Probabilistic Inference of Lossy Links Using End-to-End Data in Sensor Networks

Wei Zeng, Bing Wang, Krishna R. Pattipati
University of Connecticut

1. INTRODUCTION

Lossy links used in a sensor network affect network performance, and hence need to be detected and repaired [1, 2]. One approach to detect lossy links is that each node monitors the loss rates on its neighboring links and reports them to the sink. This approach, although straightforward, causes large amount of traffic. Another approach to detect lossy links is through end-to-end data that are transmitted periodically from sources to the sink(s) [3, 1, 2]. This end-to-end approach has the advantage of not generating any additional monitoring traffic. The challenge is, however, to develop accurate inference algorithms for lossy link detection based on end-to-end measurements.

Techniques for lossy link inference in wired networks cannot be applied directly to sensor networks since the topology in a wired network is typically a static tree while the topology in a sensor network is a reverse broadcast tree that changes over time. The studies in [3, 1, 2] develop inference algorithms for sensor networks. The techniques in [3, 1] heavily rely on a data aggregation procedure. Furthermore, their assumption of a fixed tree limits their applicability. The study in [2] considers dynamic network topologies. Their schemes, however, are based on heuristics.

In this paper, we formulate and solve an optimization problem to detect lossy links using end-to-end data in sensor networks, taking account of dynamic network topologies. Preliminary evaluation results indicate that our approach provides high detection ratio and moderate amount of false positives in a short amount of time.

2. PROBLEM SETTING

Consider a network represented as a graph $G = (V, E)$, where V and E are the set of nodes and links in the graph, respectively. A set of sources, S, transmit data to the sink t, $S \subset V, t \in V$ (our solution can be easily extended to multiple-sink settings). The routing path from a source to the sink may change over time. We assume that, for each source-sink pair, the set of paths used by this pair and the probability of using each path are known (through a path reporting service, e.g., [4]). However, the exact path used at a given point of time is unknown (since the path reporting service can only run at coarse time scales to conserve network energy). Suppose source s sends n_s packets to the sink and the sink receives r_s packets successfully, $r_s \leq n_s$. Let ψ_s denote the *transmission probability* from source s to the sink, that is, $\psi_s = r_s/n_s$.

We assume that packet losses on different links are independent. Under this assumption, only a binary performance characterization is feasible [5]. That is, a link is classified as either *lossy (bad)* or *not lossy (good)*. Existing measurement studies (e.g., [6]) have demonstrated that links in sensor networks are either good or bad. Furthermore, good and bad links are sufficiently distinct. Based on this, we assume that the transmission probability of a link is either larger than α or smaller than β, $\alpha > \beta$. A link satisfying the former is classified as good, while a link satisfying the latter is classified as bad. Our goal is to infer lossy links based on end-to-end measurements.

3. LOSSY LINK INFERENCE

Since we do not know the path used by a source-sink pair at a given point of time, we infer lossy links based on the performance measures of the source-sink pairs. We can show that, under the loss model in Section 2, for an arbitrary s-t pair, we can find a threshold, $c_s \in (0, 1)$, such that $\psi_s < c_s$ *iff* at least one link used by this pair is lossy (the c_s for different s-t pairs can be different). We say the s-t pair is *lossy* (or *bad*) if $\psi_s < c_s$, and *not lossy* (or *good*) otherwise. Let T_g and T_b denote respectively the set of good and bad pairs. Our goal is to infer the most likely set of lossy links, X, that leads to the observed good and bad source-sink pairs. That

CoNEXT'07, December 10-13, 2007, New York, NY, U.S.A.

is to solve

$$\max_{X \subseteq \bar{E}} P(X \mid T_g, T_b). \qquad (1)$$

The optimal solution of (1) is as follows (derivation omitted). Let $\bar{E}$ denote the set of links that are not used by good pairs. Let $x_l = 1$ denote that link l is lossy; $x_l = 0$ denotes otherwise. Then, the most likely set of bad links, $X = \{l \mid x_l = 1\}$, needs to satisfy the following conditions: (1) $X \subseteq \bar{E}$; (2) at least one link in X is used by a bad pair; (3) X maximizes $\sum_{l \in \bar{E}} x_l \ln(p_l/(1-p_l))$, where p_l is the prior probability that link l is lossy and is known *a priori* (e.g., based on historical data). The above is a generalized set-covering problem and is NP-hard. However, there are efficient approximate algorithms for it (e.g., [7]).

Remarks: When all the p_l's are the same and less than 0.5, the optimal solution is equivalent to finding a minimum number of bad links that cover T_b, which reduces to the Smallest Consistent Failure Set algorithm in [5] when each source uses only a single path to the sink.

For sensor networks with high fraction of lossy links, the above solution may lead to low detection ratio. Inspired by [5], we develop the following iterative algorithm to improve the detection ratio. Suppose in the i-th iteration, the set of lossy links $X_i = \{l_1, \ldots, l_{n_i}\}$, ranked from high to low in how well they explain the observation of bad source-sink pairs. We inspect the links in X_i in a greedy manner (an inspection is to compare the inferred result with the ground truth, i.e., local measurement of the link status), starting from link l_1 until finding the first false positive (i.e., a good link judged as bad), link j, $1 \le j \le n_i$; if no false positive is found, we set $j = n_i + 1$. If $j = 1$, we exclude link l_j from $\bar{E}$ and solve the optimization problem again (since the measurements are still valid) to obtain a new set of lossy links and repeat the inspection process. Otherwise (i.e., $j > 1$), we repair links l_1 to l_{j-1} (i.e., pinpoint the root causes and eliminate them), and restart measurements to obtain a new set of good and bad pairs. We repeat the iteration until all bad links are found (i.e., until all s-t pairs are classified as good).

4. PERFORMANCE EVALUATION

We consider a sensor network with 500 sources and a sink. At a certain time, the routes from the sources to the sink form a tree. In each tree, an intermediate node has 2 to 10 children (chosen uniform randomly). Each source has two paths to the sink (from two trees) and has equal probability to choose these two paths. All links have the same prior probability of being lossy. We use a loss model in [3, 1, 2], where good and bad links have transmission probabilities of 0.99 and 0.75, respectively. The fraction of lossy links varies from 0.01 to 0.20. Fig. 1 plots the results of detection ratio and false positive ratio. When not using the iterative algorithm, the detection ratio is close to 1 for very low fraction of lossy links and decreases as the fraction of lossy links increases. When using the iterative algorithm, the detection ratio is 1 with lower false positive ratios than those not using iteration; and the average number of iterations is from 3.0 to 33.5 as the fraction of lossy links increases from 0.01 to 0.20 (figure omitted).

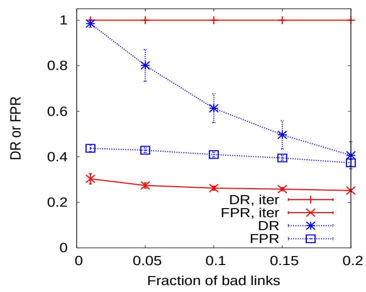

Figure 1: Detection ratio (DR) and false positive ratio (FPR) when using and not using the iterative algorithm.

5. CONCLUSIONS AND FUTURE WORK

We have formulated and solved an optimization problem to detect lossy links using end-to-end data in sensor networks. Preliminary results are encouraging. We plan to perform more comprehensive evaluation and consider other types of lossy models.

6. REFERENCES

[1] Y. Mao, F. R. Kschischang, B. Li, and S. Pasupathy, "A factor graph approach to link loss monitoring in wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, April 2005.

[2] H. X. Nguyen and P. Thiran, "Using end-to-end data to infer lossy links in sensor networks," in *Proc. of IEEE INFOCOM*, 2006.

[3] G. Hartl and B. Li, "Loss inference in wireless sensor networks based on data aggregation," in *Proc. of IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN)*, (Berkeley, CA), April 2004.

[4] T. Schmid, H. Dubois-Ferriére, and M. Vetterli, "Sensorscope: experiences with a wireless building monitoring," in *Proc. of Workshop on Real-World Wireless Sensor Networks*, 2005.

[5] N. Duffield, "Network tomography of binary network performance characteristics," *IEEE Transactions on Information Theory*, vol. 52, December 2006.

[6] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *SenSys*, 2003.

[7] J. E. Beasley and K.Jrnsten, "Enhancing an algorithm for set covering problems," *European Journal of Operational Research*, vol. 58, 1992.