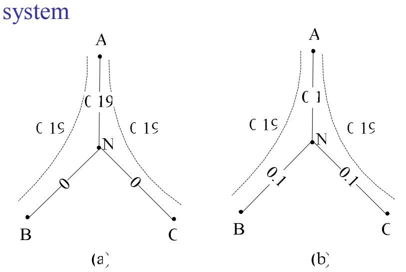
Scalable Deterministic Overlay Network Diagnosis

Yao Zhao, Yan Chen, David Bindel
{yzhao,ychen}@cs.northwestern.edu, dbindel@eecs.berkeley.edu
http://list.cs.northwestern.edu/

NORTHWESTERN UNIVERSITY

1. Motivation

- Internet measurement and diagnosis are important
- Hard to get these vital measures directly
- Solution: Do it by end user
- Difficulty: Internet is an underconstrained system



2. Previous work insufficient

- Router based approaches [SOSP03]
  - Mostly ICMP based, ICMP rate limiting
  - Unscalable for simultaneous diagnosis
  - Cannot deterministically separate forward/backward path loss
- Statistical approaches [MINC, INFOCOM03]
  - Inference based on temporal correlation in a multicast tree
  - Have to compromise for unicast, then sensitive to cross traffic
  - Optimization based on assumptions: # of lossy links are small
  - Unscalable: iterative refinement slow to converge for large networks

3. Problem formulation

- Definition of Determinability
  - Identify the properties of links or link sequences with 100% accuracy when there is no measurement noise.
- Problem
  - Given an overlay of N end hosts and $O(N^2)$ paths, to what granularity can we deterministically diagnosis the network fault?

4. Our solution

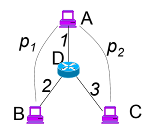
- Minimal identifiable link sequence (MILS)
  - Identifiable
  - Consecutive
  - Indecomposable
- Linear algebraic approach to achieve
  - Determinability
  - Scalability
  - Fine-grained diagnosis
  - No router support needed

5. Linear algebraic model

Path loss rate p, link loss rate l: $1-p_1=(1-l_1)(1-l_2)$

$$\log(1-p_1)=\log(1-l_1)+\log(1-l_2)=\begin{bmatrix}1 & 1 & 0\end{bmatrix}\begin{bmatrix}\log(1-l_1)\\ \log(1-l_2)\\ \log(1-l_3)\end{bmatrix}$$

$$\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}=\begin{bmatrix}b_1\\ b_2\end{bmatrix}$$

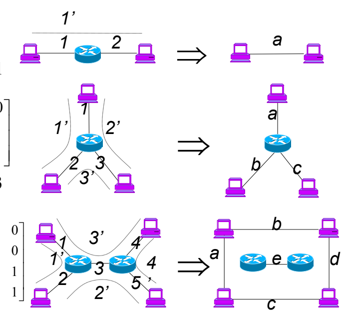


6. Examples of MILSes

$G=\begin{bmatrix}1 & 1\end{bmatrix}$
Rank(G)=1

$G=\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{bmatrix}$
Rank(G)=3

$G=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 1 & 0 & 1 & 1 & 0\\ 1 & 1 & 1 & 0 & 1\\ 0 & 0 & 0 & 1 & 1\end{bmatrix}$
Rank(G)=4

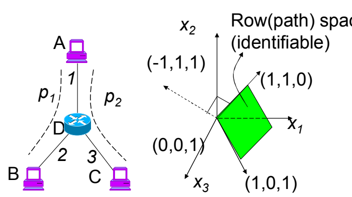


Real links (solid) and all of the overlay paths (dotted) traversing them

MILS

7. MILS in undirected graph

- An identifiable vector is in the row space of G
- Exhausted check if a link sequence a MILS
  - $O(L^2)$ potential MILSes in a path of length L

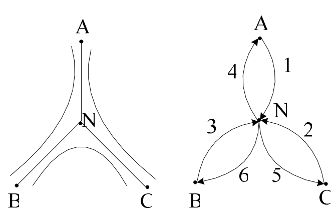


$$G=\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}$$

$$Q=\begin{bmatrix}\frac{\sqrt{2}}{2} & \sqrt{\frac{1}{6}}\\ \frac{\sqrt{2}}{2} & -\sqrt{\frac{1}{6}}\\ 0 & \sqrt{\frac{2}{3}}\end{bmatrix}$$

8. Another story for directed graph

- A MILS can't start from or end at any router
- No MILS shorter than a whole path exist
- For any interior routing node:
  - Linear combination => $\sum incoming = \sum outgoing$



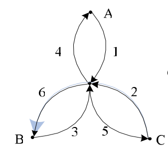
Incoming Links | Outgoing Links

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

Rank(G) = 5

9. Good path algorithm

- Undirected graph
  - Topology only
- Directed graph
  - E2E loss rate
  - Link property constrains
- Internet feature: many good paths
- Assumption:
  - All the links in a good path are good links



$$G=\begin{array}{c}\begin{array}{cccccc}1 & 2 & 3 & 4 & 5 & 6\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 & 1 & 0\end{bmatrix}\end{array} \Rightarrow G'=\begin{array}{c}\begin{array}{cccc}1 & 3 & 4 & 5\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 1 & 0\\ 0 & 1 & 0 & 1\end{bmatrix}\end{array}$$

$$\overline{G}=\begin{array}{c}\begin{array}{cccccc}1 & 2 & 3 & 4 & 5 & 6\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 1 & 0 & 0\end{bmatrix}\end{array} \Rightarrow G''=\begin{array}{c}\begin{array}{cccc}1 & 3 & 4 & 5\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 1 & 0\end{bmatrix}\end{array} \Rightarrow \overline{G'}=\begin{array}{c}\begin{array}{cccc}1 & 3 & 4 & 5\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 1 & 0\end{bmatrix}\end{array}$$

10. Internet experiments

- Planetlab
  - 135 end hosts
- Topology measured by Traceroute
  - Avg path length is 14.7
- Path loss rate by active UDP probing
  - 300 40-byte UDP packets per measured path in 90 sec
- Validation
  - Cross validation
  - IP spoofing based consistency check
- Experiment result

End-to-end Path	18,090
Avg Path Length	14.7
Avg bad path length after good path algorithm applied	9.0 (11.5)
# of MILSes	1009
Avg length of MILSes	3.0(4.3)
Avg diagnosis Granularity	2.7(4.0)