ESTIMATION OF NETWORK LINK LOSS RATES VIA CHAINING IN MULTICAST TREES

Agisilaos-Georgios P. Ziotopoulos, Alfred O. Hero III, Kimberly M. Wasserman

UNIVERISTY OF MICHIGAN ,ANN ARBOR
DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
ANN ARBOR MI 48109-2122
aziot@eecs.umich.edu,hero@eecs.umich.edu,wass@eecs.umich.edu

ABSTRACT

Of increasing importance is estimation of internal link parameters in communications networks. Multicast probes are a way to gather statistics about internal links from edge node measurements. The problem of estimating link loss probabilities for a multicast distribution tree is examined here. Our model assumes loss statistics are distributed to session participants by a network protocol such as RTCP. We propose a decentralized algorithm for ML estimation of the link loss probabilities in a chain of nodes rooted at the source node of the multicast distribution tree and terminating at a given leaf. An expression for the Cramer-Rao bound and an approximate form for the probability distribution function of the estimator are given. The performance of the algorithm is evaluated using computer simulations for a bottleneck detection application.

1. INTRODUCTION

One of the most fundamental problems in operating a computer network is measuring/predicting the traffic intensity and the probabilities of successful transmission of a packet in the network over a certain time interval. Knowledge of this information is useful for a large number of applications that are related to areas such as network design, management, access control, monitoring and pricing. The problem of estimating the traffic intensity in network links based on repeated measurements of edge node traffic has been studied recently [1]. The corresponding area of study is called "Network Tomography". A problem area closely related to the one of Network Tomography is estimating internal link loss probabilities in a network given summary statistics of all nodes in the tree. This problem is examined here using a method based on loss statistics gathered by independent transmissions of probe packets in a multicast distribution tree, where loss statistics are gathered at the leaves of the tree using the RTCP protocol.

2. NETWORK TOMOGRAPHY

The problem of Network Tomography was first proposed in [1]. The name comes from the fact that the internal link traffic rates are estimated based only on estimates of total originating and terminating traffic rates. Recent work has focused on Network Tomography using end-to-end measurements [3,2,4]. The practical implementation of tomography using end-to-end measurents has been hampered for several reasons. The first has to do with the scalability of tomography methods. In a continuously expanding network with densely connected nodes, such as the Internet, the number of internal nodes grows with a rate that makes solving the inverse problem by tomography methods virtually impossible for more than a few dozen nodes. The second reason has to do with the fact that the Internet is a complex heterogeneous network with unknown structure which is administratively diverse. Finally most tomography methods have relied on statistical independence between link loss rates, the so-called spatial independence assumption. This assumption is violated in actual networks due to factors such as "slow restart" after packet loss [5] and multiuser interference in wireless links.

From the references mentioned previously the one that is closest in spirit to this paper is [3]. In [3] a method to infer the internal single-link packet loss characteristics using end-to-end multicast probe measurements is presented. The multicast method of [3] is derived under the assumption that the transmission losses are independent for different links and different probe packets. Only leaf nodes communicate their loss rates and computations are performed at all leaf nodes of the tree to reconstruct the loss rates at internal nodes of the tree. The computational complexity of the algorithm in [3] increases proportionally to 2^k where k is the depth of the tree.

In contrast our multicast method focuses on chains of nodes rooted at the source node of the tree and ending at each leaf node. Computation sare performed at each chain *independently* of the other chains, thus the complexity of our method increases only linearly with respect to the depth of the tree. Independence assumptions are made among the transmission of different probes but no spatial independence assumption is required regarding the transmission of a single probe across subsequent links in the chain. The method described here is based on availability of statistical data on internal link loss rates. Such data is provided by the well known RTCP protocol. RTCP is the current standard for real–time multicast applications [6]. Among other data the protocol provides to the session participants is the measured loss rate for each pair of nodes in a session. A task of interest in many applications is estimation of a bottleneck link in a chain of links. The bottleneck link is defined as the link with maximum loss probability and is frequently where degradation of performance in the network begins. Identification of this link allows a protocol to use this information to take administrative measures, e.g to force rerouting of data around the bottleneck.

This work has been partially supported by the National Science Foundation under grant ANI-9734025 and by the Army Research Office under grant DAAH04-96-1-0337.

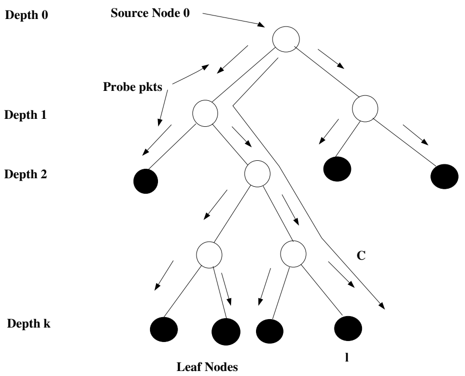

Fig. 1. Multicast Distribution Tree. All nodes in tree participate in a session during which cummulative multicast link losses are reported to all participants, e.g via RTCP. The solid circles denote edge nodes to which multicast packets are sent from the source node. The path C denotes the particular chain investigated by the ML algorithm using loss statistics along chain C only.

3. ESTIMATION OF LINK LOSS PROBABILITES IN A MULTICAST TREE

In this paper we examine the problem of estimating the probability of successful transmission for each link in a multicast distribution tree, based on loss statistics of the number of probe packets sent from the source node (sender) to the leaf nodes(receivers) of the tree. Multicast transmission provides efficient delivery of a packet to an arbitrary number of receivers by replicating the packet within the network at fan-out points along a distribution tree rooted at the transmission source. Like previous authors [3,4] we focus our attention on multicast transmission of packets because the distribution of packets across the links of a multicast tree provides us with a tractable topology on which we can perform mathematical calculations. Note that the abstraction of a multicast distribution tree masks the actual (unknown) topology of the underlying network and provides us with a set of cooperating nodes that exchange statistics. We restrict our attention to the estimation of the set of loss probabilities along the chain of links in the path from the source to a leaf node, using each leaf node separately. This allows us to develop estimation methods which do not require imposition of spatial independence assumptions. Expressions for the Maximum Likelihood (ML) estimators of these quantities are derived and are shown to be unbiased. Also a lower bound, the Cramer-Rao (CR) bound, for the covariance matrix of the estimators is calculated and it is proved that these estimators attain this bound. Finally an asymptotic density for the estimator of the link loss probability of every link is given.

4. STATISTICS OF THE MEASUREMENTS

Assume that the multicast distribution tree topology is like the one depicted in the Figure 1. By the term *path or chain C* we will mean the series of nodes from the source 0 to a specified leaf node l. Node 0 broadcasts N packets to leaf nodes. The number of packets that node i in the path successfully receives is A_i. The A_i's are decreasing monotonically with respect to i i.e $A_0 > A_1 > \dots A_k$. This information is distributed by RTCP to all session participants. The number of packets that go only down to node i and stop idle (die) there are N_i. In contrast to the A_i's the N_i's are not ordered. The reader should carefully note the difference between these two quantities because it is crucial for the rest of the derivations. The N_i packets are a subset of the A_i packets. The following relations will hold:

$$\begin{aligned} N_1 &= A_1 - A_2 \\ N_2 &= A_2 - A_3 \\ &\vdots \\ N_{k-1} &= A_{k-1} - A_k \\ N_0 &= N - \sum_{i=1}^{k} N_i \end{aligned} \tag{1}$$

The event that node i receives successfully a packet will be denoted by $I_i = 1$ (I stands for the indicator function) , and the event that node i does not successfully receive a packet will be denoted by $I_i = 0, \dots, k$. The event $\{I_0 = 1, I_1 = 1, \dots, I_i = 1, I_{i+1} = 0\}$ is the event that a packet sent from the source node will die at node i. The probabilities $\{\theta_i\}$ of these events are parameters called the probe survivor probabilities and are related to the individual link loss probabilities of interest.We note that $\sum_{i=0}^{k} \theta_i = 1$ so it is sufficient to specify the k parameters $\underline{\Theta} = [\theta_0, \dots, \theta_{k-1}]^\top$

For each source-destination path C containing $k + 1$ nodes there are k possible outcomes whenever a packet is sent from the source. Either the packet dies at the first node, or the packet dies at the 2nd node $\dots$, or the packet dies at the kth node. Let $\Delta_i = I_0 I_1 \dots I_i (1 - I_{i+1})$ denote the indicator function of the event that the packet dies at the i-th node. Then trivially, $p(\Delta_1, \dots, \Delta_k, \underline{\Theta}) = \theta_1^{\Delta_1} \dots \theta_k^{\Delta_k}$ where $\sum_{i=1}^{k} \Delta_i = 1, \Delta_i \epsilon \{0, 1\}, \theta_k = 1 - \sum_{i=0}^{k-1} \theta_i$. Under the assumption that the N transmitted probe packets are transmitted independently the joint probability distribution of the number of packets that die at each of the nodes is a multinomial.

$$p(\underline{N}, \underline{\Theta}) = \frac{N!}{N_0! \dots N_k!} \theta_0^{N_0} \dots \theta_{k-1}^{N_{k-1}} (1 - \sum_{i=0}^{k-1} \theta_i)^{N - \sum_{i=0}^{k-1} N_i} \tag{2}$$

where $(\underline{N}) = (N_0, \dots, N_k), \sum_{i=0}^{k} N_i = N$. Note that no spatial independence asssumptions are required for the validity of (2).

The form of the ML estimator for $\underline{\Theta}$ is well known for the multinomial distribution (2) and takes the form:

$$\overset{\wedge}{\underline{\Theta}} = (\frac{N_0}{N}, \dots, \frac{N_{k-1}}{N}) \tag{3}$$

The ML estimator (3) is unbiased and efficient, i.e it's covariance matrix attains the Cramer-Rao bound which is equal to the inverse of the $k \times k$ Fisher information matrix $F_{\underline{\Theta}}^{-1}$. The element of the FIM $F_{\underline{\Theta}}$ in row i and column j is given by the formula:

$$[F_\Theta]_{i,j} = -\mathcal{E}[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(\underline{N}, \underline{\Theta})] \tag{4}$$

$$\frac{\partial \ln p(\underline{N}, \underline{\Theta})}{\partial \theta_i} = \frac{N_i}{\theta_i} - \frac{N - \sum_0^{k-1} N_i}{1 - \sum_0^{k-1} \theta_i} \tag{5}$$

$$\frac{\partial^2 \ln p(\underline{N}, \underline{\Theta})}{\partial \theta_i \partial \theta_j} = \begin{cases} -\frac{N - \sum_0^{k-1} N_i}{(1 - \sum_0^{k-1} \theta_i)^2} & : \ i \neq j \\ -\frac{N_i}{\theta_i^2} - \frac{N - \sum_0^{k-1} N_i}{(1 - \sum_0^{k-1} \theta_i)^2} & : \ i = j \end{cases} \tag{6}$$

$$\mathcal{E}[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\underline{N},\underline{\Theta})] = \begin{cases} -\frac{N}{(1-\sum_0^{k-1}\theta_i)^2}\theta_k = -\frac{N}{\theta_k} & : \quad i \neq j \\ -\frac{N}{\theta_i} - \frac{N}{\theta_k} = -N\frac{\theta_i+\theta_k}{\theta_i\theta_k} & : \quad i = j \end{cases} \tag{7}$$

We observe that as $N \to \infty$ the elements of the inverse Fisher Information matrix $[F_{\underline{\Theta}}]_{i,j}$ tend to zero. This implies that $\hat{\underline{\Theta}}$ is consistent i.e it's covariance goes to zero with N.

5. ESTIMATION OF LINK LOSS PROBABILITES

The main goal of our effort is to calculate the probability of successful transmission for each link in the network. For example if we want to calculate this probability for the link that connects node $i-1$ in a path with node i we are interested in the quantity $\mathcal{V}_i \triangleq Pr(I_i = 1|I_{i-1} = 1), i = 1, \ldots, k$. Applying the law of conditional probability we get:

$$\mathcal{V}_i = \frac{\sum_{j=i}^k \theta_j}{\sum_{j=i-1}^k \theta_j}, i = 1, \ldots, k \tag{8}$$

Thus using the ML invariance property, the ML estimator for the link-loss probability $1 - \mathcal{V}_i$ is specified by the MLE of $\mathcal{V}_i$

$$\begin{aligned} \hat{\mathcal{V}}_i &= \frac{\sum_{j=i}^k \hat{\theta}_j}{\sum_{j=i-1}^k \hat{\theta}_j} = \frac{\sum_{j=i}^k N_j}{\sum_{j=i-1}^k N_j} \\ &= \frac{\sum_{j=i}^k A_j - A_{j+1}}{\sum_{j=i-1}^k A_j - A_{j+1}} = \frac{A_i}{A_{i-1}} \\ &= \frac{\sum_{j=1}^N I_i^{(j)}}{\sum_{j=1}^N I_{i-1}^{(j)}} \end{aligned} \tag{9}$$

Furthermore the CR bound on estimation error for unbiased estimators of $\underline{\mathcal{V}} = [\mathcal{V}_1, \ldots, \mathcal{V}_k]^\top$ is $F_{\underline{\mathcal{V}}}^{-1} = [\bigtriangledown_{\underline{\Theta}} g(\underline{\Theta})] F_{\underline{\Theta}}^{-1} [\bigtriangledown_{\underline{\Theta}} g(\underline{\Theta})]^\top$ where $g(\underline{\Theta}) = [\mathcal{V}_1(\underline{\Theta}) \ldots \mathcal{V}_k(\underline{\Theta})]^\top$. The following recursive formuli will be useful for computing $\bigtriangledown \mathcal{V}_k$: $\mathcal{V}_{k-1} = \mathcal{V}_k^{-1}\frac{\theta_k}{\theta_k+\theta_{k-1}}$, $\mathcal{V}_{k-i} = (\mathcal{V}_k \ldots \mathcal{V}_{k-i+1})^{-1}\frac{\theta_k}{\sum_{j=k-i}^k \theta_j}$

The family of binary random variables (rv's) $\{I_i^{(j)}\}_j$ are indicators that the jth probe has been successfully received by node i or not. We have assumed that this family consists of independent and identically distributed (iid) rv's. The assumption of independence indicates that different transmissions of packets in the network, are independent. This assumption is valid as long as the i-th probe is sent only after the (i-1)st probe has been received and the network is stable over the N probe transmissions. It is up to the protocol to choose the time-separation between subsequent packets so as to achieve temporal independence for the transmission of packets. The assumption that the $\{I_i^{(j)}\}_{j=1}^N$ are identically distributed implies that the network loss behaviour does not change over the probing interval. Although this may not hold for large time periods (there are periods of high congestion in the network and periods of low traffic) this is a reasonable assumption when the multicast transport delays are small and probes are sent in rapid succession.

The mean value of $I_i^{(j)}$ is $\mathcal{E}[I_i^{(j)}] = Pr(I_0 = 1, \ldots, I_i = 1) \triangleq p_i$ and the variance will be $var[I_i^{(j)}] = p_i(1-p_i)$. By applying the Central Limit Theorem (CLT) to the sums of the iid rv's we can approximate the distribution of the numerator and the denominator of (9). Applying the CLT we have the approximation $\frac{1}{\sqrt{N}}\sum_{j=1}^N I_i^{(j)} \sim \mathcal{N}(\sqrt{N}p_i, p_i(1-p_i))$. Under the simplifying assumption that $\mathcal{V}_i$ follows the distribution of the ratio of two independent Gaussians as N increases to infinity, it is straightforward to show that the pdf of the ratio of two indepent Gaussian rv's with means μ_1 and μ_2 respectively and variances ${\sigma_1}^2$ and ${\sigma_2}^2$ respectively, is

$$f_X(x) = \frac{e^{\frac{B^2-AC}{2A}}}{\pi\sigma_1\sigma_2}\left(\frac{e^{-\frac{B}{\sqrt{A}}}}{A} + \sqrt{\frac{2\pi}{A}}\frac{B}{A}sign(B)\left(\frac{1}{2} - Q(\sqrt{\frac{B^2}{A}})\right)\right) \tag{10}$$

where

$$A = \frac{x^2}{\sigma_1^2} + \frac{1}{\sigma_2^2}, B = \frac{\mu_1 x}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}, C = \frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} \tag{11}$$

and $Q(x)$ is the standard Gaussian integral $\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}dw$. Using $\mu_1 = \sqrt{N}p_i, \mu_2 = \sqrt{N}p_{i-1}, \sigma_1^2 = p_i(1-p_i), \sigma_2^2 = p_{i-1}(1-p_{i-1})$ in (10) and (11) we obtain the marginal distribution $f_{\hat{\mathcal{V}}_i}(x)$ for estimate $\hat{\mathcal{V}}_i$ computed from chain C. This can be used to compute the estimator bias, variance and threshold excedance probability and confidence intervals.

6. COMBINATION OF SINGLE CHAIN ESTIMATES

In order to improve the performance of the single chain method it will be necessary to fuse the estimates of common link survival probabilities (θ_i's) obtained from two different leaves (chains). Assume there are two chains C_1 and C_2 that share a common link i with survival probability θ_i. The number of packets that are transmitted down to a node i and die there for both chains $N_i^{(j)}, i = 0 \ldots k, j = 1, 2$ are dependent in a complicated way due to the fact that they share common link information. One approach that would enable us to improve our estimates of $\mathcal{V}_i$, without making any unjustified independence assumptions, would be to use the Best Asymptotic Normal (BAN) property of the ML estimator $\hat{\mathcal{V}}_i$. Let $\hat{\theta}_{C_1}$ and $\hat{\theta}_{C_2}$ be estimates of θ_i obtained from chains C_1 and C_2 terminating at leaves l_1 and l_2 respectively. The BAN property asserts that asymptotically the ML estimators are jointly Gaussian i.e $\sqrt{N}[\hat{\mathcal{V}}_{C_1} - \mathcal{V}_1, \hat{\mathcal{V}}_{C_2} - \mathcal{V}_2]^\top \sim \mathcal{N}(\underline{0}, F_{C_1C_2}^{-1})$ where $\underline{0} = [0, 0]^\top$ and $F_{C_1C_2}$ is the FIM. We can then apply ML estimation to estimate θ_i from $[\hat{\theta}_{C_1}, \hat{\theta}_{C_2}]^\top$.

Using this approach we can also compute the Fisher Information matrix for this model and compare it to the results of the single chain model. By comparing the CRB for the multiple chains to the CRB for a single chain we can estimate the additional number of probes needed for the single chain method to achieve the same performance as the multiple chain method.

7. NUMERICAL RESULTS

In order to evaluate numerically the performance of our method we ran computer simulations in C++. Link loss probabilities were assigned at random and the "bottleneck" was defined as the link with max loss probability. For each link in the chain the probe packet dies or is transmitted successfully to the next node in the node chain, according to a stationary probability over the simulation. In Fig. 2 the probability distribution function (pdf) of the estimator $\hat{\mathcal{V}}_i$ given in (10) is plotted for different numbers of probe packets. Note that as the number of packets increases the pdf concentrates around it's mean value, which supports our argument that the derived estimator is consistent. In Fig. 3 the average value of the estimator $\hat{\mathcal{V}}_i$ is given for different numbers of probe packets. The solid line represents the actual value $\mathcal{V}_i$. Note the close match between $\mathcal{V}_i$ and $\hat{\mathcal{V}}_i$. As the number of probe packets increases the distance between the estimated and the true value decreases, which supports the BAN property that $\hat{\mathcal{V}}_i$ is asymptotically unbiased. In Fig. 4 the percentage ratio of successful bottleneck localizationability events is plotted for different numbers of probe packets. The success ratio of a method that uses uniform guessing among all nodes, is plotted using crosses. We note that as the number of probe packets increases the localizability of the bottleneck link increases too. In all cases the success ratio of our method, regardless of the number of probe packets sent, is greater than the success ratio of the guessing method.

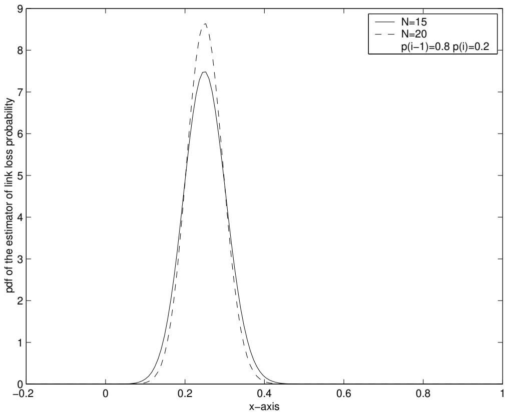

Fig. 2. pdf of the estimator for different number of probe pkts.

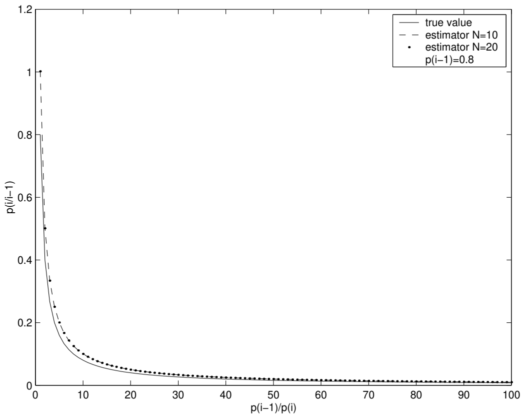

Fig. 3. Average value of the single chain estimator for different numbers of probe pkts.

8. CONCLUSIONS

In this paper we have presented a method to infer the link loss rates in a network using loss statistics gathered from nodes participating in a multicast distribution tree. The assumption that the loss

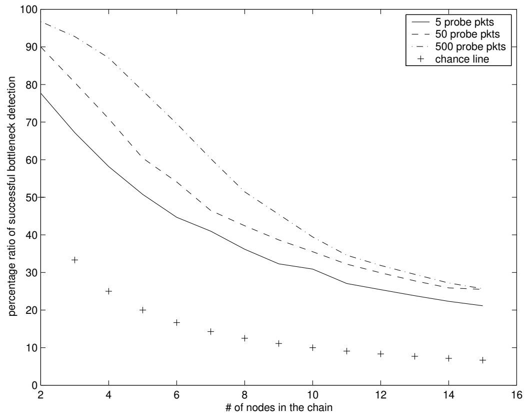

Fig. 4. Percentage of successful bottleneck localization for different values of the the number of probe pkts used.

behaviour of the network does not change over the probing interval is central to our calculations. This assumption does not hold in general for long time periods, thus it constitutes a limitation to the applicability of our method. Our method is suboptimal in performance, since we restrict our attention to estimates of link loss probabilities. However as the method uses consistent estimates, for large number of packets, the estimator variances converge to zero. The advantage of single chain methods is linear complexity with respect to the depth of the tree. In the future we will try to quantify the loss in performance (determined by the Fisher Information matrix) induced by applying our approach compared to the approach in [3] and to the optimal multichain performance. Due to space limitations we have not presented simulations of the multichain fusion method described in Sec 6. This will be presented in a later paper. Another point not taken into consideration is the fact that probe packets compete with background traffic for transmission. We plan to do more extensive simulations with tools such as *ns* so as to include such phenomena.

9. REFERENCES

[1] Y.Vardi, "Network Tomography : Estimating Source-Destination Traffic Intensities From Link Data", Journal of the American Statistical Association March 1996,Vol.91 No 433, Theory and Methods

[2] J. Cao, D. Davis, S. Vander Wiel, B. Yu "Time-Varying Network Tomography : Router Link Data", Bell Labs Tech. Memo, 29 February,2000,

[3] R. Caceres, N.G. Duffield, J. Horowitz and D. Towsley "Multicast-Based Inference of Network-Internal Loss Characteristics", IEEE Transactions on Information theory, Vol.45, No7, November 1999,

[4] S. Ratnasamy, S. McCanne "Inference of Multicast Routing Trees and Bottleneck Bandwidths using End-to-end Measurements", Proceedings of IEEE Infocom '99,

[5] V. Jacobson, "Congestion avoidance and control",, in Proc. ACM SIG-COMM,pp 314-329 Aug. 1988.

[6] IETF RFC 1889, "RTP: A Transport Protocol for Real-Time Applications", January 1996