A Signal Analysis of Network Traffic Anomalies

Paul Barford

with Jeffery Kline, David Plonka, Amos Ron

University of Wisconsin – Madison

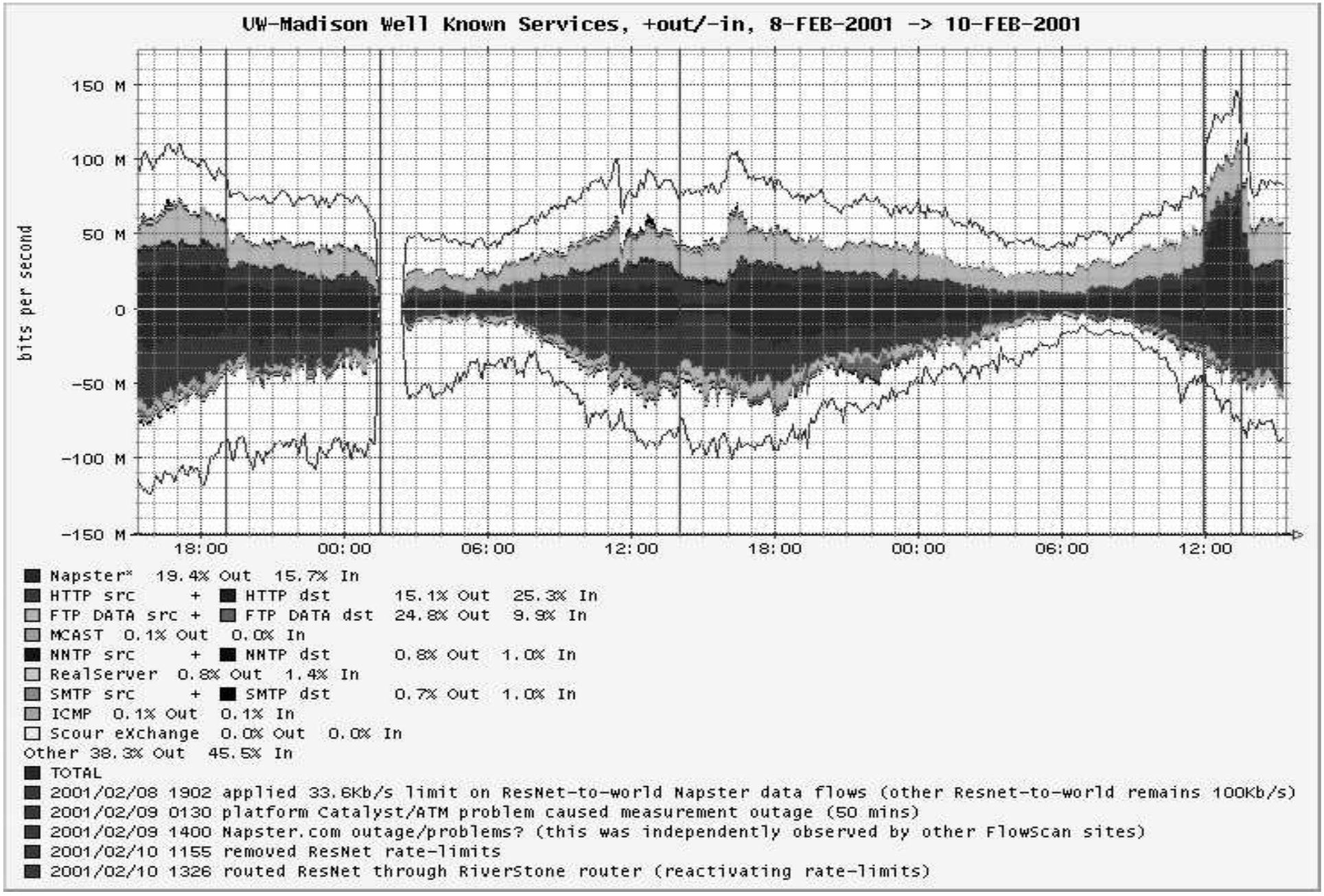
Fall, 2002

Overview

- **Motivation:** Anomaly detection remains difficult
- **Objective**: Improve understanding of traffic anomalies
- **Approach**: Multiresolution analysis of data set that includes IP flow, SNMP and an anomaly catalog
- **Method**: Integrated Measurement Analysis Platform for Internet Traffic (IMAPIT)
- **Results**: Identify anomaly characteristics using wavelets and develop new method for exposing short-lived events

Our Data Sets

- Consider anomalies in IP flow and SNMP data
  - Collected at UW border router (Juniper M10)
  - Archive of ~6 months worth of data (packets, bytes, flows)
  - Includes catalog of anomalies (after-the-fact analysis)
- Group observed anomalies into four categories
  - **Network anomalies (41)**
    - Steep drop offs in service followed by quick return to normal behavior
  - **Flash crowd anomalies (4)**
    - Steep increase in service followed by slow return to normal behavior
  - **Attack anomalies (46)**
    - Steep increase in flows in one direction followed by quick return to normal behavior
  - **Measurement anomalies (18)**
    - Short-lived anomalies which are not network anomalies or attacks


UW-Madison Well Known Services, +out/-in, 8-FEB-2001 -> 10-FEB-2001
150 M
100 M
50 M
0
-50 M
-100 M
-150 M
bits per second
18:00
00:00
06:00
12:00
18:00
00:00
06:00
12:00
Napster* 19.4% Out 15.7% In
HTTP src + HTTP dst 15.1% Out 25.3% In
FTP DATA src + FTP DATA dst 24.8% Out 9.9% In
MCAST 0.1% Out 0.0% In
NNTP src + NNTP dst 0.8% Out 1.0% In
RealServer 0.8% Out 1.4% In
SMTP src + SMTP dst 0.7% Out 1.0% In
ICMP 0.1% Out 0.1% In
Scour eXchange 0.0% Out 0.0% In
Other 38.3% Out 45.5% In
TOTAL
2001/02/08 1902 applied 33.6Kb/s limit on ResNet-to-world Napster data flows (other Resnet-to-world remains 100Kb/s)
2001/02/09 0130 platform Catalyst/ATM problem caused measurement outage (50 mins)
2001/02/09 1400 Napster.com outage/problems? (this was independently observed by other FlowScan sites)
2001/02/10 1155 removed ResNet rate-limits
2001/02/10 1326 routed ResNet through RiverStone router (reactivating rate-limits)

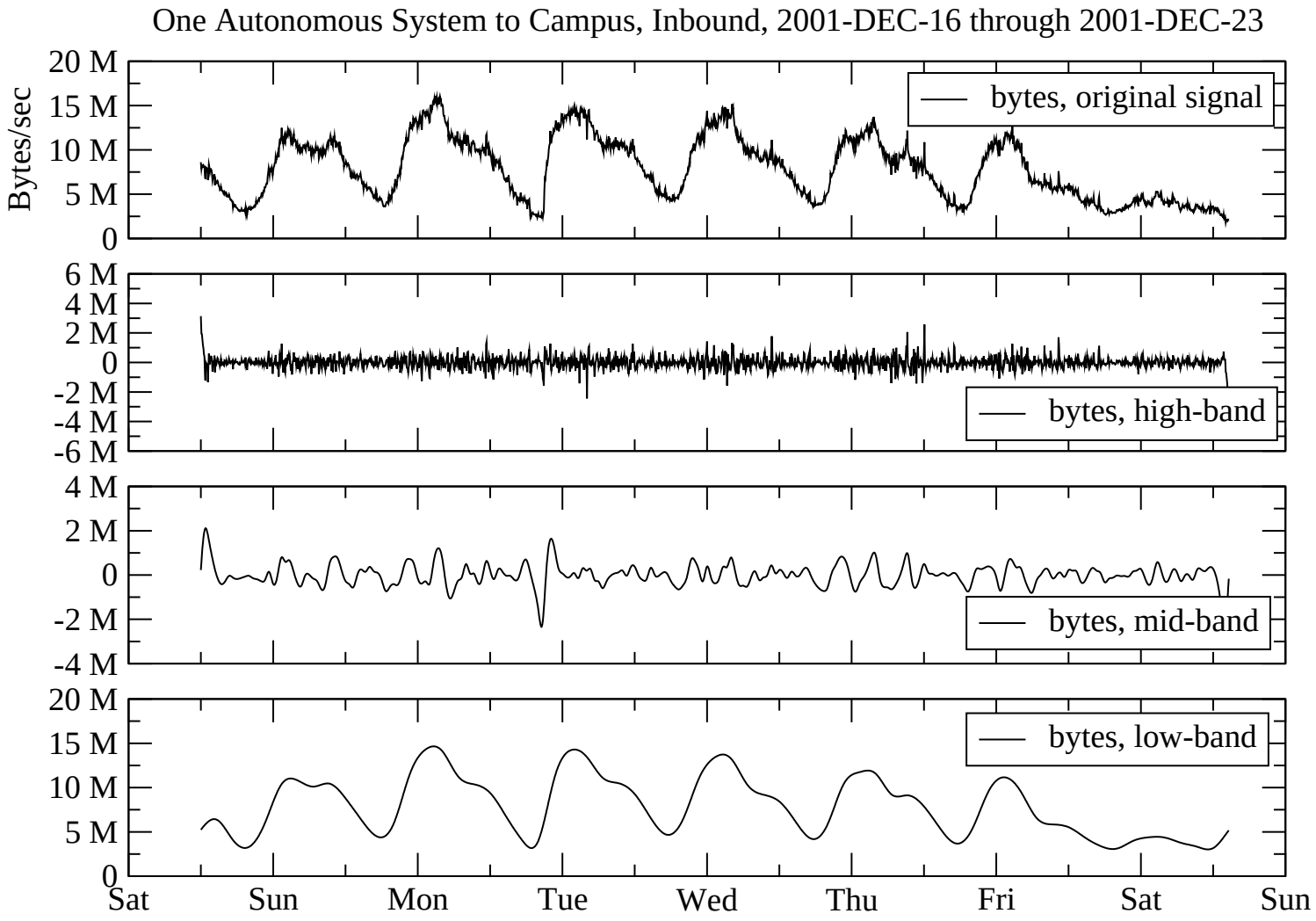
Multiresolution Analysis

- Wavelets provide a means for describing time series data that considers both *frequency* and *time*
  - Powerful means for characterizing data with sharp spikes and discontinuities
  - Using wavelets can be quite tricky
- We use tools developed at UW which together make up IMAPIT
  - FlowScan software
  - The IDR Framenet software

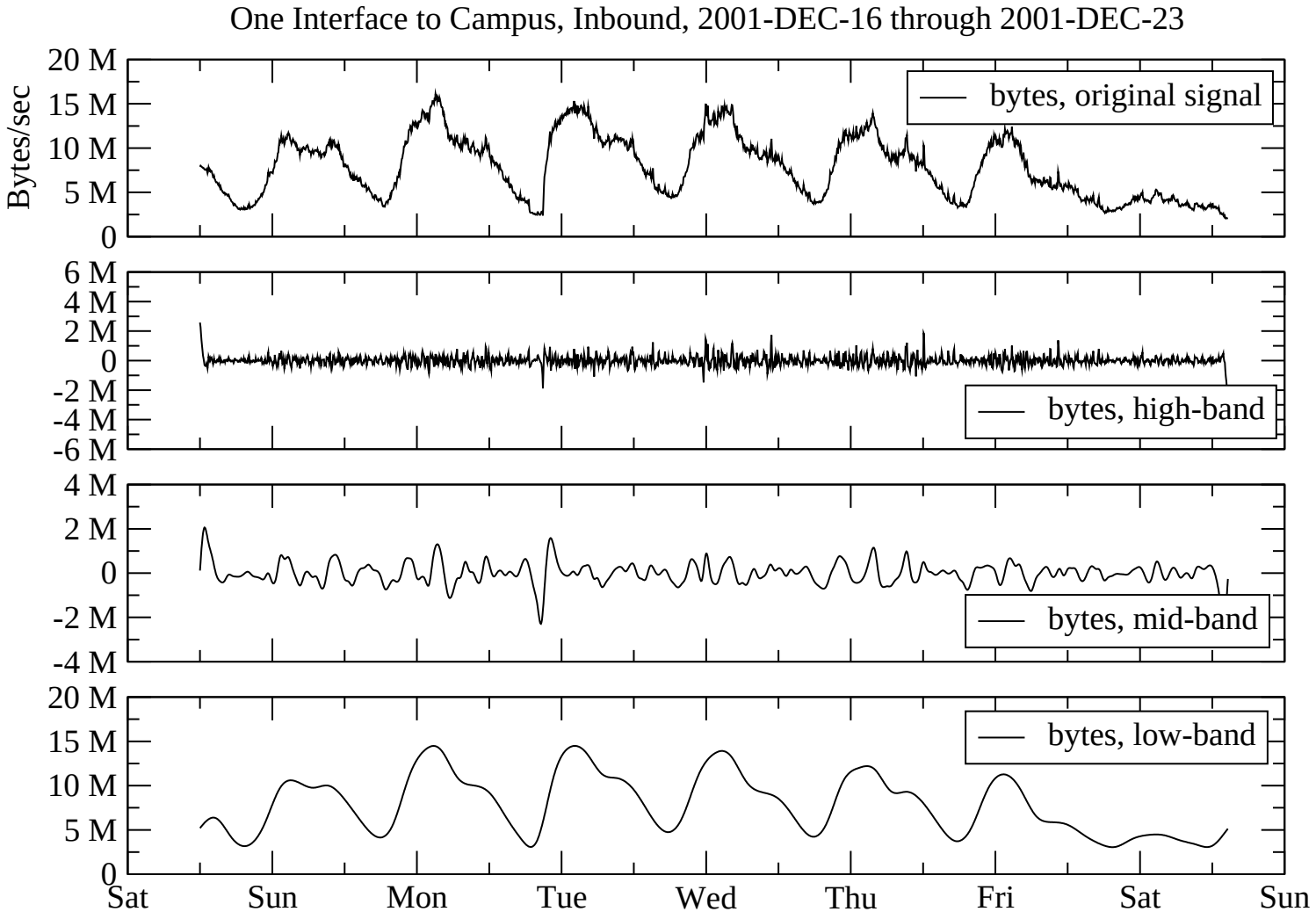
Our Wavelet System

- After evaluating different candidates we selected a wavelet system called Pseudo Splines(4,1) Type 2.
  - A *framelet* system developed by Daubechies *et al.* ‘00
  - Very good frequency localization properties
- Three output signals are extracted
  - **Low Frequency (L)**: synthesis of all wavelet coefficients from level 9 and up
  - **Mid Frequency (M)**: synthesis of wavelet coefficients 6, 7, 8
  - **High Frequency (H)**: synthesis of wavelet coefficients 1 to 5

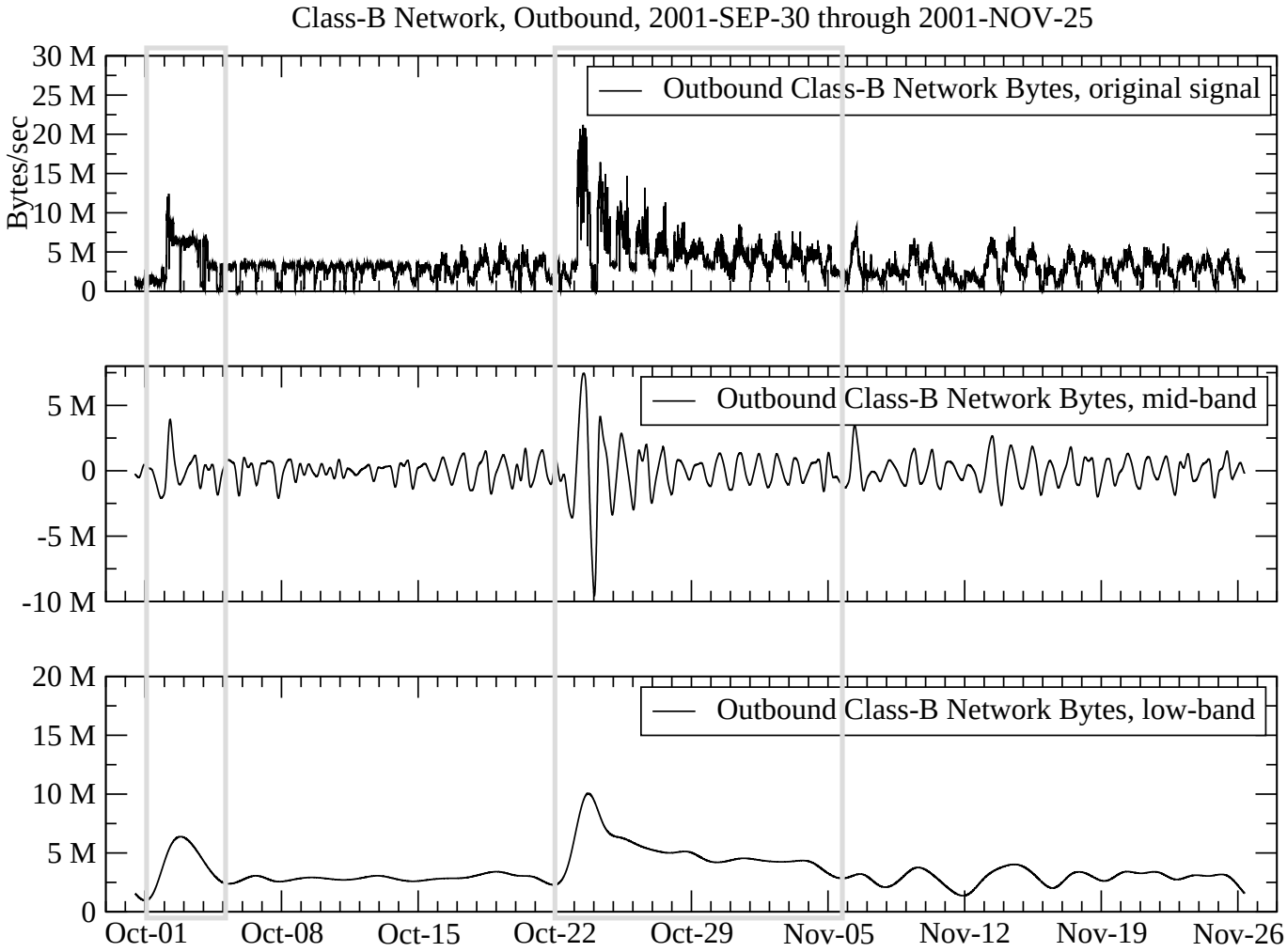
Ambient IP Flow Traffic



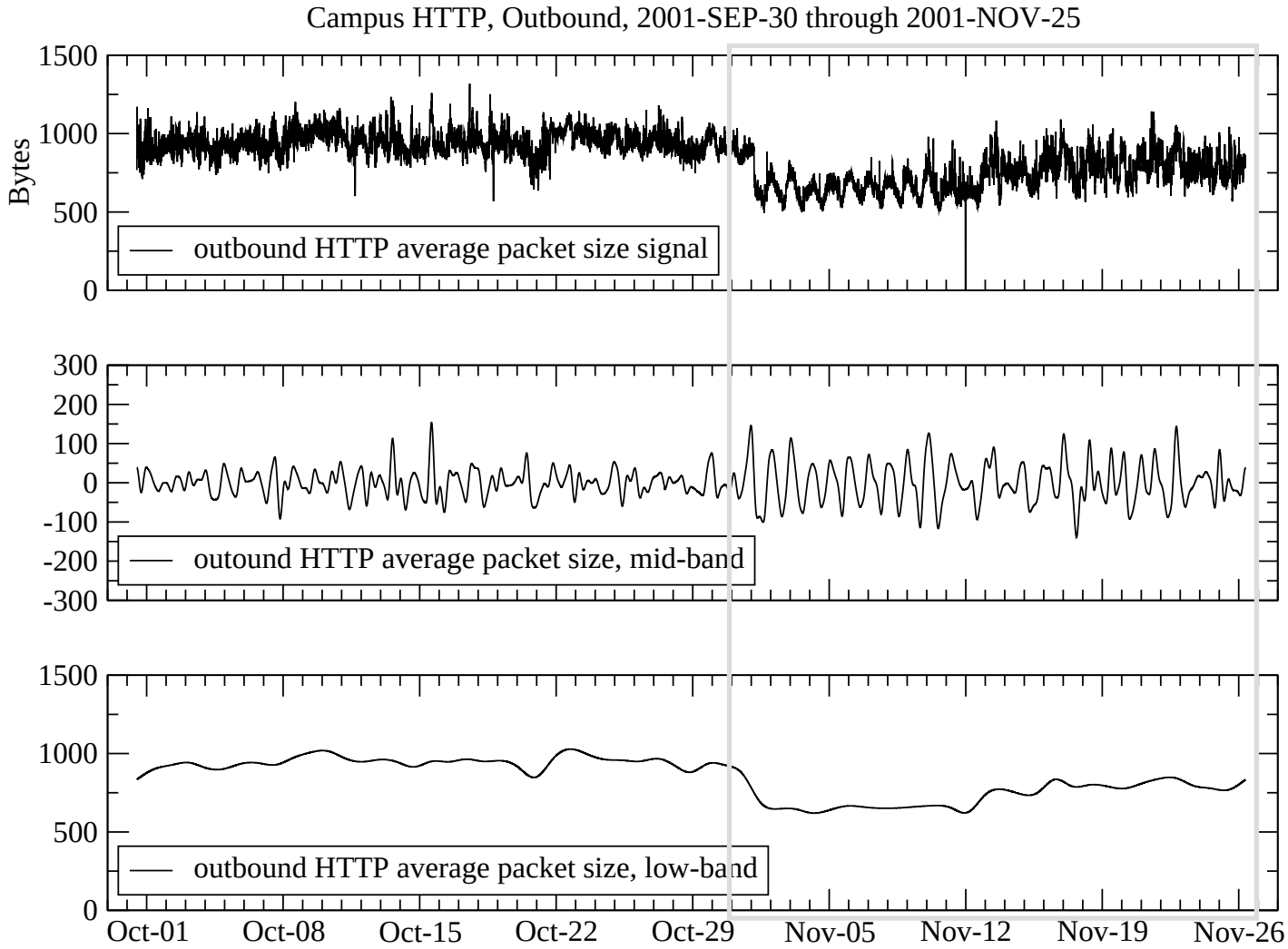
Ambient SNMP Traffic



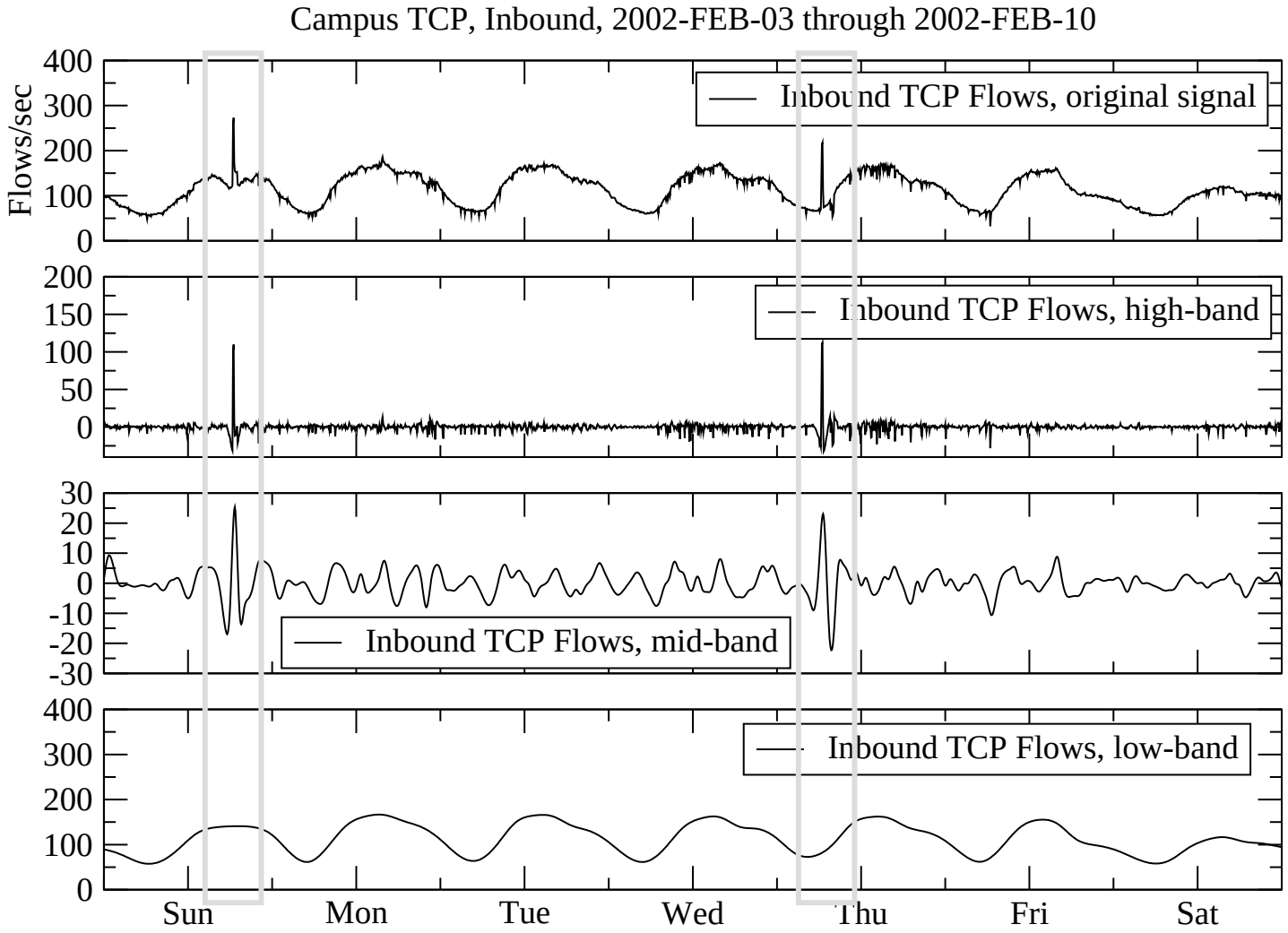
Byte Traffic for Flash Crowd



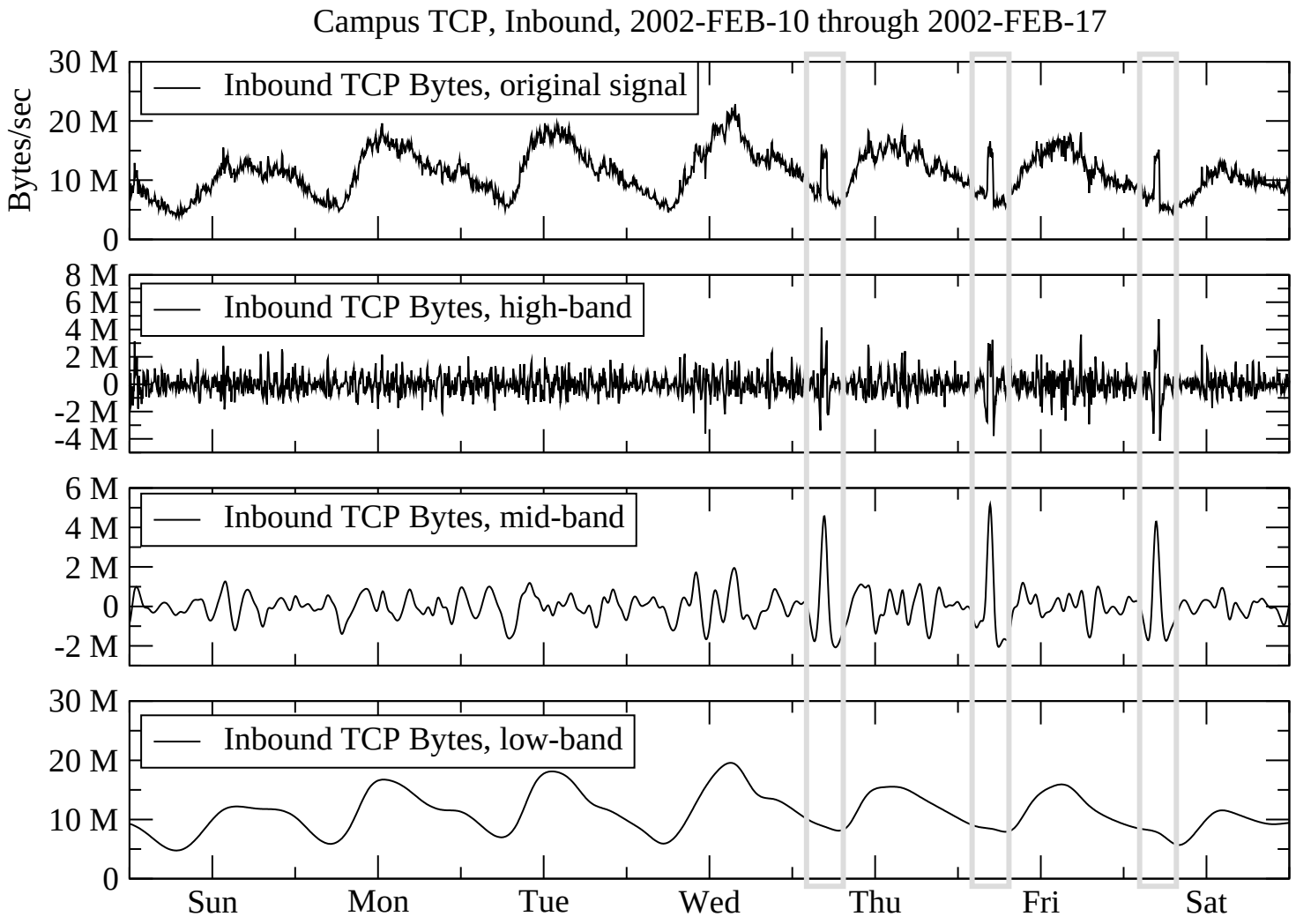
Average Packet Size for Flash Crowd



Flow Traffic During DoS Attacks



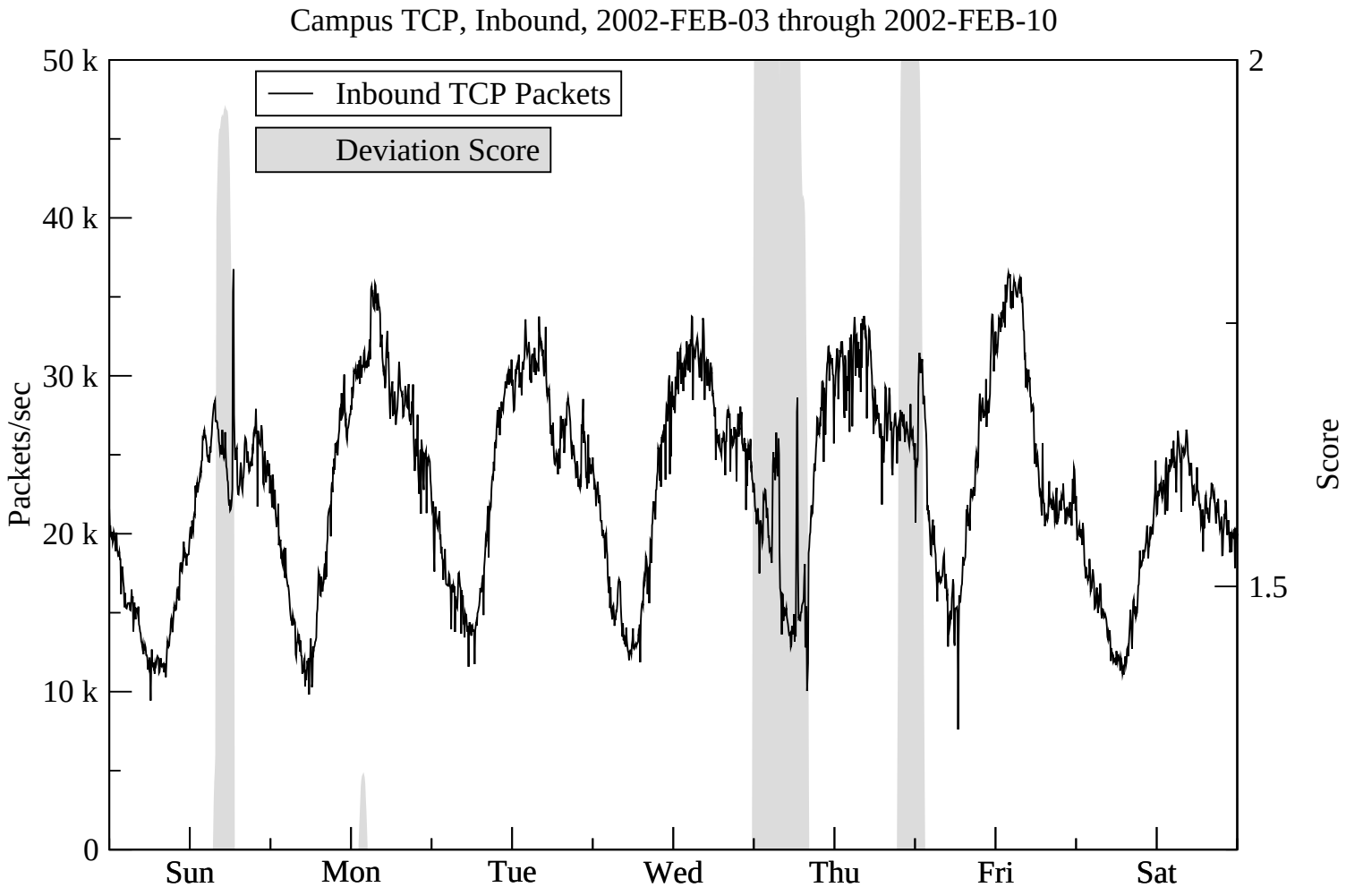
Byte Traffic During Measurement Anomalies



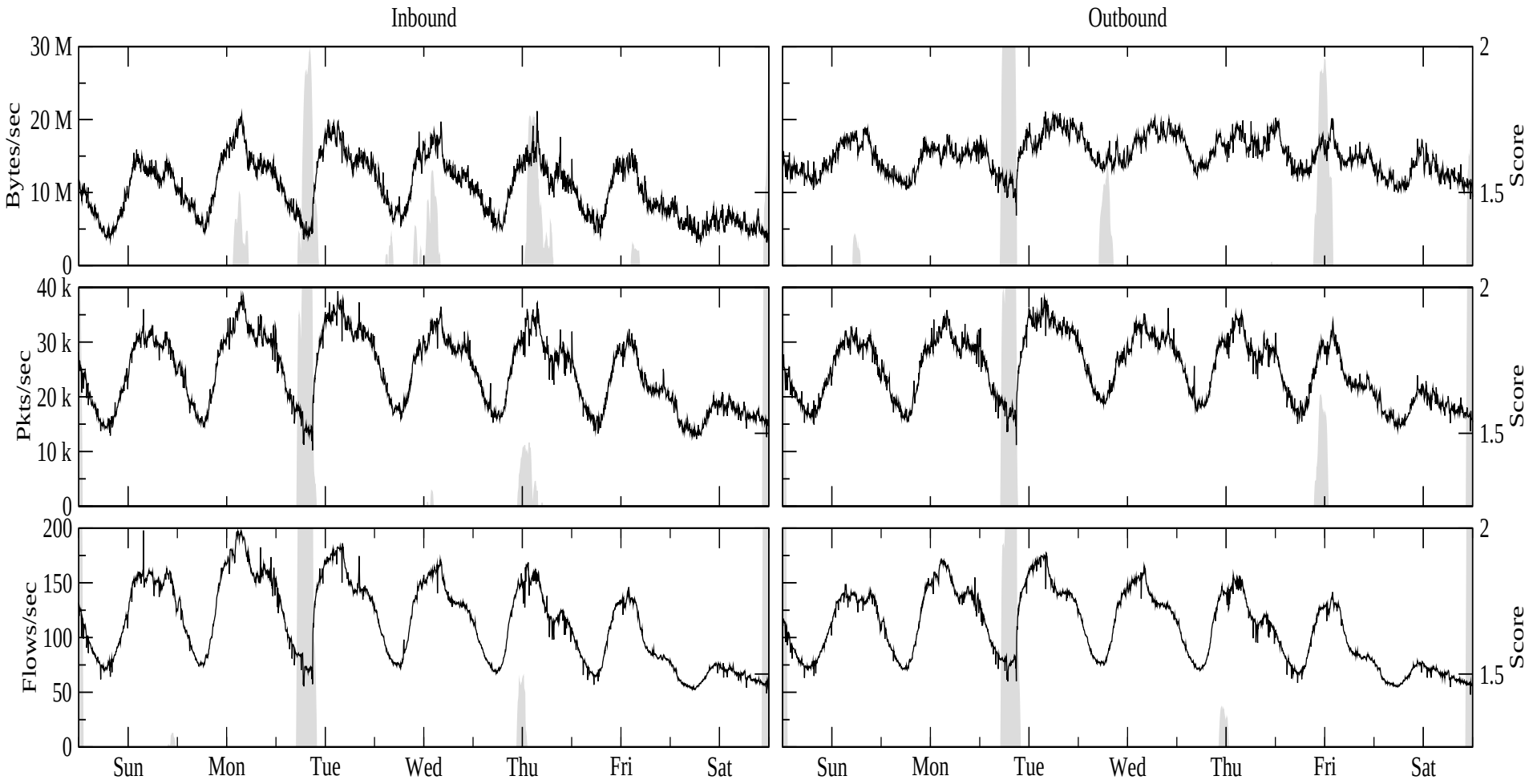
Anomaly Detection via Deviation Score

- Short-lived anomalies can be identified automatically based on variability in H and M signals
  1. Compute local variability (using specified window) of H and M parts of signal
  2. Combine local variability of H and M signals (using a weighted sum) and normalize by total variability to get deviation score V
  3. Apply threshold to V then measure peaks
- Analysis shows that V peaks over 2.0 indicate short-lived anomalies with high confidence
  - We threshold at V = 1.25 and set window size to 3 hours

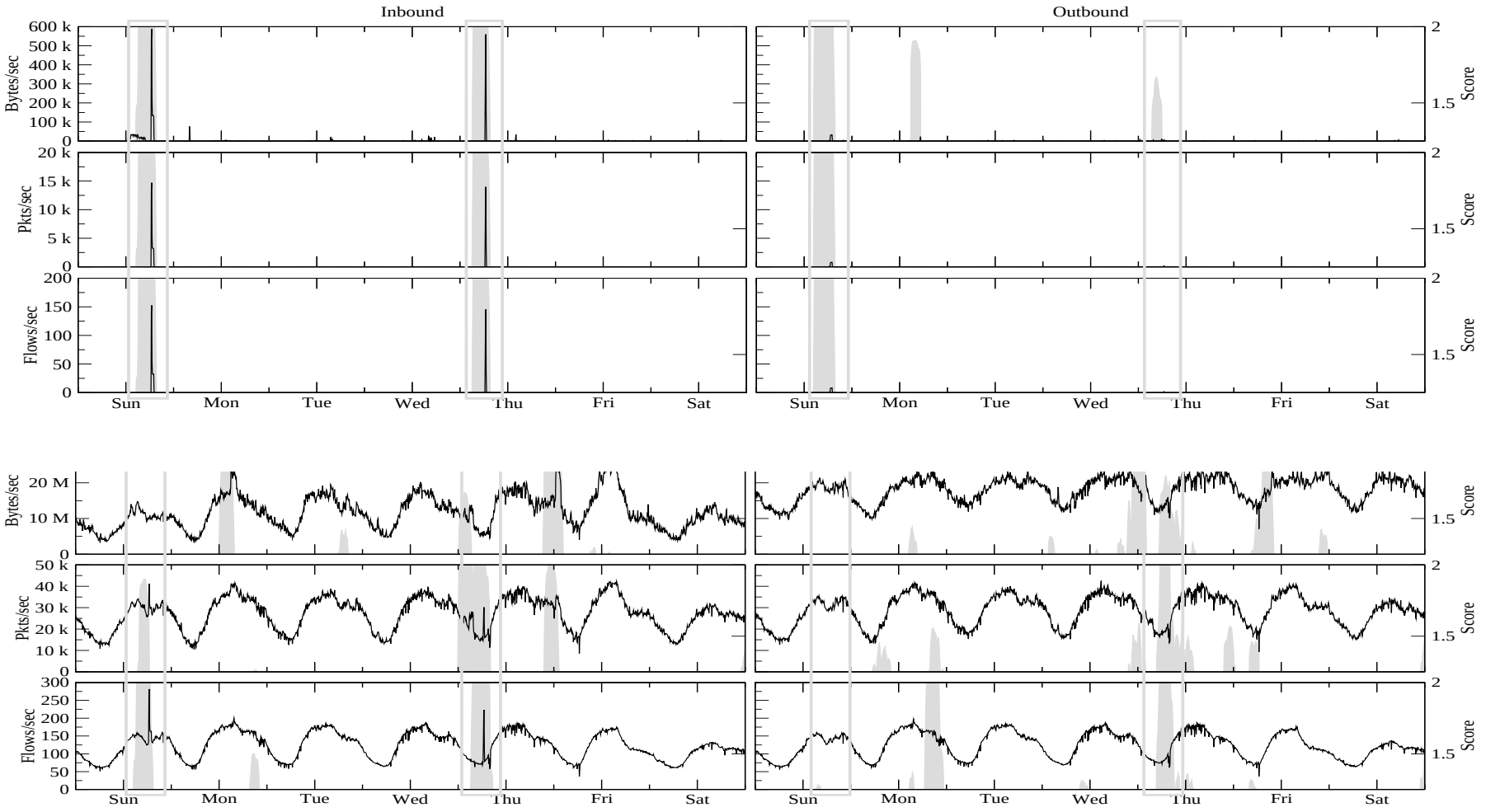
Deviation Score for Three Anomalies



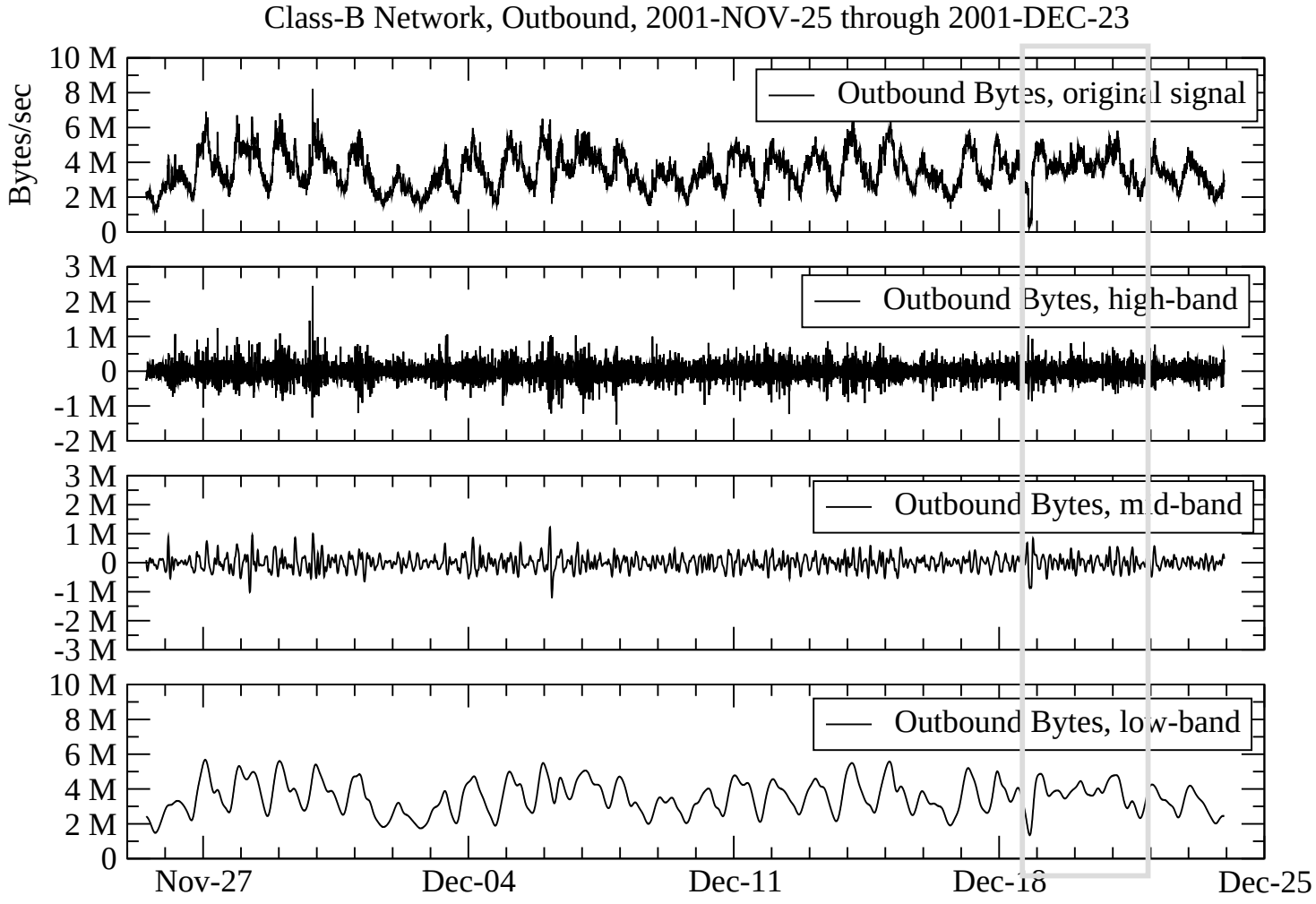
Deviation Score for Network Outage



Anomalies in Aggregate Signals



Hidden Anomalies in Low Frequency



Deviation Score Evaluation

- How effective is deviation score at detecting anomalies?
  - Compare versus set of 39 anomalies
    - Set is unlikely to be complete so we don't treat false-positives
  - Compare versus ***Holt-Winters Forecasting***
    - Time series technique
    - Requires some configuration
- Holt-Winters reported many more positives and sometimes oscillated between values

Total Candidate Anomalies	Candidates detected by Deviation Score	Candidates detected by Holt-Winters
39	38	37

Conclusion and Next Steps

- We present an evaluation of signal characteristics of network traffic anomalies
  - Using IP flow and SNMP data collected at UW border router
  - IMAPIT developed to apply wavelet analysis to data
  - Deviation score developed to automate anomaly detection
- Results
  - Characteristics of anomalies exposed using different filters and data
  - Deviation score appears promising as a detection method
- Future
  - Development of anomaly classification methods
  - Application of results in (distributed) detection systems