Scalable Fault Diagnosis in IP Networks using Graphical Models: A Variational Inference Approach

Rajesh Narasimha, Souvik Dihidar, Chuanyi Ji and Steven W. McLaughlin
Georgia Institute of Technology, Atlanta, GA 30318 USA
Email: {rajesh, dihidar, jic, swm}@ece.gatech.edu

Abstract- In this paper we investigate the fault diagnosis problem in IP networks. We provide a lower bound on the average number of probes per edge using variational inference technique proposed in the context of graphical models under noisy probe measurements. To obtain the bounds, we construct a graphical model using Bayesian networks. The advantages of the variational inference technique are the explicit choices of a simplifying conjugate function and a computationally tolerable approximation to address the intractable detection problem for large networks. We propose an entropy lower (EL) bound by drawing similarities between the coding problem over binary symmetric channel and the diagnosis problem and compare it against the variational lower bound. In addition, we discuss scalable and non-scalable scenarios in the presence of noise. Simulation results demonstrate that indeed the variational inference technique can provide a linear growth of the average number of probes per edge as a function of the network size.

1. INTRODUCTION

One of the key challenges of measurement-based network monitoring is how to accurately estimate the underlying network states using the limited amount of available network monitoring resources. For example, if the number of nodes are $V \geq 1000$, the total number of possible edges are approximately $\geq 500{,}000$, then a practical question to ask is what is the number of monitoring resources needed to determine the network states? This is the scalability issue. In addition, the other important concern is that: How rapidly would the monitoring resources required grow with the size of a network? For instance, suppose 100 monitoring resources are sufficient to obtain the accurate estimates of network states for a small network with 10 nodes. If this network becomes 200 times larger, e.g. with 2000 nodes, would the number of monitoring resources needed grow by 200, $(200)^2$ or 2^{200} times? In addition to the growth rate, the actual constant that multiplies the growth function is also of practical importance which together with the growth function provides an explicit number of measurements needed to diagnose the failures. Probe packets are intrusive and should be used only when necessary. A key challenge is how to accurately diagnose underlying network states using a limited amount of available monitoring resources, i.e., the number of probe packet measurements. A large network size poses a challenge in such probing scenario.

As a motivation, we provide here some practical scenarios where the scalability plays an important role. The Skitter tool from caida.org [1] is an example that gathers connectivity information, round trip time (RTT) and path data from source monitors scattered throughout the United States. Investigations have reported that in order to discover 131476 nodes and 279799 edges, Skitter [1] uses approximately 54 million probes (193 probes per edge). Ideally, Skitter requires that the packet overhead on the existing traffic and the number of packets sent be minimized. Another example is the Resilient Overlay Network (RON) [2] architecture that detects and recovers from path outages. RON suffers from scalability as the traffic growth is of $O(N^2)$ where N represents the nodes on the overlay network. Hence, the ideal goal is to be able to reliably localize all faulty links in a large network with little overhead, i.e., the number of probes inserted into the traffic.

In this work we focus on two aspects of failure diagnosis, namely detection and localization. We consider an approach based on path-based active probing which is to send probe packets from sources through network routes (path), and collect measurements as either received or lost probes. The measurements are then used to infer potential failures within the network. Simple examples of probes to analyze the status of a path are traceroute/ping and bandwidth monitoring tool such as the pipechar [3]. The downside of passive probing is that the routers have to be polled to collect network statistics such as delay, loss and available bandwidth. In an inter-domain setting, since the domains are managed by multiple ISPs and the service providers do not disclose confidential information about their domains. Hence active probing may be the only convenient choice when the interior of a network is not assessable directly [4] and thus avoids polling from internal nodes. The diagnosis can be easier using end-end measurements as they contain global information of a network.

In learning theory, the number of measurements/samples needed to obtain a desired generalization error is well studied and documented [5]. In general, application of learning techniques to real-time diagnosis in communication networks under rapidly changing stochastic environment has been investigated little. In networking, failure diagnosis based on event correlation [6], Bayesian belief networks [7] and heuristic algorithms [8] have been explored, but so far the scalability issue has not been studied sufficiently. In a recent work by Wen *et al* [9], an efficient fault diagnosis algorithm for all-optical networks is proposed based on sequential probing for both single and multiple failures. The focus of the work in [9] is on linear optical networks under noiseless conditions. Another related work is on failure diagnosis [10] which develops an algorithm for active probing and discusses mini-bucket approximations and their diagnosis complexity. As these related works are for focused conditions/networks, they motivate this work to provide a general approach for scalability of path-based diagnosis for IP networks.

From source coding, the minimum number of bits required to encode the message equals to the source entropy if all the messages of the source are equiprobable. The minimum average number of probes required to diagnose the network is equivalent to the minimum average number of bits needed to represent the source [9]. In our earlier work, using coding approach we have provided fundamental limits on the relationship between the number of probe packets, size of the network and the ability to perfectly identify all congested nodes [11] in the presence of noiseless measurements. The noisy probing scenario is similar to sending bits through a noisy channel and hence can be mapped to a binary symmetric channel (BSC) with the respective cross-over probabilities. Also, the mapping provides a way to analytically obtain the constant that multiplies the growth function.

1-4244-0353-7/07/$25.00 ©2007 IEEE

The main contributions of our work are: (a) Scalability of measurement-based network monitoring is formulated as a machine learning problem based on graphical models that characterize the spatial and statistical dependence in path-based probe measurements. (b) Variational inference [12, 13] and source-coding techniques are applied to derive a lower bound on the average number of observations *(m)* required in the presence of *noisy* probe outcomes given the prior probability of faults ρ and the network size n. If the number of observations is less than m, we are guaranteed not to achieve a zero diagnosis error asymptotically, where the error is quantified in terms of the *most probable explanation (MPE)* error and the bit-error rate (BER). We further simplify the bound by imposing an additional constraint that all probe lengths are equal and show that the average number of measurements can be lower bounded as $m \geq cn$, where $c = \log(1/\phi)/c_1; \phi = \max(\rho, 1-\rho)$ and c_1 is a scaling constant which is explicitly characterized in both our approaches. Thus, the lower bound provides a linear growth rate with respect to the size of the network, and an expression for the constant c. Note that the constant is generally intractable in most prior results in learning theory. (c) Next, we examine the growth rate of the average number of measurements as the noise parameters increase and investigate scalable and obtain non-scalable scenarios. (d) By mapping the noisy network probing scenario to a binary symmetric channel in source coding, we develop a lower bound on the average number of probe measurements in terms of the entropy of the prior probability of link failure ρ and the noise parameters. We then compare the lower bound obtained using variational method against the entropy lower bound. (e) In the end, we provide an algorithm based on maximizing the information gain of the network to localize single failures for path based active probing and provide insights on the optimum probe length.

2. PROBLEM FORMULATION

Given a network, the problem here is to determine the average number of probes per network state (i.e., link) needed so that the detection error is less than a given performance measure. From source coding, the minimum number of bits required to encode the message equals to the source entropy if all the messages of the source are equiprobable. The problem of fault diagnosis is similar to that of coding in the sense that the minimum average number of probes required to diagnose the network is similar to the minimum average number of bits needed to represent the source [9].

Consider a given network topology *G(V,E)*, where V and E are a set of nodes and links respectively. Let the status of link j be X_j, where X_j is a binary label, $X_j = 1$ if the link is working and $X_j = 0$ if the link is faulty, for $1 \leq j \leq n$, n being the total number of links in the network, i.e., $n = |E|$. Thus the vector $\boldsymbol{X} \in \{0,1\}^n$ represents the states of all the links in the network. Let m be the number of probes that are sent. Let $D_{m\times n}$ denote a matrix whose ij^{th} element is d_{ij}, where d_{ij}=1 if the i^{th} probe passes through j^{th} link, and d_{ij}=0, otherwise. Let Y_i be the i^{th} probe measurement, $Y_i = 1$ if the probe is received, and $Y_i = 0$, otherwise, for $1 \leq i \leq m$. Evidently, a probe is received if all links are working, i.e., Y_i is a logical AND operation $Y_i = \prod_j d_{ij}$. $\boldsymbol{Y} = (Y_i)_{i=1}^m$ is the measurement vector. Let $pa(Y_i)$ represents the i^{th} row of $D_{m\times n}$ matrix be the parent set of Y_i which consists of all the links on the path of Y_i.

Probe measurements are spatially dependent. For example, two probe measurements are "lost" if they intercept a broken link. Such a spatial dependence can be used to infer link/node states for intermediate nodes/links where measurements are not directly available. The complexity of exact inference is of $O(2^k)$ where $k = |pa(Y_i)|$, which is inefficient for large networks with a large parent set [13]. Consider Fig. 1(a), where a percentage of the links are faulty and the status of links $\boldsymbol{X}$ are considered to be random due to uncertain failure and network conditions. $\boldsymbol{Y}$ is also random due to measurement noise. Then as shown in the Fig. 1(a), the probing measurements and the links in the corresponding path form a directed graph [14]. In the graph, the links in the probed path are the parents $pa(Y_i)$ that influence the end-to-end measurement Y_i. The number of probe paths corresponds to the number of measurements *m*. The problem of diagnosis is to obtain the *most probable explanation (MPE)*, $x^*(\boldsymbol{Y}) = \arg\max_{\boldsymbol{x}} P(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{Y} = \boldsymbol{y})$, of the network states given the measurement set $\boldsymbol{Y}$. $x^*(\boldsymbol{Y})$ is the inferred status of the links given observation vector $\boldsymbol{Y}$. The performance in terms of the *MPE error* can be evaluated as follows,

$$P(\boldsymbol{X} \neq \boldsymbol{X}^*(\boldsymbol{Y})) = 1 - \sum_{\boldsymbol{y}} \max_{\boldsymbol{x}} P(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}) = 1 - \sum_{\boldsymbol{y}} P(\boldsymbol{X} = \boldsymbol{x}^*, \boldsymbol{Y} = \boldsymbol{y}) \quad (1)$$

The performance can also be represented by bit-error rate *(BER)* defined as $BER = \sum_{j=1}^{n} P(X_j \neq x_j^*(\boldsymbol{Y}))/n$, where $x_j^*(\boldsymbol{Y})$ is the most-likely assignment of X_j given the observations $\boldsymbol{Y}$. Similarly as in the case of (1), we can obtain an expression for the bit-error rate for the j^{th} link as follows,

$$BER(X_j) = P(X_j \neq x_j^*(\boldsymbol{Y})) = 1 - \sum_{\boldsymbol{y}} P(X_j = x_j^*, \boldsymbol{Y} = \boldsymbol{y}) \quad (2)$$

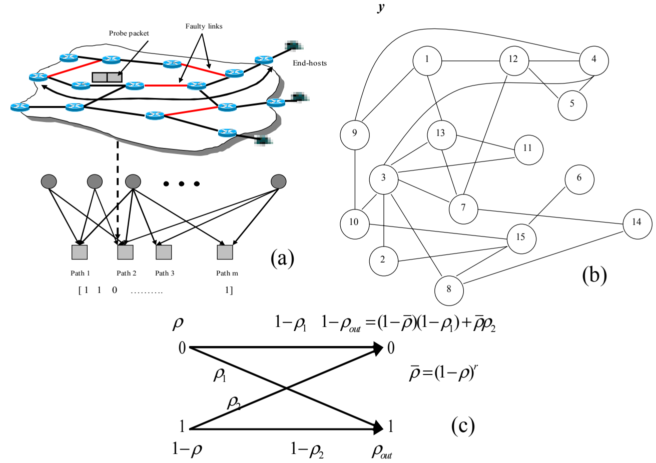

Fig. 1. (a) Representation of Fault in the network, (b) Sample Network topology and (c) Binary Symmetric Channel.

Given a desired performance measure, the scalability can be defined as follows.

Definition 1: *Let $[\delta_1, \delta_2]$ be the desired performance measures, i.e., $P(\boldsymbol{X} \neq \boldsymbol{X}^*(\boldsymbol{Y})) \leq \delta_1$ and $BER \leq \delta_2$. The scalability is the number of probe measurements m that are required to infer the status of the links with the desired performance as the size of the network n grows. If the number of measurements needed grows linearly with respect to n, the diagnostic approach is scalable.*

3. VARIATIONAL LOWER BOUNDS ON THE DIAGNOSIS ERROR

Computing x^* which is needed to calculate the error in terms of the network parameters is NP-hard [13] and hence a bound on the magnitude of the performance would provide useful information.

3.1. Noise Model

In order to obtain the bound, we make the following assumptions: (a) the underlying IP topology and the network size are known, (b) the prior probability of a faulty link (ρ) is known apriori and the links are assumed to fail independently and identically (*iid*), and (c) the noise in the probe measurements are *iid*. In reality the assumption (b) is valid since the diversity of traffic and links makes large and long-lasting spatial (correlated) link failures unlikely in a real network such as the Internet. Furthermore, the introduction of Random Early Detection (RED) policies in routers will help break such dependencies. Then, the joint distribution of $\boldsymbol{X}$ and the observation vector $\boldsymbol{Y}$, where lower case letters denote the realizations is given as,

$$P(\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y})=\left(\prod_{i=1}^{m}P(y_i\mid \boldsymbol{x})\right)\left(\prod_{j=1}^{n}P(x_j)\right)=\left(\prod_{i=1}^{m}P(y_i\mid \boldsymbol{x}_{pa(y_i)})\right)\left(\prod_{j=1}^{n}P(x_j)\right) \quad (3)$$

From Fig. 1(a), the probe outcomes are only dependent on the links in its path, i.e., $P(y_i\mid \boldsymbol{x})=P(y_i\mid \boldsymbol{x}_{pa(y_i)})$. Let us denote '+' as the positive evidence, $P(y_i=1)$ and '-' as the negative evidence, $P(y_i=0)$ [13]. Noise in the observations are modeled as [13]

$$P(y_i^+\mid \boldsymbol{x}_{pa(y_i)})=(1-p_{i0})\prod_{j\in\pi_i}(1-p_{ij})^{1-x_j} \quad (4)$$

where $p_{ij}=P(y_i=0\mid x_j=0); p_{i0}=P(y_i=0\mid L)$ and hence represents NOISY-OR Bayesian network. Here p_{ij} is the probability that the probe i fails given that the link j in the path fails. p_{i0} is the leak probability, where $L=\{x_j=1; j\in pa(Y_i)\}$, i.e., the probe fails although all the links in the network are operational (due to measurement errors). The second kind of noise is when the probe succeeds with a probability $1-p_{ij}$ even if a single link x_j on its path fails due to the casual independence assumption [13]. For example, the IS-IS routing protocol automatically recomputes alternate routes in case of IP link failures if such a route exists [15]. The re-routing noise, $1-p_{ij}$ is a function of the route length and hence depends on the physical network topology unlike the measurement noise, p_{i0}. We let $\theta_{ij}=-\log(1-p_{ij})$ [13].

3.2. MPE Error Evaluation

Exact MPE: In this section we derive an expression for the *exact MPE* evaluates this against the *approximate MPE* and *variational MPE*. From equation (3), $\prod_{j=1}^{n}P(x_j)=\prod_{j=1}^{n}(1-\rho)^{x_j}(\rho^{(1-x_j)})$ and

$$\prod_{i=1}^{m}P(y_i\mid x)=\prod_{i=1}^{m}\left[P(y_i^+\mid \boldsymbol{x}_{pa(y_i)})\right]^{y_i}\left[P(y_i^-\mid \boldsymbol{x}_{pa(y_i)})\right]^{1-y_i} \quad (5)$$

Plugging in the conditional densities in (5) and maximizing with respect to $\boldsymbol{x}$ and using equation (1) we have the expression for *exact MPE*. One should note that this is intractable for large networks.

Variational MPE: We now derive a lower bound on the *MPE error* using variational inference technique. For any concave function $f(x)$ the tangent plane serves a bound; $f(x)\le \xi^T x-f^*(\xi)$ [13]. The parameter ξ is known as the variational parameter. The bound is better for some values of the variational parameters than for others, and for a particular value of ξ the bound is exact, $f(x)=\min_{\xi}\{\xi^T x-f^*(\xi)\}$ [13]. A duality exists between $f(x)$ and the concave function $f^*(\xi)$.

Theorem 1: The lower bound on the MPE error using variational method is given as

$$1-\sum_{\boldsymbol{y}}\max_{\boldsymbol{x}}P(\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y})\ge 1-\sum_{\boldsymbol{y}}\phi^n\left(\prod_{i=1}^{m}e^{\beta_i y_i}e^{\alpha_i}\right) \quad (6)$$

where $\phi=\max_j(\max(\rho_j,1-\rho_j))$, $\alpha_i=\xi_i(\theta_0+\theta r_i)-f^*(\xi_i)$, $\theta_0=-\log(1-p_0)$, $\theta=-\log(1-p)$ are the noise parameters, r_i *is the route length of the* i^{th} *path* , $\xi_i=e^{-(\theta_0+\theta r_i)}/(1-e^{-(\theta_0+\theta r_i)})$ *is the optimal variational parameter,* $f^*(\xi_i)=-\xi_i\log(\xi_i)+(1+\xi_i)\log(1+\xi_i)$, *and* $\beta_i=-(\theta_0+\alpha_i)$.

Proof: Since noise parameters are *iid*, using the tangent plane as a bound and the conjugate function for NOISY-OR network [13] we obtain,

$$Q(X,Y)=\max_x P(\boldsymbol{X},\boldsymbol{Y})\le\phi^n\left(\prod_{i=1}^{m}\left[e^{-\theta_{i0}}\right]^{y_i}\left[e^{\xi_i(\theta_{i0}+\sum_{j\in\pi_i}\theta_{ij})-f^*(\xi_i)}\right]^{1-y_i}\right). \quad (7)$$

Assuming the network to be homogenous we have, $\alpha_i=\xi_i(\theta_0+\theta r_i)-f^*(\xi_i)$, and hence

$$Q(X,Y)\le\phi^n\left(\prod_{i=1}^{m}e^{-\theta_0 y_i}e^{\alpha_i(1-y_i)}\right)=\phi^n\left(\prod_{i=1}^{m}e^{-(\theta_0+\alpha_i)y_i}e^{\alpha_i}\right)=\phi^n\left(\prod_{i=1}^{m}e^{\beta_i y_i}e^{\alpha_i}\right) \quad (8)$$

where $\beta_i=-(\theta_0+\alpha_i)$. The expression optimum variational parameter can be derived as

$$\xi_i=e^{-(\theta_0+\theta r_i)}/(1-e^{-(\theta_0+\theta r_i)}) \quad (9)$$

Substituting (9) in (8) and using equation (1) the expression for *MPE error* is obtained. We assume that the network is uniform, i.e., $1-p_{ij}=1-p$ and $p_{i0}=p_0$ for simplicity. In (6) ϕ^n is dependent on the network topology *(n)* and fault prior whereas the term inside the parentheses is function of the path length, number of paths, and the noise parameters. Under noiseless condition, equation (6) is just a function of the network topology *(n)* and fault prior since the term $\left(\prod_{i=1}^{m}e^{\beta_i y_i}e^{\alpha_i}\right)$ simplifies to a constant as shown in (11).

Corollary 1: *Assume that the network topology with n links is symmetric (end hosts are equidistant from source nodes) and hence all the probe lengths are equal. The lower bound on the number of observations for asymptotic error free diagnosis is given as*

$$m/n\ge\frac{\log(1-\delta_1)/n-\log(\phi)}{(e^{-\theta_0}+e^{\alpha})}=\frac{\log(1-\delta_1)/n-\log(\phi)}{\log\left[(1-p_0)+(1-p_0)^{-\xi}(1-p)^{-r\xi}\xi^{\xi}(1+\xi)^{-(1+\xi)}\right]} \quad (10)$$

Proof: Under the equal probe length assumption we have, $\xi_i=\xi,\beta_i=\beta,\alpha_i=\alpha$ and by bounding the *MPE error* by the performance measure as $MPE\le\delta_1$, (10) is obtained.

Result 1: Noise-free case- In the absence of noisy observations the lower bound on the number of observations is given as

$$m\ge\left(\log(1-\delta_1)-n\log(\phi)\right)/\log(2) \quad (11)$$

When the fault priors are equal to 0.5 which is a worst-case scenario, i.e., $\phi=0.5$ then, $m\ge n$ for δ_1=0. Hence, we require the number of observations to be equal to the number of nodes on the bipartite graph, i.e., $m=n$ for the *MPE error* to be equal to zero asymptotically. In the case when both the noise parameters are large, i.e., $p_0=1-\varepsilon_1$, $1-p=1-\varepsilon_2(r)$ where $\varepsilon_2(r)$ is a function of the route length and $\varepsilon_1,\varepsilon_2(r)$ are negligible we have the following result.

Result 2: Non-scalable case- If $p_0=1-\varepsilon_1$, $1-p=1-\varepsilon_2(r)$, and

$\varepsilon_2(r) \ll \varepsilon_1$, since $\varepsilon_2(r)$ is assumed to be a multiplicative function of route length and hence depends on *n*, then (10) reduces to non-scalable case, i.e.

$$m \geq \left(\log(1-\delta_1) - n\log\phi\right) / \log(1+O(\varepsilon_2(r)) = O(n^k); k > 2 \quad (12)$$

Note that larger the network size, larger will be the route length which implies that if each link has a noise value ε_2, the denominator of (10) is a function of the longest route, $(\varepsilon_2)^{r_{\max}}$ for variable length probing and $(\varepsilon_2)^{r}$ for fixed length probing scheme as shown in Fig.2(c).

3.3. Bit-Error Rate (BER) Evaluation

We now provide the result in equation (2) using BER as a performance measure by relaxing the fixed probe length assumption.

Theorem 2: *The lower bound on the bit-error rate using variational method is given as*

$$BER \geq 1 - \phi(\eta_0 + \eta_1)^{\max_j |ch_j|} \quad (13)$$

where $\phi = \max_j(\max(\rho_j, 1-\rho_j))$,

$\eta_z = \max_{i\in\{1,2,..m\}} \max_{pa_{y_i}} P(y_i = z \mid pa_{y_i}); z \in \{0,1\}$, $\eta_1 = e^{-\theta_{i0} - r_{\max}\theta_{ij}}$; $r_{\max}$ is the maximum route length, $\eta_0 = e^{-\theta_{i0}}$, $\theta_{ij} = -\log(1-p_{ij})$ and $|ch_j|$ is the cardinality of the set of children of link *j*.

Proof: For the j^{th} link X_j, the joint probability can be bound as follows,

$$P(X_j = x_j^*, Y = y) = \max_{x_j}\left\{P(X_j)\prod_{y_i\in ch_j}\left[P(y_i^+ \mid pa_{y_i^+}\right]^{y_i}\left[P(y_i^- \mid pa_{y_i^-}\right]^{1-y_i}\right\}$$
$$\leq \phi\left\{\prod_{\substack{i=1\\ y_i\in ch_j}}^{|ch_j|}[\eta_1]^{y_i}[\eta_0]^{1-y_i}\right\} \quad (14)$$

As before, using *iid* noise assumption, the tangent plane as a bound, and the conjugate function for NOISY-OR network, we have,

$$P(X_j = x_j^*, Y = y) \leq \phi\left\{\prod_{\substack{i=1\\ y_i\in ch_j}}^{|ch_j|}\left[e^{-\theta_{i0}}\right]^{y_i}\left[1-e^{\xi_i(\theta_{i0}+r_{\max}\theta_{ij}) - f^*(\xi_i)}\right]^{1-y_i}\right\} \quad (15)$$

Assuming homogenous conditions, (15) simplifies to

$$P(X_j = x_j^*, Y = y) \leq \phi\left\{\prod_{\substack{i=1\\ y_i\in ch_j}}^{|ch_j|}\left[e^{-\theta_0}\right]^{y_i}\left[1-e^{\alpha}\right]^{1-y_i}\right\} = \phi\sum_{k=0}^{|ch_j|}\binom{|ch_j|}{k}\eta_0^k\eta_1^{|ch_j|-k} = \phi(\eta_0+\eta_1)^{|ch_j|} \quad (16)$$

Now the average *BER* can be lower bounded by the minimum *BER* over all the link *j* as follows,

$$BER = \frac{1}{n}\sum_{j=1}^{n}P(X_j \neq x_j^*(Y)) \geq \min_j BER(X_j) \geq 1 - \phi(\eta_0+\eta_1)^{\max_j |ch_j|} \quad (17)$$

Corollary 2: *If we bound the bit-error rate by the performance measure* δ_2, *the lower bound on the average number of observations for asymptotic error free diagnosis can be given as*

$$m/n \geq \frac{\log(1-\delta_2) - \log(\phi)}{r_{\max}(e^{-\theta_0} + e^{\alpha})}$$
$$= \frac{\log(1-\delta_2) - \log(\phi)}{r_{\max}\log\left[(1-p_0) + (1-p_0)^{-\xi^{opt}}(1-p)^{-r\xi^{opt}}\xi^{opt\,\xi^{opt}}(1+\xi^{opt})^{-(1+\xi^{opt})}\right]} \quad (18)$$

where $\xi^{opt} = \dfrac{e^{-(\theta_0+\theta r_{\max})}}{(1-e^{-(\theta_0+\theta r_{\max})})}$ and $\alpha = \xi^{opt}\left(\theta_0 + r_{\max}\theta\right) - f^*(\xi^{opt})$.

Proof: Since there are *n* links and *m* observations in the bipartite graph, *m/n* denotes the average number of observations per edge and hence $\max_j |ch_j| \simeq r_{\max}(m/n)$. The main difference between equations (10) and (18) is that the assumption on the fixed probe length is relaxed and (18) is valid for variable probing scenario. Next, if we assume that the route lengths (r) and the number of children of each parent node are equal then we have $\max_j |ch_j| = r(m/n)$. Hence, the lower bound simplifies in (17) to $BER \geq 1 - \phi(\eta_0+\eta_1)^{r(m/n)}$. The relation between average number of probes per edge required to obtain asymptotic error free diagnosis in the fixed length probing scenario can be given as $(m/n)_{BER} = (m/n)_{MPE}/r$.

4. ENTROPY LOWER BOUND

4.1. Scalability Analysis

We now consider how the number of measurements needed relates to the entropy of failure distribution. The minimum average number of probes required to diagnose the network is equivalent to the minimum average number of bits needed to represent the source [9]. The probing path in a noisy network is similar to a noisy channel and hence can be mapped to a binary symmetric channel (BSC) with the respective noise parameters as shown in Fig. 1(c). Entropy bound provides a basis for comparing the average number of probes needed for asymptotic error free diagnosis to that obtained by variational approach in (10) . Let the prior probability of link failure be ρ. ρ_1 be the probability that the probe succeeds even though a link on the path has failed which is equivalent to *1-p* and ρ_2 be the probability that the probe fails even due to measurement errors and is equivalent to p_0. For a BSC we have, $1-\rho_{out} = (1-\bar{\rho})(1-\rho_1) + \bar{\rho}\rho_2$ where $\bar{\rho} = (1-\rho)^r$ and *r* is the route length. In the absence of noise we have $\rho_{out} = \bar{\rho}$.

Lemma 1: *For a network of n links, the information gained about the link state* $\boldsymbol{X}$ *by a single probe* Y *of length r in the presence of measurement and re-routing noise and the prior probability of link failure* ρ *is given by* $I(\boldsymbol{X};Y) \leq H(\rho_{out})$.

Proof: From the relation, $I(\boldsymbol{X};Y) = H(Y) - H(Y\mid\boldsymbol{X})$, $H(Y\mid X)$ is zero since given the status of the links on the path of *y* the outcome is deterministic if we assume that the network is noise free. Hence $I(\boldsymbol{X};Y) = H(\rho_{out})$. In the presence of noise we have, $I(\boldsymbol{X};Y) \leq H(Y)$ and hence the result can be proved.

Lemma 2: *For a network of n links, the upper bound on the information gained about the link state* $\boldsymbol{X}$ *by m probes each of length* $\{k_i\}_{i=1}^m$ *is given by* $I(\boldsymbol{X};\boldsymbol{Y}) \leq \sum_{i=1}^{m}H((1-(1-\rho)^{k_i})(1-\rho_1) + (1-\rho)^{k_i}\rho_2)$.

Proof: For *m* probes the information gained can be bounded as,

$$I(\boldsymbol{X};Y_1,Y_2,.....Y_m) = H(Y_1,Y_2,.....Y_m) - H(Y_1,Y_2,.....Y_m\mid\boldsymbol{X})$$
$$\leq \sum_{i=1}^{m}H(Y_i) = \sum_{i=1}^{m}H((1-\rho)^{k_i}) \quad (19)$$

In the above result, if all the probes have the same lengths, then $I(\boldsymbol{X};Y_1,Y_2,...Y_m) \leq \sum_{i=1}^{2}I(\boldsymbol{X};Y_i) = mH((1-\rho)^k)$. In the presence of equal length probes and noisy observations it can be shown that $I(\boldsymbol{X};\boldsymbol{Y}) \leq mH(\rho_{out})$ and the equality holds for non-noisy case.

Theorem 3: *For a noisy BSC, the lower bound on the average number of probes per edge (m/n) for asymptotic error free diagnosis is given as* $m/n \geq H(\rho)/H(\rho_{out})$, *where* ρ *is the prior probability of link failure and* $1-\rho_{out} = (1-\bar{\rho})(1-\rho_1) + \bar{\rho}\rho_2$.

Proof: If the prior probability of failure ρ is known we would require on an average $nH(\rho)$ bits to represent $\boldsymbol{X}$. Hence from *lemma 2* we have,

$$nH(\rho) \le mH(\rho_{out}) \Rightarrow m/n \ge H(\rho)/H(\rho_{out}) \quad (20)$$

In the absence of noise, we have $m/n \ge H(\rho)$. The optimal probe length for noisy case is given as

$$r^{opt} = \max_r \{ H(\bar{p}(1-\rho_2) + \bar{p}\rho r/(1-\rho)\rho_1) - \bar{p}[H(p_0)] - \bar{p}\rho r/(1-\rho)[H(p)] \} \quad (21)$$

as shown in Fig. 3(a) & (b) and in the noiseless case, (21) reduces to $r^{opt} = \max_r H((1-\rho)^r)$ which is compared with the result from [9] for varying prior in Fig. 2(h) where the plot depicts an exact match.

Proposition 1: The expression for the information gained by sending a probe of length *r* on a network with *n* links in a single fault case is $I(\boldsymbol{X};Y) = H(1 - r\rho/n)$ as shown in Fig. 3(c).

Proof: Follows from similar argument as in *lemma 1*. Based on *proposition 1* an algorithm for single fault localization can be formulated based on *path-splitting* as follows,

START:

1. $r^{opt} = \left\lceil \max_r H(1 - r\rho/n) \right\rceil$

2. *destination* = r^{opt}

   if *(probe_outcome (destination) =0)*

      probe the first half of the links until fault is localized iterating step 1 and 2

   else

      probe the second half (n- r^{opt} *) of the links until fault is localized iterating step 1 and 2*

  end

END

5. LOWER BOUND COMPARISON AND NUMERICAL RESULTS

To compare the tightness of the variational lower bound, we generate a synthetic random topology of 15 nodes and 24 links using BRITE topology generator [16] where the routes were calculated using shortest path algorithm. We then employ greedy heuristic algorithm for obtaining the probing paths since this is well-known NP-hard problem [17]. The goal of this setup is obtain minimum number of paths that has maximum network coverage. In our simulations for the exact inference case, we averaged over 1000 runs and in each run we randomly generated the link states and the *iid* noise parameters p_0 and *1-p*. Fig. 2(g) demonstrates that the bound is tight for large noise values, p_0*=1-p=0.99.* Nonetheless, the result is insufficient to conclude about the tightness of the variational lower bound for large networks which is the limitation of this work. The difficulty is the exponential computational cost in evaluating the exact MPE error and BER for large networks for comparison. This is also the problem encountered in the related works [9, 10]. Nevertheless, the bounds provided for fault diagnosis are valid for any size network. Fig. 2(e) depicts the decrease in the *MPE error* when the number of observations are increased as a function of the network size for a variable routing scheme with a prior $\phi = 0.2$ and noise values, p_0*=1-p=0.01* and p_0*=0.75, 1-p=0.5* respectively. The results were averaged over 100 independent runs. Fig. 2(f) shows the effect of re-routing noise parameter *(1-p)* on the *MPE error* for fixed *(r=4 and 7)* and variable routing schemes. The MPE error decreases with the increase in probe length and number of probes since more information is gathered about the network state. In all the cases we chose the number of observations (*m=24),* p_0*=0.01* and fault prior, $\phi = 0.5$. Fig. 2(a-d) shows a comparison of the entropy lower bound from and variational MPE lower bound from on the average number of probes per edge for a network size of n_f*=2000,* varying prior and noise values. A fixed routing scheme *(r=1, 4, 7 and 12)* was used for the comparison. For noiseless case, the bounds intersect at $\phi = 0.5$ and as the noise parameters increase the intersection point shifts to the left as seen from the figure. Hence, we can observe that the entropy bound is tighter for small faulty priors and low noise values whereas the variational bound becomes tighter compared to the EL bound as noise values increases. The comparison of the constants is provided in Table 1. For low noise case, the growth rate of the measurements as a function of the network size is linear when compared to the non-scalable case in Fig. 2(d). The fixed routing scheme finds application in topology discovery scenarios where source routing is implemented and in routing policies used in dumbbell topology.

6. CONCLUSIONS AND FUTURE WORK

A lower bound on the average number of probes for asymptotic error free diagnosis under noisy measurements is provided using variational methods supported by simulation results. Motivated from source coding, entropy bound was formulated and compared against the variational bound in the presence of noise. The number of path-based measurements needed is shown to grow linearly with respect to the number of links in the network. The constant that multiplies the growth function is derived explicitly and shown to depend on the failure and the noise probabilities. As part of further investigation, we will assess the bounds using measurements from operational networks.

REFERENCES

[1] http://www.caida.org/tools/measurement/skitter/.

[2] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient Overlay Networks," *Proceedings of the 18th ACM SOSP, Banff, Canada*, October 2001.

[3] "http://dsd.lbl.gov/NCS/generic/net-tools.html."

[4] A. Habib, M. Khan, and B. Bhargava, "Edge-to-edge measurement-based distributed network monitoring," *Computer Networks*, vol. 44, pp. 211-233, Feb 2004.

[5] S. R. Kulkarni, G. Lugosi, and S. S. Venkatesh, "Learning Pattern Classification-A Survey," *IEEE Transactions on Information Theory*, vol. 44, pp. 2178-2206, October 1998.

[6] S. Klinger, S. Yemini, Y. Yemini, D. Ohsie, and S. Stolfo, "A coding approach to event correlation," *Proceedings of the fourth international symposium on Integrated network management IV*, pp. 266 - 277, 1995.

[7] M. Steinder and A. S. Sethi, "Probabilistic fault localization in communication systems using belief networks," *IEEE Transactions on Networking*, vol. 5, pp. 809-822, October 2004.

[8] I. Katzela and M. Schwartz, "Schemes for Fault identification in communication network," *IEEE Transactions on Networking*, vol. 3, December 1995.

[9] Y. Wen, V. W. S. Chan, and L. Zheng, "Efficient Fault Diagnosis Algorithms for All-Optical WDM Networks," *IEEE/OSA Journal of Lightwave Technology Special Issue on Optical Networks*, October 2005.

[10] I. Rish, M. Brodie, S. Ma, N. Odintsova, A. Beygelzimer, G. Grabarnik, and K. Hernandez, "Adaptive diagnosis in distributed systems," *IEEE Transactions on Neural Networks*, vol. 16, pp. 1088-1109, September 2005.

[11] R. Narasimha, S. Dihidar, C. Ji, and S. W. McLaughlin, "A Scalable Probing-based Approach for Congestion Detection using Message Passing," *Proceedings of ICC*, 2006.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

[12] M. I. Jordan, *Learning in Graphical Models*: The MIT Press (Adaptive Computation and Machine Learning), 1998.

[13] T. S. Jaakkola, "Variational Methods for Inference and Estimation in Graphical Models," in *Department of Brain and Cognitive Sciences*, vol. Ph.D. Cambridge: Massachusetts Institute of Technology, 1997.

[14] M. Brodie, I. Rish, and S. Ma, "Intelligent Probing: A cost-effective approach to fault diagnosis in computer networks," *IBM Systems Journal*, vol. 41, 2002.

[15] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C.-N. Chuah, and C. Diot, "Characterization of Failures in an IP Backbone," *IEEE INFOCOM*, 2004.

[16] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: Universal Topology Generation from a user's Perspective," *Technical Report (User Manual) BU-CS-TR-2001-003*, April 2001.

[17] V. V. Vazirani, *Approximation Algorithms*: Springer, 2004.

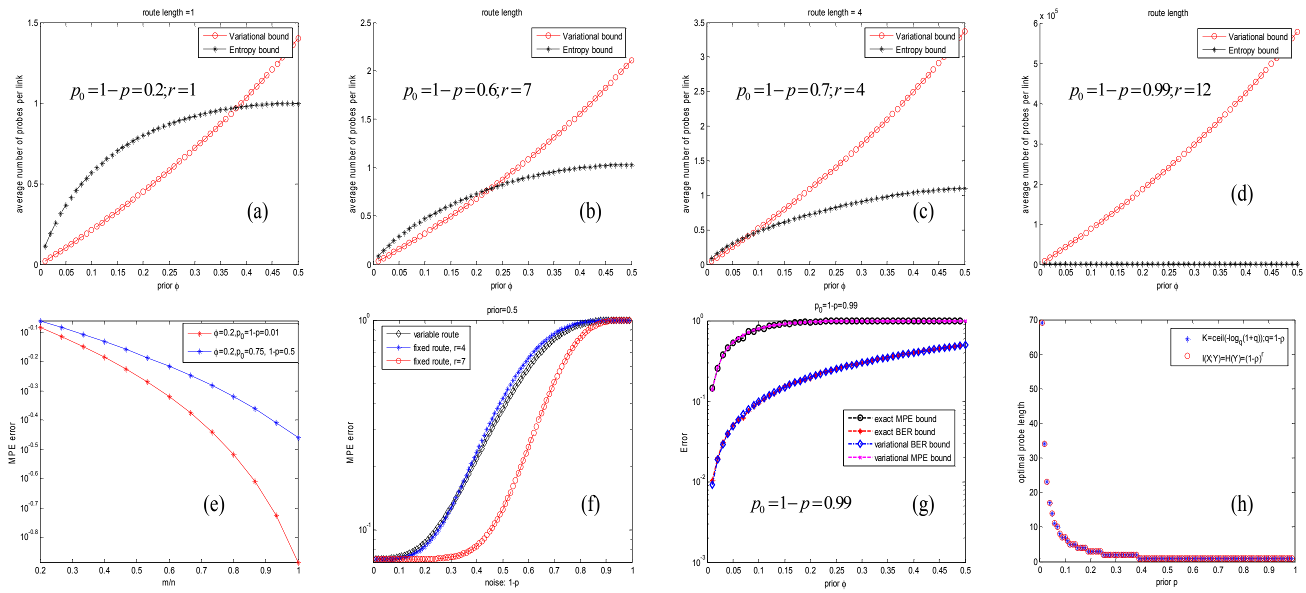

Fig. 2. (a-d) Comparison of the variational and the entropy lower bound on the average number of probes per edge for varying noise values, (e) variational *MPE error* lower bound as a function of (*m/n*), (f) variational *MPE error* lower bound as a function of the re-routing noise, (g) Comparison of the exact *MPE* error/*BER* vs. variational *MPE* error/*BER* lower bounds, and (h) optimal probe length comparison.

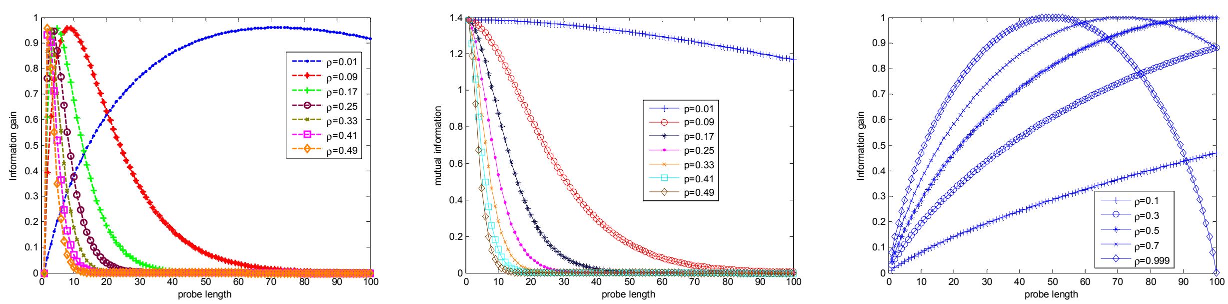

Fig. 3. (a) Optimal probing length for varying prior ρ with noise ($1-p=p_0=0.01$), (b) Optimal probing length for varying prior ρ with noise ($1-p=p_0=0.99$) and (c) Optimal probing length ($r^{opt}=\left\lceil \max_r H(1-r\rho/n) \right\rceil$).

VB Bound	*Constants*
MPE Error	$c_1=\left(\log\left[(1-p_0)+(1-p_0)^{-\xi}(1-p)^{-r\xi}\xi^{\xi}(1+\xi)^{-(1+\xi)}\right]\right)^{-1}$; $\delta_1=0$
BER	$c_2=\left(r_{\max}\log\left[(1-p_0)+(1-p_0)^{-\xi^{opt}}(1-p)^{-r\xi^{opt}}\xi^{opt\,\xi^{opt}}(1+\xi^{opt})^{-(1+\xi^{opt})}\right]\right)^{-1}$; $\delta_2=0$
EL Bound	$c_3=\left(H(\rho_{out})\right)^{-1}; 1-\rho_{out}=(1-\bar{\rho})(1-\rho_1)+\bar{\rho}\rho_2$

Table 1. Comparison of the constants for various lower bounds.