Measurement-Based Network Monitoring and Inference: Scalability and Missing Information

Chuanyi Ji, *Member, IEEE* and Anwar Elwalid

***Abstract*—Using measurements collected at network monitors to infer network conditions is a promising approach for network-centric monitoring. In this context, an important question arises: given the number and locations of network monitors, how much network management resources (e.g., the number of measurements) are needed to obtain an accurate estimate of network states? We define the scalability of measurement-based network monitoring as the growth rate of the number of measurements required for accurate network monitoring/inference with respect to the size of a network. We develop a framework for investigating the scalability in the context of multicast inference with the monitors at the edges of a network. In such a framework, network monitoring/inference can be formulated as probability density estimation of network states. The growth rate is characterized through the sample complexity, which is the number of measurements needed to accurately estimate the density. The missing data framework is introduced to estimate the growth rate, where the missing data reflect unavailable measurements at the unobservable nodes without resident monitors, and the underlying nodal packet losses. We show that when the missing information is mainly due to the number of unobservable nodes, the number of measurements needed grows linearly with the size of the network, and the measurement-based inference approach is, thus, scalable. When the missing information is mainly due to the underlying nodal packet losses, the number of measurements needed grows faster than linear with the size of the network, and the measurement-based inference approach is, thus, not scalable. Our results provide guidelines for accessing feasibility of the measurement-based inference approach, and the number of probes required. We give numerical examples to illustrate some of our results.**

***Index Terms*—Measurements, missing information, multicast inference, networks monitoring, sample complexity, scalability.**

I. Introduction

SCALABLE NETWORK management approaches are essential for providing reliable services over large Internet protocol (IP) networks. A basic method for achieving desired network reliability and performance is network monitoring [5], [12], [15], where states of a network are frequently monitored. Monitored information can be used to infer quality-of-service (QoS) at a network relating to congestion [16], [5], potential faults [15] and attacks [12]. Inferred QoS provides valuable information for reducing severe degradation of services or potential network failures.

Network centric monitoring approaches [14], where states of the network are monitored rather than individual nodes in isolation, are important for network management. Network states can be considered as a collection of nodal states which may correspond to packet losses or delays at the nodes. Network monitors can reside at all or subset of network nodes. A goal of network-centric monitoring is to infer states of an entire managed network using measurements collected locally at network monitors.[1] In this context the following questions arise: Given the number and locations of monitors, how much network management resources (e.g., the number of measurements and the computation time) are needed to obtain an accurate estimate of network states? How rapidly would the needed resources grow when the size of a network gets larger? For instance, suppose 100 measurements are sufficient to obtain accurate estimates of network states for a small network with ten nodes and two monitors at given locations. If the network becomes 100 times as large (e.g., with 1000 nodes), would the number of measurements needed grow 100 or 2^{100} times?

The main focus of this work is to develop a framework to formally investigate the scalability of measurement-based network monitoring. At the high level, the scalability characterizes the growth rate of the required network management resources with respect to the size of the network, where the resources include the number of measurements needed to obtain an accurate estimate of network performance, and the time spent on computing the estimate. The number of measurements affects network resources spent on management such as the available transmission bandwidth and processing capacity. The growth rates of the number of measurements and the computation time together determine the feasibility of measurement-based network monitoring in large networks. Once known, the growth rate can also be used to estimate on the number of measurements needed to monitor a given network.

The above scalability study is challenging due to the complexity of large IP networks. Thus, this work chooses multicast networks for our initial investigation. We adopt the approach developed in the prior work [5] which establishes a rigorous framework for network inference. Specifically, probe packets are multicasted through a network. Monitors are located only at the multicast receivers (on the edges of a network), recording receptions or losses of probe packets. If a packet is lost at an intermediate node, multicast receivers would not receive the packet. Measurements, thus, contain information on spatially correlated

Manuscript received March 8, 2001; revised December 20, 2001. This work was supported in part by the National Science Foundation under Contract ECS-9908578.

C. Ji is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: jic@ece.gatech.edu).

A. Elwalid is with the Department of Mathematics of Networks and Systems, Bell-Labs Lucent Technologies, Murray Hill, NJ 07974-0636 USA (e-mail: anwar@research.bell-labs.com).

Publisher Item Identifier S 0733-8716(02)05074-6.

[1]Such an inference is possible since nodal packet losses/delays are spatially correlated, and measurements contain information on network states.

nodal packet losses within a multicast tree,[2] and can be used to infer the internal packet losses.

How to formally define the scalability? We show that the measurement-based network monitoring can be formulated as density estimation. The density is the joint probability of network states. The unknown parameters of the density correspond to probabilities of nodal packet losses within a multicast network. The estimation is on the unknown parameters using measurements. The performance of network monitoring can then be defined through the accuracy of a density estimator [1]. The simplest such a performance measure is the estimation error on the unknown parameters which we choose in this work. A sufficient number of samples (measurements) are needed to achieve a desired performance, and called the sample complexity [1]. The sample complexity provides a basis for defining the scalability: the linear growth rate of the number of measurements with respect to the size of a network is the lowest possible. Thus, if the number of measurements grows linearly with respect to the size of a network, the measurement-based network monitoring is scalable; otherwise, it is nonscalable.

How to evaluate the growth rate? The framework of density estimation does not provide conditions on when the linear rate is achievable. We derive the growth rate using the missing data framework [4], where the missing data corresponds to states of internal nodes without monitors. Specifically, the missing information is determined by both the complexity of a multicast network (e.g., the number of internal nodes without monitors residing), and the underlying network losses. The observed information is determined by the number of monitors, and the number of measurements (probe packets). When the size of a multicast network grows, both the missing and observed information grows accordingly but their relative growth rate is determined by the underlying losses. We derive conditions based on underlying losses for uniform multicast networks on the scalability/nonscalability. We show that, asymptotically (for large networks), there exists a critical value for underlying losses. When the loss probability is below this value, the amount of missing information is dominated by the size of the network, and the number of measurements needed scales linearly. When the loss probability exceeds this value, the amount of missing information is dominated by the underlying losses, and the number of measurements grows faster than linear, resulting in nonscalability. These cases are verified by analysis for small and large losses, and by simulation for moderate losses. We also show that the growth rate of the number of measurements is tightly coupled with the computation time. The computation time is characterized by the convergent rate of the expectation-maximization (EM) algorithm, and determined by the amount of missing information.

A closely related work is the multicast inference developed in [5]. This work focuses on developing a multicast inference approach which provides a feasible model for our analysis of scalability. The work also suggests the possibility that the inference might be difficult when the losses are large. This motivates our investigation on the scalability. The scalability is general, and related to fundamental limits of, measurement-based network monitoring. To our knowledge, the scalability has neither been defined nor investigated previously for measurement-based network monitoring/inference. The density estimation introduced in this work establishes a formal setting to address the performance and complexity for measurement-based network monitoring. The sample complexity of density estimation provides the base line (e.g., the best growth rate) for defining the scalability, but is not directly applicable to evaluating the growth rate of the number of measurements. The context of density estimation does not include computation time either. The missing data framework makes the analysis of scalability possible. Specifically, the likelihood function used in the prior work [5] has no close-form expression and, thus, cannot be used to deriving the number of measurements needed. The likelihood function resulting from the missing data formulation has a simple analytical form and, thus, makes it possible to derive the number of measurements needed and a relation to the convergence rate. The missing data formulation also provides a framework to investigate how the missing information affects the performance and complexity of network monitoring, resulting in basic understanding on fundamental limitation of network monitoring. Missing data formulation has been used, with a different focus, in other related work which includes network tomography on estimating traffic matrices [2], approximating multicast through unicast [3], and multicast inference with missing measurements [6].

The rest of the paper is organized as follows. Section II describes the problem of network monitoring in a specific context of multicast inference. Section III presents monitoring in the missing data framework. Section IV introduces density estimation, and defines the scalability. Section V investigates the scalability analytically and numerically. Section VI concludes the paper.

II. Problem

We adopt the same problem set up as in [5]. Consider a full multicast tree[3] shown in Fig. 1. Probe packets are sent from the source on the top of the tree to the rest of the nodes through multicast. Upon receiving a probe packet, an intermediate node replicates the packet and sends it to its children. If the probe packet is lost at an intermediate node, none of its descendants receive the probe. Monitors we use assume a simple functionality of recording the reception of probes packets and, thus, are called passive monitors. They are assumed to be located at all receivers, i.e., edge monitors, and possibly at internal nodes.

A node with a monitor is observable. If a probe is received, the measurement recorded by the monitor[4] is an "1." A node without a monitor is unobservable, i.e., no measurement can be obtained. A set of n measurements, $D_{\mathbf{inc}} = \{X_j^{(q)}, j \in O\}_{q=1}^{n}$, consists of observations of the states of the observable nodes

[2]The use of probe packets also synchronizes the local measurements at monitors.

[3]Each node is assumed to have a unique parent, and there are at least two children for each parent.

[4]We assume that there is a protocol coordinating such a process as in [5], and details are omitted in this paper.

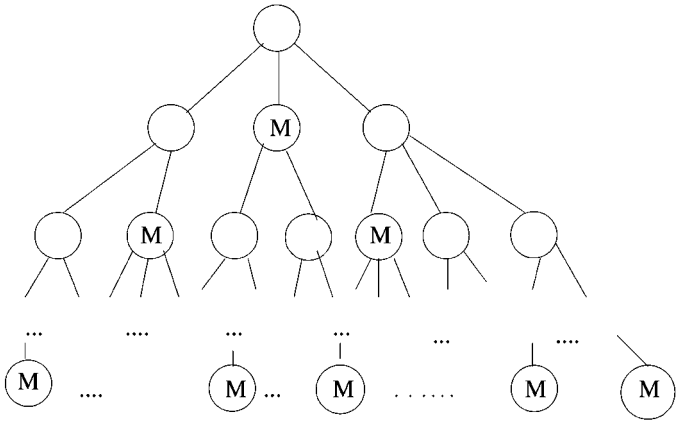

Fig. 1. A multicast tree.

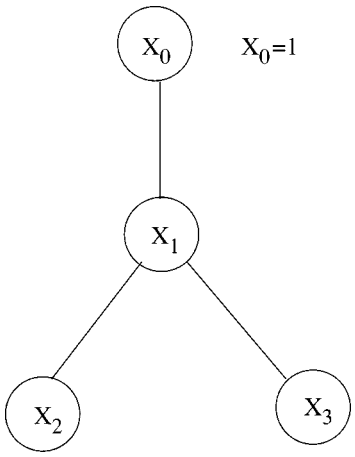

Fig. 2. A simple multicast tree.

using n probes, where O is a set of all observable nodes. The conditional probability, $\alpha_j = \Pr(X_j = 1 \mid X_{f(j)} = 1)$, characterizes the state of node j given that of it parent $f(j)$. A vector $\alpha = (\alpha_j)$ determines the state of a network for all $j \in U \bigcup O$, where U is a set of all unobservable nodes. The problem of network monitoring in the context of multicast inference involves estimating $\alpha_j = \Pr(X_j = 1 \mid X_{f(j)} = 1)$ at the unobservable nodes using a set (D_{inc}) of measurements. Losses of the probe packets are assumed to be conditionally independent, i.e., given a probe is received at a parent node, it is received by its children independently.

III. Missing Data Formulation

Missing data framework [4] which is widely used in statistics is a natural fit to network monitoring with partially observable information. The key idea is to introduce the so-called missing random variables whose information can not be obtained directly but estimated through measurements of a set of observable variables [2], [18].

A. Missing and Complete Variables

In this work, the missing random variables are the states of unobservable nodes, X_j for $j \in U$, and observable variables are the states of observable nodes, X_j for $j \in O$. The set of complete random variables is $X_{\text{com}} = \{X_j, \forall j \in U \bigcup O\}$. The set of actual measurements, D_{inc}, made at the monitors corresponds to the incomplete data in the missing data framework, while the set of fictitious measurements, $D_{\text{miss}} = \{X_k^{(q)}, \forall k \in U\}_{q=1}^{n}$, of the missing variables corresponds to the missing data. Thus, the complete data set is $D_{\text{com}} = \{X_j^{(q)}, \forall j \in O \bigcup U\}_{q=1}^{n}$.

The tasks of monitoring and inference of network states consist of two steps: (a) developing a model, $\Pr(X_{\text{com}})$, called the complete likelihood, and (b) estimating the unknown parameters of the model, α_j for $j \in U$.

B. Model

The complete likelihood function, $\Pr(X_{\text{com}})$, is essentially a model of network states. To determine the form of $\Pr(X_{\text{com}})$, we begin by considering a simple example with one source, one unobservable node and two edge monitors as in Fig. 2. Monitors are located at the receivers (edges), and X_1 is the only missing variable for an internal node without a monitor. The state vector of the network is $X_{\text{com}} = (X_1, X_2, X_3)^t$, where t denotes the transpose of a matrix. Then the probability model is

$$\Pr(X_j = x_j \mid X_{f(j)} = 1) = \alpha_j^{x_j}(1-\alpha_j)^{(1-x_j)} \tag{1}$$

where $x = (x_1, x_2, x_3)^T$.

Extending the idea to an arbitrary size tree shown in Fig. 1, we can obtain a general expression for the complete likelihood

$$\Pr(X_{\text{com}} = x) = \prod_{j=1}^{d'} \left\{ \alpha_j^{x_{f(j)}x_j} \left[(1-\alpha_j) \times \prod_{k \in de(j)} (1-x_k) \right]^{x_{f(j)} - x_{f(j)}x_j} \right\} \tag{2}$$

where $de(j)$ is the set of all descendants of node j; d' is the total number of nodes of the multicast tree, $d' = \mid O \bigcup U \mid$, d is the number of unobservable nodes; and x is a d'-dimensional binary vector.

This general expression for the complete likelihood is essentially a modified multinomial probability belonging to a general class of the exponential family [4]. As the complete likelihood function consists of factors of Bernoulli probabilities, it has the simplest form possible. This simple model is applicable to both internal and edge monitors in general, with the latter as a special case.

C. Estimation of Missing State Parameters

The inference or monitoring is to estimate the unknown parameters using measurements. The estimation is done through the EM algorithm resulting from the missing data formulation.

1) Expectation (E) Step: The E-step consists of evaluating the expected log complete likelihood. For i.i.d. probe packets, the log likelihood of the complete data D_{inc} for n probes is the product of the joint probability of a single probe: $P_c = \prod_m^n \Pr(X_{\text{com}} = x^{(m)})$, where $x^{(m)} \in D_{\text{inc}}$ for $1 \le m \le n$.

The expected complete log-likelihood $Q(\alpha \mid \alpha^{(p)})$ can then be obtained by the expectation over the missing variables, $Q(\alpha \mid \alpha^{(p)}) = 1/n\mathbf{E}[\log P_c \mid D_{\text{inc}}, \alpha^{(p)}]$. The product form of the multinomial type of probabilities for the complete likelihood

function renders a simple expression for the expected complete log likelihood

$$Q(\alpha \mid \alpha^{(p)}) = \sum_{j \in U \bigcup O} \left[f_{1j}^{(p)} \ln \alpha_j + f_{2j}^{(p)} \ln(1-\alpha_j) + C_j^{(p)} \right] \tag{3}$$

where $C_j^{(p)}$ is a term irrelevant to α. $f_{1j}^{(p)}$ is the average conditional expectation of the state for node j given measurements

$$f_{1j}^{(p)} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}[X_j \mid y^{(i)}, \alpha^{(p)}] \tag{4}$$

where $y^{(i)}$ is a vector consisting of measurements at the observable nodes for the ith probe, i.e., $y^{(i)} = (X_j^{(i)}, j \in O)$. $f_{2j}^{(p)}$ has a similar expression

$$f_{2j}^{(p)} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}[X_{f(j)} \mid y^{(i)}, \alpha^{(p)}] - \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}[X_j \mid y^{(i)}, \alpha^{(p)}]. \tag{5}$$

2) Maximization (M) Step: The M-step maximizes the expected log likelihood, resulting in a new set of conditional probabilities indexed by $p+1$

$$\alpha_j^{(p+1)} = \frac{f_{1j}^{(p)}}{f_{1f(j)}^{(p)}} \tag{6}$$

for a missing random variable $j \in U$. By enumerating on j and using the fact $\Pr(X_j = 1) = \alpha_1 \ldots \alpha_{f(j)} \alpha_j$, we have for $j \in U$

$$\Pr^{(p+1)}(X_j = 1) = f_{1j}^{(p)} \tag{7}$$

where $\Pr^{(p+1)}(X_j = 1)$ is the estimated probability that node j receives a packet. $f_{1j}^{(p)}$, as the conditional expectation of a binary state X_j, is the posterior probability (averaged over n probes) that node j receives a packet given measurements. That is, for an unobservable node ($j \in U$)

$$\Pr^{(p+1)}(X_j = 1) = \frac{1}{n} \sum_{i=1}^{n} \Pr(X_j = 1 \mid y^{(i)}, \alpha^{(p)}) \tag{8}$$

where the probability on the right hand side is evaluated using the parameters $\alpha^{(p)}$. This expression provides an intuitive explanation: when the state of a node is inaccessible, the probability that the node receives a probe packet should be estimated by the posterior probability given the measurements. Since all measurements are assumed equally likely, the average posterior probability results in a valid estimate of the true probability.

The average posterior probability can be further evaluated for an unobservable node, leading to recursive relations for parameter estimation.

Let $O(j)$ be a subset of the monitors which are the descendants of node j. For conditionally independent losses, only the measurements obtained by monitors in $O(j)$ are relevant to determining the posterior probabilities. Let γ_j be the probability that at least one monitor in $O(j)$ receives a packet,[5] and $\hat{\gamma}_j$ be the sample estimate of γ_j, where $\hat{\gamma}_j = n_j/n$ with n_j being the number of probes received by at least one monitor in $O(j)$.

For an unobservable node $j (j \in U)$, the probabilities at the $p+1$th step can be estimated by those at the previous (p) step through the following recursive equations:

$$\alpha_j^{(p+1)} = \frac{\hat{\gamma}_j}{\Pr^{(p+1)}(X_{f(j)} = 1)} + \frac{\Pr^{(p)}(X_{f(j)} = 1)}{\Pr^{(p+1)}(X_{f(j)} = 1)} \alpha_j^{(p)} \frac{(1-\hat{\gamma}_j)}{(1-\gamma_j^{(p)})} \prod_{k \in C(j)} \left(1 - \frac{\gamma_k^{(p)}}{\alpha_j^{(p)} \Pr^{(p)}(X_{f(j)} = 1)} \right) \tag{9}$$

where $C(j)$ is a set of children of node j. $\gamma_j^{(p)}$ is the probability that at least one monitor in $O(j)$ receives a probe packet determined by parameters $\alpha^{(p)}$. $\gamma_k^{(p)} (k \in C(j))$ is the probability that at least one monitor in $O(k)$ (a set of monitors of node k) with $k \in C(j)$. $\gamma_k^{(p)}$ is also determined by $\alpha^{(p)}$.

The E-step and M-step alternate until the convergence is achieved.

IV. Density Estimation and Scalability

The missing data formulation shows that the network monitoring can be considered as density estimation [1].

A. Performance and Sample Complexity

The density in this work is the complete likelihood $\Pr(X_{\text{com}})$. The estimation is on the unknown parameters of the density, $\Pr(X_j = 1)$ or equivalently α_j, for $j \in U$. The number of unobservable nodes equals to the number of unknown parameters. Measurements correspond to the so-called samples in density estimation.

The performance of density estimation can be measured by a distance[6] between the estimated and the true density [1]. Such a distance is related to a simpler measure, which is the mean-square error between the estimated ($\hat{\alpha}$) and the true ($\alpha*$) parameters, $\mathbf{E}[\| \alpha^* - \hat{\alpha} \|^2]$ [1]. For simplicity, we choose to use this mean-square error to define the scalability.

Given a desired performance δ, the number of measurements needed to achieve such a performance (e.g., $\mathbf{E}[\| \alpha^* - \hat{\alpha} \|^2] < \delta$) is the so-called sample complexity. The sample complexity depends on the number of unknown parameters to be estimated, i.e., the larger the number of unknown parameters, the larger the sample complexity. Such a relation can be characterized for the sample complexity n as

$$n \sim c \frac{d}{\delta} \tag{10}$$

for a large n. The dependence of n to δ is easy to understand. The smaller the δ, the larger the n. c is an important scaling factor: it is the number of samples normalized by the number of unknowns. c determines the growth rate of n. When c is a constant, n grows linearly with respect to d, which is the best growth rate achievable [1]. When c also grows with d, n increases at a faster rate than linear with respect to d, which is un-desirable. Such a growth rate on the sample complexity provides a baseline for defining the scalability in terms of the number of probe packets.

[5] γ_j is determined by α. See [5] for details.

[6] i.e., Kullback–Leibler Divergence.

B. Definition of Scalability

We use the square error as the performance measure but with a slight modification: the true parameters are regarded as random variables from a Bayesian's view [9] for feasibility of analysis. That is, $\mathbf{E}[|| \alpha^* - \hat{\alpha} ||^2 | D_{inc}]$ is the estimation error between $\hat{\alpha}$ and α^* given a set of measurements D_{inc}, where the expectation is done over α^*.

Definition of Scalability: Let n be the number of probes needed to achieve a desired estimation error, i.e., $\mathbf{E}[|| \alpha^* - \hat{\alpha} ||^2 | D_{\text{inc}}] < \delta$. If n grows linearly with respect to the number of unobservable nodes d, the measurement-base network monitoring in the context of multicast loss inference is scalable. If n grows faster than linear in d, the approach is not deemed scalable.

V. Analysis

To access the growth rate of the number of measurements involves evaluating the constant c. c depends on network topology and underlying losses. The results on sample complexity are general, and do not offer approaches for obtaining c. Therefore, we turn to missing data formulation. We assume that monitors are only located at multicast receivers for simplicity of analysis.

A. Missing Information and Estimation Error

A key concept we use is the missing information. Two factors contribute to missing information: the number of unobservable nodes, and the underlying losses. The more the unobservable nodes, the more measurements would be needed for maintaining a small estimation error. Similarly, the larger the underlying losses, the more probe packets need to be sent for a sufficient number of them to arrive at the multicast receivers. Quantitatively, the missing information can be characterized by the missing information matrix [9], I_{miss}, where $I_{\text{miss}} = \mathbf{E}\{\frac{\partial^2 \log P(X_{\text{miss}}|D_{\text{inc}},\alpha)}{\partial\alpha\partial\alpha'}\}$.

The so-called complete information is characterized by the complete information matrix [9], $I_{\text{com}} = [c_{ij}]$ with $c_{ij} = \mathbf{E}\{\frac{\partial^2 \log P(X_{\text{com}}|D_{\text{inc}},\alpha)}{\partial\alpha_i\partial\alpha_j}\}$. The difference between the complete and missing information is the observed information, which can be characterized by the observed information matrix, I_{obs}, where $I_{\text{obs}} = I_{\text{com}} - I_{\text{miss}}$.

The estimation error can be expressed as $\sum_{j\in U} \mathbf{E}[|| \alpha^* j - \hat{\alpha}_j ||^2 | D_{\text{inc}}] = \frac{1}{n} Tr\{[I_{\text{com}} - I_{\text{miss}}]^{-1}\}$, with $Tr(A)$ being the trace of matrix A. This relation shows that the estimation error decreases as the number of measurements n increases, and the constant c can be estimated through the trace of $[I_{\text{com}} - I_{\text{miss}}]^{-1}$.

B. Missing Information and Convergence Rate

I_{com} is diagonal [9] and can be easily obtained through the complete likelihood function. I_{miss} is related to the convergence rate of the EM algorithm, where the EM recursive relation can be expressed as $\alpha^{(p+1)} = M(\alpha^{(p)})$ (see (9). Then the derivatives of $M(\alpha^{(p)})$ with respect to parameter $\alpha^{(p)}$ evaluated at the solution $\hat{\alpha}$ result in the Jacobian matrix $DM(\hat{\alpha})$. The missing information matrix can then be obtained as $I_{\text{miss}} = DM(\hat{\alpha}) I_{\text{com}}^{-1}$ [9]. This leads to $\sum_{j\in U} \mathbf{E}[|| \alpha^* j - \hat{\alpha}_j ||^2 | D_{\text{inc}}] = \frac{1}{n} Tr\{[I - DM(\hat{\alpha})]^{-1} I_{\text{com}}^{-1}\}$ where I is the identity matrix; $I - DM(\hat{\alpha})$ is the so-called speed matrix S of the EM algorithm [9], and the smallest eigenvalue of S characterizes the convergence rate.

The convergence rate determines the computation time. For example, if the convergence rate is close to 1, the convergence is exponentially fast, and a small number of iterations would lead to convergence for all parameters. Otherwise, if the convergence rate is a small number, convergence is exponentially slow, and a large number of iterations is needed to reach a fix point. Therefore, convergence time needs to be included in accessing the scalability.

C. Bounds

How to evaluate the number of measurements needed, and the convergence rate using the information and speed matrices? $DM(\hat{\alpha})$ usually has nonzero off-diagonal elements, and is therefore too complex to render a feasible evaluation. But when the off-diagonal elements are neglected, the recursive relations become decoupled similar to that used in [5] (see Appendix IV for details). This decoupling results in a speed matrix with only diagonal elements so that the convergence rate can be obtained from each recursive equation independently. This suggests that if the missing information is contained mainly in the diagonal elements of the missing information matrix, the estimation error may be characterized by a simple expression.

Theorem: Let I_{missd} consist of diagonal elements of I_{miss} and zero off-diagonal elements. Let $I_{\text{misso}} = I_{\text{miss}} - I_{\text{missd}}$ consist of off-diagonal elements of I_{miss}, and zero diagonal elements. If I_{misso} has the spectrum radius less than 1

$$\frac{1}{n}\sum_{j=1}^{d} \frac{1}{\frac{\hat{\lambda}_j}{\hat{\sigma}^2_{\text{jcom}}}} \le \sum_{j\in U} \mathbf{E}[|| \alpha_j^* - \hat{\alpha}_j ||^2 | D_{\text{inc}}] \tag{11}$$

where for the multicast tree, $\hat{\sigma}^2_{\text{jcom}} = \frac{\hat{\alpha}_j(1-\hat{\alpha}_j)}{\prod_{l=1}^{f(j)} \hat{\alpha}_l} \hat{\sigma}^2_{\text{jcom}} = C_{jj}$. $\hat{\lambda}_j$ is the convergence rate of the jth decoupled EM recursion.

Then the number of measurements satisfies

$$n \ge \frac{1}{\delta}\sum_{j=1}^{d} \frac{1}{\frac{\hat{\lambda}_j}{\hat{\sigma}^2_{\text{jcom}}}}. \tag{12}$$

The proof of the theorem is given in Appendix IV.

Intuitively, the condition on the spectrum radius of I_{misso} quantifies the "concentration" of missing information on the diagonal elements. But when this condition is satisfied is not completely clear yet due to the difficulty of evaluating the complex speed matrix.

To provide an intuitive understanding of this result, we note that for the j's missing variable, $\frac{1}{\hat{\sigma}^2_{\text{jcom}}}$ characterizes the complete information, and $\hat{\lambda}_j$ represents the percentage of the observed information. The sum $\sum_{j=1}^{d} \frac{1}{\frac{\hat{\lambda}_j}{\hat{\sigma}^2_{\text{jcom}}}}$, roughly speaking, is inversely proportional to the amount of observed information at the network.

Therefore, the smaller the amount of observed information (i.e., the larger the amount of missing information), the larger the estimation error.

D. Evaluation of Convergence Rate

The convergence rate of the simplified EM algorithm, $\hat{\lambda}_j$, can be easily obtained by completing the derivative of $\alpha^{(p+1)}$ with respect to $\alpha^{(p)}$

$$\hat{\lambda}_j = \hat{v}_j - \sum_{k \in C(j)} \hat{v}_k \prod_{q \in C(j), q \neq k} (1 - \hat{v}_q) \tag{13}$$

where each term in the expression has an intuitive interpretation. In particular

$$\hat{v}_j = \frac{\hat{\gamma}_j}{\Pr(X_j = 1 \mid \hat{\alpha})} \tag{14}$$

is an estimate of the probability that at least one monitor in $O(j)$ receives a probe given node j receives the probe, $\hat{v}_s = (\hat{\gamma}_s)/(\Pr(X_j = 1 \mid \hat{\alpha}))$ is an estimate of the probability that at least one monitor in $O(s)$ receives a probe given node[7] j receives the probe for $s \in C(j)$ and, therefore

$$\sum_{k \in C(j)} \hat{v}_k \prod_{q \in C(j), q \neq k} (1 - \hat{v}_q) \tag{15}$$

corresponds to the probability that given node j receives a probe, there is at least one monitor receives the probe. Those monitors which receive the probe are all located in the same subset of monitors associated with one child of node j.

It follows that the approximated convergence rate, $\hat{\lambda}_j$, is an estimate of the probability that given node j receives a probe there are at least *two* monitors in $O(j)$ which receive the probe (and each monitor belongs to a different subset of monitors associated with a child of node j). This is consistent to the intuitive meaning of the minimum $\hat{\lambda}_j$, the convergence rate, which is the percentage of the observed information, given in Section V-B.

A close form expression can be obtained for special multicast trees which are layer-wise uniform, i.e., each node at layer $l (1 < l \leq L-1)$ has the same number k_{cl} of children, and the conditional probability of each node within the same layer is α_l. Then the convergence rate of the EM algorithm for an unobservable node at layer l is

$$\hat{\lambda}_l = \hat{v}_l - k_{cl}(1 - \hat{v}_l)^{(k_{cl}-1)/(k_{cl})} + k_{cl}(1 - \hat{v}_l) \tag{16}$$

where $\hat{v}_l = (\hat{\gamma}_l)/(\prod_{m=l}^{1} \hat{\alpha}_m)$.

If the multicast tree is completely uniform with $k_{cl} = k_c$, and $\hat{\alpha}_l = \hat{\alpha}'$

$$\hat{\lambda}_l = \hat{v}_l - k_c(1 - \hat{v}_l)^{(k_c-1)/(k_c)} + k_c(1 - \hat{v}_l). \tag{17}$$

The derivations of (16) and (17) can be found in Appendix I. As will be shown in the following sections, these simple expressions for $\hat{\lambda}_j$ and $\hat{\sigma}^2_{jcom}$, when substituted into the lower bound given by (11), allow us to derive explicit scalability results for the uniform networks.

E. Best Case of Edge Monitors: Small Losses

For the feasibility of analysis, we consider extreme cases and uniform tree. We start from the best case when the nodal loss probability is small relative to the size of the network, resulting in negligible amount of missing information due to underlying losses. Mathematically, we derive a condition on the underlying losses relative to the depth of a multicast tree, and a bound for the estimation error.

Corollary: If the conditional packet loss probability, $1 - \hat{\alpha}'$, is small so that $L(1 - \hat{\alpha}') = o(1)$, the estimation error satisfies

$$\mathbf{E}[|| \alpha - \hat{\alpha} ||^2 | D_i] \geq \frac{(1 - \hat{\alpha}')d}{n}(1 + O(L(1 - \hat{\alpha}'))). \tag{18}$$

The convergence rate is the same for all nodes and is close to one, i.e.,

$$\hat{\lambda}_j = 1 - O(k_c(1 - \hat{\alpha}')) \tag{19}$$

for all j, and $k_c(1 - \hat{\alpha}') = o(1)$ *with* k_c being the number of children of each node.[8]

If the desired estimation error is δ, the number of required probes,[9] n

$$n \approx \frac{(1 - \hat{\alpha}')d}{\delta}. \tag{20}$$

Therefore, n grows linearly with the number of internal nodes, and the scaling factor c is a constant, $c = 1 - o(1)$. In addition, the convergence rate corresponds to the exponentially fast convergence time at all internal nodes. Therefore, the multicast inference using only edge monitors is scalable for small losses.

This result is consistent to that given in [5], which shows that the error approaches zero when the underlying losses approach zero. We provide the rate of convergence, and the condition on the underlying losses. Such a condition suggests, how small the loss probability is, should be relative to the size of the network. The larger the network (the deeper the tree), the smaller the losses would have to be for the scalability to hold. This is consistent to our intuition since more losses may occur when a tree has more depth.

F. Worst Case For Edge Monitors: Large Losses

We now consider the case when a multicast network has large nodal losses everywhere. This corresponds to a pathological case of missing information. The underlying losses exceed the relative value (see the previous section), and the amount of missing information dominates the complete information. This serves as the worst case of scalability as shown below.

Corollary: If the conditional probability of receiving a packet, $\hat{\alpha}'$, is small ($k_c\hat{\alpha}' = o(1)$), and the depth of the multicast tree (L) is large ($k_c^L \gg 1$ and $L \gg 1$), we have

$$\mathbf{E}[|| \alpha - \hat{\alpha} ||^2 | D_i] \geq \Omega\left(\left(\frac{1}{\hat{\alpha}'}\right)^{L-1} \frac{d}{n}\right). \tag{21}$$

Furthermore, the convergence rate for a node at layer l *is*

$$\hat{\lambda}_l = (k_c\hat{\alpha}')^{L-l} + O((k_c\hat{\alpha}')^{L-l+1}) \tag{22}$$

for $1 \leq l \leq L-1$.

The proof of the corollary is given in Appendix II. The quantity, $\Omega((1/\hat{\alpha}')^{L-1}d/n)$ is, thus, a lower bound of the estimation error. Comparing this lower bound with the result for the best

[7] s is a child of node j. $O(s)$ is a subset of the monitors associated with node s.

[8] The proof of the corollary is based on algebraic manipulation of the lower bound, and omitted.

[9] This is not completely rigorous yet as an upper bound needs to be shown to be $O(d)$ and will be completed in the sequel.

case given in (18), we observe a common factor d/n for the estimation error but different scaling factors. The scaling factor, c, is essentially the normalized number of probes required to achieve a desired estimation error. For large losses

$$c \geq \left(\frac{1}{\hat{\alpha}'}\right)^L \tag{23}$$

which grows exponentially with the depth of the multicast tree; whereas for small losses, c takes a constant value of $1-\hat{\alpha}'$.

Intuitively, $\hat{\alpha}'^L$ is the probability for a probe to reach an edge monitor, and $(1/\hat{\alpha}')^L$ is roughly the average number of probes needs to be sent so that one is received. When $\hat{\alpha}'$ is small and L is large, probes are lost with a high probability (i.e., the missing information is large), a lot of probes need to be sent to maintain a small estimation error.

This result shows that the case of large losses results in the worst scaling property for the number of probes, i.e.,

$$n \geq \Omega\left(\left(\frac{1}{\hat{\alpha}'}\right)^L d\right). \tag{24}$$

By definition, the inference approach using only edge monitors is not scalable when nodal packet losses are large.

Furthermore, the convergence rate (22) is exponentially small in the depth of the tree at layers away from the source, leading to an exponentially large number of iterations.

In the prior work [5], the behavior of the maximum variance of the estimated conditional probabilities (α_js) is investigated with respect to the depth of a uniform multicast tree. Specifically, for small losses, the prior work [5] shows analytically that the maximum variance is a constant; and for large losses, the prior work [5] demonstrates empirically that the maximum variance increases rapidly with respect to the depth of the tree. In fact, the maximum variance corresponds to the largest estimation error (among all α_js) when the number of probes is infinite.

The results here quantify the exact growth rate of the number of probes, which is exponential for the case considered in this section.

G. Benign Case of Edge Monitors: Local Losses

We consider now the case when large losses occur in parts of the uniform network. This is a more realistic case than the uniformly large losses discussed in Section V-B. Such a case may correspond to network bottlenecks, local congestion states, or partial network anomalies caused by either hardware or software problems.

We assume, without loss of generality, that large losses occur at layer l_0 of the multicast tree[10] for $1 \ll l_0 \leq L-1$. The conditional probability is the same for all nodes at layer l_0 and takes a small value $\hat{\alpha}_0$ ($\hat{\alpha}_0 = o(1)$). Both the depth of the tree, L, and the number of children of each node, k_c, are large ($L \gg 1$, and $k_c \gg 1$). The degree of the local losses is characterized by a constant n_0 through a relation, $k_c^{n_0}\hat{\alpha}_0 = O(1)$, for $1 \ll n_0 \leq L-1$. The rest of the nodes on the tree assume to have the same conditional probability $\hat{\alpha}'$ and small losses, i.e., $\hat{\alpha}' = 1 - o(1)$.

[10]This assumes that the layer with large losses is away from the source, for feasibility of analysis.

Corollary: The estimation error is lower bounded as

$$\mathbf{E}[|| \alpha^* - \hat{\alpha} ||^2 | D_{inc}] \geq \Omega\left(\frac{d^g}{n}\right) \tag{25}$$

where $1 \leq g \leq 3$, and g depends on n_0, L and l_0. The convergence rate is

$$\hat{\lambda}_l = \begin{cases} k_c^{2(l_0-1)}\hat{\alpha}_0^2 + o(1), & \text{for } 1 \leq l \leq l_0 - 1; \\ 1 - o(1), & \text{for } l_0 \leq l \leq L-1. \end{cases} \tag{26}$$

The proof of the corollary is given in Appendix III. The number of probes needed can then be obtained from (25)

$$n \sim O(d^g). \tag{27}$$

Therefore, the power g characterizes the growth rate of the number of probes with respect to d. In particular, g depends on two factors: the degree of losses in terms of n_0, and the relative location of layer l_0 on the tree. For a given location l_0, g grows with respect to n_0, suggesting that more probes are needed to achieve a desired estimation error when losses get large. When the degree of losses n_0 is fixed, g grows with l_0. This is because the number of nodes on layer l_0 with large losses increases with l_0. When both l_0 and n_0 take the maximum value of $L-1$ and L, respectively, g takes the maximum value of three. The corresponding number of probes needed is $O(d^3)$. This gives the worst scaling property of the number of probes for the case of local losses. Furthermore, the convergence time (26) is exponentially fast at nodes below layer l_0, since those nodes consist of an equivalent uniform subtree with small losses.

The convergence time is exponentially slow at nodes above layer l_0, since layer l_0 with large losses makes the inference difficult. These results suggest that when local losses are large and sufficiently close to the edge monitors, the multicast inference is nonscalable.

VI. Numerical Investigations

The above analysis is asymptotic in the depth of the multicast tree, and for cases when the loss is either small or large. In this section, we report numerical investigations for the cases when the packet losses are moderate and the size of the tree is finite. Specifically, the set of figures address the following questions pertaining to the use of only edge monitors. 1) at what packet loss probabilities, is the inference technique nonscalable? 2) Given the underlying packet loss probabilities, how would the size of the network affect the scalability? We also investigate the accuracy of the lower bound of the estimation error.

A. Simulation Setup

We consider a multicast tree with L layers, and three children for each node. The underlying packet loss probabilities at all nodes are assumed to be the same; L is varied in the simulation. The probes are generated independently, and multicasted through the network. At the intermediate nodes, they either get transmitted or lost according to the given nodal loss probability. Receivers as edge monitors collect probes.

The EM algorithm estimates the internal conditional loss probabilities, $\hat{\alpha}_j$ for $j \in U$. If $| \alpha_j^{(p+1)} - \alpha_j^{(p)} | \leq 0.000\,01$ for all $j \in U$, the EM algorithm terminates. Ten runs are performed, each of which generates a distinct set of measurements.

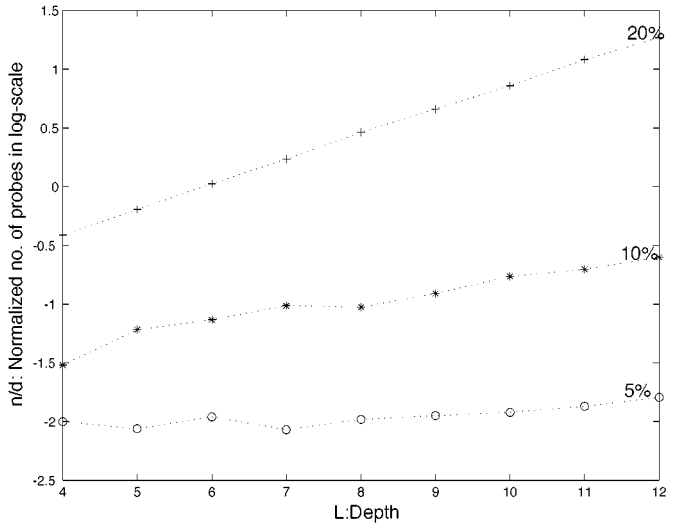

Fig. 3. The normalized number of probes versus the depth L of a multicast tree.

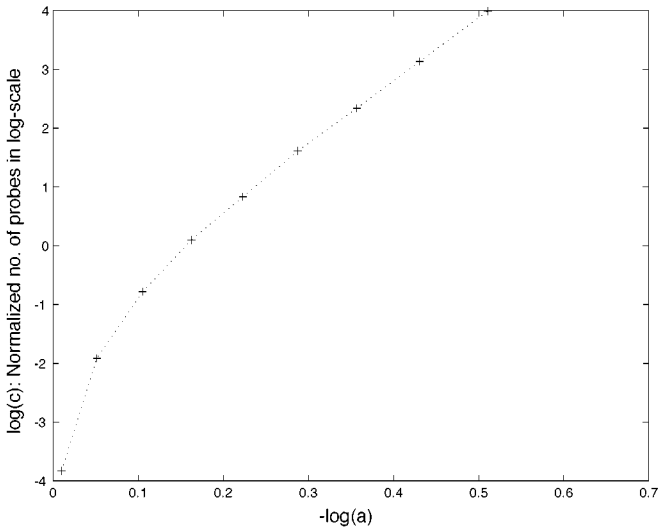

Fig. 4. The normalized number of probes (c) versus $\ln(1(\hat{\alpha}))$ in log scale.

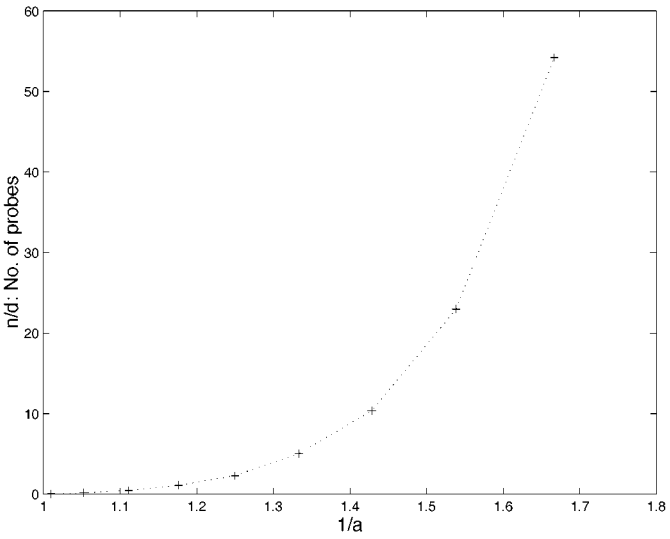

Fig. 5. The normalized number of probes versus $1/(\hat{\alpha})$.

The performance is given by the estimation error, e, averaged over both the total number of internal nodes (d) and the number of runs (R)

$$e = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{d}\sum_{j=1}^{d}(\alpha' - \hat{\alpha}_{rj})^2 \tag{28}$$

where α' is the common underlying conditional probabilities for all nodes, and $\hat{\alpha}_{rj}$ is the underlying conditional probability for the jth internal node estimated at the rth run.

B. Number of Probes

We begin by investigating the impact of the nodal losses on the normalized number of required probes c. In Fig. 3, we plot $\log c$ as a function of the depth, L, of a binary multicast tree for various values of conditional nodal loss probabilities.[11] The nodes of the tree are assumed to have a common conditional loss probability $(1-\alpha)$, which is varied in the figure from 0.05 to 0.2.

In the simulations, probes are multicasted from the source to the receivers, and are subject to loss in the intermediate nodes according to the underlying loss probabilities. Each of the three curves in the figure which corresponds to a particular loss probability is obtained through averaging the results from ten different runs. The curve corresponding to 0.2 conditional loss probability is straight, demonstrating that c grows exponentially with the depth of the tree L.

It is interesting to observe that the exponential growth extends from a small value of L up to twelve. The slope of the straight line characterizes the rate of the exponential growth of c, and is measured as 0.2181 from the figure. This is in a good agreement with $\ln(1/\alpha^*) = 0.2231$ where $\alpha^* = 0.8$ is the true conditional probability.

This suggests that the asymptotic result derived for large networks with large losses holds even for a tree with four layers and nodal loss probability of 0.2. The other two curves corresponding to 0.05 and 0.1 loss probabilities are not quite as straight. In fact, the curve corresponding to 0.05 loss is insensitive to the depth of the tree until L =10. This motivates us to conduct another simulation by changing the underlying nodal losses but keep the depth L fixed ($L = 10$).

Fig. 4 shows the log of the normalized number of probes versue $\ln(1/\hat{\alpha})$ for L fixed to be ten. It shows that for conditional packet loss probabilities around 0.12–0.18 the curve becomes a straight line, indicating a polynomial-like relationship $c = (1/\hat{\alpha})^s$. Interestingly, the exponent s measured from the figure is between 10.7 and 12, while the actual depth of the tree is ten. This again suggests that the asymptotic relation developed for large losses seems to hold to a tree of depth 10 with a moderate packet loss rate of 0.15. For losses smaller than 0.15, the normalized number of probes grows at a slower rate than linear, which can be easily seen from a plot of c versus $1/\hat{\alpha}$ in Fig. 5.

The above examples suggest the existence of a "critical" packet loss probability associated with a given tree depth. At this critical packet loss probability, the normalized number of probes begins to increase significantly.

[11] $(1-\alpha_j)$s.

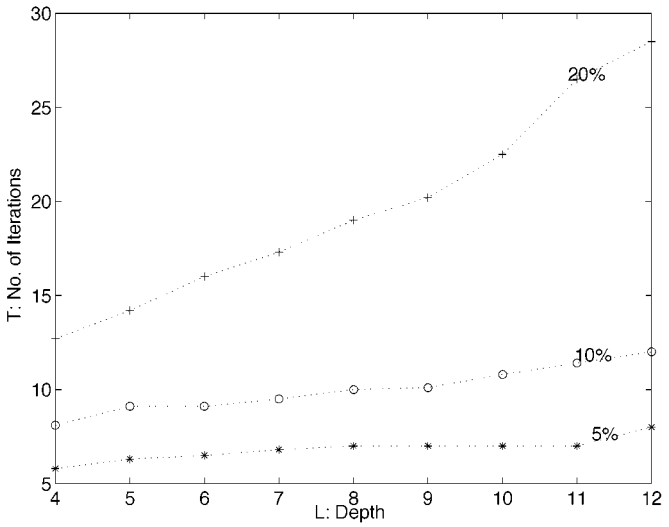

Fig. 6. Number of iterations versus the depth (L).

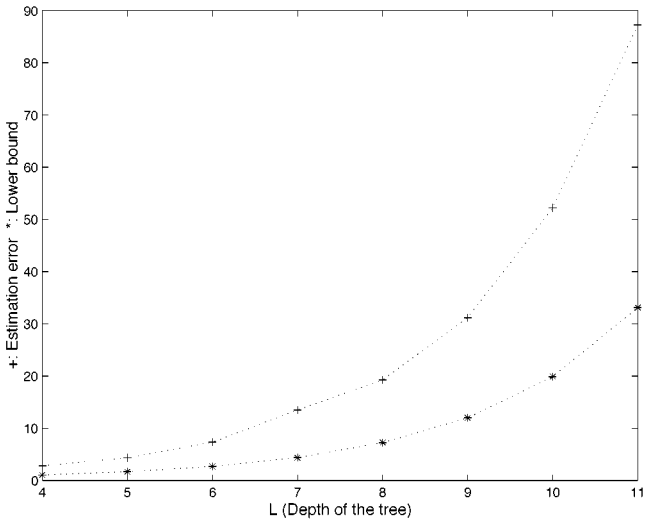

Fig. 7 The mean-square-error (estimation error) and the lower bound verus the depth L.

C. Convergence Rate

To demonstrate how computation time scales with the size of the tree and the underlying packet losses, we plot in Fig. 6 the number of iterations of the simplified EM algorithm as a function of the depth of the tree for loss probabilities ranging from 0.05 to 0.2. Specifically, this number of iterations is the maximum number of iterations for the simplified EM algorithm to converge at all unobservable nodes. Similar to the simulations used to generate Fig. 5, each point in Fig. 6 is also obtained through averaging over ten runs. The curve corresponding to 0.2 loss probability grows at least at a linear rate. For 0.1 losses, the number of iterations grows at a slower rate than linear for small values of L but increases significantly starting at $L = 10$. For 0.05 losses, the number of iterations is almost constant up to an L value of eleven. The different behaviors exhibited in the figure are consistent with the results of our analysis: the convergence rate is nearly constant for small losses, and it is exponentially small for large losses.

D. Lower Bound

To illustrate the lower bound for the estimation error, we plot in Fig. 7 both the mean-square-error (averaged over ten runs) and the lower bound (11) as a function of the depth L of the multicast tree.

The underlying tree which generates the measurements is binary and uniform with conditional nodal loss probability of 0.2. The figure demonstrates the validity of the lower bound on the estimation error. In particular, the bound tracks the actual error as L increases. This is important to observe as the bound has been used above to derive asymptotic results.

In terms of the relative value to the mean-square-error, the bound seems to be more accurate for small and moderate L than for large L.

VII. Practical Implications

We have so far investigated the scalability in terms of the growth rate of the number of probes and the computation time of the EM algorithm. We have developed asymptotic analysis for extreme cases, and numerical results for finite-size trees with moderate losses.

Our scalability results provide guidelines for accessing the feasibility of the inference approach. In particular, our results imply that the use of edge monitors is most feasible for networks with small nodal losses in the order of a few percent. This corresponds to the current internet in normal operation. For networks with nodal loss probability 0.15 or above, the resources needed grow rapidly with the size of a network. This suggests that the inference approach should be used with caution for large networks with either a heavy load or wide-spread congestion: a large number of probes and the long computation time may be needed to estimate the state of the network. For networks with local congestion/faults resulting in packet losses, our results suggest that a large number of probes and iterations of the EM algorithm may also be needed unless local losses are moderate. As one may be wondering whether such an accessment is useful since underlying network conditions are unknown, we note that certain prior knowledge is often available about a network from network managers to help accessing the feasibility of the measurement-based inference approach.

The scalability results in the growth rate of the number of probes can also provide an estimate on the number of probes required. For instance, assume the desired estimation error, δ, is given as 0.01.

From (10), we have the number of probes required is $n \sim c/\delta d \sim 100 * c * d$. Since the normalized number of probes, c, depends on the underlying losses which are unknown, it is impossible to obtain an exact value of n. However, when certain prior knowledge is available, the range of n may be estimated using the scalability results. For instance, if the conditional loss probabilities at nodes are from 0.01 to 0.1, we can obtain an upper and a lower bound on c

$$0.01 \leq c \leq \frac{1}{(0.9)^L} \tag{29}$$

where the lower bound is for small losses, and the upper bound is for large losses.

Assume the depth of the tree $L = 10$. The number of probes can be estimated as

$$d \leq n \leq 300 * d. \tag{30}$$

VIII. Internal Monitors: Preliminary Investigations

When the edge monitoring is nonscalable, can additional monitors placed at the interior of a network improve the scalability? As the goal for placing internal monitors is to reduce the amount of missing information, the first question arises as to what types of monitors can reduce missing information.

We first consider the use of the passive monitors at the interior of the multicast tree. We focus on the case where large losses occur at one layer (l_0) of the tree, and assume that the internal monitors are placed exactly at layer l_0. As the nodes on this layer become observable, the computation time reduces to zero at those nodes. That is, the use of internal monitors improves the computation time. But the internal monitors have no impact on the number of probes needed, since these monitors are passive and can not regenerate probes lost at layer l_0. Therefore, the use of passive monitors internally at the multicast tree can not improve the scalability in the growth rate of the number of probes.

How about other types of monitors? Consider a monitor with two functionalities: the one assumed for the passive monitor for recording the reception of probes from a source, and the other assumed for a source for sending probes to the nodes down the tree. Such a monitor can be considered as an active monitor, and placed at the interior of a network. The framework and results given in this work provide guidelines and ingredients for investigating how active monitors improve the scalability, and how to place them optimally using measurements. We are investigating these questions in our on-going work.

IX. Conclusion

We have focused on developing a framework to investigate the scalability of measurement-based network monitoring. The framework is drawn from missing data formulation and sample complexity of density estimation. Using the missing data framework, we have shown that measurement-based network monitoring can be cast as a problem of density estimation. Based on sample complexity of density estimation, we have defined the scalability of measurement-based network monitoring as the growth rate of the number of measurements with respect to the size of a multicast network. Specifically, if the growth rate is linear, which is the best achievable, the approach is scalable.

We have applied the missing data framework to derive the growth rate. Our results show that the scalability can be characterized by the average number (np) of measurements received at edge monitors, where n is the number of probe packets sent, and p is the probability that a probe packet arrives at edge monitors. As a rule of thumb, when $np \approx \frac{d}{\delta}$, the estimation error is bounded by δ for a multicast tree with d unobservable nodes. When the underlying losses at internal nodes are relatively small (i.e., p is almost 1), the size of the network dominates the missing information, and the monitoring technique using edge monitors is scalable. When the underlying losses are relatively large (i.e., p decreases rapidly with respect to the size of the network), the internal losses dominate the missing information, and the monitoring technique is non-scalable.

Future work includes refining some of our results to better deal with complex information matrices; and investigating how to extend this work to a more general network setting.

Appendix I
Derivations of Equation (16)

For the layer-wise uniform networks, the convergence rate at all nodes within a layer is the same and can be denoted as $\hat{\lambda}_l$ for layer l ($1 \leq l \leq L-1$).

Then (13) reduces to

$$\hat{\lambda}_l = \hat{v}_l - k_{c(l+1)}(1-\hat{v}_l)\frac{\hat{v}_{k'}}{1-\hat{v}_{k'}} \tag{31}$$

where $k' \in C(l)$, $C(l)$ is the set of children of a node at layer l, and $k_{c(l+1)}$ is the number of children of a node at layer $l+1$. By definition

$$\begin{aligned}\hat{v}_l =& \frac{\hat{\gamma}_l}{\prod_{m=l}^{1}\hat{\alpha}_m} \\ =& 1-(1-\hat{\beta}_{l+1})^{k_{cl}}\end{aligned} \tag{32}$$

where $\beta_{l+1} = \Pr(\Omega(l+1) \mid X_{f(l+1)} = 1)$, and $\Omega(l+1)$ is the event that at least one monitor of unobservable node $l+1$ receives a probe. Similarly

$$\hat{v}_{k'} = \frac{\hat{\gamma}_{k'}}{\prod_{m=l}^{1}\hat{\alpha}_m} \tag{33}$$

$$= \hat{\beta}_{k'} \tag{34}$$

where

$$\hat{\beta}_{k'} = \hat{\alpha}_{l+1}\left[1 - \prod_{j\in C(l+1)}(1-\hat{\beta}_j)\right]. \tag{35}$$

Comparing (35) with (32), we have

$$\hat{v}_{k'} = \hat{\alpha}_{l+1}\hat{v}_{l+1}. \tag{36}$$

Furthermore, from [5], we have

$$\hat{v}_l = 1-(1-\hat{\beta}_{l+1})^{k_c} \tag{37}$$

where $\hat{\beta}_{l+1} = \hat{\alpha}_{l+1} + \hat{\alpha}_{l+1}(1-\hat{\beta}_{l+1})^{k_{cl}}$, and k_{cl} is the number of children of a node at layer l. This results in a recursive relation

$$\hat{v}_l = 1-(1-\hat{\alpha}_{l+1}\hat{v}_{l+1}). \tag{38}$$

Combining (31), (36), and (38), we have

$$\hat{\lambda}_l = \hat{v}_l - k_{cl}(1-\hat{v}_l)^{(k_{cl}-1)/(k_{cl})} + k_{cl}(1-\hat{v}_l). \tag{39}$$

Appendix II
Proof of Corollary 2 (Large Losses)

We start from evaluating the convergence rate. Inserting $k_c\hat{\alpha} = o(1)$ into (17), we have

$$\hat{v}_l = k_c\hat{\alpha}\hat{v}_{l+1} + O((k_c\hat{\alpha}\hat{v}_{l+1})^2). \tag{40}$$

Using the recursion on l staring from $l = L-1$, and $\hat{v}_{L-1} = k_c\hat{\alpha} + O((k_c\hat{\alpha})^2)$, we have

$$\hat{v}_l = (k_c\hat{\alpha})^{L-l} + O((k_c\hat{\alpha})^{L-l+1}) \tag{41}$$

for $1 \le l \le L-1$. Using the fact $k_c\hat{\alpha} = o(1)$, (16) is reduced to

$$\begin{aligned}\hat{\lambda}_l =& C\hat{v}_l^2 + O(\hat{v}_l^3)\\ =& C(k_c\hat{\alpha})^{2(L-l)} + O(k_c\hat{\alpha})^{2(L-l)+1}\end{aligned} \tag{42}$$

where $C = 1/2(1 - 1/k_c)$. Using (42) and $\hat{\sigma}^2_{icom} = (1-\hat{\alpha})/\hat{\alpha}^{l-1}$, we can obtain the dominant term for the estimation error through algebraic manipulations

$$\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} = \frac{C_1}{n}\left[\frac{1}{(k_c\hat{\alpha})^{2(L-1)}} + \frac{k_c^L}{\hat{\alpha}}\right] \tag{43}$$

where $C_1 = (1-\hat{\alpha})/(\mid k_c^3\hat{\alpha} - 1 \mid)$. If $k_c\hat{\alpha} = o(1)$, and $k_c^3\hat{\alpha} > 1$

$$\begin{aligned}\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} =& O\left(\frac{1}{\hat{\alpha}^{L-1}}\frac{k_c^{L-1}}{n}\right)\\ =& O\left(\frac{1}{\hat{\alpha}^{L-1}}\frac{d}{n}\right)\end{aligned} \tag{44}$$

where L is assumed large, and $d = O(k_c^L)$. If $k_c\hat{\alpha} = o(1)$, and $0 < k_c^3\hat{\alpha} < 1$, using similar algebraic manipulations, we can obtain

$$\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} = O\left(\frac{1}{n(k_c\hat{\alpha})^{2L}}\right). \tag{45}$$

Since $(1/\hat{\alpha}) > k_c^3$, and $d = O(k_c^L)$, we have $1/((k_c\hat{\alpha})^{2L}) > \Omega(d/\hat{\alpha}^L)$, and

$$\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} > \Omega\left(\frac{1}{\hat{\alpha}^L}\frac{d}{n}\right). \tag{46}$$

Combining both cases, we have

$$\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} \ge \Omega\left(\frac{1}{\hat{\alpha}^L}\frac{d}{n}\right). \tag{47}$$

Q.E.D.

Appendix III
Proof of Corollary 3 (Local Losses)

We begin proving the corollary by evaluating the convergence rate. The convergence rate at layer l for $l_0 < l \le L-1$ is the same as that for small losses

$$\hat{\gamma}_l = 1 - (1-\hat{\alpha})^{k_c} \tag{48}$$

where $(1-\hat{\alpha})^{k_c} = o(1)$ since $\hat{\alpha} = 1 - o(1)$. For $1 \le l < l_0$, the convergence rate at layer l can be obtained using the results, on uniform tree.

Equation (16) for layer-wise uniform networks, and $k_c^{l_0}\hat{\alpha}_0 = o(1)$

$$\hat{\lambda}_l = k_c^{2(l_0-l)}\hat{\alpha}_0^2 + o(\cdot) \tag{49}$$

where $o(\cdot)$ is a smaller order term compared to $k_c^{2(l_0-l)}\hat{\alpha}_0^2$. Combining the two cases, we have the convergence rate

$$\hat{\lambda}_l = \begin{cases} k_c^{2(l_0-l)}\hat{\alpha}_0^2 + o(\cdot), & \text{for } 1 \le l < l_0\\ 1 - o(1), & \text{otherwise.}\end{cases} \tag{50}$$

We then evaluate the variance due to complete information

$$\hat{\sigma}_l^2 = \begin{cases} \frac{1-\hat{\alpha}}{\hat{\alpha}^{l-2}}, & \text{for } 1 \le l < l_0\\ \frac{\hat{\alpha}_0(1-\hat{\alpha}_0)}{\hat{\alpha}^{l-1}}, & \text{for } l = l_0\\ \frac{1-\hat{\alpha}}{\hat{\alpha}_0\hat{\alpha}^{l-3}}, & \text{for } l_0+1 \le l \le L-1\end{cases}. \tag{51}$$

Inserting (50) and (51) into the lower bound in (11), we have

$$\begin{aligned}\frac{1}{n}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} =& \sum_{l=1}^{l_0-1}\frac{k_c^l(1-\hat{\alpha})}{n\hat{\alpha}^{l-2}k_c^{2l_0-l}\hat{\alpha}_0^2} + \frac{k_c^{l_0}\hat{\alpha}_0(1-\hat{\alpha}_0)}{n\hat{\alpha}^{l_0-1}}\\ &+ \sum_{l=l_0}^{L}\frac{k_c^l(1-\hat{\alpha})}{n\hat{\alpha}^{l-3}\hat{\alpha}_0} + o()\\ =& O\left(\frac{1-\hat{\alpha}}{n\hat{\alpha}_0^2 k_c^{5-l_0}}\right)\\ &+ O\left(\frac{(1-\hat{\alpha})k_c^{L-l_0}}{n\hat{\alpha}_0}\right).\end{aligned} \tag{52}$$

If $(1-\hat{\alpha}/n\hat{\alpha}_0 k_c^{5-l_0}) \le k_c^{L-l_0}$, for $n_0 \ge l_0$, which is $2l_0+n_0 \le L$, under the condition $k_c^{n_0}\hat{\alpha}_0 = O(1)$, we have

$$\begin{aligned}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} =& O\left(\frac{k_c^{Ll_0+n_0}}{n}\right)\\ =& O\left(\frac{d^g}{n}\right)\\ &\cdot \begin{cases} g = 1, & \text{if } n_0 \le l_0\\ g = 1 + \frac{n_0-l_0}{L}, & \text{if } n_0 > l_0\end{cases}.\end{aligned} \tag{53}$$

If $(1-\hat{\alpha})/(n\hat{\alpha}_0 k_c^{5-l_0}) > k_c^{L-l_0}$, which is $2l_0+n_0 > L$, we have similarly

$$\begin{aligned}\sum_j \frac{\hat{\sigma}^2_{\text{jcom}}}{\hat{\lambda}_j} =& O\left(\frac{k_c^{L-l_0+n_0}}{n}\right)\\ =& O\left(\frac{d^g}{n}\right)\begin{cases} g = 1, & \text{if } n_0 \le l_0\\ g = \frac{2l_0+n_0}{L}, & \text{if } n_0 > l_0\end{cases}.\end{aligned} \tag{54}$$

Combining the above two cases, we prove the corollary.

Q.E.D.

Appendix IV
Proof of Theorem 1

We first evaluate the (nonzero) diagonal elements of I_{missd}. Let $I_{\text{missd}} = [q_{ij}]$. Then

$$q_{ii} = \frac{dm_i(\alpha_i^{(p)})}{d\alpha_i^{(p)}}\Big|_{\alpha_i^{(p)}=\hat{\alpha}_i} \tag{55}$$

where $m_j(\alpha_j^{(p)})$ is defined by the right hand side of the decoupled EM recursion given below

$$\alpha_j^{(p+1)} = \frac{\hat{\gamma}_j}{\Pr^{(p+1)}(X_{f(j)}=1)} + \frac{\Pr^{(p)}(X_{f(j)}=1)}{\Pr^{(p+1)}(X_{f(j)}=1)} \alpha_j^{(p)} \prod_{k\in C(j)} \left(1 - \frac{\hat{\gamma}_k^{(p)}}{\alpha_j^{(p)} \Pr^{(p)}(X_{f(j)}=1)}\right) \tag{56}$$

for $j \in U$. When the spectrum radius of I_{misso} is less than 1

$$[I - I_{\text{misso}}]^{-1} = I + I_{\text{misso}} + I_{\text{misso}}^2 + \ldots \tag{57}$$

so that

$$[I - I_{\text{misso}}]^{-1}[I_{\text{com}} - I_{\text{missd}}]^{-1} = [I_{\text{com}} - I_{\text{missd}}]^{-1} + I_{\text{misso}}[I_{\text{com}} - I_{\text{missd}}]^{-1} + I_{\text{misso}}^2[I_{\text{com}} - I_{\text{missd}}]^{-1} + o(\cdot)$$

where $o(\cdot)$ represents a higher order term. By definition, I_{misso} has zero diagonal elements. $[I_{\text{com}} - I_{\text{missd}}]$ is diagonal. Then the diagonal terms are zero for $I_{\text{misso}}[I_{\text{com}} - I_{\text{missd}}]^{-1}$, and positive for $I_{\text{misso}}^2[I_{\text{com}} - I_{\text{missd}}]^{-1}$. Then

$$Tr\{[I - I_{\text{misso}}]^{-1}[I_{\text{com}} - I_{\text{missd}}]^{-1}\} \geq Tr\{[I_{\text{com}} - I_{\text{missd}}]^{-1}\}.$$

Since $I_{\text{com}} - I_{\text{missd}}$ is diagonal, using the expressions of their diagonal elements, we can arrive at the conclusion of the theorem.

Q.E.D.

ACKNOWLEDGMENT

The authors would like to thank C. Liu for helpful discussions on the EM algorithm and providing references, D. Mitra for emphasizing the importance of network-centric monitoring, B. Yu, J. Cao, and S. Vander Wiel for sharing with us their work on Network Tomography using the EM algorithm, K. Ramanan for helpful discussions on multicast, V. Chan for intuitive explanations of the results, and anonymous reviewers for helpful comments. CJ gratefully acknowledges the hospitality and support from D. Mitra and Department of Mathematics of Networks and Systems at Bell-Labs during her sabbatical when most of this work was performed.

REFERENCES

[1] A. Barron and T. M. Cover, "Minimum Complexity Density Estimation," *IEEE Trans. Inform. Theory*, vol. 37, no. 4, pp. 1034–1054, July 1991.

[2] J. Cao, S. Vander Wiel, D. Davis, and B. Yu, "Time-Varying Network Tomography: router link data," J. Amer. Statist. Assoc., to be published.

[3] M. Coates, A. Hero, R. Nowak, and B. Yu, "Internet Tomography," IEEE Signal Processing Mag., May 2002, to be published.

[4] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm," *J. Royal Statistical Society, B*, vol. 39, pp. 1–38.

[5] R. Caceres, N. G. Duffield, J. Horowitz, and D. Towsley, "Multicast-based Inference of Network-Internal Loss Characteristics," *IEEE Trans. Inform. Theory*, pp. 2462–2480, Nov. 1999.

[6] N. Duffield, J. Horowitz, D. Towsley, W. Wei, and T. Friedman, "Multicast-based loss inference with missing data," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 700–713, May 2002.

[7] C. Ji and A. Elwalid, "Measurement-Based Network Monitoring: Missing Data Formulation and Scalability Analysis," in *Proc. Int. Symp. Information Theory*, Italy, June 2000.

[8] ——, "Measurement-Based Network Monitoring: Missing Data Framework and Scalability," Bell Labs, 2000.

[9] R. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. New York: Wiley, 1987.

[10] C. Liu, D. B. Rubin, and Y. Wu, "Parameter Expansion to Accelerate EM: the PX-EM Algorithm," *Biometrika*, pp. 755–770, 1998.

[11] C. Liu and J. Nannenmacher, "Broadcast Audience Estimation," in *Proc. IEEE Infocom*, 2000, pp. 952–960.

[12] M. Medard, D. Marquis, and S. R. Chinn, "Attack Detection Methods for All-Optical Networks," in *Proc. Symp. Network and Distributed System Security*, Mar. 1998, pp. 272–273.

[13] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan, "ATM Network Design and Optimization: A Multirate Loss Network Framework," *IEEE/ACM Trans. Networking*, vol. 4, pp. 531–543, Aug. 1996.

[14] D. Mitra, Personal communication.

[15] M. Thottan and C. Ji, "Proactive Anomaly Detection Using Distributed Intelligent Agents," *IEEE Network, Special Issue Network Manage.*, Oct 1998.

[16] I. Widjaja and A. Elwalid, "MATE: MPLS Adaptive Traffic Engineering," IETF Draft, 1999.

[17] O. Wolfson, S. Sengupta, and Y. Yemini, "Managing communication networks by monitoring databases," *IEEE Trans. Software Eng.*, vol. 17, pp. 944–53, Sept. 1991.

[18] Y. Vardy, "Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data," *J. Amer. Statist. Assoc.*, 1996.

Chuanyi Ji (M'02) received the B.S. degree (honors) from Tsinghua University, Beijing, China, in 1983, the M.S. degree from University of Pennsylvania, Philadelphia, PA, in 1986, and the Ph.D. degree from California Institute of Technology, Pasadena, CA, in 1992, all in Electrical Engineering.

She is an Associate Professor at the Department of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA. She was on the faculty at Rensselaer Polytechnic Institute from 1991 to 2001. She spent her sabbatical at Bell Labs, Lucent Technologies, in 1999, and was a visiting faculty at Massachusetts Institute of Technology (MIT), Cambridge, MA, in the Fall 2000. Her research lies in the areas of network management and control, and adaptive learning systems. Her research interests are in understanding and managing complex networks, applications of adaptive learning systems to network management, learning algorithms and theory.

Dr. Ji received the CAREER award from NSF, in 1995, and the Early Career award from Rensselaer Polytechnic Institute, Troy, NY, in 2000.

Anwar Elwalid received the B.S. degree in Electrical Engineering from Polytechnic Institute of New York, Brooklyn, and the M.S. and Ph.D. degrees in Electrical Engineering from Columbia University, NY.

He is with Bell Labs, Lucent Technologies, Murray Hill, NJ, where he is a Distinguished Member of Technical Staff (DMTS). Since 1991, he has been with the Mathematics of Networks and Systems Research Department at Bell Labs, Murray Hill, NJ, where he developed theory and algorithms for network resource management and QoS support, and for the analysis and engineering of multimedia traffic. He holds several patents. His current research interests include IP and optical network architectures, traffic engineering, and stochastic systems. He has been active in the Traffic Engineering and MPLS Working Groups of the Internet Engineering Task Force (IETF). He served on the executive and technical program committees of several conferences.

Dr. Elwalid is a member of Tau Beta Pi (National Engineering Honor Society) and Sigma Xi.