Towards Deterministic Network Diagnosis

Yao Zhao, Yan Chen
EECS Department
Northwestern University
{yzhao, ychen}@cs.northwestern.edu

David Bindel
Division of Computer Science
University of California at Berkeley
dbindel@eecs.berkeley.edu

Categories & Subject Descriptors

C.2.3 [Computer-Communication Networks]:
Network Operations - Network monitoring

General Terms

Measurement, Experimentation

Keywords

Internet diagnosis, measurement, loss rate

1. MOTIVATION

> "When something breaks in the Internet, the Internet's very decentralized structure makes it hard to figure out what went wrong and even harder to assign responsibility."
>
> – "Looking Over the Fence at Networks: A Neighbor's View of Networking Research", by Committees on Research Horizons in Networking, National Research Council, 2001.

Internet fault diagnosis is important to end users, overlay network service providers (like Akamai), and Internet service providers (ISPs). For example, with Internet fault diagnosis tools, users can choose more reliable ISPs. However, The modern Internet is heterogeneous and largely unregulated, which renders the Internet diagnosis an increasingly challenging problem.

Though several router-based Internet diagnosis tools have been proposed [1, 2], these tools generally need special support from routers. For example, Tulip [1] requires the routers to support continuous IP-ID for generated ICMP packets. Also these ICMP-based tools are subject to ICMP rate limiting and are sensitive to cross-traffic. In contrast, many recently-developed tools for *Internet Tomography* use signal processing and statistical approaches to infer link level properties [3, 4, 5, 6] or shared congestion [7] based on end-to-end measurements of IP routing paths.

The advantage of statistical inference is their fine diagnosis granularity, *i.e.*, they infer the property of each virtual link (a consecutive subpath of an IP path with no branches [5]). However, the accuracy of the predicted link properties is subject to uncertainty in the model assumptions. Since true multicast does not exist in the Internet, many mulitcast-based tomography approaches [3, 4] have to use unicast for approximation, such as [8] and [9]. Thus the accuracy of the probe measurements heavily depends on the cross traffic in the network, and there is no guarantee of their accuracy. Recently, the statistically-based tools introduced in [5] and [6] use uncorrelated end-to-end measurements to identify lossy network links. However, these approaches can roughly identify the most likely lossy virtual links with certain false positive. Thus it remains an open problem to find which links or sequences of links can be *uniquely* characterized from end-to-end measurements, for which we will tackle in this paper.

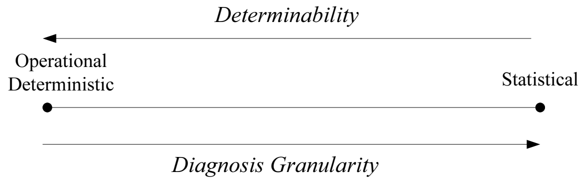


Figure 1: The spectrum of network diagnosis methods.

2. OPERATIONAL DETERMINISTIC DIAGNOSIS

In this paper, we introduce a new framework for network diagnosis: deterministic diagnosis. Further, when combined with statistical inference, this framework gives a full spectrum of network diagnosis methods with smooth tradeoff between determinability and diagnosis granularity. Here, we define the *diagnosis granularity* as the length of the smallest consecutive link sequences whose properties are inferred.

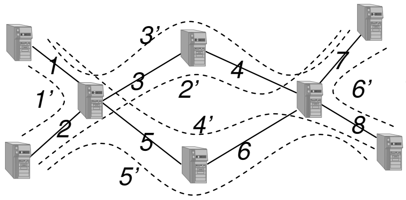
Mathematically speaking, the determinability means we can identify link-level properties like loss rates with 100% accuracy and without any statistical assumptions. However, any network measurement itself involves a random and stochastic process. Furthermore, the linear system model described before is a fundamental connection between path- and link- level properties, but assumes independence between link-level performance. Thus, it is virtually impossible to achieve the strict *mathematical determinability* for network measurement, letting alone network tomography.

On the other hand, the network end-to-end measurements, when well designed, can give accurate results on path properties. Thus if we do not introduce *any* extra statistical assumptions but can still identify the link-level properties, they will be highly accurate. We define such capability of practical estimation without any extra statistical assumptions as *operational determinability*, which is the main focus of this paper.

Compared with existing work with statistical assumptions, there is tradeoff between operational determinability and diagnosis granularity as shown in Figure 1. For example, with more and more statistical assumptions, we can reduce the diagnosis granularity while sacrificing the determinability. Basically, these statistical assumptions are often biased, and thus introduce extra inaccuracies for inferring link-level properties.

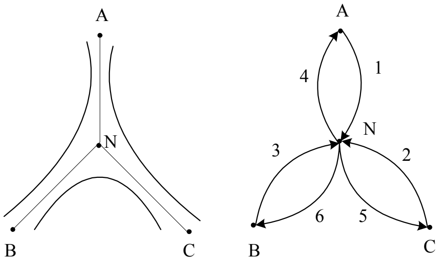
Thus our goal is to achieve as fine diagnosis granularity as possible without sacrificing determinability. In addition, our Operational Deterministic Diagnosis (ODD) system desires the following prop-

SIGMetrics/Performance'06, June 26–30, 2006, Saint Malo, France.
ACM 1-59593-320-4/06/0006.



$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Figure 2: Sample network with 8 links and 6 paths (*e.g.*, $1'$).



$$G = \left[\begin{array}{ccc|ccc} \multicolumn{3}{c}{\text{Incoming Links}} & \multicolumn{3}{c}{\text{Outgoing Links}} \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right]$$

$\text{Rank}(G) = 5$

Figure 3: Undirected graph vs Directed graph.

erties:

- *Scalability*: Both the measurement and the inference computation should impose low overhead even for large networks.
- *No special router support needed.*

To solve this problem, we define a *minimal identifiable link sequence* (MILS) as a link sequence of minimal length whose properties can be uniquely identified from end-to-end measurements. A MILS can be a virtual link, which is usually the finest diagnosis granularity of Internet tomography, or a sequence of virtual links. Note that the "virtual link" defined in [10] is not a physical link or a subpath (different definition to this paper and other tomography papers), which cannot be used to do link level diagnosis. Then we apply an algebraic approach to separate the identifiable and unidentifiable components of each path to find the MILSes. For networks modeled as undirected graphs, we can use network topology to find out the MILSes which are uniquely defined by the inherent path sharing of the Internet; and we propose efficient algorithms to find all such MILSes and infer their loss rates. Fig. 2 shows an example how we use the linear algebraic model to find identifiable MILSes. Here G denotes the path matrix, in which $G_{ij} = 1$ means the jth link is in the ith path. All identifiable MILSes belong to the row space of G, i.e., a subspace containing all the vectors can be written as linear combination of the row vectors of G. For example, link 1 is a MILS, of which the loss rate can be inferred through the linear combination of path success rates as $(b_1 + b_3 - b_2)$.

However, the real Internet has asymmetric link properties (*e.g.*, loss rate), and so must be modeled as a directed graph. But to find the MILSes in a directed graph is significantly more challenging. *We found and proved that for a directed graph, each path itself has to be a MILS, and no path segment can have its property operationally determined given only the topology information.* For example, Fig. 3 shows a simple star topology in both undirected graph and directed graph. It is easy to verify that although the directed graph contains all the 6 end-to-end paths, none of single link in the graph is identifiable.

In this work, we make the following contributions.

- We propose the operationally deterministic diagnosis framework and introduce the concept of MILS.
- Taking a network as a directed graph, when only topology information is used, we prove that each path is a MILS: no path segment smaller than an end-to-end path has properties which can be uniquely determined by end-to-end measurements.
- To address the problem above, we observe that in practice, there are many good paths with almost zero loss rates. Then as a fact rather than a statistical assumption, we know all the links on such paths must also have almost no losses. Based on such observation, we propose a "good path" algorithm, which uses both topology and measurement snapshots to find MILSes with the finest granularity.
- We show that our approach complements other tomography techniques – it helps significantly reduce their complexity and improves their inference accuracy.
- To validate our estimates, we propose a novel method of link-level loss rate inference based on IP spoofing which enables a limited form of source routing.

3. EVALUATION

We evaluate the ODD system through extensive simulations, and further validate our results through Internet experiments. Both give promising results. We define the diagnosis granularity of a path as the average of the lengths of all the lossy MILSes contained in the path. For the experiments with 135 PlanetLab hosts (each from a different organization), the average diagnosis granularity is only four hops for all the lossy paths. This can be further improved with larger overlay networks, as shown through our simulation with a real router-level topology from [11]. This suggests we can do very fine-level deterministic diagnosis with reasonably large overlay networks.

In addition, the loss rate inference on the MILSes is highly accurate, as verified through the cross-validation and IP spoof-based validation schemes. This accuracy is due to the inherent determinism in our design. The ODD system is also highly efficient. For the PlanetLab experiments with 135 hosts, the average setup (monitoring path selection) time is 109.3 seconds, and the online diagnosis of 18,090 paths, 3,714 of which are lossy, takes only 4.2 seconds on a 2.8GHz P4 machine. For more details, please refer to `http://list.cs.northwestern.edu/measurement.html`.

REFERENCES

[1] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson, "User-level internet path diagnosis," in *ACM SOSP*, 2003.

[2] K. Anagnostakis, M. Greenwald, and R. Ryger, "cing: Measuring network-internal delays using only existing infrastructure," in *IEEE INFOCOM*, 2003.

[3] A. Adams et al., "The use of end-to-end multicast measurements for characterizing internal network behavior," in *IEEE Communications*, May, 2000.

[4] T. Bu, N. Duffield, F. Presti, and D. Towsley, "Network tomography on general topologies," in *ACM SIGMETRICS*, 2002.

[5] V. Padmanabhan, L. Qiu, and H. Wang, "Server-based inference of Internet link lossiness," in *IEEE INFOCOM*, 2003.

[6] N. Duffield, "Simple network performance tomography," in *ACM SIGCOMM Internet Measurement Conference (IMC)*, 2003.

[7] D. Rubenstein, J. F. Kurose, and D. F. Towsley, "Detecting shared congestion of flows via end-to-end measurement," *ACM Transactions on Networking*, vol. 10, no. 3, 2002.

[8] M.J. Coates and R. Nowak, "Network loss inference using unicast end-to-end measurement," in *ITC Conf. IP Traffic, Modeling and Management*, 2000.

[9] N.G. Duffield, F.L. Presti, V. Paxson, and D. Towsley, "Inferring link loss using striped unicast probes," in *IEEE INFOCOM*, 2001.

[10] Y. Chen, D. Bindel, H. Song, and R. H. Katz, "An algebraic approach to practical and scalable overlay network monitoring," in *ACM SIGCOMM*, 2004.

[11] R. Govindan and H. Tangmunarunkit, "Heuristics for Internet map discovery," in *IEEE INFOCOM*, 2000.