interest, the Directed Minimum Spanning Tree and the Shortest Path Tree (rooted at a randomly chosen node) on this edge-weighted graph after pruning it using different blacklisting policies. Each policy *blacklists* a subset of the edges in the graph based on some metric. The objective is to compare the cost of these tree structures for the new graph with the original ones. These trees were constructed using Dijkstra's algorithm for Shortest Path Tree and Edmonds' algorithm for Directed Minimum Spanning Tree ([11], [12]).

The percentage of links that were blacklisted was taken as the common metric across the four different policies. This was calculated as the ratio of the number of links that were blacklisted to the total number of links. Links with a cost of over 1000.0 were not considered in calculating this percentage.

Since one possible drawback of blacklisting is disconnected graphs, the probability of this event is another quantity of interest. This was obtained as the ratio of the number of runs in which the graph got disconnected to the total number of runs.

B. Blacklisting Policies

We now discuss the four blacklisting strategies that were considered in this study.

1) *Global Blacklisting*: In this strategy, a cost threshold is set up globally. Each node blacklists *all* of its outgoing links with cost greater than this threshold.
2) *Local Blacklisting*: In this strategy, each node *locally* blacklists a certain percentage of its outgoing links, starting with the higher cost links.
3) *Hybrid Blacklisting*: This strategy combines the first two methods. Here, starting with the higher cost links, each node blacklists a certain percentage of links only if their cost is more than the globally set threshold.
4) *All-Except-Best-K Blacklisting*; Here, each node keeps its best K links, blacklisting all others.

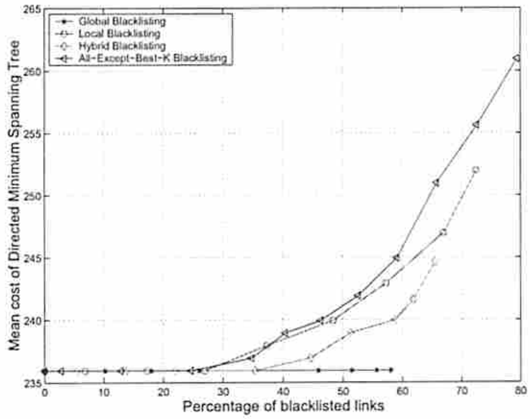

Fig. 6. Mean cost of Directed Minimum Spanning Tree (DMST) for different Blacklisting policies. The DMST is rooted at a randomly chosen node on 100 node graphs with randomly chosen node locations

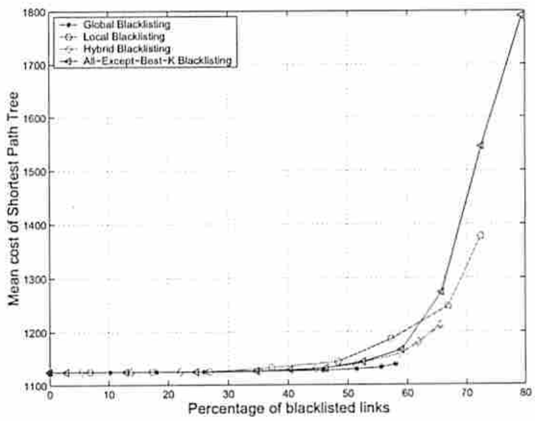

Fig. 7. Mean cost of Shortest Path Tree (SPT) for different Blacklisting policies. The SPT is rooted at a randomly chosen node on 100 node graphs with randomly chosen node locations

C. Results and Discussion

Fig. 6 and 7 show the mean cost of the DMST and SPT respectively for 100 node graphs over all the runs for the four strategies. The plots were obtained up to those levels of blacklisting beyond which the probability of getting disconnected exceeds 0.5. Fig. 8 shows the probability of getting disconnected at different levels of blacklisting for these strategies.

It can be observed in Fig. 6 that in the Global blacklisting

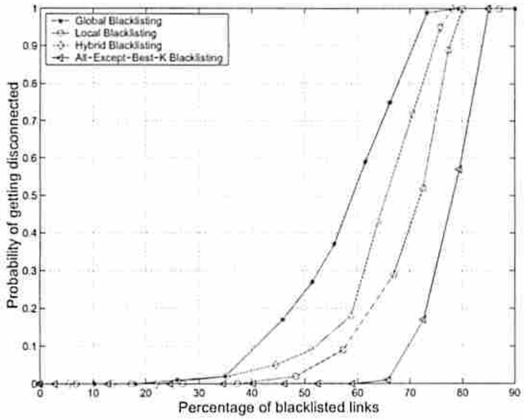

Fig. 8. Probability of getting disconnected on 100 node graphs for different Blacklisting policies

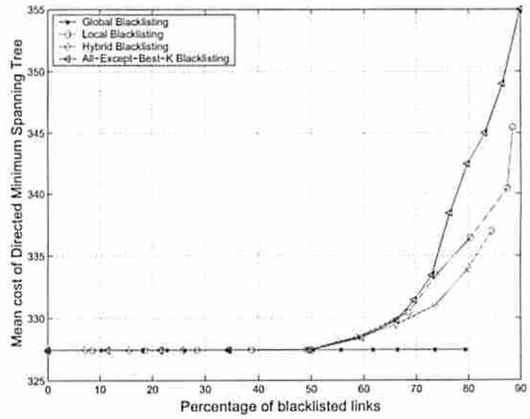

Fig. 9. Mean cost of Directed Minimum Spanning Tree (DMST) for different Blacklisting policies. The DMST is rooted at a randomly chosen node on 200 node graphs with randomly chosen node locations

scheme, the mean cost of DMST remains unchanged. This is similar to the behavior of the undirected MST over GRG. In the other schemes, the mean cost starts to increase with increased levels of blacklisting. This increase is most prominent in the All-Except-Best-K blacklisting scheme, followed by Local blacklisting and then Hybrid blacklisting.

Fig. 7 for SPT shows similar trends, albeit with a subtle difference. We see that with increased levels of blacklisting, even the Global blacklisting scheme causes suboptimal SPTs. However, the cost of the SPT is much better than those obtained under other schemes.

If we observe Fig. 8, we find that the connectivity properties exhibit a reverse trend. Given a level of blacklisting, the All-Except-Best-K blacklisting scheme shows highest probability of remaining connected, followed by Local, Hybrid and Global schemes.

This shows that there exists a tradeoff between the degree of blacklisting and optimality of the tree structures under these blacklisting schemes. While the Global scheme promises best performance in terms of cost of the tree, it is poorest in terms of connectivity guarantees. The All-Except-Best-K blacklisting scheme shows best probability of staying connected, at the cost of sub-optimal trees which can have substantially higher costs.

Fig. 9, 10 and 8 show the observations for 200 node graphs. While the nature of observations is similar, it should be noted that the blacklisting levels for optimality and connectivity higher than the 100 node case. This possibly represents an optimistic trend: we can blacklist an increasing percentage of links with increasing densities.

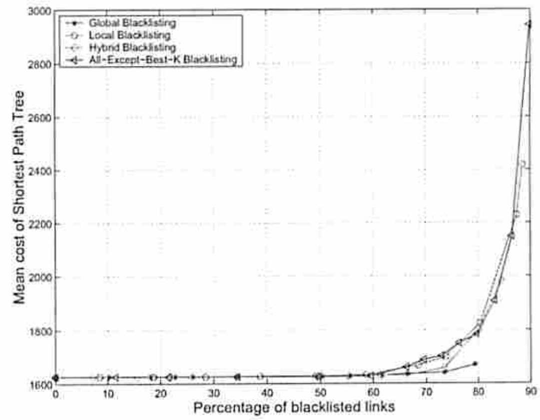

Fig. 10. Mean cost of Shortest Path Tree (SPT) for different Blacklisting policies. The SPT is rooted at a randomly chosen node on 200 node graphs with randomly chosen node locations

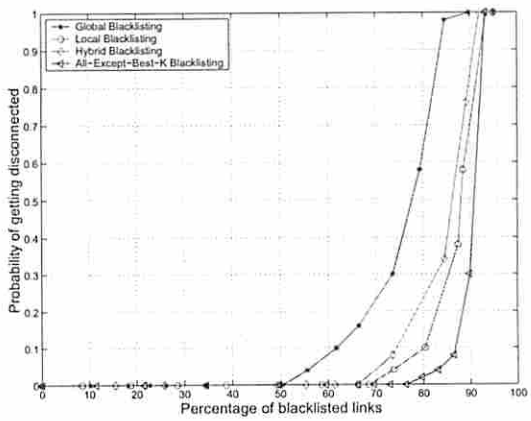

Fig. 11. Probability of getting disconnected on 200 node graphs for different Blacklisting policies

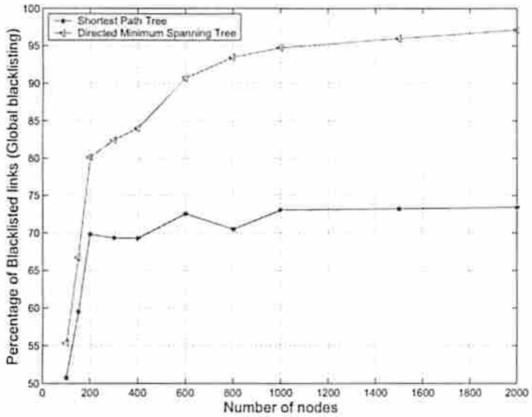

Fig. 12. Percentage of Blacklisting to maintain optimality of DMST and SPT for varying number of nodes

To explore this possibility, we plot the blacklisting levels for optimality of DMST and SPT under Global policy for increasing node densities (Fig. 12). Note that the curve for connectivity is same as the curve for DMST optimality.

Fig. 12 shows an interesting trend. While the percentage of blacklisting that still maintains optimality of the DMST *increases* with node density, the percentage of blacklisting that maintains optimality of the SPT remains almost constant with increasing densities (70% in Fig. 12). This observation is not very optimistic in terms of scalability of neighborhood table management. However, we must point out that if we are ready to tradeoff optimality to some extent, we can achieve levels of blacklisting that scale with node density. It would be an interesting future work to quantify this tradeoff.

V. Related Work

Experimental studies in recent papers [1], [2], [3], [4], [5] suggest that the link layer characteristics in a real wireless medium can deviate substantially from the ideal *connectivity within R* model. Specifically, these non-ideal characteristics become even more prominent in sensor networks with their low-power radio devices and high node densities. The authors of [1], [5] propose a new metric which captures the effect of link layer dynamics on routing decisions. This metric denotes the *expected* number of transmissions (including retransmissions) required for delivering a packet over a link and is thus a good indicator of the energy-cost of transmission over a link. We use this metric as the weight of an edge of the network graph in our simulations.

The issue of neighborhood table management has been addressed in [1]. This issue becomes important in sensor networks where high node densities and non-ideal link characteristics mean a potentially large number of neighbors per node. The problem considered is how to determine a subset of neighbors that are most useful for routing. The authors propose a combination of *insertion, eviction and reinforcement* policy for this purpose. Specifically, they fix the neighborhood table size and compare different policies based on the fraction of time they yield good neighbors (see [1] for details). However, the tradeoff between the degree of blacklisting and the connectivity and optimality of the underlying graph structure is not considered. The blacklisting approach to neighborhood table management proposed in this paper is different from [1] because here, a link once blacklisted, is not considered for insertion into the neighborhood table.

The idea of blacklisting has been considered in Dynamic Source Routing (DSR) [8], a routing protocol proposed for mobile ad-hoc networks. In DSR, to deal with asymmetric links, each node maintains a *blacklist*, which lists immediate neighbors with unidirectional links to the node. These are links over which the node might receive broadcast requests, but which are unsuitable for unicast traffic. Nodes are added to or removed from this list based on the success of packet forwarding.

VI. Conclusions and Future Work

In this paper, we examined the impact of blacklisting-based approaches to data gathering trees in wireless sensor networks. These schemes exploit the fact that it is possible to construct low-cost tree structures that are same as or close to the optimal structure even if a substantial number of high-cost links are removed from the network graph. The main advantages of these techniques are reduced complexity of constructing and maintaining such trees besides scalable neighborhood table management. Specifically, we considered two important standard tree structures of interest, namely the Directed Minimum Spanning Tree (DMST) and the Shortest Path Tree (SPT). In the context of data-aggregation, these assume even more importance.

Using a realistic model of link loss statistics, we showed interesting tradeoffs between the connectivity and optimality of these structures for different blacklisting techniques which differ in the way they decide on the set of links to be blacklisted. It was found that while the

Global policy performs best in terms of optimality, its connectivity guarantees are the poorest. The All-Except-Best-K blacklisting shows best connectivity properties, though the tree structures have higher costs than the optimum ones. However, even for the Global policy, we showed that up to 70% of the links can be blacklisted with negligible impact on the optimality and connectivity of the SPT with even higher degree of blacklisting possible for the DMST. Furthermore, if we are ready to tradeoff optimality to some extent, we can achieve levels of blacklisting that scale with node density.

Our study provides interesting insights into the optimality-connectivity tradeoffs for the two tree structures considered using simulations. Quantifying these tradeoffs analytically is a challenging future work. We would also like to incorporate effects of temporal dynamics of the link layer in future models.

ACKNOWLEDGEMENTS

We would like to thank Marco Zuniga, a member of the Autonomous Networks Research Group, EE-Systems, USC for his help with the empirical link model.

REFERENCES

[1] A. Woo, T. Tong, and D. Culler, "Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks," *The First ACM Conference on Embedded Networked Sensor Systems (Sensys'03)*, Los Angeles, CA, USA. November, 2003.

[2] 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," UCLA CS Technical Report UCLA/CSD-TR 02-0013, 2002.

[3] D. Lal, A. Manjeshwar, F. Herrmann, E. Uysal-Biyikoglu, A. Keshavarzian, "Measurement and Characterization of Link Quality Metrics in Energy Constrained Wireless Sensor Networks," *IEEE Globecom* 2003

[4] J. Zhao and R. Govindan, "Understanding Packet Delivery Performance In Dense Wireless Sensor Networks," *The First ACM Conference on Embedded Networked Sensor Systems (Sensys'03)*, Los Angeles, CA, USA. November, 2003.

[5] Douglas S.J De Couto, D. Aguayo, J. Bicket, and R. Morris, "A High-Throughput Path Metric for Multi-Hop Wireless Routing," *MOBICOM 2003*, San Diego, Sept. 2003.

[6] M. Sanchez, P. Manzoni, and Z. Haas, "Determination of Critical Transmission Range in Ad-Hoc Networks," *Multiaccess, Mobility and Teletraffic for Wireless Communications (MMT'99)*, 1999.

[7] R. Ramanathan and R. Rosales-Hain, "Topology Control of Multihop Wireless Networks using Transmit Power Adjustments," *IEEE INFOCOM*, 2000.

[8] David B. Johnson, David A. Maltz, and Yih-Chun Hu, "The Dynamic Source Routing protocol for mobile ad hoc networks (DSR)," Internet draft (work in progress), IETF, April 2003. http://www.ietf.org/internet-drafts/ draft-ietf-manet-dsr-09.txt.

[9] P. Gupta and P. R. Kumar, "Critical Power for Asymptotic Connectivity in Wireless Networks," *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming, W. M. McEneaney, G. Yin, and Q. Zhang (Eds.)*, 1998

[10] Mathew Penrose, "The Longest Edge of the Random Minimal Spanning Tree," *Annals of Applied Probability*, vol. 7, pp. 340-361, 1997

[11] R. Gallager and D. Bertsekas, *Data Networks*, 2nd edition, PrenticeHall, 1995

[12] E. Lawler, *Combinatorial Optimization: Networks and Matroids*, Saunders College Publishing, 1976