Binary vs Analogue Path Monitoring in IP Networks

Hung X. Nguyen and Patrick Thiran

School of Computer and Communication Sciences, EPFL
CH-1015 Lausanne, Switzerland
{hung.nguyen, patrick.thiran}@epfl.ch

Abstract. Monitoring systems that can detect path outages and periods of degraded performance are important for many distributed applications. Trivial pair-wise probing systems do not scale well and cannot be employed in large networks. To build scalable path monitoring systems, two different approaches have been proposed in the literature. The first approach [1], which we call the continuous or analogue model, takes real measurement values and infers the performance metrics of unmeasured paths using traditional $(+, \times)$ algebra. The second approach [2], which we call the Boolean model, takes binary values from measurements (e.g., whether the delay/loss of an end-to-end path is above a given threshold) and infers the performance quality of unmeasured paths using Boolean algebra. Both approaches exploit the fact that end-to-end paths share network links and hence that the measurements of some paths can be used to infer the performance on others. In this work, we are only interested in detecting whether the performance of a path is below an acceptable level or not. We show that when the number of *beacons* (nodes that can send probes and collect monitoring information) is small, the Boolean model requires fewer direct measurements; whereas for a large number of beacons the continuous model requires fewer direct measurements. When the number of beacons is significantly large, however, there is no difference in terms of the number of paths that we need to measure directly in both models. We verify the results by simulations on inferred network topologies and on real measurement data.

1   Introduction

Network dynamics may significantly affect the performance of distributed applications such as distributed system management, replicated services, and application layer multicast [3]. Robust and efficient distributed systems therefore need to adapt their behavior to environment changes. Loss rate monitoring systems that can detect path outages and periods of degraded performance can both facilitate distributed system management (such as virtual private network (VPN) or a content distribution network), and help build adaptive overlay applications, e.g., streaming media [4].

Monitoring systems that target small networks [5] usually employ pair-wise probing where each node probes the paths from itself to all other nodes. For

a network monitoring system with n_B beacons, the number of possible end-to-end measurements is $O(n_B^2)$. Therefore, active end-to-end measurement in such settings does not scale well and as a result cannot be deployed for complete network-wide measurement. Furthermore, this approach generates highly redundant measurements where many links in the network are repeatedly measured. It is therefore important to have a scalable overlay monitoring system that does not generate redundant information. Existing scalable network estimation systems determine network characteristics by measuring end-to-end paths periodically. To monitor a path, a node at one end of the path periodically sends probe packets to the node at the other end. From the delay characteristics and delivery status of these probe/acknowledgement packet pairs, the sending node can infer the quality of the path. This method is similar to network tomography approaches that infer the internal network characteristics based on end-to-end observations. Network tomography has been extensively studied in the literature([6] provides a detailed survey). However, to the best of our knowledge, none of the existing tomography works deal with minimizing the number of probes that need to be sent in a tomography system.

Chen et al. [1] have shown that it is possible to reconstruct complete end-to-end path properties exactly from the measurements of only a subset of paths. The results in [1] are based on a linear algebraic analysis of routing matrices of the monitoring systems where a routing matrix is the binary matrix that specifies the links that occur in a given path. Since a maximal set of independent paths can be used to recover any other path in the network, it is enough to monitor only this set. The number of independent paths in a monitoring system, which is the rank of the routing matrix, tends to be much smaller than the total number of paths. A similar approach is given in [7] where only bounded estimations for network paths can be achieved. Chua et al. [8] show that a significantly smaller number of direct path measurements than that are required by the monitoring systems of [1] can be used to approximate some network wide properties.

Padmanabhan et al. [9] studied the end-to-end packet loss rate experienced by clients of the Web server at *microsoft.com*. They report that the correlation between end-to-end loss rate and hop count is weak, which suggests that end-to-end paths are dominated by a few lossy links. Furthermore, the end-to-end loss rate is stable for several minutes. A notable feature of the model considered in [9] is that its parameters (the loss rates on the logical links) are not statistically identifiable from the data (the server-to-clients loss rate), meaning that there exist different sets of parameters that give rise to the same statistical distribution of data. Although the model is not statistically identifiable, some methods proposed in [9] are quite successful in identifying the lossiest links, both in simulated and real networks. The underlying reasons behind the success of the methods in [9] are the nature of link performance in the studied networks [10]. In such networks, suppose that we can classify links as "good" or "bad" with performance measures sufficiently far apart, then the performance experienced along a network path will be bad only if one of its constituent links is bad. Duffield et al. [10] calls this kind of link performance a *separable perfor-*

mance and identifies many separable performance metrics of network links such as connectivity, high-low loss model, and delay spike model.

A special case of separable performance, where links are either up or down, has been widely studied in the literature [2, 11, 12]. In our previous work [2], we showed that analogously to the work of Chen et al. [1], it is sufficient to monitor only a subset of end-to-end paths to infer the connectivity of all end-to-end paths in the network. The difference between [1] and our model is that our model relies on Boolean $(\max, \times)$ algebra instead of traditional $(+, \times)$ algebra (which we also refer to as the "continuous model"). The end-to-end paths that need to be monitored are those that form the basis of the row space of the routing matrix in this Boolean algebra.

In this work, we compare the efficiency of the continuous and Boolean path monitoring systems for separable performance metrics. Since we are only interested in the classification of links as "good" or "bad", depending on whether the metric of interest has exceeded or not a given threshold, the continuous and Boolean models bring the same information. To obtain this same information, however, they will need different sets of end-to-end measurements. Thus we want to determine the system that uses fewer direct path measurements. Specifically, we are interested in comparing the dimension of the basis of the vector space in Boolean algebra for the Boolean model with the dimension of the basis of the same vector space in traditional $(+, \times)$ algebra for the continuous model. Our main contributions are as follows.

- First, we show that the Boolean model in [2] can be used for other separable performance metrics.
- Second, by simulations on Rocketfuel topologies [13], we show that for a separable performance when the number of beacons is small, it is better to use the Boolean model. On the contrary, when the number of beacons is large, it is better to use the continuous model. However, when the number of beacons is significantly large, both models result in the same number of paths that need to be measured directly. We also provide some intuitive explanations for the simulation results.
- Finally, we verify our results on the data set gathered by the NLANR's AMP infrastructure [14].

The remainder of this paper is organized as follows. We introduces the network models and the basic algorithms in Section 2. We presents the numerical comparison of the two models on Rocketfuel ISP topologies in Section 3. Evaluations of the performance of the models on the NLANR's AMP active measurement infrastructure are given in Section 4. Finally, we conclude the paper in Section 5.

2   Network Model and Basic Algorithms

The network is modelled as an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where the graph nodes, $\mathcal{V}$, denote the network components and the edges, $\mathcal{E}$, represent the communication links connecting them. The number of nodes and edges is denoted by $n = |\mathcal{V}|$

and $e = |\mathcal{E}|$, respectively. Suppose there are n_B beacons that belong to a single or confederated overlay monitoring system. They cooperate to share an overlay monitoring service, and are instrumented by a central authority. The set of all beacons is denoted by $\mathcal{V}_B$. Furthermore, we use $P_{s,t}$ to denote the path traversed by an IP packet from a source node s to a destination node t. Let $\mathcal{P}$ be the set of all paths between the beacons on the network and let $n_p = |\mathcal{P}|$.

For a known topology $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a set of paths $\mathcal{P}$, we can compute the routing matrix D of dimension $n_p \times e$ as follows. The entry $D_{ij} = 1$ if the path $P_{s,t} \equiv P_i$, with $i = (s,t)$, contains the link e_j and $D_{ij} = 0$ otherwise. A row of D therefore corresponds to a path, whereas a column corresponds to a link. Note here that if a column contains only zero entries, the link corresponding to that column does not have any effect on the performance of the paths in $\mathcal{P}$. We drop these columns from the routing matrix to obtain a matrix of dimensions $n_p \times n_l$, where $n_l \le e$ is the number of links that are covered by at least one path in $\mathcal{P}$.

Our performance model is as follows. During some measurement period, each beacon sends a set of packets to each destination (chosen among other beacons). When traversing link e_j, each packet is subject to a performance degradation (e.g. loss or delay) according to a distribution specified by a parameter ϕ_{e_j}. If the path P_i comprises links $e_1, ..., e_m$, the performance degradation along the path follows a composite distribution described by the parameters $\phi_i = \{\phi_{e_1}, ..., \phi_{e_m}\}$.

2.1 Continuous $(+, \times)$ algebraic model

In the continuous model [1], the performance parameters ϕ take values in $\mathbb{R}$. Let $y \in \mathbb{R}^{n_p}$ be a vector that represents a metric measured on all paths $P_i \in \mathcal{P}$. y is linearly related to the value $x \in \mathbb{R}^{n_l}$ of that same metric over the links $e_j \in \mathcal{E}$.

For example, letting ϕ_i denote the packet loss probability on path P_i and ϕ_{e_j}, the corresponding probability on link e_j, and assuming independence among loss-events on links, the relation between the path-wise and link-wise loss probability becomes

$$y = Dx = \left[\sum_{j=1}^{n_l} x_j D_{ij} \right]_{1 \le i \le n_p}, \tag{1}$$

where $y_i = \log(1 - \phi_i)$ and $x_j = \log(1 - \phi_{e_j})$.

There are $n_p = O(n_B^2)$ equations in (1). However, in general the matrix D is rank deficient, i.e., $k = rank(D) < n_p$. For the sake of simplicity, assume that the first k rows of D form a basis of the row vector space $\mathcal{R}(D)$ of D. Because every row vector d_i, $i > k$, of D can be represented as a linear combination of the first k independent row vectors, we can write that $d_i = \sum_{j=1}^{k} \alpha_j d_j$ for some α_j, $1 \le j \le k$. The metric y_i of the path P_i can be obtained from $y_1, ..., y_k$ as: $y_i = \sum_{j=1}^{k} \alpha_j y_j$. Therefore, only k independent equations of the n_p equations in (1) are needed to compute all elements of y, and as a result we only need to measure k paths, which form a basis of $\mathcal{R}(D)$, to obtain the loss rate on all paths [1].

2.2 Boolean (max, ×) algebraic model

In the Boolean model [2], the values of the performance parameters ϕ are partitioned into two subsets that we call "good" and "bad". We call the link e_j bad if and only if its parameter ϕ_{e_j} is bad and we call the path $P_i = \{e_1, ..., e_m\}$ bad if and only if ϕ_i is bad. The partitions are called separable when a path is bad if and only if at least one of its constituent links is bad. For example, in the loss model LM_1 of [9], good links have loss rates ϕ_{e_j} uniformly distributed between 0% and 1%; bad links have loss rates uniformly distributed between 5% and 10%. Taking the threshold between good and bad path transmission rates as 0.95, this model is separable if each path does not contain more than 5 links.

In a separable model, a path is bad if and only if at least one of its constituent links is bad. If we use the variable y_i to represent whether the path P_i is good ($y_i = 0$) or bad ($y_i = 1$) and the variable x_j is used to represent whether the network link e_j is good ($x_j = 0$) or bad ($x_j = 1$), we then have:

$$y_i = \bigvee_{j=1}^{n_l} x_j \cdot D_{ij} \text{ for all } i, \tag{2}$$

where "$\vee$" denotes the binary max operation, and "$\cdot$" denotes the usual multiplication operation.

Let us now introduce some concepts of Boolean vector spaces that are useful for the analysis of the Boolean model. Let $\mathcal{D} = \{d_i\}_{1 \leq i \leq h}$ be a set of binary vectors of equal length, and let $I = \{1, ..., h\}$ be the index set of $\mathcal{D}$. A vector span $\mathcal{S}$ can be defined on $\mathcal{D}$ as follows.

Definition 1. *[Vector span] The vector span of $\mathcal{D}$ is*

$$\mathcal{S} = <\mathcal{D}> = \{\bigvee_{i \in I} \alpha_i \cdot d_i \mid \alpha_i \in \{0, 1\}, d_i \in \mathcal{D}\}$$

Vectors in $\mathcal{D}$ are called the *generator vectors* of $\mathcal{S}$.

It was shown in [2] that each vector span $<\mathcal{D}>$ has a unique basis $\mathcal{B}$, which is the smallest set of vectors in $\mathcal{D}$ such that all other vectors in $<\mathcal{D}>$ can be written as a linear combination of vectors in $\mathcal{B}$. Let $b = |\mathcal{B}|$, b is called the dimension of $<\mathcal{D}>$. Without loss of generality, we assume that the first b rows of D form a basis of the row span $<\mathcal{D}>$ of D. Because every row vector d_i, $i > b$, of D can be represented by a linear combination of the first b independent row vectors, we can write that $d_i = \bigvee_{j=1}^{b} \alpha_j \cdot d_j$ for some α_j. The value y_i of the path P_i can be obtained from $y_1, ..., y_b$ as: $y_i = \bigvee_{j=1}^{b} \alpha_j \cdot y_j$. Therefore, only b independent equations of the n_p equations in (2) are needed to compute all elements of y, and hence we only need to measure b paths, which form a basis of $<\mathcal{D}>$, to determine whether any path in the overlay system is good or bad.

2.3 A brief comparison of the two models

If we are only interested in the classification of end-to-end paths as "good" or "bad", the continuous and Boolean models bring the same information. However,

as we will show in this section, the two models use different algebraic structures, and as a result, they usually need different sets of end-to-end measurements to obtain this same information.

General properties relating network topologies with the dependency between measurement paths in the continuous and Boolean models are kept for future work. In this section, we are interested in the conditions under which a set of linearly dependent/independent vectors in the continuous model is dependent/independent in the Boolean model and vice versa. The observations in this section are useful to explain the results of our simulation and experimental studies in Sections 3 and 4. We first show that a linearly dependent set of vectors in $(+, \times)$ algebra is not necessarily linearly dependent in Boolean algebra and vice versa. This assertion can be verified in the following examples. The set of four vectors: $d_1 = \{1, 1, 1, 1\}, d_2 = \{0, 0, 1, 1\}, d_3 = \{1, 1, 1, 0\}$ and $d_4 = \{0, 1, 0, 1\}$ is linearly dependent in Boolean algebra as $d_1 = d_2 \vee d_3 \vee d_4$, but is linearly independent in $(+, \times)$ algebra; whereas the set of four vectors: $d_1 = \{1, 0, 0, 1\}, d_2 = \{0, 0, 1, 1\}, d_3 = \{1, 1, 0, 0\}$ and $d_4 = \{0, 1, 1, 0\}$ is linearly dependent in $(+, \times)$ algebra as $d_1 = d_2 + d_3 - d_4$ but not in Boolean algebra. It is not difficult to verify that for the Boolean model, the necessary and also sufficient condition for a set of vectors to be dependent is that one vector has entries of 1s at all the positions where other vectors have entries of 1s. Clearly, this statement does not apply for the continuous model as shown in the first example.

3 Numerical evaluations

We conducted a series of numerical studies in order to obtain a preliminary comparison of the efficiency of the continuous and Boolean models in monitoring end-to-end network properties. We perform our investigations on three backbone ISP topologies with sizes ranging from small (Exodus: 80 nodes and 147 links) to medium (Telstra: 115 nodes and 153 links), and large (Tiscali: 164 nodes and 328 links). For the sake of simplicity, we assume that all the ISPs use shortest path routing to route traffic. In this section, we summarize our findings and provide explanations for the results in the context of the Rocketfuel topologies.

Recall that n is the number of nodes in the network. In our experiments, the number of beacons $|\mathcal{V}_B|$ is varied from $n/50$ to $n/2$. We select the beacon candidates randomly by picking a random permutation of the set of nodes in the network. After building the routing matrix as in Section 2, we first calculate the rank of the routing matrix D to obtain the number of end-to-end measurements for the continuous model, and then use the PS algorithm in [2] to find the number of end-to-end measurements for the Boolean model. For each topology, we plot the percentage of independent paths returned by the PS algorithm (for the Boolean model) and the rank of the matrix D (for the continuous model) for different numbers of beacons.

Fig. 1 shows the results for the Exodus topology. We observe that for a small number of beacons $|\mathcal{V}_B|$ (less than 5%) the Boolean model requires fewer direct
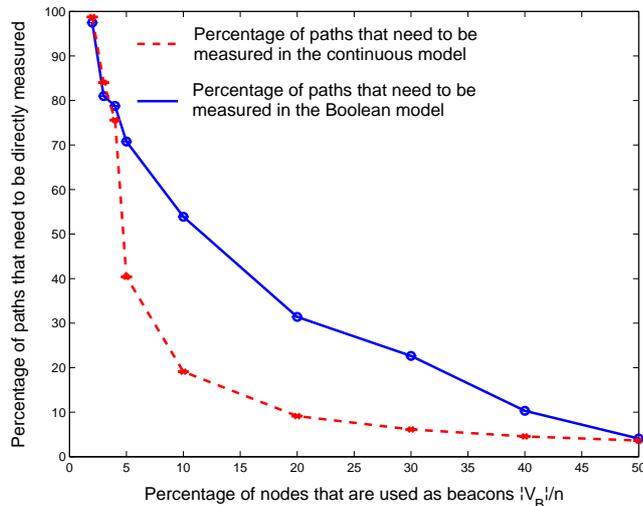
Fig. 1. Percentage of paths that need to be measured directly for complete determination of the quality of all paths in the Exodus topology with 80 nodes and 147 links.

measurements, whereas for a larger number of beacons $|\mathcal{V}_B|$ (between 10%-40%) the continuous model requires fewer direct measurements. However, when the percentage of beacons is above 50%, both models require the same number of direct measurements.

Similar observations can be found for the Telstra and Tiscali topologies, as shown in Fig. 2 and 3, even though the exact percentages at which the two curves representing the binary and continuous model cross are different in each topology. The results suggest that although all topologies exhibit the same tendencies for the number of direct measurements in the continuous and Boolean model, exact results in each network are sensitive to the network topology. We also observe that for all three topologies, the percentages of required direct measurements for both continuous and Boolean models drop rapidly with the number of beacons when the latter is small. However, these percentages decrease slowly when the number of beacons is large. This observation suggests that the benefit of introducing additional beacons lessens with the increase of the number of beacons.

The above observations can be explained by the structure of a basis of a vector space in $(+, \times)$ algebra and the basis of a vector span in Boolean algebra. In the Boolean model, and contrary to the continuous model, if all links in a path are also present in some other paths, then the first path is redundant because it is a linear combination of the other paths. When the number of paths is small, this situation frequently occurs in the studied ISP networks. In $(+, \times)$ algebra, the rank of a matrix is upper bounded by the minimum of the number of rows and columns in the matrix. Hence, when we initially increase the number of
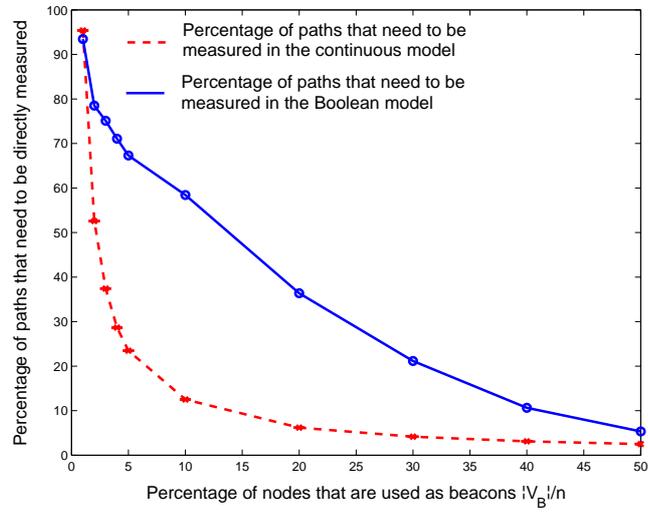
Fig. 2. Percentage of paths that need to be measured directly for complete determination of the quality of all paths in the Telstra topology with 115 nodes and 153 links.
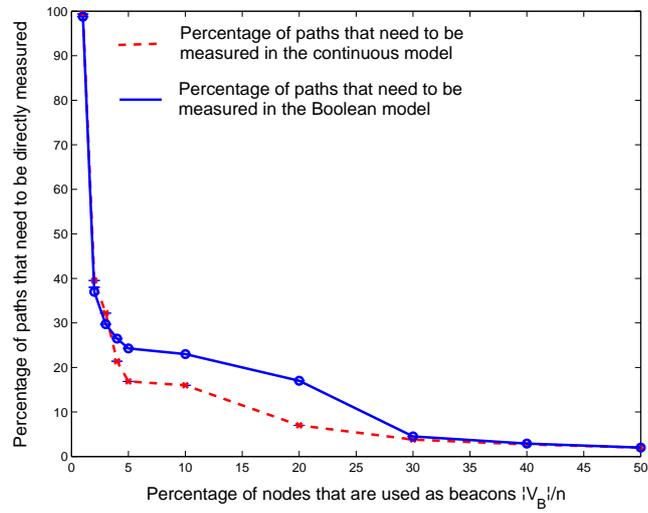

Fig. 3. Percentage of paths that need to be measured directly for complete determination of the quality of all paths in the Tiscali topology with 164 nodes and 328 links.

paths (i.e., the number of rows) the dimension of the basis rapidly increases. But when the rank of the matrix approaches the number of columns, which stays almost constant, it increases only slowly. When there are a lot of paths, both the Boolean vector span and continuous vector space have the same basis that contains mostly unit vectors (vectors that have only one 1 entry).

4   Internet evaluation

In this section, we compare the performance of the continuous and the binary models on the data collected from the NLANR's AMP (Active Measurement Project) [14] measurement infrastructure, which performs site-to-site active measurements between campuses connected by high performance networks (most of them at NSF-funded HPC-sites). The NLANR measurement infrastructure consists of approximately 140 monitors and about 15,000 end-to-end paths. Each monitor performs RTT, topology and throughput tests to all other monitors. More details can be found at [14]. We have analyzed the data collected by AMP on the 23rd of January, 2004. After grouping the common edges and nodes in all end-to-end paths, we have a sample of 133 monitors and 17556 end-to-end paths that cover 9405 nodes and 36674 directed edges. Note here that the number of available monitors (133 monitors) is significantly smaller than the number of nodes (9405 nodes) in the systems. The AMP system is therefore operating in the regime where only a small percentage (below 1.4%) of nodes are used as beacons. To evaluate the effect of the number of beacons on the two models, we vary the number of beacons from 1 to 133, which corresponds to 0.01%-1.4% of the total number of nodes. We then construct the routing matrix D and calculate the rank of D to obtain the number of end-to-end measurements for the continuous model. We also calculate the number of independent paths in the Boolean model using the PS algorithm in [2]. The results are plotted in Fig. 4.

We observe that the results reflect the behaviors that we have already seen for the Rocketfuel topologies. That is, in the regime where the percentage of nodes that are used as beacons is very small (below 1% in this case) the Boolean model requires fewer direct measurements if we are only interested in separable "good" or "bad" performance of paths. However, the difference between the two models is very small for this range, which can be explained by the fact that in this case most of the end-to-end paths in the network are independent in both models. We suspect that the difference becomes more important, and in favor of the continuous model, for larger numbers of beacons. Furthermore, for both the continuous and Boolean models the percentage of direct measurements drops rapidly as the number of beacons increases.

We also evaluate the accuracy of the two approaches on predicting the quality of unmeasured paths. For this purpose, we proceed as follows. First, we fix the number of beacons. We then determine the set of independent paths in the continuous and Boolean models. We called these independent paths the *directly measured paths*. Using performance data from these directly measured paths, we calculate the performance on the unmeasured paths. That is, in the continuous
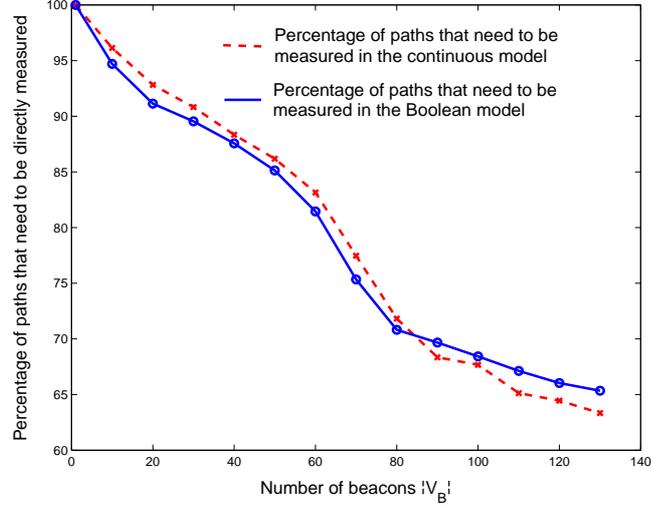
Fig. 4. Percentage of paths that need to be measured directly for complete determination of the quality of all paths in the NLANR's AMP monitoring system.

model, we calculate the loss rates on all unmeasured paths and compare them against a loss threshold of 5%, which is the threshold between "tolerable loss" and "serious loss" as defined in [15], to determine whether the path is good or bad. In the Boolean model, we first determine the quality of the directly measured paths and then calculate the quality on the unmeasured paths. We compare the inferred results of the two models against the real measurement data. The results are given in Table 1. We observe that both models achieve

Table 1. Accuracy of the continuous and Boolean models. CV is the coverage rate and FP is the false positive rate

Number of beacons	The continuous model	The Boolean model
2	CV = 99.7%, FP = 0.3%	CV = 98.7%, FP = 0.9%
100	CV = 97.2%, FP = 3.4%	CV = 95.6%, FP = 4.7 %

a high coverage rate and a low false positive rate. The Boolean model is less accurate than the continuous model because the Boolean model relies on the assumption that the loss rates of network paths and links are separable, which sometimes does not hold in practice.

5 Conclusion

In this paper, we have compared the performance of two end-to-end path monitoring approaches. We show that when the number of beacons is small, the Boolean model requires fewer direct measurements; but the difference appear to be very small on real measurements. For a large number of beacons the continuous model requires fewer direct measurements, and the different can be quite significant. However, when the number of beacons is significantly large, there is no difference in terms of the number of paths that we need to measure directly in both models. We verify the results by simulations on existing ISP topologies and on real measurement infrastructure.

We are currently working on various extensions of this work. First, we are investigating the influence of the structure of the routing matrix on the differences between the number of probes required for the continuous and Boolean models. Second, so far in this work we have taken the restriction that nodes in the monitoring systems can send probes only to other nodes in the monitoring systems. However, since probes can be sent to many other nodes in the network, we are investigating how sending probes to other nodes in the network would affect the efficiency of the monitoring system in both the continuous and Boolean models.

Acknowledgements

References

1. Chen, Y., Bindel, D., Song, H., Katz, R.H.: An algebraic approach to practical and scalable overlay network monitoring. In: Proceedings of the ACM SIGCOMM, Portland (2004)
2. H.X.Nguyen, Thiran, P.: Active measurement for failure diagnosis in IP networks. In: Proceedings of the Passive and Active Measurment Workshop, Juan-les-Pins, France (2004) 185–194
3. Braynard, R., Kostic, D., Rodriguez, A., Chase, J., Vahdat, A.: Opus: an overlay peer utility service. In: Proceedings of the 5th International Conference on Open Architectures and Network Programming (OPENARCH). (2002)
4. Chen, Y.: Toward a Scalable, Adaptive and Network-aware Content Distribution Network. PhD thesis, University of carlifornia at Berkeley (2003)
5. Andersen, D.G., Balakrishnan, H., Kaashoek, M.F., Morris, R.: Resilient overlay networks. In: Proceeding of the 18th ACM Symp. on Operating System Priciples. (2001) 131–145
6. Coates, M., Hero, A., Nowak, R., Yu, B.: Internet tomography. IEEE Signal Processing Magazine **19** (2002)

7. Tang, C., McKinley, P.: On the cost-quality tradeoff in topology-aware overlay path probing. In: Proceedings of the IEEE ICNP. (2003)
8. Chua, D.B., Kolaczyk, E.D., Crovella, M.: Efficient monitoring of end-to-end network properties. private communication (2004)
9. Padmanabhan, V.N., Qiu, L., Wang, H.J.: Server-based inference of internet performance. In: Proceedings of the IEEE INFOCOM'03, San Francisco, CA (2003)
10. N.Duffield: Simple network perormance tomography. In: Proceedings of the IMC'03, Miami Beach, Florida (2003)
11. Bejerano, Y., Rastogi, R.: Robust monitoring of link delays and faults in IP networks. In: Proceedings of the IEEE INFOCOM'03, San Francisco (2003)
12. Horton, J., Lopez-Ortiz, A.: On the number of distributed measurement points for network tomography. In: Proceedings of IMC'03, Florida (2003)
13. Spring, N., Mahajanand, R., Wetherall, D.: Measuring ISP topologies with Rocketfuel. In: Proceedings of the ACM SIGCOMM. (2002)
14. AMP web site: http://watt.nlanr.net/. (Accessed January 2005)
15. Zhang, Y., Duffield, N., Paxson, V., Shenker, S.: On the constancy of internet path properties. In: Proceedings of ACM SIGCOMM Internet Measurement Workshop, San Francisco (2001)