Network Anomography

Yin Zhang
yzhang@cs.utexas.edu

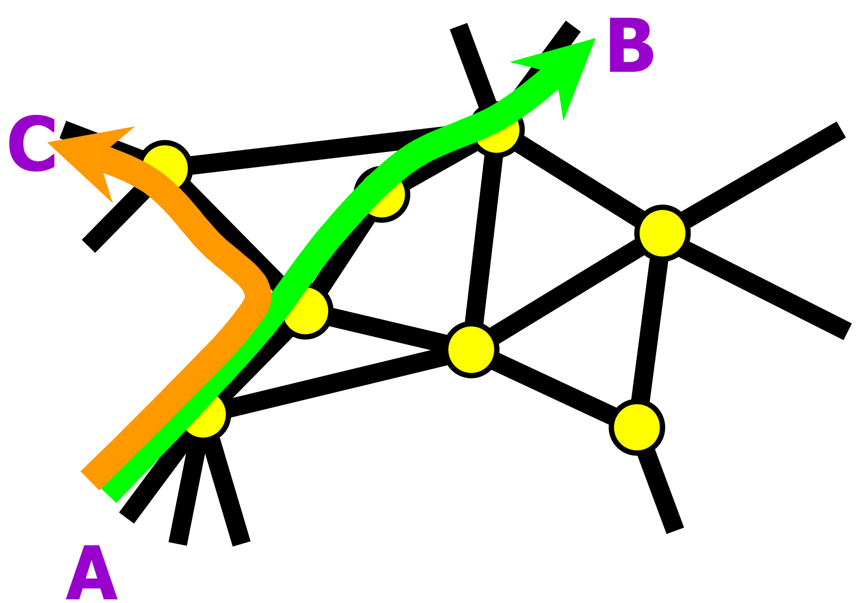
Joint work with
Zihui Ge, Albert Greenberg, Matthew Roughan

Internet Measurement Conference 2005
Berkeley, CA, USA

Network Anomaly Detection

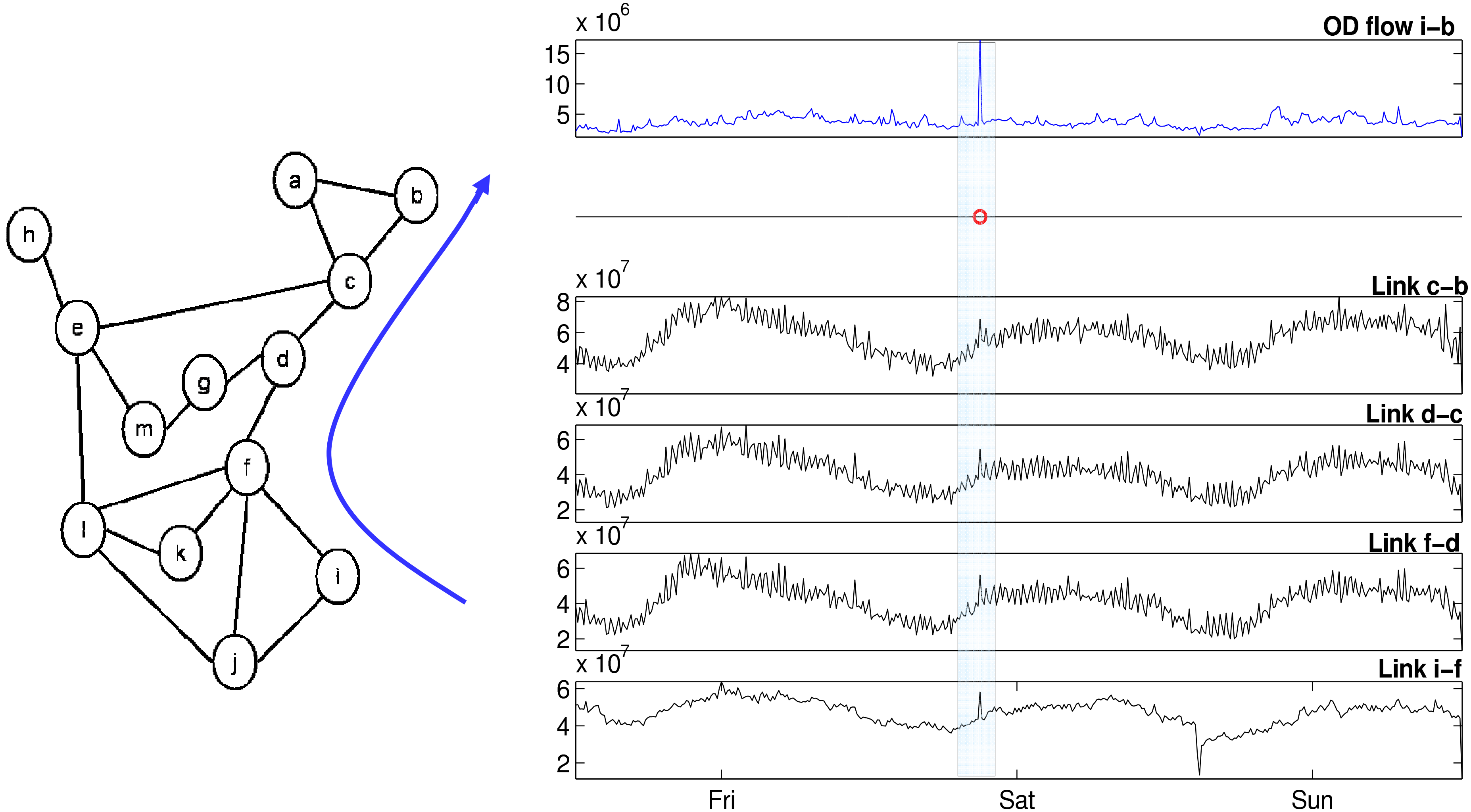
- Is the network experiencing unusual conditions?
  - Call these conditions anomalies
  - Anomalies can often indicate network problems
    - DDoS, worms, flash crowds, outages, misconfigurations …
  - Need rapid detection and diagnosis
    - Want to fix the problem quickly
- Questions of interest
  - Detection
    - Is there an unusual event?
  - Identification
    - What's the best explanation?
  - Quantification
    - How serious is the problem?

Network Anomography



- What we want
  - Volume anomalies [Lakhina04]
    Significant changes in an Origin-Destination flow, i.e., traffic matrix element
- What we have
  - Link traffic measurements
  - It is difficult to measure traffic matrix directly
- Network Anomography
  - Infer volume anomalies from link traffic measurements

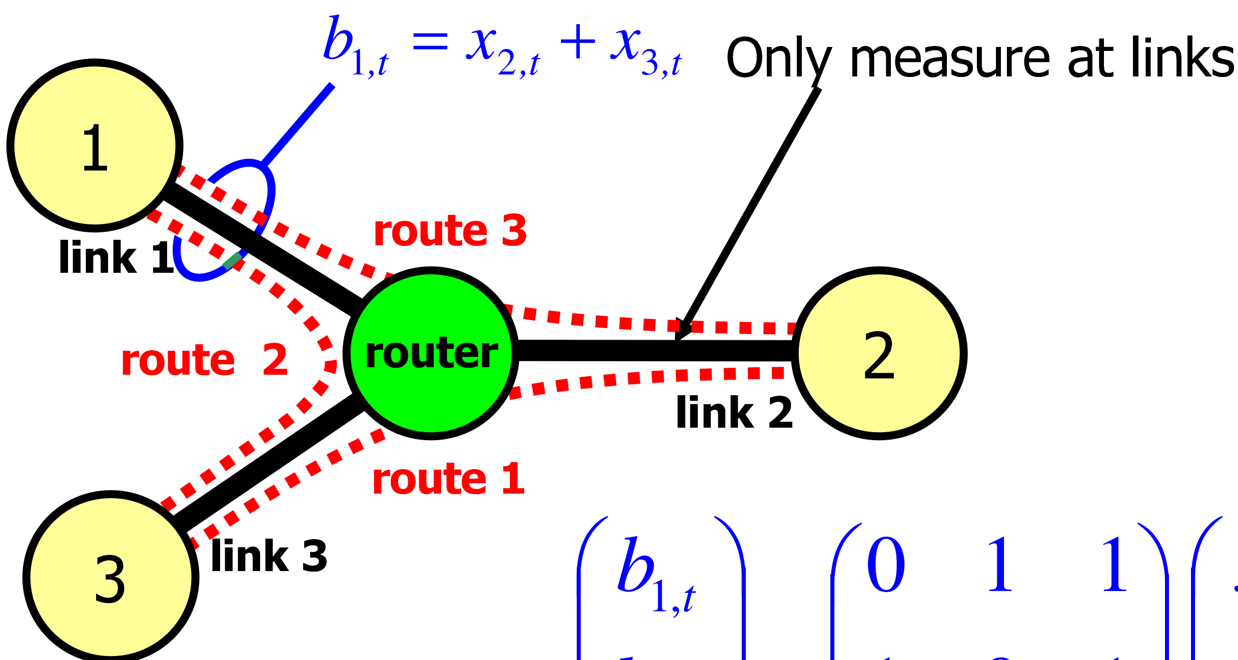
An Illustration



Courtesy: Anukool Lakhina [Lakhina04]

Anomography =
Anomalies + Tomography

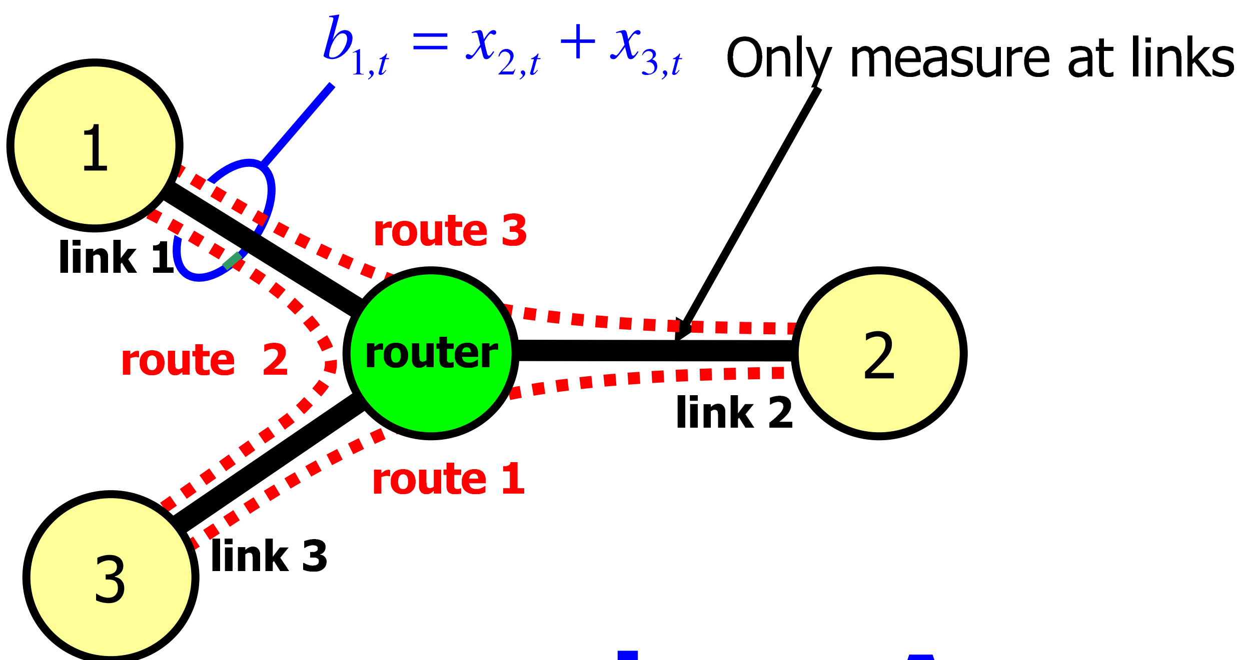
Mathematical Formulation



$$\begin{pmatrix} b_{1,t} \\ b_{2,t} \\ b_{3,t} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix}$$

Problem: Infer changes in TM elements (x_t) given link measurements (b_t)

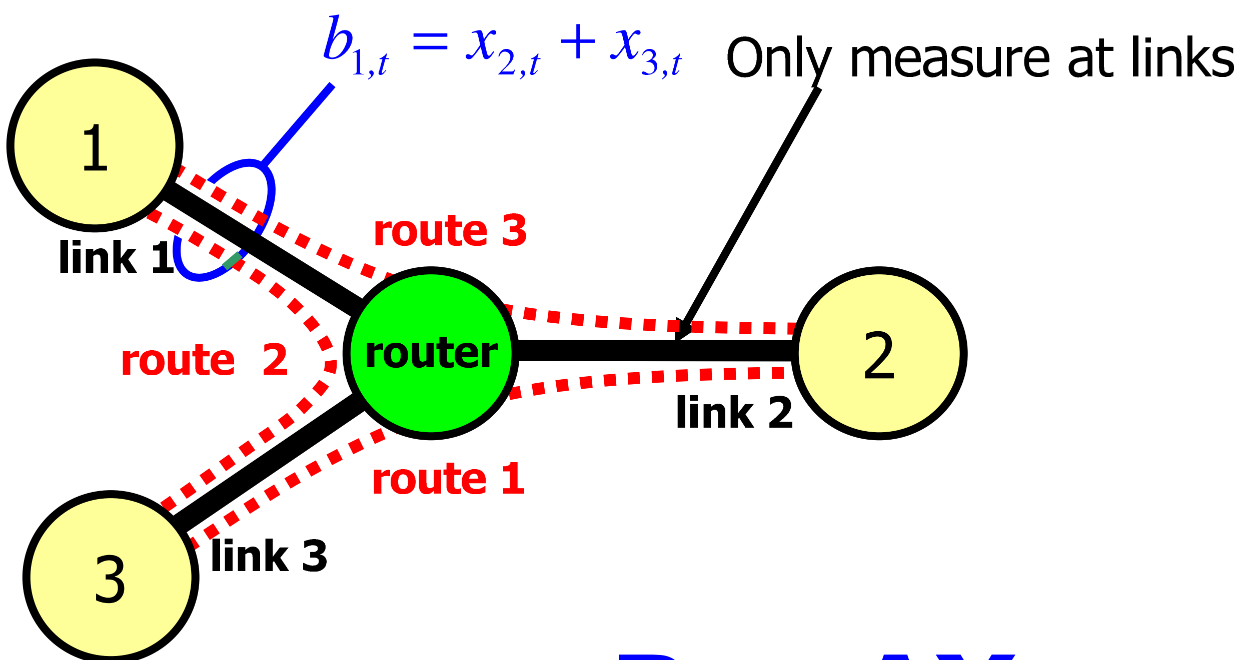
Mathematical Formulation



$$\mathbf{b_t = A_t\, x_t} \quad (t=1,\ldots,T)$$

Typically massively under-constrained!

Static Network Anomography



B = AX

Time-*invariant* A_t (= A), $B=[b_1 \ldots b_T]$, $X=[x_1 \ldots x_T]$

Anomography Strategies

- Early Inverse
  1. Inversion
     - Infer OD flows X by solving $b_t = Ax_t$
  2. Anomaly extraction
     - Extract volume anomalies $\tilde{X}$ from inferred X

  Drawback: errors in step 1 may contaminate step 2

- Late Inverse
  1. Anomaly extraction
     - Extract link traffic anomalies $\tilde{B}$ from B
  2. Inversion
     - Infer volume anomalies $\tilde{X}$ by solving $\tilde{b}_t = A\tilde{x}_t$

  Idea: defer "lossy" inference to the last step

Extracting Link Anomalies $\tilde{B}$

- Temporal Anomography: $\tilde{B} = BT$
  - ARIMA modeling
    - Diff: $f_t = b_{t-1}$ $\tilde{b}_t = b_t - f_t$
    - EWMA: $f_t = (1-\alpha)\, f_{t-1} + \alpha\, b_{t-1}$ $\tilde{b}_t = b_t - f_t$
  - Fourier / wavelet analysis
    - Link anomalies = the high frequency components
  - Temporal PCA
    - PCA = Principal Component Analysis
    - Project columns onto principal link column vectors

- Spatial Anomography: $\tilde{B} = TB$
  - Spatial PCA [Lakhina04]
    - Project rows onto principal link row vectors

Extracting Link Anomalies $\tilde{B}$

- Temporal Anomography: $\tilde{B} = BT$
  - Self-consistent
    - Tomography equation: $B = AX$
    - Post-multiply by T: $BT = AXT$

$$\tilde{B} = A\tilde{X}$$

- Spatial Anomography: $\tilde{B} = TB$
  - No longer self-consistent

Solving $\tilde{b}_t = A\,\tilde{x}_t$

- Pseudoinverse: $\tilde{x}_t = \mathrm{pinv}(A)\,\tilde{b}_t$
  - Shortest minimal L_2-norm solution
    - Minimize $|\tilde{x}_t|_2$ subject to $|\tilde{b}_t - A\,\tilde{x}_t|_2$ is minimal

- Maximize sparsity (i.e. minimize $|\tilde{x}_t|_0$)
  - L_0-norm is not convex $\Rightarrow$ hard to minimize
  - Greedy heuristic
    - Greedily add non-zero elements to $\tilde{x}_t$
    - Minimize $|\tilde{b}_t - A\,\tilde{x}_t|_2$ with given $|\tilde{x}_t|_0$
  - L_1-norm approximation
    - Minimize $|\tilde{x}_t|_1$ (can be solved via LP)
    - With noise $\Rightarrow$ minimize $|\tilde{x}_t|_1 + \lambda\,|\tilde{b}_t - A\tilde{x}_t|_1$

Dynamic Network Anomography

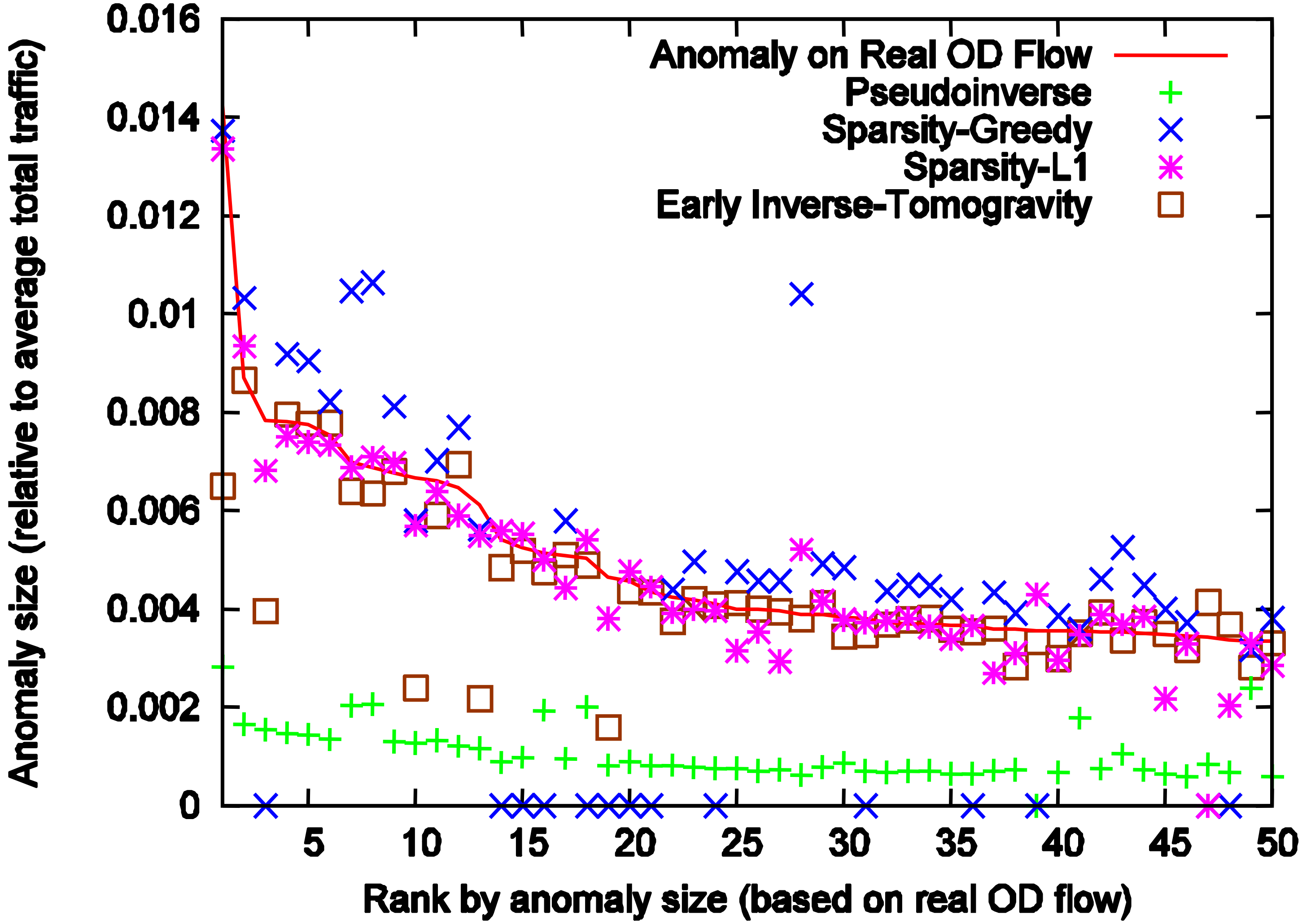
- Time-varying A_t is common
  - Routing changes
  - Missing data
    - Missing traffic measurement on a link $\Leftrightarrow$ setting the corresponding row of A_t to 0 in $b_t = A_t x_t$
- Solution
  - Early inverse: Directly applicable
  - Late inverse: Apply ARIMA modeling
    - L_1-norm minimization subject to link constraints
      - minimize $|\tilde{x}_t|_1$
        subject to $\tilde{x}_t = x_t - x_{t-1}$, $b_t = A_t x_t$, $b_{t-1} = A_{t-1} x_{t-1}$
    - Reduce problem size by eliminating redundancy

Performance Evaluation: Inversion

- Fix one anomaly extraction method
- Compare "real" and "inferred" anomalies
  - "real" anomalies: directly from OD flow data
  - "inferred" anomalies: from link data
- Order them by size
  - Compare the size
- How many of the top N do we find
  - Gives detection rate: $| \text{top } N_{\text{"real"}} \cap \text{top } N_{\text{inferred}} | / N$

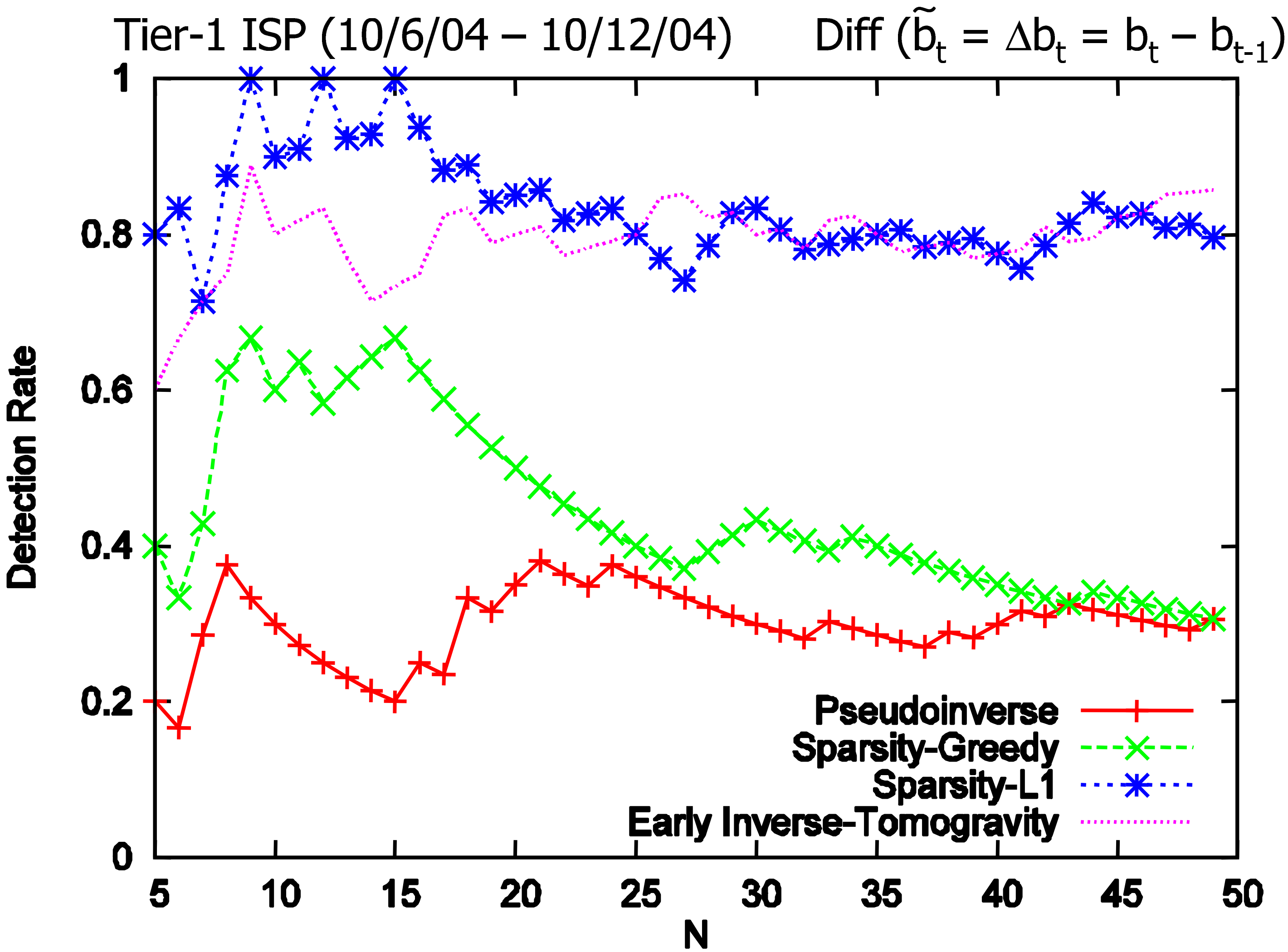
Inference Accuracy

Tier-1 ISP (10/6/04 – 10/12/04) Diff ($\tilde{b}_t = \Delta b_t = b_t - b_{t-1}$)



Sparsity-L1 works best among all inference techniques

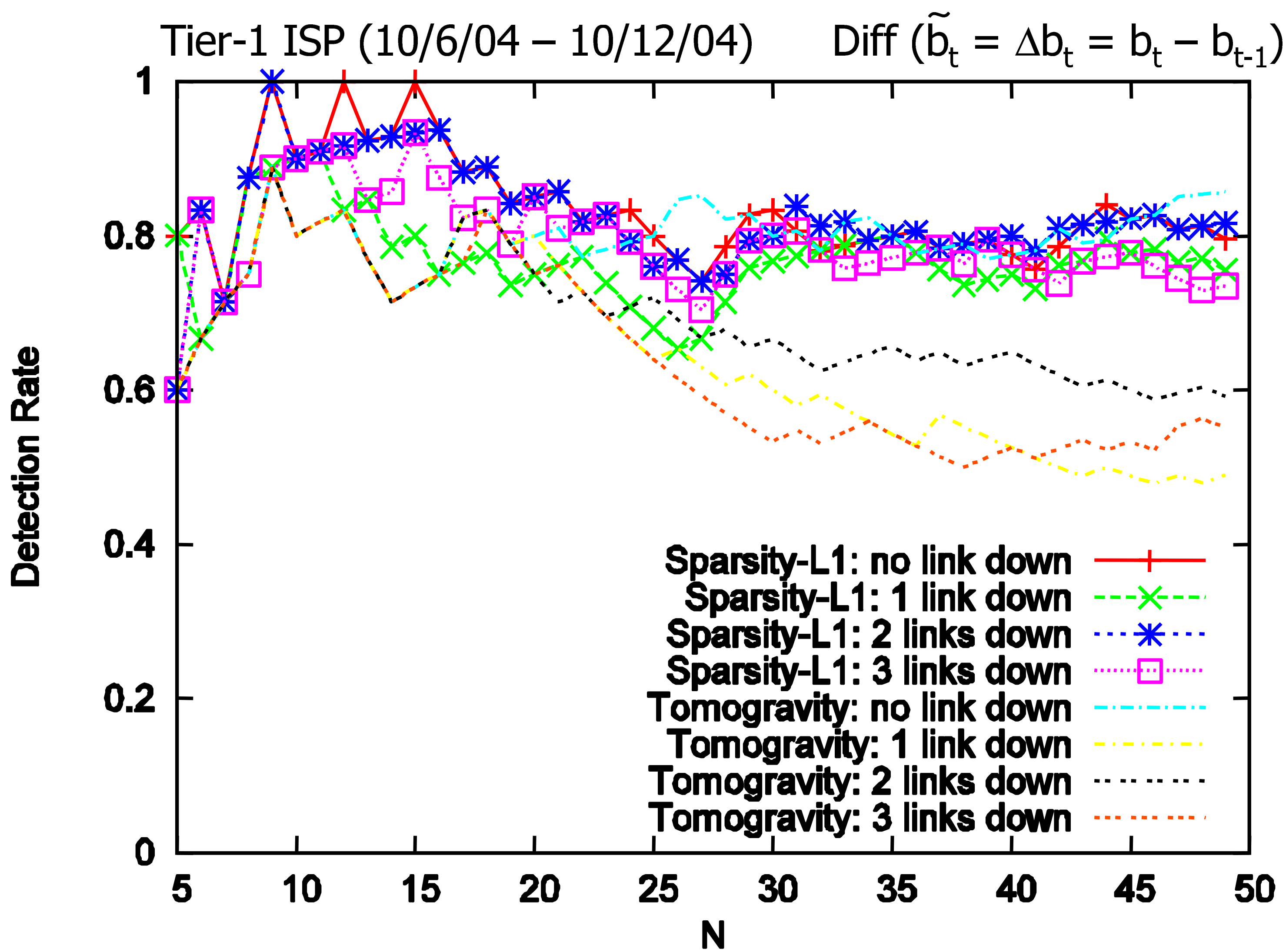
Inference Accuracy



detection rate = | top N_{real} $\cap$ top $N_{inferred}$ | / N

Sparsity-L1 works best among all inference techniques

Impact of Routing Changes



Late inverse (sparsity-L1) beats early inverse (tomogravity)

Performance Evaluation: Anomography

- Hard to compare performance
  - Lack ground-truth: what is an anomaly?
- So compare events from different methods
  - Compute top M "benchmark" anomalies
    - Apply an anomaly extraction method directly on OD flow data
  - Compute top N "inferred" anomalies
    - Apply another anomography method on link data
  - Report $\min(M,N) - |\text{top } M_{benchmark} \cap \text{top } N_{inferred}|$
    - $M < N \Rightarrow$ "false negatives"

      # big "benchmark" anomalies not considered big by anomography
    - $M > N \Rightarrow$ "false positives"

      # big "inferred" anomalies not considered big by benchmark method
  - Choose M, N similar to numbers of anomalies a provider is willing to investigate, e.g. 30-50 per week

Anomography: "False Negatives"

Top 50 Inferred	"False Negatives" with Top 30 Benchmark							
	Diff	EWMA	H-W	ARIMA	Fourier	Wavelet	T-PCA	S-PCA
Diff	**0**	0	1	1	5	5	**17**	**12**
EWMA	0	**0**	1	1	5	5	**17**	**12**
Holt-Winters	1	1	**0**	0	6	4	**18**	**12**
ARIMA	1	1	0	**0**	6	4	**18**	**12**
Fourier	3	4	8	8	**1**	7	**19**	**18**
Wavelet	0	1	2	2	5	**0**	**13**	**11**
T-PCA	**14**	**14**	**14**	**14**	**19**	**15**	**3**	**15**
S-PCA	**10**	**10**	**13**	**13**	**15**	**11**	1	**13**

1. Diff/EWMA/H.-W./ARIMA/Fourier/Wavelet all largely consistent
2. PCA methods not consistent (even with each other)
   - PCA cannot detect anomalies in the "normal" subspace
   - PCA insensitive to reordering of $[b_1 \ldots b_T] \Rightarrow$ cannot utilize all temporal info
3. Spatial methods (e.g. spatial PCA) are not self-consistent

Anomography: "False Positives"

Top 30 Inferred	"False Positives" with Top 50 Benchmark							
	Diff	EWMA	H-W	ARIMA	Fourier	Wavelet	T-PCA	S-PCA
Diff	**3**	3	6	6	6	4	**14**	**14**
EWMA	3	**3**	6	6	7	5	**13**	**15**
Holt-Winters	4	4	**1**	1	8	3	**13**	**10**
ARIMA	4	4	1	**1**	8	3	**13**	**10**
Fourier	6	6	7	6	**2**	6	**19**	**18**
Wavelet	6	6	6	6	8	**1**	**13**	**12**
T-PCA	**17**	**17**	**17**	**17**	**20**	**13**	**0**	**14**
S-PCA	**18**	**18**	**18**	**18**	**20**	**14**	1	**14**

1. Diff/EWMA/H.-W./ARIMA/Fourier/Wavelet all largely consistent
2. PCA methods not consistent (even with each other)
   - PCA cannot detect anomalies in the "normal" subspace
   - PCA insensitive to reordering of $[b_1 \ldots b_T] \Rightarrow$ cannot utilize all temporal info
3. Spatial methods (e.g. spatial PCA) are not self-consistent

Summary of Results

- Inversion methods
  - Sparsity-L1 beats Pseudoinverse and Sparsity-Greedy
  - Late-inverse beats early-inverse
- Anomography methods
  - Diff/EWMA/H-W/ARIMA/Fourier/Wavelet all largely consistent
  - PCA methods not consistent (even with each other)
    - PCA methods cannot detect anomalies in "normal" subspace
    - PCA methods cannot fully exploit temporal information in $\{x_t\}$
      - Reordering of $[b_1 \ldots b_T]$ doesn't change results!
  - Spatial methods (e.g. spatial PCA) are not self-consistent
    - Temporal methods are
- The method of choice: ARIMA + Sparsity-L1
    - Accurate, consistent with Fourier/Wavelet
    - Robust against measurement noise, insensitive to choice of λ
    - Works well in the presence of missing data, routing changes
    - Supports both online and offline analysis

Conclusions

- Anomography = Anomalies + Tomography
  - Find anomalies in $\{x_t\}$ given $b_t=A_tx_t$ $(t=1,\ldots,T)$
- Contributions
  1. A general framework for anomography methods
     - Decouple anomaly extraction and inference components
  2. A number of novel algorithms
     - Taking advantage of the range of choices for anomaly extraction and inference components
     - Choosing between spatial vs. temporal approaches
  3. The first algorithm for dynamic anomography
  4. Extensive evaluation on real traffic data
     - 6-month Abilene and 1-month Tier-1 ISP
- The method of choice: ARIMA + Sparsity-L1

Future Work

- Correlate traffic with other types of data
  - BGP routing events
  - Router CPU utilization
- Anomaly response
  - Maybe with an effective response system, false positives become less important?
- Anomography for performance diagnosis
  - Inference of link performance based on end-to-end measurements can be formulated as $b_t = Ax_t$
- Beyond networking
  - Detecting anomalies in other inverse problems
  - Are we just reinventing the wheel?

Thank you !