Active Network Tomography: Design and Estimation Issues

George Michailidis
Department of Statistics
The University of Michigan

November 18, 2003

Collaborators on Network Tomography Problems:
Vijay Nair, Bowei Xi, Earl Lawrence
Outline

● Background
● Active Network Tomography Problem
● Probing Experiments and Identifiability Issues
● Inference
● Performance Issues
● Concluding Remarks
1. Large-Scale Network Monitoring

**Broad objectives:**
assess network performance and plan its capacity, estimate QoS parameters, detect anomalous behavior, and design efficient networks/routing.

Specifically:
- Estimate network parameters such as link loss rates and delay distributions
- Determine/characterize spatial/temporal variation
- Identify bottleneck nodes
- Detect connectivity failures and routing faults
- Detect malicious/unusual behavior (e.g. DDoS)
Large-Scale Network Monitoring
Challenging Problem

Because

Difficult to quantitatively assess network performance – need data and new models for characterizing complex network behavior, there are data measurement and technology limitations, would like to have scalable, distributed algorithms, ...
2. **Active Network Tomography**

**Objective:**

Estimate network performance characteristics based on a limited set of end-to-end measurements.

**Mechanism:** Test probes are sent from a source node to a destination node and their loss and delay characteristics are observed and recorded at the destination nodes.

**Problem:** Infer individual link characteristics from the end-to-end measurements.
Two types of measurement schemes:

- **Passive monitoring**
  Main drawback: requires cooperation from network nodes

- **Active probing**
  Drawbacks:
  - multicast transmission mechanism not widely deployed
  - unicast probing based methods require some infrastructure plus additional assumptions for identifiability purposes
Active Network Tomography: the Main Idea
An Open Issue:

How well do active probing measurements capture the QoS experienced by users?
Active Network Tomography
Probing Schemes

- **Unicast**: send a probe packet to one of the receiver nodes at a time
- **Multicast**: send copies of each probe packet to all the receiver nodes *simultaneously*
The Unicast Mechanism

Consider possible outcomes for packet losses:

Receiver 2: 1 (successful transmission) or 0 (loss)
Receiver 3: 1 (successful transmission) or 0 (loss)
The Multicast Mechanism

possible outcomes

1 1
0 1 → loss on link 2
1 0 → loss on link 3
0 0
Unicast vs Multicast Experiments

- Unicast probes fail to identify the parameters of interest (e.g. loss rates).

  **Solution:** Use closely spaced in time probe packets, thus mimicking the multicast mechanism.

- Multicast probes lead to an explosive growth in the number of potential outcomes with the size of the network.

  **Solution:** Use several multicast packets destined for small subsets of the receivers to cover all the receiver nodes.
3. **Modeling Framework for Active Network Tomography**

1. Topology
2. Performance measures of interest
   - (loss rates and discrete delay distributions)
3. Stochastic assumptions
4. Type of Experiments
3.1. **Examples of Network Topologies**
Focus on tree topologies

Some notation:

Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote a tree with root node $\{0\}$ and receiver set $\mathcal{R}$.

Let $\mathcal{I}$ be the set of internal nodes and let $\mathcal{P}(s, u)$ denote the path between nodes $s$ and $u$.

Finally, $f(i)$ denotes the parent node of node $i$ and $\mathcal{D}(i)$ the set of children nodes.
Key assumption:
The logical topology of the probing experiment is known.
3.2. **Performance Measures of Interest**

- Link loss probabilities
- Delay distributions

The focus in the literature has been on discrete delay distributions; i.e., on each link delay falls in the set \( \{0, q, 2q, \ldots, Dq\} \) (\( D \) some positive integer). Hence, for a path containing \( k \) links, the end-to-end delay takes values in \( \{0, q, 2q, \ldots, kDq\} \).
End-to-end delay = \{0, 1, 2, ..., 4\}

End-to-end loss = \{0, 1\}
3.3. **Stochastic Model for Link Losses**

- Suppose a probe packet $t$ is sent from node 0 to node $i$. Let $Z_i(t) = 1$ if the probe reached node $i$ and 0 otherwise.

- Only $Z_r(t)$, $r \in \mathcal{R}$ can be observed.

- Let $\alpha_i(t) = P(Z_i(t) = 1 \mid Z_{f(i)}(t) = 1)$.

- Assume:

1. $\alpha_i(t) \equiv \alpha_i$.

2. $P(Z_j(t) = 1, \forall j \in \mathcal{D}(i) \mid Z_i(t) = 1) = \prod_{j \in \mathcal{D}(i)} \alpha_j$

   In other words, the children’s loss behaviors are conditionally independent given that of the parent.
This model has been used in several studies (see Caceres et al. (1999), Xi et al. (2003) and Michailidis et al. (2003))

Extensions of the model with spatial dependence can be found in Caceres et al. (1999), Dinwoodie et al. (2003))
The Statistical Problem

Simplest possible setting - binary 2-layer tree.
Under the posited stochastic model we obtain the following observed outcomes:

1. $(1, 1)$ w.p. $\gamma_{11} = \alpha_1 \alpha_2 \alpha_3$
2. $(1, 0)$ w.p. $\gamma_{10} = \alpha_1 \alpha_2 \bar{\alpha}_3$
3. $(0, 1)$ w.p. $\gamma_{01} = \alpha_1 \bar{\alpha}_2 \alpha_3$
4. $(0, 0)$ w.p. $\gamma_{00} = 1 - \cdots$

where $\vec{\gamma}$ can be estimated from the end-to-end measurements.

This is a multinomial experiment where $\vec{\gamma} = h(\vec{\alpha})$. 
3.4. **Stochastic Model for Link Delays**

Main assumption: individual link delays are independent and stationary.

This model has been used in Lo Presti et al. (2002), Liang et al. (2003), Tsang et al. (2003), Lawrence et al. (2003), ...
3.5. Probing Experiments

Let $\mathcal{C} = \bigcup_{j=1}^{M} C_j$ be a collection of $M$ multicast schemes.

Main Issues:

1. Identifiability (for what collections $\mathcal{C}$ is $\vec{\alpha}$ estimable)
2. Estimation and Inference
3. Design and Efficiency Issues
One multicast packet

to receivers 4, 5 and 6

One multicast packet to
receivers 3 and 7

\[ C = \{<4,5,6>, <3,7>\} \]
3.6. Identifiability

Question: Under what conditions can we estimate (recover) $\vec{\alpha}$ from $\{\vec{\gamma}_m^M\}_{M=1}^M$ obtained from the collection $C$ of $M$ multicast probing schemes?

Proposition: Consider a probing experiment $C$ that covers all the receiver nodes. The parameters of all the link parameters are estimable using $C$ if and only if: Every internal node in the tree topology corresponds to a splitting node for at least one of the multicast probes used in $C$. 
C=\{<4,5,6>, <3,7>\}

Condition satisfied by C

C=\{<4,5,6,7>, <3>\}

Condition NOT satisfied by C
Remark: The condition holds for both the loss and the discrete delay problem, under multicast and back-to-back unicast probing.

Remark: The multicast scheme (Caceres et al., 1999) automatically satisfies the condition. But unicast experiments do not.

Remark: The probing experiment with the smallest number of outcomes is the one comprised of $E$ bicast schemes ($E =$ number of internal nodes) and additional unicast schemes to cover the remaining receiver nodes (Xi et al. 2003).
3.7. Inference

Link loss problem


(II) For multicast scheme, a clever algorithm for approximate MLEs (Caceres et al. 1999). Requires solving a polynomial equation for obtaining the estimates. But variance computations still hard.

(III) Generalized linear model $\Rightarrow$ Approximate log-linear model formulation $\Rightarrow$ Fast least-squares algorithms for estimation and inference (Michailidis et al. (2003)).
Link delay problem

(I) Maximum likelihood estimates through EM for combination of multicast probing schemes (Lawrence et al. (2003)).

(II) Fast recursive algorithm (Lo Presti et al. (2002)).

(III) Pseudo-likelihood formulation, where the multicast information is processed for pairs of receivers (Liang et al. (2003)).
1. Simple EM algorithm for the loss problem for the special case:
Combination of bicasts and unicasts

E-step

(1) Get the updated path probabilities $\pi^{(k)}(0, s_b), \pi^{(k)}(s_b, i_b), \pi^{(k)}(s_b, j_b)$ and $\gamma^{b, (k)}_{00}$. 

(2) For each node $\ell \in \mathcal{P}(0, s_b) \cup \mathcal{P}(s_b, i_b) \cup \mathcal{P}(s_b, j_b)$, compute $V^{(k+1)}_{\ell, b} = E_{\alpha^{(k)}}[V_{\ell} N^b]$ as follows. 

For node $\ell \in \mathcal{P}(0, s_b)$, 

$$V^{(k+1)}_{\ell, b} = N^b - N^b_{00} \frac{1 - \alpha^{(k)}_{\ell}}{\gamma^{b, (k)}_{00}}.$$ 

For link $\ell \in \mathcal{P}(s_b, i_b)$, 

$$V^{(k+1)}_{\ell, b} = N^b - N^b_{01} \times \frac{1 - \alpha^{(k)}_{\ell}}{1 - \pi^{(k)}(s_b, i_b)} - N^b_{00} \frac{(1 - \alpha^{(k)}_{\ell})(1 - \pi^{(k)}(0, j_b))}{\gamma^{b, (k)}_{00}}.$$ 

For link $\ell \in \mathcal{P}(s_b, j_b)$, 

$$V^{(k+1)}_{\ell, b} = N^b - N^b_{01} \times \frac{1 - \alpha^{(k)}_{\ell}}{1 - \pi^{(k)}(s_b, j_b)} - N^b_{00} \frac{(1 - \alpha^{(k)}_{\ell})(1 - \pi^{(k)}(0, i_b))}{\gamma^{b, (k)}_{00}}.$$ 

(1a) For node $\ell \in \mathcal{P}(0, u)$ for unicast transmissions to receiver node $u$, compute 

$$V^{(k+1)}_{\ell, u} = N^u - N^u_0 \times \frac{1 - \alpha^{(k)}_{\ell}}{1 - \pi^{(k)}(0, u)}.$$ 

M-step The $(k+1)$-th update for the M-step is then:

$$\alpha^{(k+1)}_{\ell} = \frac{\sum_{b \in \mathcal{B}_\ell} V^{(k+1)}_{\ell, b} + \sum_{u \in \mathcal{U}_\ell} V^{(k+1)}_{\ell, u}}{\sum_{b \in \mathcal{B}_\ell} N^b + \sum_{u \in \mathcal{U}_\ell} N^u},$$

where $\mathcal{B}_\ell$ is the set of bicast pairs that includes the node $\ell$ in its path and $\mathcal{U}_\ell$ is the set of all unicast schemes that includes node $\ell$ in its path.
2. Algorithm in Caceres et al. (1999)

Consider the following logical tree topology.
• Step 1: From the end-to-end multicast measurements calculate

\[ \hat{\gamma}_k = P(\text{at least 1 receiver in the subtree of node } k \text{ got one copy of the probe}) \]

e.g. \( \hat{\gamma}_2 = N(1, +, +, +)/N \).

• Step 2: Calculate path probabilities \( \hat{A} \).

For receiver nodes, e.g. \( \alpha_1 \hat{\alpha}_2 = \hat{A}_2 = \hat{\gamma}_2 \), while for internal nodes solve a polynomial equation, e.g.

\[
(1 - \frac{\hat{\gamma}_3}{\hat{A}_3}) = (1 - \frac{\hat{\gamma}_4}{\hat{A}_3})(1 - \frac{\hat{\gamma}_5}{\hat{A}_3})(1 - \frac{\hat{\gamma}_6}{\hat{A}_3})
\]
• Step 3: Solve for $\hat{\alpha}$s.

e.g. $\hat{\alpha}_1 = \hat{A}_1$, $\hat{\alpha}_2 = \hat{A}_2 / \hat{A}_1$, ..., $\hat{\alpha}_4 = \hat{A}_4 / \hat{A}_3$, etc.
3. Approximate log-linear model

- Step 1: Calculate from the end-to-end measurements

\[ \hat{\gamma}_{11} = \frac{N(1, 1)}{N} \]
\[ \hat{\gamma}_{1+} = \frac{N(1, 1) + N(1, 0)}{N} \]
\[ \hat{\gamma}_{+1} = \frac{N(1, 1) + N(0, 1)}{N} \]
• Step 2: Notice that

\[ \hat{\gamma}_{11} = \alpha_1 \alpha_2 \alpha_3, \quad \hat{\gamma}_{1+} = \alpha_1 \alpha_2 \text{ and } \hat{\gamma}_{+1} = \alpha_1 \alpha_3. \]

• Step 3: Write the following model

\[
\log(\hat{\gamma}) = X \log(\bar{\alpha}) + \epsilon,
\]

where

\[
X = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1
\end{pmatrix}
\]

and where \( \epsilon \) has an explicitly given covariance matrix \( V(\hat{\gamma}) \).
For a collection of $M$ multicast schemes $\mathcal{C}$ we can similarly write

$$Y = X\beta + \epsilon,$$

where

$$Y = [\log(\hat{\gamma}^1) \cdots \log(\hat{\gamma}^M)]'$$

$\beta = \log(\bar{\alpha})$, $X$ an appropriately defined design matrix, and

$$\text{Cov}(\epsilon) = V$$

a block-diagonal matrix since the various multicast probing schemes are independent.
Remarks:

1. In this parameterization of the problem, every $k$–cast scheme (multicast transmission reaching $k$ receiver nodes) in $C$ contributes $2^k - 1$ rows to the vector $Y$. Hence, a multicast scheme with $R$ receivers leads to an exponential complexity regression model ($\mathcal{O}(2^R)$). On the other hand, a collection comprised of bicast and unicast schemes leads to at most a quadratic complexity model ($\mathcal{O}(R^2)$).

2. However, this is not the case with an alternative parameterization (through the likelihood function) that even for an $R$–cast scheme leads to quadratic complexity.

3. The advantage of this formulation is that the covariance matrix of $\alpha$ has a clean form, unlike that obtained from the likelihood formulation.
3.8. **Estimation methods**

Given the regression model one can use OLS, one-step GLS, or iteratively re-weighted least squares (IRWLS) procedures to estimate $\alpha$.

1. **OLS**: \[ \log(\vec{\alpha}) = (X'X)^{-1}X'\log(\vec{Y}) \]

2. **GLS**: \[ \log(\vec{\alpha}) = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\log(\vec{Y}) \], where $\hat{V}^{-1}$ is some plug-in estimate

3. **IRWLS**: Like GLS, but iterate on $V$ also until convergence
3.9. Some Properties:

- IRWLS and GLS are consistent, asymptotically normal and asymptotically equivalent to the MLEs (and hence efficient).
- In finite samples, GLS can have poor behavior but IRWLS is very close to the MLE.
- OLS is consistent and asymptotically normal.
- Variance of \( \hat{\alpha} \) increases as \( \alpha \) gets smaller (high loss rate).
- Variance behavior of \( \hat{\alpha} \)'s at different nodes is complex: depends on probing scheme, the level of the node, and true values of other \( \alpha \)'s.
Structure of Variance

3-layer symmetric binary tree topology
Bicast probing experiment
Asymptotic of Variances at nodes 1, 2, and 4.
Bias Comparisons for Multicast Experiments

OLS vs GLS vs IRWLS
3-layer symmetric binary tree topology
Probe size = 4000 total
Bias Comparisons for Collections of Bicast Probing Schemes

OLS vs GLS vs IRWLS
3-layer symmetric binary tree topology
Probe size = 4000 total
Some Plots for Delay Distributions using the EM algorithm

3-layer symmetric binary tree topology

Using a bicast probing scheme and the EM algorithm

Probe size = 20000 total

Using a geometric distribution for the underlying delays.
Using 50,000 multicast (4-cast) probes
Comparisons between the EM algorithm and the recursive algorithm of Lo Presti et al. (2002) on 3-layer binary symmetric tree with a 4-bin geometric distribution on each link for 100 replications.
3.10. **Design Issues**

1. Constructing the collection of probing schemes $\mathcal{C}$. Corresponds to a set covering problem. Can be solved by an integer program (see Michailidis et al. (2003) and Adler et al. (2001)).

2. Allocating number of probes to members of $\mathcal{C}$. Can use a D-optimality (minimize $\det(\text{Cov}(\bar{\alpha}))$) or A-optimality (minimize $\text{trace}(\text{Cov}(\bar{\alpha}))$) criterion for addressing this issue (see Xi et al. (2003)).
4. **The ns-package**

The network simulator (ns) is a discrete event simulator targeted at networking research. It provides substantial support for simulation of the TCP protocol, routing, and multicast protocols over wired and wireless (local and satellite) networks.

It allows the user to specify the network topology, the transport protocol, the queue management and packet scheduling mechanisms, the size of packets, and their inter-arrival times, etc.
**1st ns scenario**

1. Topology: 3-layer symmetric binary tree
2. All links have 1.5 Mb/sec bandwidth, 10 ms propagation delay and served by a FIFO queue of length 10
3. Background traffic: Constant Bit Rate (CBR); 1-2 flows/link; 500 byte packets with a uniform inter-packet distribution
4. Probing traffic: CBR; 40 byte packets; less than 5% of total traffic
2nd ns scenario

Same as 1st scenario but traffic consists of TCP and CBR mix.
3rd ns scenario

1. Topology: 4-layer symmetric binary tree
2. All links have 1.5 Mb/sec bandwidth, 10 ms propagation delay and served by a FIFO queue of length 4
3. Background traffic: TCP; 7-8 flows/link; 1000 byte packets
4. Probing traffic: CBR; 40 byte packets; less than 2% of total traffic
5. A First Look at the Change Point Detection Problem with Active Tomography

1. Setting: Some elements of $\vec{\alpha}$ exhibit a change (e.g. increase in the loss rate).

2. Can investigate such changes using ideas from statistical process control.

3. For this purpose variances come in handy.

4. Used Exponential Weighted Moving-Average Control Chart.

5. Scenarios investigated on a 4-layer symmetric binary topology: (i) sudden change, (ii) gradual change, (iii) many links on a path almost broken, (iv) minuscule change.

6. Used a collection of bicast schemes with 300 probes/scheme/time period.
Sudden change on some links.
Sudden change on some links.
A succession of links on a path almost broken.
A minuscule change on two successive links.
6. **Concluding Remarks and Future Research Directions**

- Provided overview of the active network tomography problem.
- Provided necessary and sufficient condition for identifiability of the parameters of interest.
- Discussed several estimation methods.
- Extensions include: Spatio-temporal models and Detection of malicious activities.
- **Validation of tomography models and methods.**
Some References


• Lawrence, E., Michailidis, G. and Nair, V.N. (2003), Inference of network delay distributions using the EM algorithm, (Technical Report)


• Dinwoodie, I.H. and Mosteig, E. (2003), Statistical Inference for Internal Network Reliability with Spatial Dependence, *SIAM Journal on Discrete Mathematics*

• Multicast based Inference of Network internal Characteristics (MINC), [http://www.research.att.com/minc](http://www.research.att.com/minc)