IP Forwarding Anomalies and Improving their Detection using Multiple Data Sources

Matthew Roughan (Univ. of Adelaide)

Tim Griffin (Intel Research Labs)

Z. Morley Mao (Univ. of Michigan)

Albert Greenberg, Brian Freeman

(AT&T Labs-Research)

Network Anomalies

- What are network anomalies?
  - Any unexpected behavior in networks
  - Likely indications of network problems
- Networks anomalies occur all the time
  - Require better understanding
  - Require fast and accurate detection
- Anomalies are not known in advance
  - Cannot match signatures, no stable signatures
  - Need to detect *anomalous* behavior
  - Need to define what is normal and anomalous
  - Difficult to prevent them from happening

Automated Anomaly Detection

- Automated techniques are rarely perfect
  - Two types of errors
    - Failure to detect (false negative)
    - False alarm (false positive)
  - Tradeoff between the two
- Single source
  - Improvements must reduce both errors
  - Quickly reach the point of diminishing returns
  - Current approach of most researchers
- However, two sources can — Our approach
  - Reduce the false alarm rate dramatically
  - Not much reduction in detection rate

Two Sources with Independent Errors

- Probability of detection in source i is
  p_i=0.99
- Probability of detection in both sources (AND) is
  $p_1 \times p_2 = 0.99 \times 0.99 = 0.98$
- False alarm probability in source i is
  q_i =0.02
- False alarm probability in AND is
  $q_1 \times q_2$ =0.02 x 0.02 = 0.0004

Large reduction in false alarms, slight reduction in detection accuracy.

IP Forwarding Anomalies are Important

- Severe disruption in forwarding
  - High impact events
  - Typically affecting more than one router or link
  - Affecting end to end performance of customers
- Causes:
  - Control plane failures
  - Implementation bugs
  - Configuration errors
- Typical symptoms:
  - Packet drops, reordering, high delays
  - Unreachable destinations
  - Fluctuating routes
  - Changes in traffic volume

Detection Methodology

- Methodology:
  1. Use two data sources: routing and traffic
  2. Individually process each source
  3. Combining anomaly signals
     - signal alarms when **both** indicate anomalies concurrently
- Advantages:
  - Uncorrelated errors, correlated anomalies
  - Low false alarm rate, high detection rate
  - Simple and robust
  - Scalable, automated, self-training

Traffic Analysis

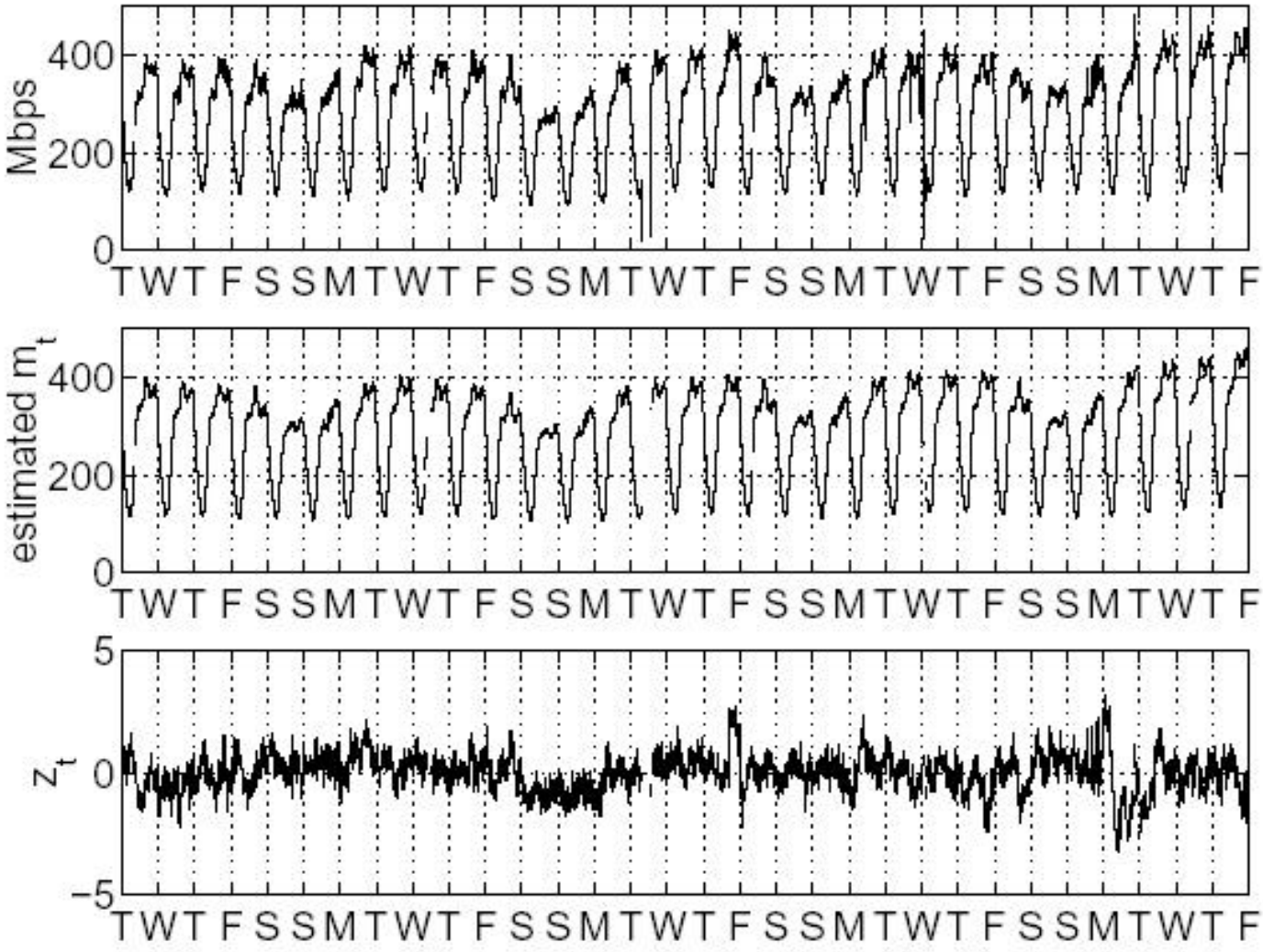
- SNMP (Simple Network Management Protocol)
  - Traffic volume per time interval
  - Coarse grained
  - Ubiquitous

1. Basic anomaly detection algorithm
   1. Holt-Winters
2. Decomposition-based algorithm
   - Decompose into 4 components:
     1. Seasonal/Periodic Component: S_t
     2. Long term trend: T_t
     3. Normal stochastic component: W_t
     4. Anomalies, (impulse functions): I_t
   - X_t: traffic at time t,
   - a: peakedness parameter,
   - m_t: regular, predictable mean ($m_t = S_t * T_t$)

$$x_t = m_t + \sqrt{a m_t} W_t + I_t$$

SNMP traffic data processing

Link 1 output traffic (May 2001)



Original SNMP traffic data

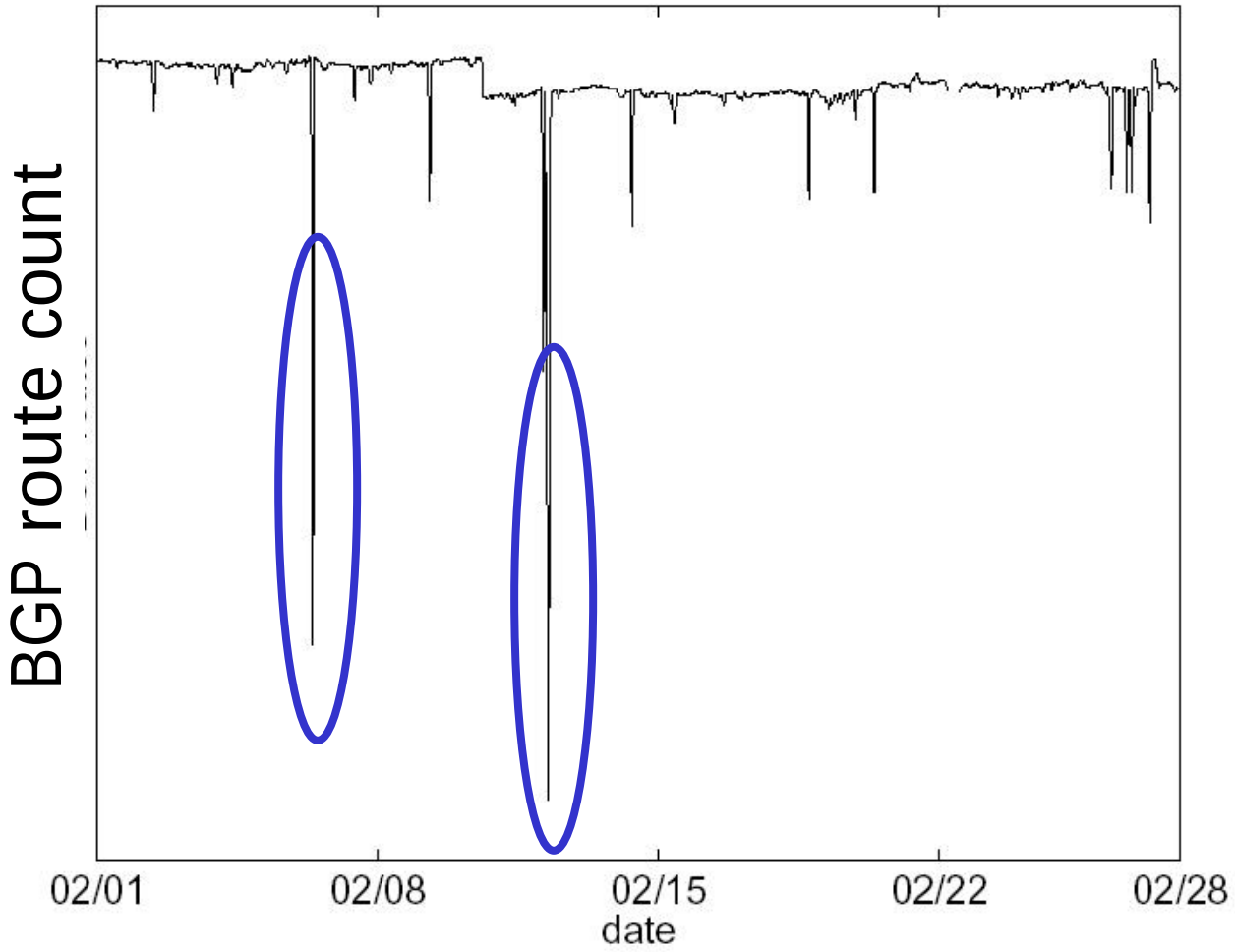
Regular component

Stochastic component

Routing Analysis

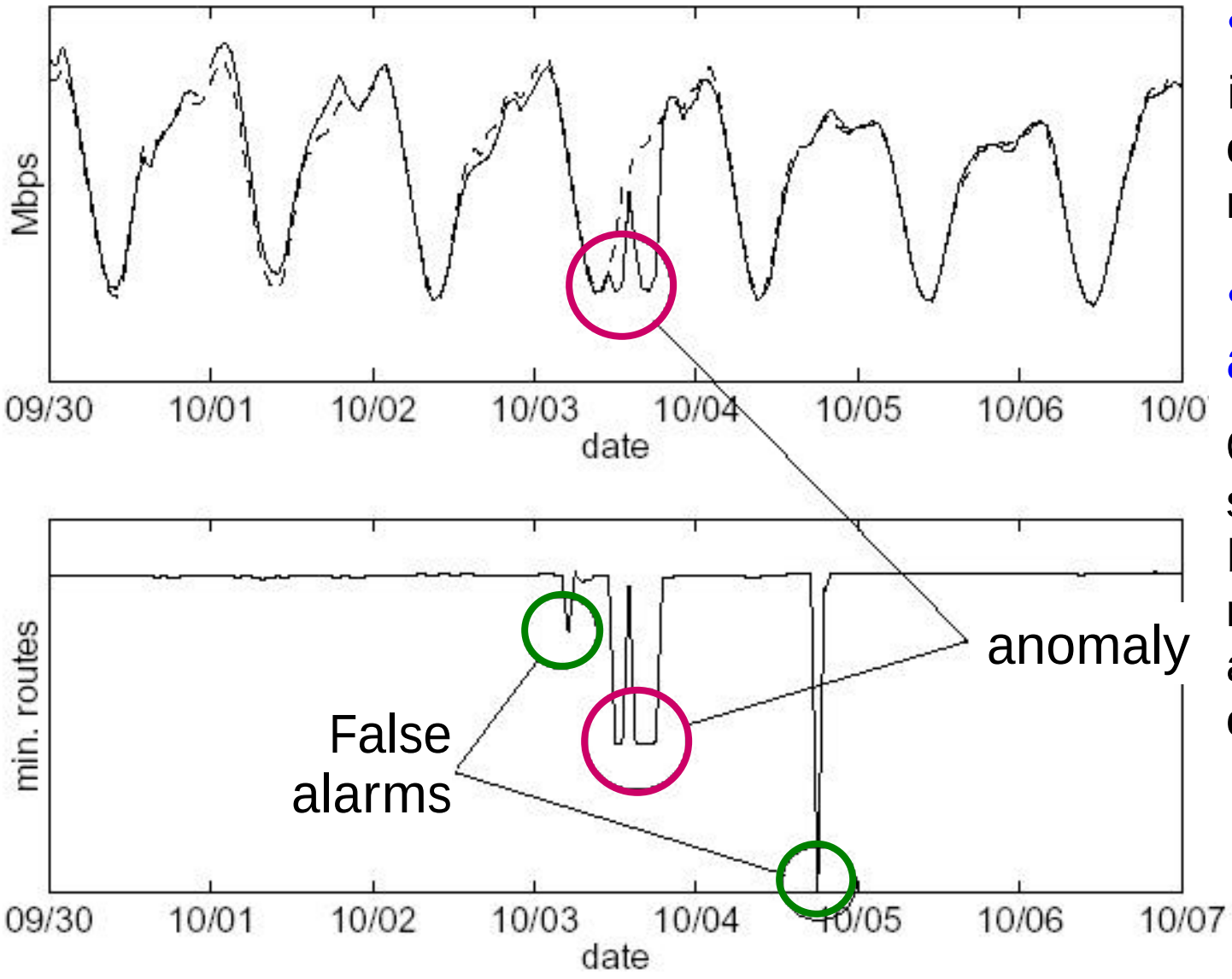
- BGP: Interdomain routing protocol
- Internal route monitor to all route reflectors
- Aggregate routes based on exit point
  - Look for route fluctuations
  - Appear as differences in route counts
- BGP data processing
  - EWMA (Exponentially Weighted Moving Average)
  - Exclude anomalies from moving average

BGP Data Analysis:



- Number of BGP routes from a major PoP
- Anomalies (of interest) are short-lived, steep drops in the number of routes

Example 1



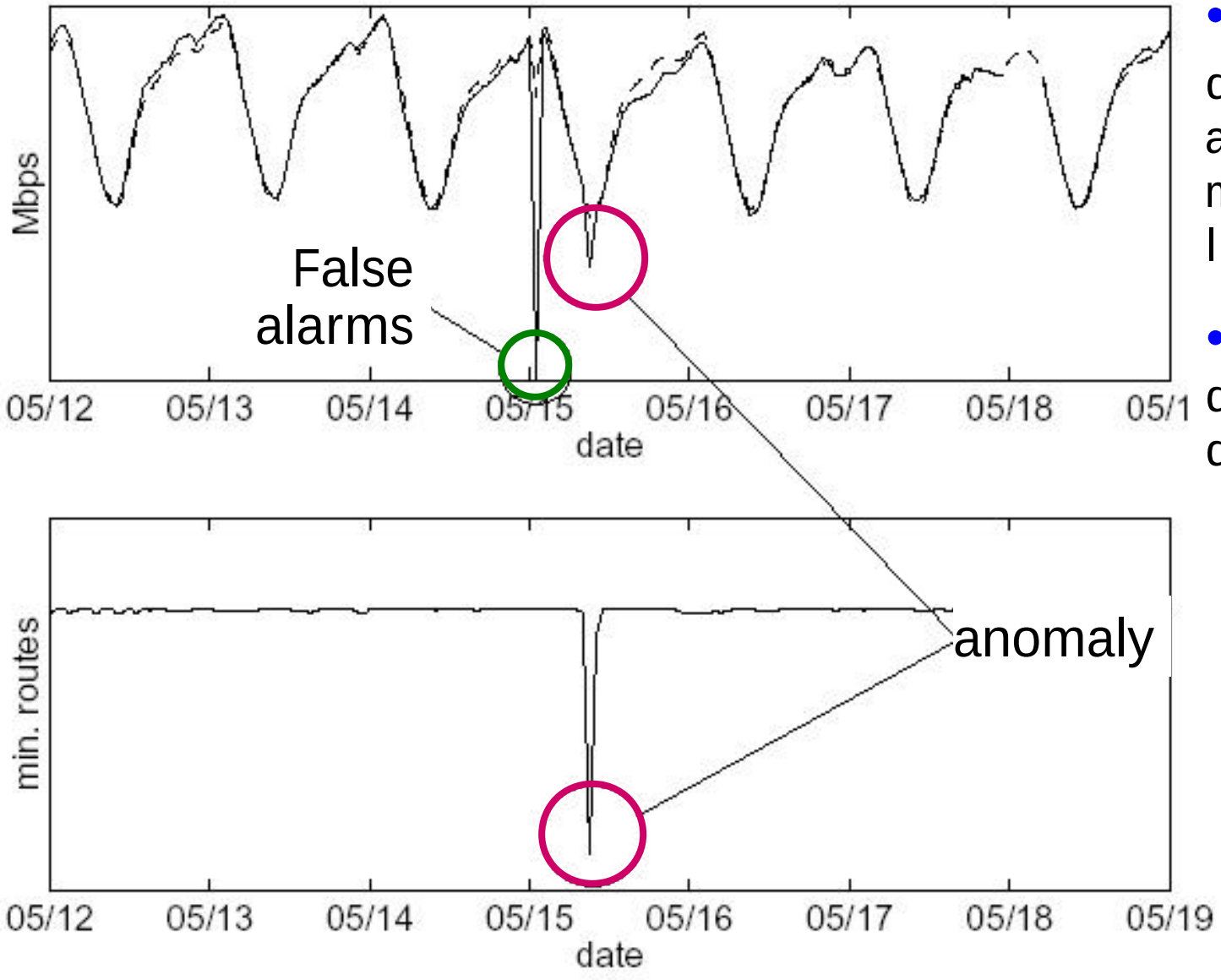
- Anomaly

is due to failure of a major network peer

- BGP false alarm

Commonly due to session reset between the monitoring router and the operational router

Example 2



- Anomaly

due to outage of a router with many peering links

- False alarm

due to missing data

Evaluation using Known Events

- List of known events over a year
  - Notified by operations
  - Considered important
  - Perfect detection accuracy

Evaluation of Individual Algorithms

Data set	Algorithm	False alarm rate	Expected false alarms per day
SNMP	Decomposition	3.4%	78
SNMP	Holt-Winters	4.3%	99
BGP	EWMA	0.5%	12

Evaluation using Fault Tickets

- Feb to May 2003
  - Take all the detections
  - Identify root cause analysis based on detailed fault tickets

Root cause	Decomposition on SNMP EWMA on BGP
Edge node/link outage	67%
Simultaneous outages	11%
Unknown cause	22%
False alarms	0%

Conclusions

- Powerful idea of combining multiple data sources for anomaly detection
  - Significantly lower false alarms
  - Little degradation in detection accuracy
  - Simple and robust (e.g. missing data)
  - Scalable, automated, self-training
- Applied to detecting forwarding anomalies
  - Important to network operations
  - Discovered SNMP and BGP features with the right properties

Future Work

- Multi-Dimensional Event Correlation
  - Extension to include additional data sources (ongoing work by Ramana Kompella et. al.)
  - OSPF, SONET PM data, Router Syslogs, Trouble Tickets, etc.
- Algorithm improvement
  - Statistical techniques (e.g., Wavelets)
  - Bayesian networks to combine different data
- Automated fault diagnosis and troubleshooting (root cause analysis)