July 12, 2012

Recovery of Low-Rank Plus Compressed Sparse Matrices with Application to Unveiling Traffic Anomalies

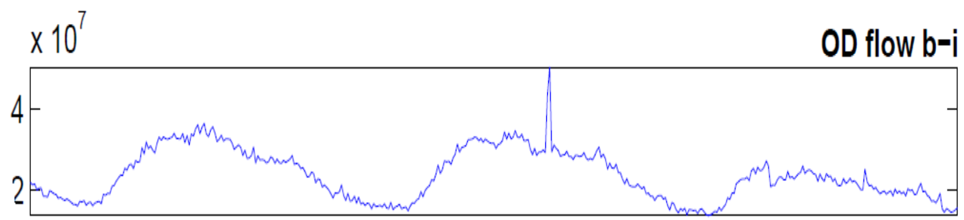
Morteza Mardani

Dept. of ECE, University of Minnesota, Minneapolis, MN 55455

Acknowledgments: Dr. Mateos and Prof. Giannakis
MURI Grant No. AFOSR FA9550-10-1-0567

Context

- Backbone of IP networks
- Traffic anomalies: changes in origin-destination (OD) flows
  - Failures, transient congestions, DoS attacks, intrusions, flooding



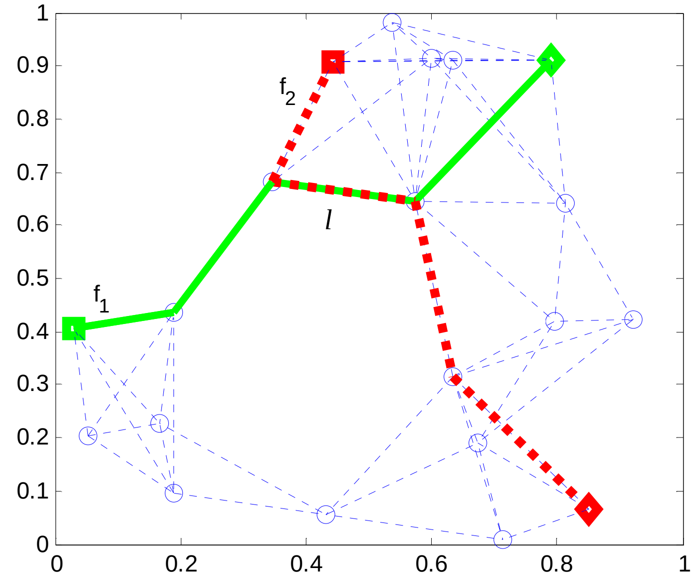
- Motivation: Anomalies→ congestion→ limits end-user QoS provisioning

Goal: Measuring superimposed OD flows per link, identify anomalies by leveraging sparsity of anomalies and low-rank of traffic.

Model

- Graph $\mathcal{G}(\mathcal{N},\mathcal{L})$ with N nodes, L links, and F flows ($F >> L$)

(as) Single-path per OD flow $x_{f,t}$



- Packet counts per link l and time slot t

Anomaly

$$\boxed{y_{l,t} = \sum_{f\in\mathcal{F}} \underbrace{r_{l,f}}_{\in\{0,1\}}(x_{f,t} + \overbrace{a_{f,t}}^{\text{Anomaly}}) + v_{l,t}}$$

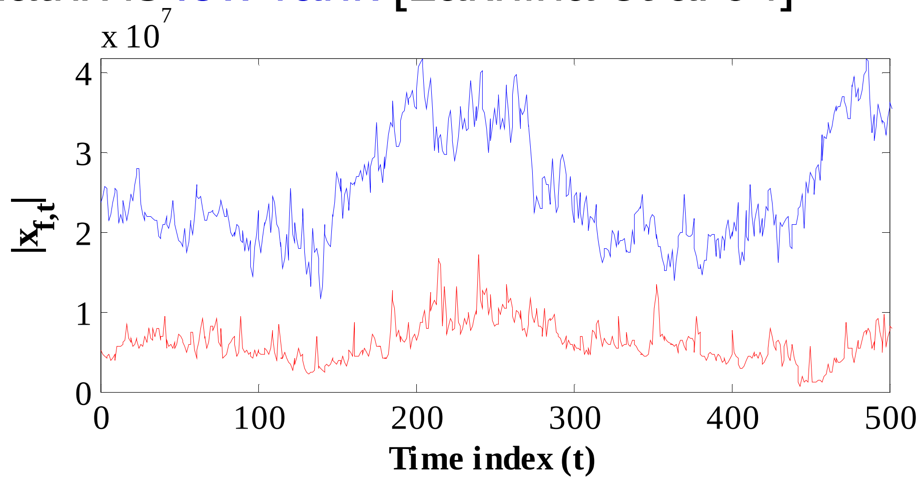
- Matrix model across T time slots

$$\boxed{\underbrace{\mathbf{Y}}_{LxT} = \underbrace{\mathbf{R}}_{LxF}(\mathbf{X}+\mathbf{A}) + \mathbf{V}}$$

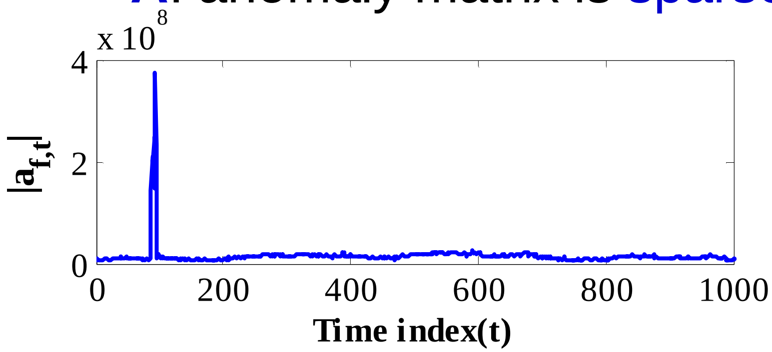
Low rank and sparsity

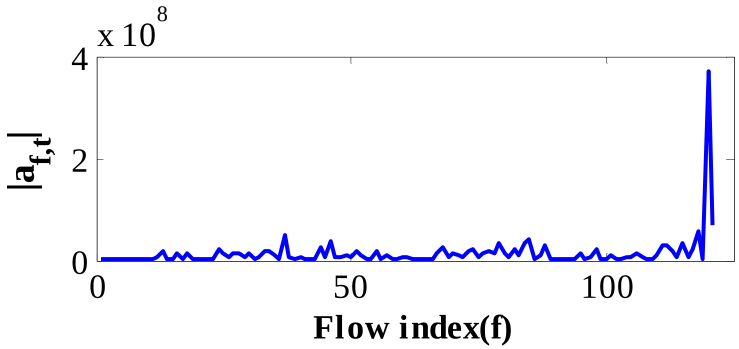
$$\mathbf{Y} = \mathbf{R}\left(\mathbf{X} + \mathbf{A}\right) + \mathbf{V}$$

- **X**: traffic matrix is low-rank [Lakhina et al'04]



- **A**: anomaly matrix is sparse across both time and flows





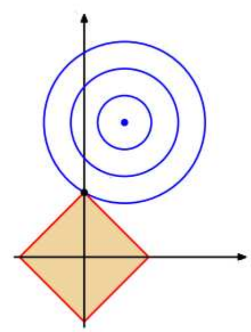
Objective and criterion

$$\mathbf{Y} = \underbrace{\mathbf{RX}}_{:=\mathbf{X}_R} + \mathbf{RA} + \mathbf{V}$$

- Given $\mathbf{Y}$ and routing matrix $\mathbf{R}$, identify sparse $\mathbf{A}$ when $\mathbf{X}$ is low rank
  - ***R*** fat but $\boldsymbol{X}_R$ still low rank
- Low-rank → sparse vector of SVs → nuclear norm $\|\ \|_*$ and l_1 norm

$$\boxed{(\hat{\mathbf{X}}_R, \hat{\mathbf{A}}) = \arg \min_{(\mathbf{X}_R, \mathbf{A})} \frac{1}{2}\|\mathbf{Y} - \mathbf{X}_R - \mathbf{RA}\|_F^2 + \lambda_* \|\mathbf{X}_R\|_* + \lambda_1 \|\mathbf{A}\|_1} \quad \text{(P1)}$$

$\sum_i \sigma_i(\mathbf{X}_R)$ $\qquad$ $\sum_{i,j} |a_{i,j}|$



Prior art

- Anomaly identification (ID)
  - Change detection on per link time series [Brutlag'00], [Casas et al'10]
  - Spatial principal component analysis (PCA) [Lakhina et al'04]
  - Network anomography [Zhang et al'05]

- Suboptimal ID of anomalies across flows and time

- Rank minimization with the nuclear norm [Recht-Fazel-Parrilo'10]
  - Matrix decomposition [Candes et al'10], [Chandrasekaran et al'11]

$$\mathbf{M} = \mathbf{L}_0 + \mathbf{S}_0$$

$$\boxed{\begin{array}{l} \min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{s. to} \quad \mathbf{M} = \mathbf{L} + \mathbf{S} \end{array}}$$

  - Principal Components Pursuit (PCP)

Challenges and importance

$$\mathbf{Y} = \mathbf{X}_R + \mathbf{R}\mathbf{A} + \mathbf{V}$$

- $\mathbf{R}\mathbf{A}$ not necessarily sparse and $\mathbf{R}$ fat → PCP not applicable

- $\underbrace{LT}_{X_R} + \underbrace{FT}_{A} \gg \underbrace{LT}_{Y}$ **Seriously underdetermined**

- Important special cases
  - ***R*** = ***I*** : matrix decomposition with PCP
  - ***X*** = ***0*** : compressive sampling with basis pursuit
  - ***R*** = ***I*** and ***A*** = ***0*** : PCA

Exact recovery

- Noise-free model: $\mathbf{Y} = \mathbf{X}_{R,0} + \mathbf{R}\mathbf{A}_0$

$$\boxed{\begin{aligned}(\hat{\mathbf{X}}_R, \hat{\mathbf{A}}) = \arg \min_{(\mathbf{X}_R, \mathbf{A})} \quad & \|\mathbf{X}_R\|_* + \lambda \|\mathbf{A}\|_\mathbf{1} \\ \text{s.to} \quad & \mathbf{Y} = \mathbf{X}_R + \mathbf{R}\mathbf{A}\end{aligned}}$$

Q: Can one recover sparse $\mathbf{A}_0$ and low-rank $\mathbf{X}_{R,0}$ exactly?

A: Yes**!** Under certain conditions on $\{\mathbf{A}_0, \mathbf{X}_{R,0}, \mathbf{R}\}$

- Both rank($\mathbf{X}_{R,0}$) and support of $\mathbf{A}_0$ generally unknown

Identifiability

$$\mathbf{Y} = \mathbf{X}_{R,0} + \mathbf{RA}_0 = \underbrace{\mathbf{X}_{R,0} + \mathbf{RH}}_{\mathbf{X'}_{R,0}} + \mathbf{R}\underbrace{(\mathbf{A}_0 - \mathbf{H})}_{\mathbf{A'}_0}$$

- Problematic cases
  - $\mathbf{RH} \neq \mathbf{0}$ but $\mathbf{X}'_{R,0}$ low-rank and $\mathbf{A}'_0$ sparse
  - $\mathbf{RH} = \mathbf{0}, \mathbf{H} \neq \mathbf{0}$

- For $\mathbf{X}_{R,0} = \mathbf{U\Sigma V}'$ and r = rank($\mathbf{X}_{R,0}$), low-rank-preserving matrices ***RH***

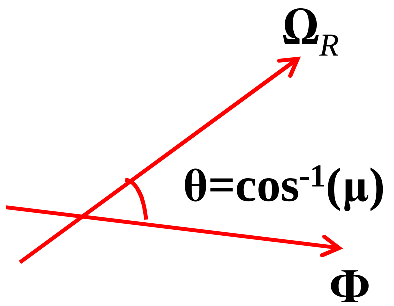
$$\Phi(\mathbf{X}_{R,0}) := \{\mathbf{Z} \in \mathbb{R}^{L\times T} : \mathbf{Z} = \mathbf{UW}_1' + \mathbf{W}_2\mathbf{V}', \; \mathbf{W}_1 \in \mathbb{R}^{T\times r}, \; \mathbf{W}_2 \in \mathbb{R}^{L\times r}\}$$

- Sparsity-preserving matrices ***RH***

$$\Omega_R(\mathbf{A}_0) := \{\mathbf{Z} \in \mathbb{R}^{L\times T} : \mathbf{Z} = \mathbf{RH}, \; \mathrm{supp}(\mathbf{H}) \subseteq \mathrm{supp}(\mathbf{A}_0)\}$$

Incoherence measures

$$\mu(\Omega_R, \Phi) := \max_{\mathbf{Z}\in\Omega_R\setminus\{\mathbf{0}\}} \frac{\|\mathcal{P}_\Phi(\mathbf{Z})\|_F}{\|\mathbf{Z}\|_F}$$



- Identifiability requires $\Omega_R(\mathbf{A}_0) \cap \Phi(\mathbf{X}_R) = \{\mathbf{0}\} \leftrightarrow \mu(\Omega_R, \Phi) < 1$, $\delta_k(\mathbf{R}) < 1$
- Incoherence among columns of ***R*** → $\delta_k(\mathbf{R})$, $\theta_{s_1,s_2}(\mathbf{R})$
- Exact recovery requires $\mu(\Omega_R, \Phi) \ll 1$, $\delta_k(\mathbf{R}) \ll 1$
- Incoherence between $\boldsymbol{X}_{R,0}$ and ***R***

$$\mu_R(\mathbf{U}) := \max_i \frac{\|\mathcal{P}_U(\mathbf{R}\mathbf{e}_i)\|}{\|\mathbf{R}\mathbf{e}_i\|}, \ \mu(\mathbf{V}) := \max_i \|\mathcal{P}_V(\mathbf{e}_i)\|_F$$

$$\mu_R(\mathbf{U}, \mathbf{V}) := \|\mathbf{R}'\mathbf{U}\mathbf{V}'\|_\infty$$

Main result

$$\mathbf{Y} = \mathbf{X}_{R,0} + \mathbf{R}\mathbf{A}_0 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}' + \mathbf{R}\mathbf{A}_0$$

$$r := \text{rank}(\mathbf{X}_{R,0}) \qquad s := \|\mathbf{A}_0\|_0$$

Theorem: Given $\mathbf{Y}$ and $\mathbf{R}$, if every row and column of $\mathbf{A}_0$ has at most *k* non-zero entries and $\mathbf{R}$ has full row rank, then

$$(1 - \mu(\Omega_R, \Phi))^2 (1 - \delta_k(\mathbf{R})) > \omega_{\max}$$

$$(1 + \alpha_{\max}) \left(\frac{1 + \beta_{\max}}{1 - \beta_{\max}} \right) \mu_R(\mathbf{U}, \mathbf{V}) \sqrt{s} + \mu(\Omega_R, \Phi)(1 + \delta_k(\mathbf{R}))^{1/2} (1 + \alpha_{\max}) \sqrt{r} < 1$$

imply $\exists \;\; \lambda > 0$ for which (P1) exactly recovers $\{\mathbf{X}_{R,0}, \mathbf{A}_0\}$

M. Mardani, G. Mateos, and G. B. Giannakis,``Unveiling network anomalies across flows and time via sparsity and low rank," *IEEE Trans. Information Theory*, 2012 (submitted).

Intuition

- Exact recovery if
  - r and s are sufficiently small
  - Nonzero entries of $\mathbf{A}_0$ are "sufficiently spread out"
  - Columns and rows of $\mathbf{X}_0$ not aligned with canonical basis
  - $\mathbf{R}$ behaves like a "restricted" isometry
- Interestingly
  - Amplitude of non-zero entries of $\mathbf{A}_0$ irrelevant
  - No randomness assumption
- Satisfiability for certain random ensembles w.h.p

Validating exact recovery

- Setup

  L=105, F=210, T = 420

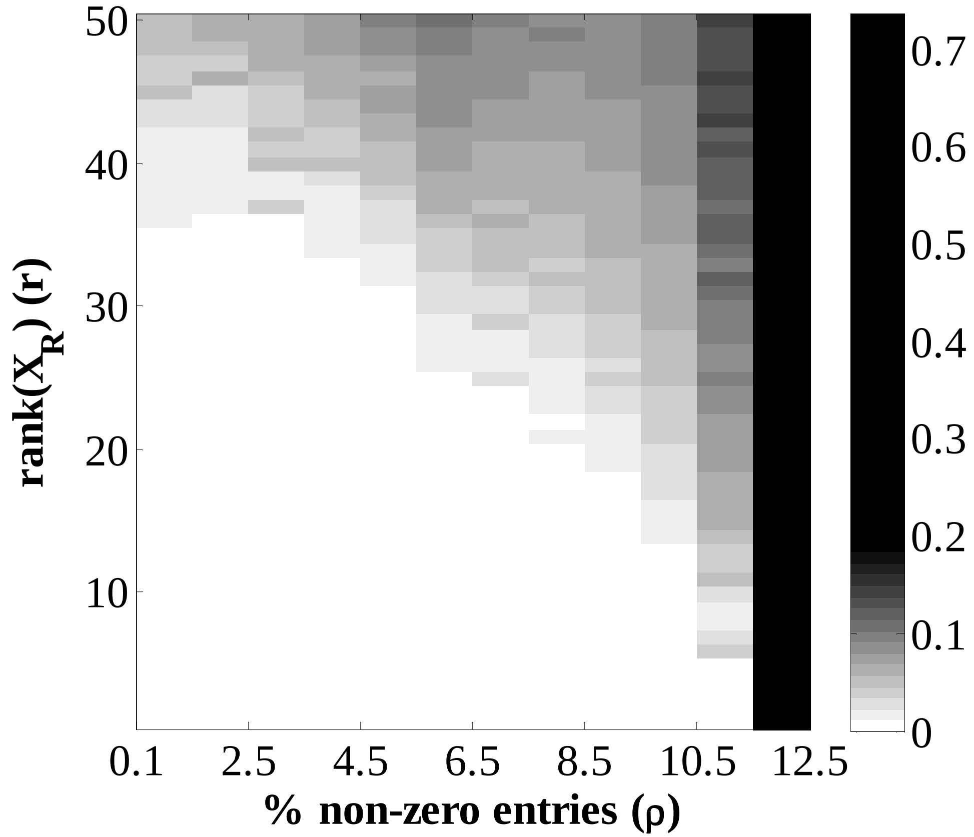
  R ~ Bernoulli(1/2)

  $\mathbf{X}_R = \mathbf{RWZ}'$, **W**, **Z** ~ N(0, $10^4/FT$)

  $a_{ij} \in \{-1,0,1\}$ w. prob. $\{\rho/2, 1-\rho, \rho/2\}$

- Relative recovery error

$$e = \frac{\|\hat{\mathbf{A}} - \mathbf{A}_0\|_F}{\|\mathbf{A}_0\|_F}$$



Centralized algorithm

$$\boxed{(\hat{\mathbf{X}}_R, \hat{\mathbf{A}}) = \arg\min_{(\mathbf{X}_R,\mathbf{A})} \frac{1}{2}\|\mathbf{Y} - \mathbf{X}_R - \mathbf{R}\mathbf{A}\|_F^2 + \lambda_*\|\mathbf{X}_R\|_* + \lambda_1\|\mathbf{A}\|_1}$$

- Accelerated proximal gradient method for $\mathbf{R} = \mathbf{I}_F$ [Lin et al'09]
- Idea: minimize a sequence of overestimates of (P1)

$$\mathbf{Z}[k] := \arg\min_{\{\mathbf{X}_R,\mathbf{A}\}} \left\{ \frac{L_f}{2}\|[\mathbf{X}_R\ \mathbf{A}]' - \underbrace{\mathbf{G}[k]}_{\mathbf{T}[k] - \frac{1}{L_f}\nabla f(\mathbf{T}(k))}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \lambda_*\|\mathbf{X}_R\|_* \right\}$$

Separable wrt X_R and A

- Intelligent choice of $\boldsymbol{T}[k]$ and $t[k]$

$$\mathbf{T}[k+1] = \mathbf{Z}[k] + \frac{t[k-1]-1}{t[k]}\left(\mathbf{Z}[k] - \mathbf{Z}[k-1]\right)$$

$$t[k] = \left[1 + \sqrt{4t^2[k-1]+1}\right]/2$$

- Converges fast with accuracy $O(1/k^2)$ [Nesterov'83]

Thresholded updates

- Traffic matrix update

$$\mathbf{X}_R[k+1] := \arg\min_{\mathbf{X}_R} \left\{ \frac{L_f}{2} \|\mathbf{X}_R - \mathbf{G}_X[k]\|_F^2 + \lambda_* \|\mathbf{X}_R\|_* \right\}$$

Sol. $\boxed{\mathbf{X}_R[k+1] = \mathbf{U}\mathcal{S}_{\underbrace{\lambda_*/L_f}}[\mathbf{\Sigma}]\mathbf{V}'}$

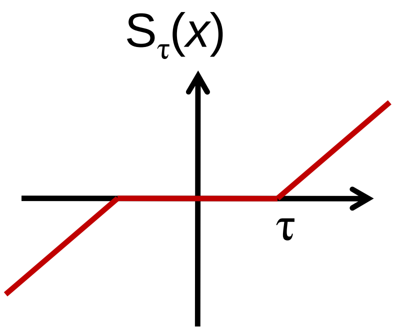
Singular-value thresholding

- Anomaly matrix update

$$\mathbf{A}[k+1] := \arg\min_{\mathbf{A}} \left\{ \frac{L_f}{2} \|\mathbf{A} - \mathbf{G}_A[k]\|_F^2 + \lambda_1 \|\mathbf{A}\|_1 \right\}$$

Soft thresholding

Sol. $\boxed{\mathbf{A}[k+1] = \mathcal{S}_{\underbrace{\lambda_1/L_f}}[\mathbf{G}_A[k]]}$



- Low rank $\mathbf{X}_R[k+1]$ and sparse $\mathbf{A}[k+1]$

Benchmark: PCA-based method

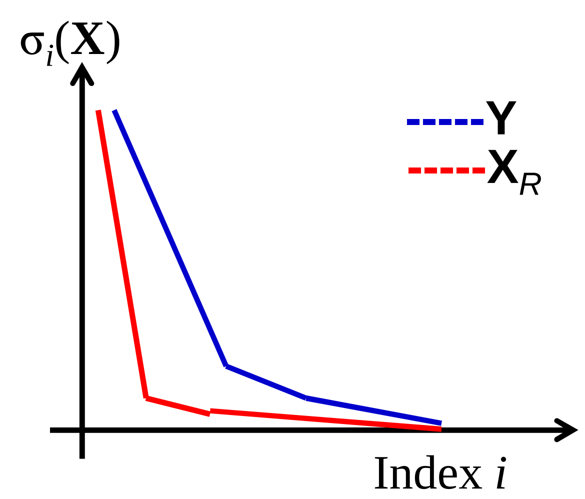
- Idea: anomalies increase considerably rank(**Y**)

Algorithm [Lakhina et al'04]

i) Form subspace S via r-dominant left singular vectors of **Y** (resp. S^c)

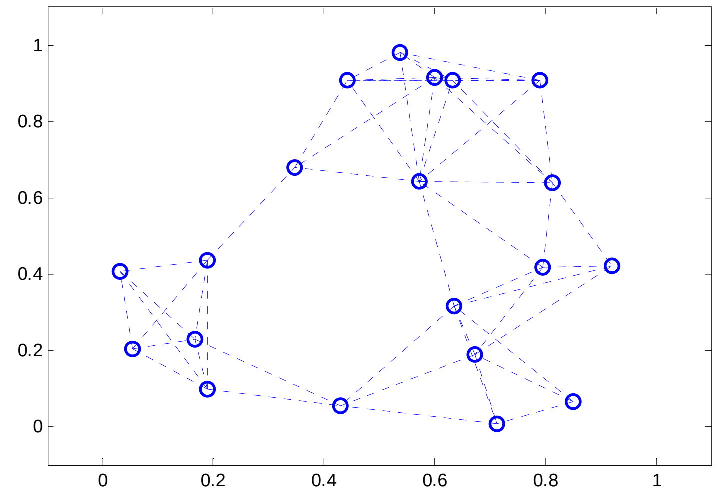
ii) Test: $\|\mathcal{P}_{S^c}(\mathbf{y}_t)\|_2 \gtrless_{H_0}^{H_1} \tau$, for $t = 1, \ldots, T$

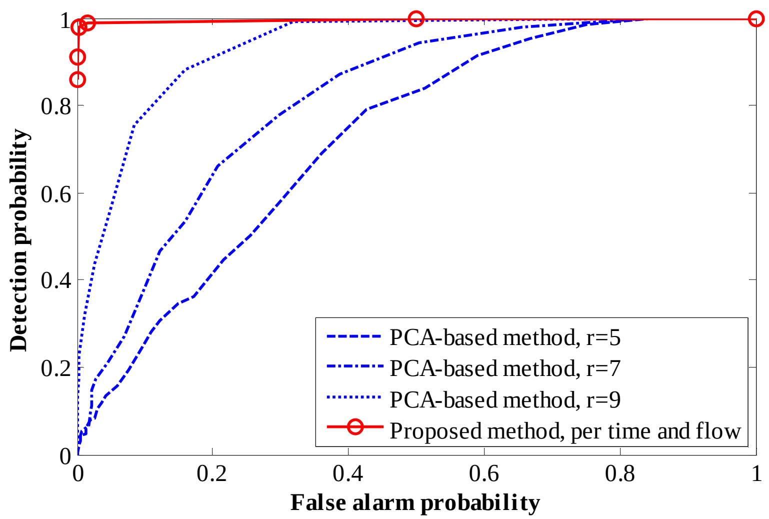
- Limitations
  - Unable to identify flows
  - Assumes knowledge of rank($\mathbf{X}_R$)



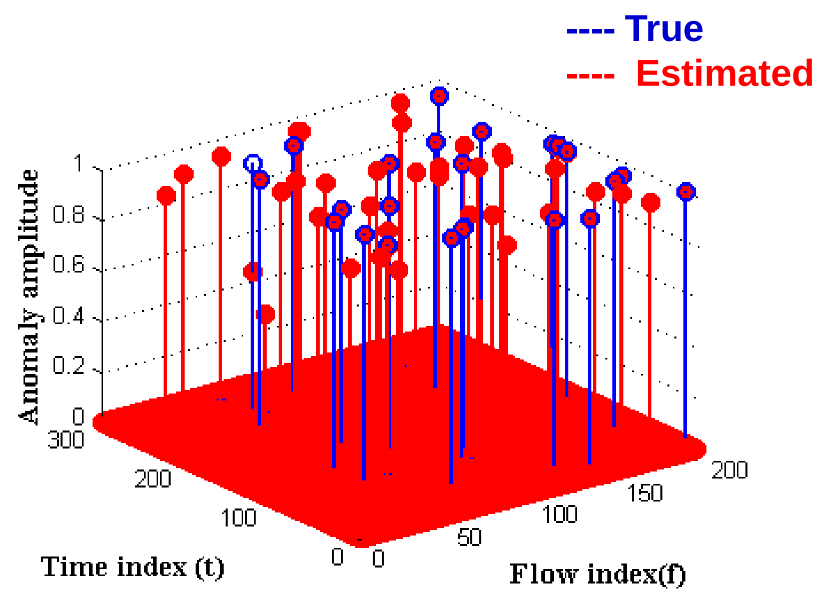
Synthetic data

- Random network topology
  - N=20, L=108, F=360, T=760
  - Minimum hop-count routing



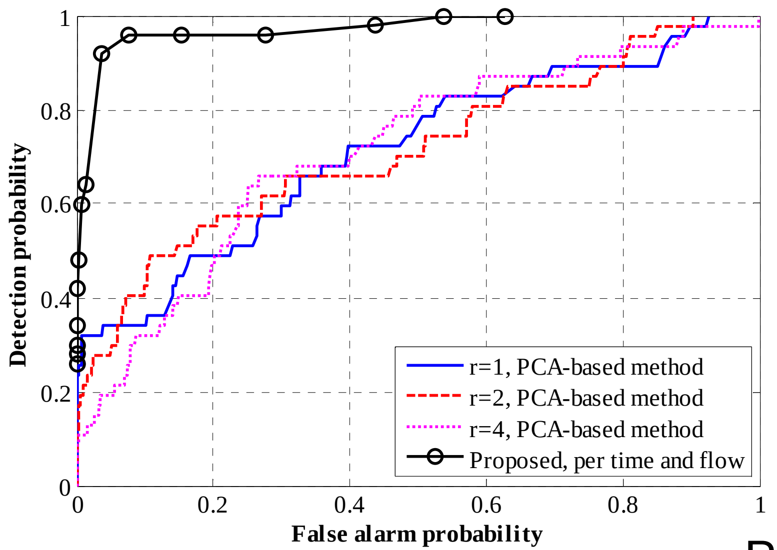


P_f=10^{-4}
P_d = 0.97

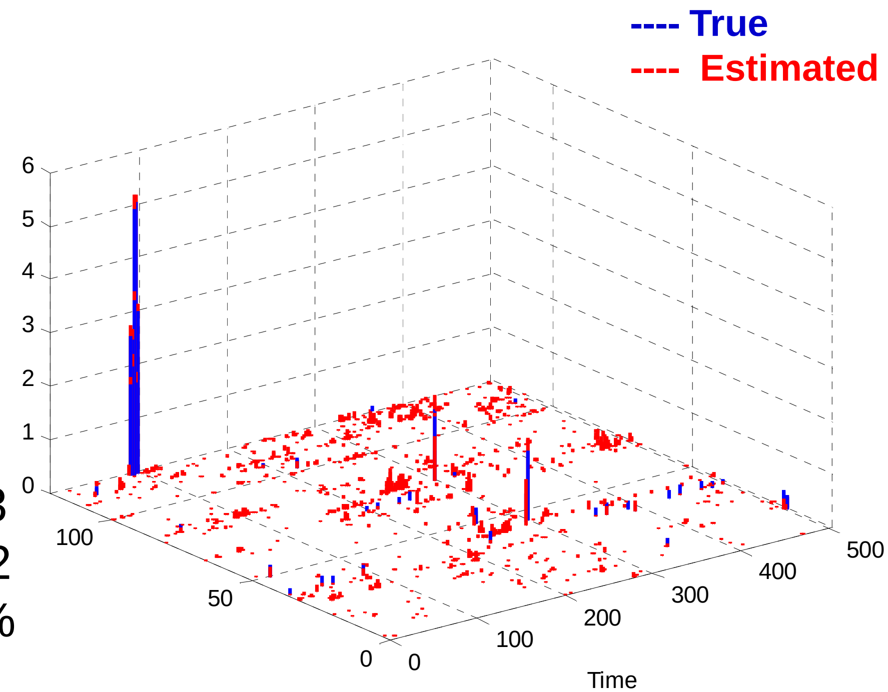


Real data

- Abilene network data
  - Dec. 8-28, 2008
  - N=11, L=41, F=121, T=504



$P_f = 0.03$
$P_d = 0.92$
$Q_e = 27\%$



Concluding summary

- Exact recovery of low-rank plus compressed matrices
- Important task of network anomaly detection

- Detailed discussion
  - Identifibaility and exact recovery

M. Mardani, G. Mateos, and G. B. Giannakis,``Recovery of Low-Rank Plus Compressed Sparse Matrices with Application to Unveiling Traffic Anomalies," *IEEE Trans. Info. Theory*, submitted Apr. 2012. **arXiv: 1204.6537**

  - Distributed optimization with missing data

M. Mardani, G. Mateos, and G. B. Giannakis, ``In-network Sparsity-regularized Rank Minimization: Algorithms and Applications," *IEEE Trans. Sig. Proc.*, submitted Feb. 2012. **arXiv: 1203.1570**

Thank You!