Signal Processing Methods for Network Anomaly Detection

Lingsong Zhang

Department of Statistics and Operations Research

Email: *LSZHANG@email.unc.edu*

March 7, 2005; March 9, 2005

Outline

- Introduction and Background
- Single Time Series Methods
  - Spectral Analysis
  - Wavelet Analysis
  - Singular Value Decomposition
- Multiple Time Series and Multivariate Methods
  - Multivariate Outlier Detection
  - Principal Component Analysis
- Further Work and Comments

Part I – Introduction

Introduction and Background

- DoS is popular now, with possible catastrophe, by consuming finite resource
- Detection of and response to DoS is essential for network
- Network Traffic itself is hard to be analyze
  - Non-Gaussian, Non-Stationary, Long Range Dependence, Heavy tailed
- Attackers will try to make the attack traffic hard to be distinguished from normal traffic
- Not sure which measurements best fit for anomalies detection
- Not sure which method is best

Major issues for Detection

- Stand-alone intrusion detection appliances should automatically recognize the network is under attack and adjust its traffic flow to ease the attack impact downstream
- The detection and response techniques should be adaptable to a wide range of network environments, without significant manual tuning
- False negative and false positive should be as small as possible
- Attack response should employ intelligent packet discard mechanisms to reduce the downstream impact of the flood while preserving and routing the non-attack packets
- The detection method should be effective against a variety of attack tools available today and also robust against future attempts by attackers to evade detection

Part II – Analysis Methods

Analysis Methods

- Single Time Series Analysis Methods
  - Spectral analysis
  - Wavelet analysis
  - Singular Value Decomposition
  - other methods?
- Multiple Time Series or Multivariate Analysis Methods
  - Multivariate Outlier Detection Method
  - Principal Component analysis (Singular Value Decomposition) Method

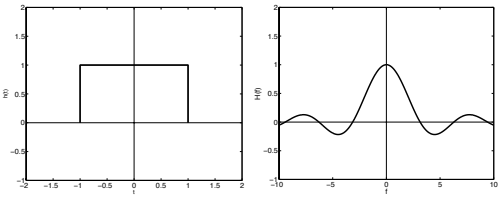
Spectral analysis in Defense Against DoS Attacks

- Motivation
  - normal TCP flows must exhibit periodicity in packet transport associated with round-trip times.
  - fourier transform is a good tool to test the periodicity.
- Spectral Analysis
  - Fourier transform is a frequency-domain representation of a function. This allows us to examine the function from another point of view, the transformed (frequency) domain.
  - A good reference
    Brigham, E. Oran, (1988) "*The fast Fourier transform and its applications*", Prentice-Hall, Inc.

Outline of Spectral Analysis

- An simple example
- Properties of TCP traffic
  - Simulated traffic
  - Real traffic
- Discussion

An example



where

$$h(t) = \begin{cases} 1 & t \in [-1,1] \\ 0 & o.w. \end{cases}, \quad H(f) = \frac{2sin(f)}{f}$$

Fourier Transform and inverse Fourier Transform

- Let $h(t)$ as the time series of one measurement, we have a corresponding fourier transform as

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-ift}dt \tag{1}$$

under certain conditions, the inverse equation holds

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(f)e^{itf}df \tag{2}$$

We have a bijection between $h(t)$ and $H(f)$, corresponding to *time* and *frequency* domain representation

Fourier Transform

- $H(f)$ can be written as

$$H(f) = \int_{-\infty}^{\infty} h(t)\cos(ft)dt - i\int_{-\infty}^{\infty} h(t)\sin(ft)dt$$

we then get

Real Part: $R(f) = \mathrm{Re}(H(f)) = \int_{-\infty}^{\infty} h(t)\cos(ft)dt$

Imaginary Part: $I(f) = \mathrm{Im}(H(f)) = -\int_{-\infty}^{\infty} h(t)\sin(ft)dt$

Fourier Spectral (Amplitude): $|H(f)| = \sqrt{R(f)^2 + I(f)^2}$

Phase angle: $\theta(f) = (tan)^{-1}[\frac{I(f)}{R(f)}]$

Fourier Transform

- Usually we check the properties of $|H(f)|$, and then get the corresponding properties of $h(t)$.
- Usually a time series with periodicity will get spikes in frequency domain.
- For time series analysis, we might analyze the fourier transform of the autocovariance function or autocorrelation function, i.e. the power spectral density(PSD).

Power Spectral Density

- Assume $R_{XX}(k)$ as the autocorrelation function of a time series X, we have the power spectral density of X as

$$S_X(f) = \sum_{k=-\infty}^{\infty} R_{XX}(k) e^{-i2\pi fk}$$

- Periodogram is a commonly used PSD estimate technique, which captures the "power" that a signal contains at a particular frequency. In the following example, the authors use Welch's periodogram to compute PSD estimates.

Application in Defense Against DoS Attacks

- Spectral analysis for identifying normal TCP traffic
  - Cheng et al. (2002) describe a novel use of spectral analysis in identifying normal TCP traffic
  - They exploit the fact that normal TCP flows must exhibit periodicity in packet transport associated with RTT.
- Spectral analysis for detecting attacks
  - can complement existing DoS defense mechanisms that focus on identifying attack traffic
  - rule out those candidates which are deemed to be normal TCP traffic, reduce the impact of false positives of other methods.

Power Spectral Density of Packet Process

- Packet conservation principal
  - every arriving data packet at the receiver allows the departure of an ACK packet, and every arriving ACK packet at the sender enables the injection of a new data packet into the network.
- TCP flows exhibit periodicity
  - If we see a TCP packet at any point in the network, then chances are that after (/no more than) one round-trip time(RTT), we will see another packet belonging to the same TCP flow passing through the same point.

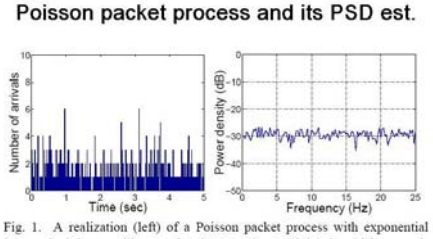


Fig. 1. A realization (left) of a Poisson packet process with exponential inter-arrival times and its associated PSD estimate (right), in which power is evenly spread across all frequencies due to lack of periodicity

Poisson process and its PSD

- Data is from
  - counting the number of arrivals in each of the 10ms bins, with the inter-arrival times independently drawn from an exponential distribution, and with mean arrival rate equal to 200 arrivals per second.
- Power density estimation
  - The PSD estimate has a rather flat power distribution, which corresponds to that of a white noise process.

Pareto interarrival and its PSD est.

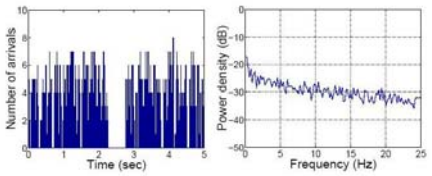


Fig. 2. A realization (left) of a heavy-tailed packet process with Pareto inter-arrival times and its associated PSD estimate (right): the PSD contains more power at low frequencies, compared with Fig. 1

Heavy-tailed process and its PSD

- Pareto Distribution
  - inter-arrival times are drawn from a Pareto distribution $\alpha = 1.3$, and $k = 0.001$. the pdf of a Pareto distribution is

$$f(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, x \geq k$$

- resulting PSD estimate
  - more power at low frequencies than last figure.

Exp. inter-arrival with deterministic arrivals

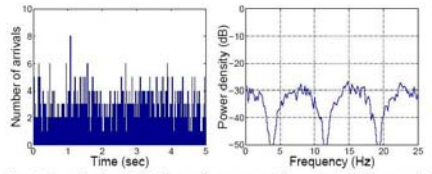


Fig. 3. A realization (left) of a packet process with exponential inter-arrival times mixed with deterministic arrivals and its associated PSD estimate (right): the first peak is located at the frequency corresponding to the time lag of probabilistic arrivals and the triggered deterministic arrivals

Periodicity – Deterministic arrivals

- Deterministic arrivals
  - generating deterministic arrivals interleaved with probabilistic arrivals
  - probabilistic arrivals have exponentially distributed inter-arrival times and each of them further triggers a deterministic arrival after 130ms
- Resulting PSD estimate
  - peaks at approximately 7.7 Hz. which converts to 130ms
  - not like a band-limited signal, no decay. Gaussian-like process, flat PSD

Pareto int-arrival mixed random arrivals

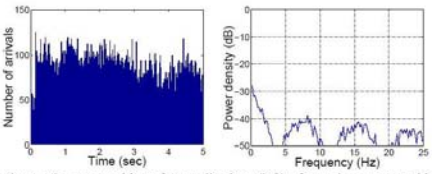


Fig. 4. The superposition of 16 realizations (left) of a packet process with Pareto inter-arrival times mixed with slightly perturbed deterministic arrivals and its associated PSD estimate (right): note that the PSD has lower peaks due to superposition and RTT variation, as well as a decaying envelope, compared with that in Fig. 3, due to long-range dependence

Periodicity - Non-Deterministic arrivals

- (Semi-)Deterministic arrivals
  - periods are drawn from a uniform distribution in 130±10% ms.
  - Pareto inter-arrival
- Resulting PSD
  - show periodicity

Network Simulations

- simulation to validate the idea of using spectral analysis to identify TCP traffic.
- simulation topology
  - binary tree of depth $d = 10$.
  - S_0 is the traffic sink, which sits behind a 100Mbps link, and other links are 1Gbps.
  - internal links $L_1 - L_{511}$ have a propagation delay of 10ms, leaf links $L_{512} - L_{1023}$ have propagation delays $\sim U[10, 20]$ ms, Resulting RTT $\sim U[200, 220]$ ms.
  - 750-packet Random Early Detection(RED) queue. threshold (125, 375), `gentle_bit` set.

Topology of the simulation

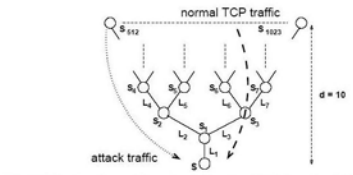


Fig. 5. The topology of the simulation, in which the attack flow is from the leftmost leaf node, whereas legitimate traffic comes from the rest of the leaf nodes

Network Simulations (continued)

- simulation of attacks
  - from node S_{513} to node S_0
  - constant bit rate UDP packet process with randomized inter-packet times and an average bit rate of 10Mbps
  - long FTP sessions between all other leaf nodes, S_{513} through S_{1023}, and the sink node. All packets contain 1000 bytes.
- Aim: to validate that large-volume TCP flows exhibit periodicity around RTT.

Network Simulations (continued)

- Real vs. Simulation
  - In the simulation, they deliberately make TCP flows operating in congestion avoidance phase without experiencing many retransmission timeouts (RTO).
  - In real traffic, however, some of the TCP flows could experience quite a number of RTOs from time to time, but such flows are unlikely to pose serious threats in terms of bandwidth usage

PSD est. for single TCP flow

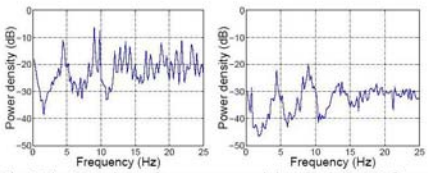


Fig. 6. The PSD estimates for aggregates consisting of a single TCP flow at S_{513} (left) and 128 TCP flows at S_8 (right): the height of the peak at fundamental frequency decrease as the level of aggregation increases

Network Simulations (continued)

- TCP packet processes show strong periodicity
  - peaks at frequency 4.7Hz, corresponding to 210ms
  - periodicity is preserved after aggregation. Power gets spread out as the degree of statistical multiplexing increases.

fundamental frequencies of estimates

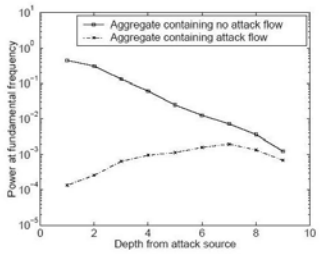


Fig. 7. The relative power under fundamental frequencies of PSD estimates for two groups of TCP aggregates; the one with UDP attack traffic has lower power at fundamental frequencies

Network Simulations (continued)

- The height of the first peak decreases as level of statistical multiplexing in an aggregate of TCP flows increases.
- Whether the aggregate has been contaminated by attack flows.

Validation using real traces

- data
  - May 6 and 7, 1999, at a 100Mbps link that connected the Harvard faculty of Arts and Sciences to the rest of the campus.
  - 66% of the packets are carried by flows that last longer than 40 seconds, suitable for spectral analysis
  - variation in RTTs from one trip to another is quite small: 88% of the flows have relative standard deviation of less than 50%.

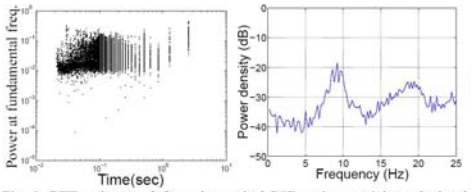


Fig. 8. RTT estimates (left) and a typical PSD estimate (right) calculated from the 1999 Harvard trace; it confirms that most of the long TCP flows exhibit periodicity, which can be seen in PSD

Result

- Use simple Heuristics to find the power at fundamental frequency
  - out of the highest five peaks from the PSD estimate, select the one with the lowest frequency and use it to estimate RTT
- For most of the TCP flows, the relative power at the fundamental frequency is above 5×10^{-3}. Use it to determine whether a flow is TCP.

Harvard trace result

	TCP flows	Non-TCP flows
Identified as periodic traffic	81.8%	15.7%
Identified as non-periodic traffic	18.2%	84.3%

Discussion

- Not a independent anomalies detection method
  - After we select candidate anomalies (connection, trace), the spectral analysis can help to say whether it is a normal TCP trace or not.
  - do not provide when the attack exists. Need to examine more.
- Attackers mimic the periodicity of normal TCP flows?
  - Consider return paths along with forward paths, similar periodicity
  - Use closed-loop protocols to launching attacks? attackers have to consume an amount of resources comparable to that of a normal TCP sender.

Discussion (continued)

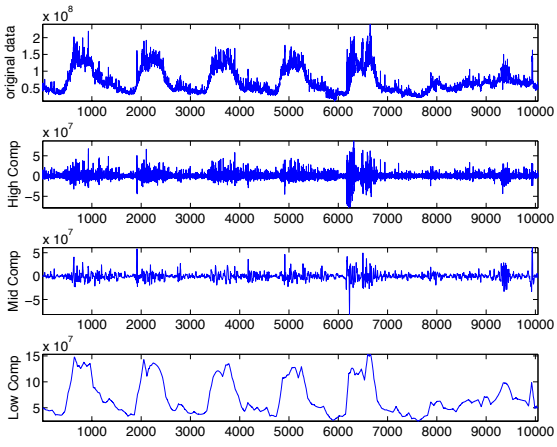
- Best deal with long TCP flows, For short TCP flows the effect of their statistical multiplexing may outweigh their intrinsic periodicity.
  - Fortunately, short TCP flows usually represent a small percentage of the total TCP load to a network in terms of packet counts

Wavelet analysis

- Motivation
- Outline
  - An example of wavelet analysis
  - Definition of wavelet transform and its properties
  - Wavelet methods for detection, results.
  - Discussion

An example

- Data
  - UNC network data, Monday June 23, 2003-Sunday July 27, 2003
  - One minute bin size, packet count data
  - This data contains some special features, might be anomalies
  - Will visit later using SVD method



UNC data

- low frequency component shows a clear daily pattern
- low frequency component shows that Sunday had an increasing (long term) network usage
- high frequency component shows that Friday had (short term) anomalies.

Good References

- Daubechies, I. "*Ten Lectures on Wavelets*"
- Mallat, S. "*A Wavelet tour of Signal Processing*"
- Percival, D.B. and Walden, A.T. "*Wavelet Methods for Time Series Analysis*"

The following introduction is mainly from Prof. Taqqu's lecture notes on "*Long Range Dependence*", 2003, SAMSI.

Wavelet Transform

- Wavelet $\psi(t)$

  A *wavelet* is a function $\psi(t)$, $t \in \mathbb{R}$, such that

  $$\int_{\mathbb{R}} \psi(t)dt = 0$$

  which satisfies some integrability conditions, for instance $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$

- N *zero moments* (also called *vanishing moments*)
  - Wavelet ψ is said to have N zero moments if

  $$\int_{\mathbb{R}} t^k \psi(t)dt = 0, k = 0, 1, \cdots, N-1 \quad (3)$$

Typical examples of wavelets

- Derivatives of the standard normal density

  $$\psi(t) = \frac{d^n}{dt^n}\left(\frac{1}{2\pi}e^{-\frac{t^2}{2}}\right)$$

- *Haar wavelets*

  $$\psi(t) = \begin{cases} 1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases}$$

- *Daubechies* wavelets
  - multiresolution analysis, a family of wavelets which is indexed by their number of vanishing moments; orthonormal wavelet basis.

Dilation and translation of wavelets

- The functions

  $$\psi_{j,k}(t) = \frac{1}{2^{j/2}}\psi(2^{-j}t - k) = 2^{-j/2}\psi(2^{-j}(t - 2^j k)), j \in \mathbb{Z}, k \in \mathbb{Z}$$

  are "dilations" and "translations" of ψ. The factors 2^j and j are called respectively the *scale* and the *octave*.

  The normalization factor $2^{j/2}$ ensures that for all $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$

  $$\int_{\mathbb{R}} \psi^2_{j,k}(t)dt = \int_{\mathbb{R}} \psi^2(t)dt$$

Discrete wavelet transform

- Using the function $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ as set of filters, we can now define the *discrete wavelet transform* DWT of a function (or of the sample path of a stochastic process) $\{X(t), t \in \mathbb{R}\}$ as

  $$d_{j,k} = \int_{\mathbb{R}} X(t)\psi_{j,k}(t)dt, j, k \in \mathbb{Z}$$

  The coefficients $d_{j,k}$ are called *wavelet coefficients* or *details*

Multiresolution analysis

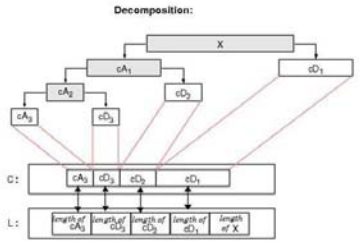
- Multiresolution wavelets
  - the wavelet ψ ("*mother wavelet*") is defined through a *scaling function*, ϕ.
  - both ϕ and ψ satisfy so-called two-scale equations

  $$\begin{aligned} \phi(t/2) &= \sqrt{2}\Sigma_n u_n \phi(t-n) \\ \psi(t/2) &= \sqrt{2}\Sigma_n v_n \phi(t-n) \end{aligned}$$

  - the approximation coefficients $a_{j,k}$ is defined as

  $$a_{j,k} = \int_{\mathbb{R}} X(t)\phi_{j,k}(t)dt, j \in \mathbb{Z}, k \in \mathbb{Z}$$

  where $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k)$.



Using wavelet to decompose signal

- set the details of larger scale as zeros, we can get high frequency representation of the original process. This will help to find the short-lived, small-scale variabilities
- set the details of smaller scale as zeros, we can get low frequency representation of the original process. This will help to find the long-lived, large-scale variabilities, periodicity etc.

A Signal Analysis of Network Traffic Anomalies

- data
  - two types of data, SNMP data and IP flow data
  - Main link of University of Wisconsin-Madison and outside world
  - five minute sampling interval
  - byte and packet counts for each direction of each wide-area link

A Signal Analysis of Network Traffic Anomalies (continued)

- Analysis Environments and Methods
  - Use wavelet to decompose the signal into three components: L(ow frequency)-part, M(id frequency)-part, and H(igh frequency)-part of the signal
  - Normalized the H- and M-parts to have variance one, compute local variability of the (normalize) H- and M- parts by computing the variance of the data falling within a moving window of specified size
  - experiments focuses on anomalies of duration 1-4 hours, and uses a moving 3-hour local deviation window

  - combining the local variability of the H-part and M-part of the signal using a weighted sum. The result is the V(ariable)-part of the signal
  - Apply threshold to the V-signal, we can find the anomalies if the V-signal exceeds the threshold.

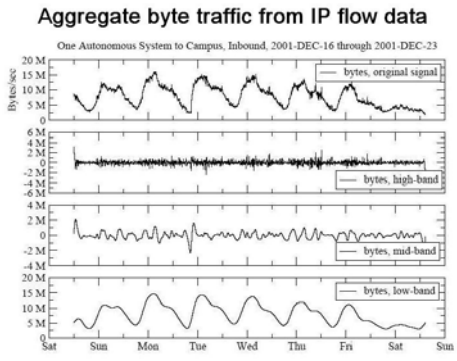


Fig. 1. Aggregate byte traffi c from IP flow data for a typical week plus high/mid/low decomposition.

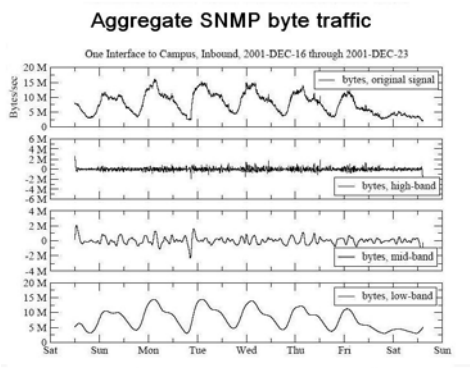


Fig. 2. Aggregate SNMP byte traffi c for the same week as Figure 1 plus high/mid/low decomposition.

Characteristics of Ambient Traffic

- Byte counts of inbound traffic, IP flow data
  - regular daily component of the signal is clear in the low band
- Byte traffic for the same week, SNMP data
  - nearly indistinguishable from the IP flow data

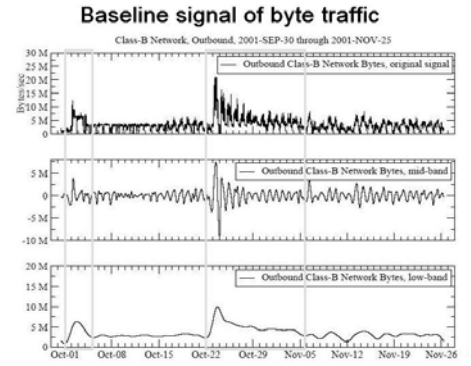


Fig. 3. Baseline signal of byte traffi c for a one week on either side of a flash crowd anomaly caused by a software release plus high/mid/low decomposition.

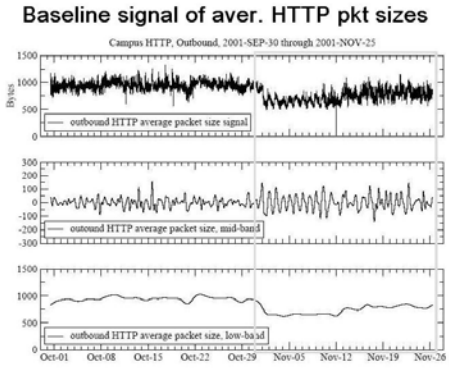


Fig. 4. Baseline signal of average HTTP packet sizes (bytes) for four weeks on either side of a flash crowd anomaly plus mid/low decomposition.

Characteristics of Flash Crowds

- flash crowds
  - long-lived features which should be exposed by the mid- and low-band filters
- outbound traffic (class-B network which contains an ftp mirror server)
  - increasing at the low-band signal
- HTTP byte data
  - *more stable* in the mid-band signal

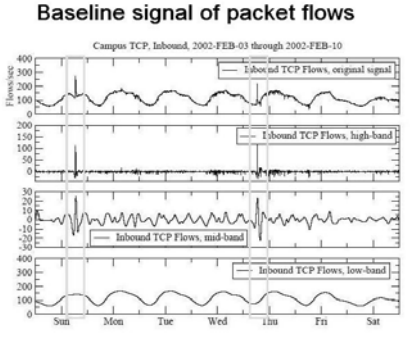


Fig. 5. Baseline signal of packet flows for a one week period highlighting two short-lived DoS attack anomalies plus high/mid/low decomposition.

Baseline signal of byte traffic from flow data

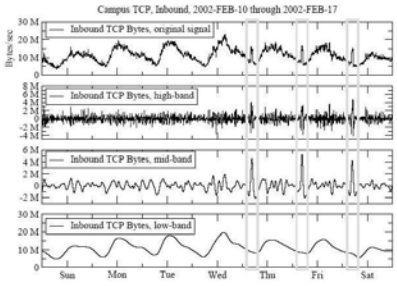


Fig. 6. Baseline signal of byte traffi c from fbw data for a one week period showing three short-lived measurement anomalies plus high/mid/low decomposition.

Characteristics of short-term Anomalies

- Short-term anomalies
  - attacks, network outages, measurement anomalies.
  - should be best exposed by mid-band and high-band filters which isolate short-timescale aspects of signals.
- two inbound DoS attacks (Figure 5)
  - floods of 40-byte TCP SYN packets destined for the same campus host
  - the flood packets had dynamic source addresses and TCP port numbers

Characteristics of short-term Anomalies

- Another type of short-term Anomalies (Figure 6)
  - periodic sequence of three measurement anomalies
  - a host in the outside world performing nightly backups to a campus backup server

Deviation analysis - Pkt count data

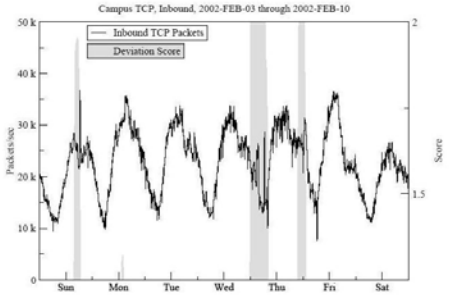


Fig. 7. Deviation analysis exposing two DoS attacks and one measurement anomaly in for a one week period in packet count data.

A Discriminator for short-term Anomalies

- Deviation Score
  - deviation scores of 2.0 or higher, high-confidence anomalies
  - below 1.25 as "low-confidence".
- Result of Figure 7
  - Two of the anomalies are DoS floods
  - The third band marks an measurement anomaly unrelated to the DoS attacks

Multi-day network abuse anomaly

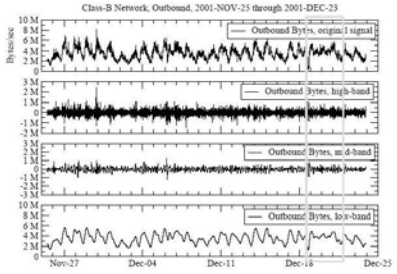


Fig. 10. Example of three-band analysis exposing a multi-day network abuse anomaly.

Hidden Anomalies

- able to identify a number of "hidden" anomalies in our data sets
- outbound traffic from one of the campus' class-B networks during a four week period
- network abuse
  - four campus hosts had their security compromised and were being remotely operated as peer-to-peer file servers

Deviation Score Evaluation

- Select a set of anomalies logged in the network operator journal as a baseline and evaluate deviation score detection capability.
- Comparing it with Holt-Winters Forecasting

Table II Comparison of Anomaly Detection Methods

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

Holt-Winters results

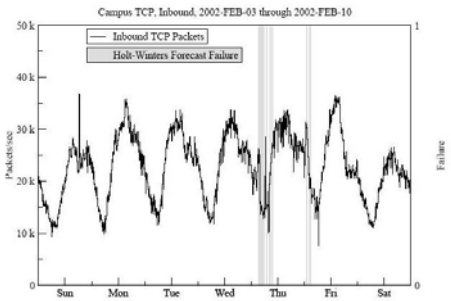


Fig. 11. Holt-Winters results corresponding to Figure 7

Deviation Score Evaluation (continued)

- Deviation Score vs. Logged Anomalies
  - 38 of them with significantly confident score of 1.7 or higher
  - Visual inspection of the plot of this signal showd that this was due to a more prominent anomaly which was detected earlier in the week
  - smaller window might help.
- Holt-Winters Forecasting vs. Logged Anomalies
  - Holt-Winters method for comparison because it is perhaps the most sophisticated technique that is being used currently by network operators.
  - larger window size will report the rest anomalies

Deviation Score Evaluation (continued)

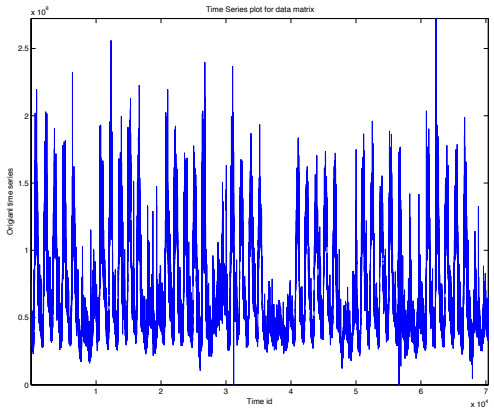
- Deviation Score vs. Holt-Winters Forecasting
  - Holt-Winters method is more sensitive to potential anomalies than deviation method, that induces more false-positives.
  - the deviation score technique tends to "blur" signal features by widening them with respect to time, making it less likely for a single anomaly to be erroneously reported multiple times.
  - deviation score method more readily reported small amplitude features in the signal than did the Holt-Winters method

Conclusions and Future Work

- Wavelet system effectively isolates both short and long-lived traffic anomalies.
- Deviation score is extremely effective at isolating anomalies
- To investigate machine learning methods to evaluate the impact of additional features in deviation scores.
- To investigate how well the deviation score method can be implemented to detect anomalies in real time.
- Intend to pursue the idea of coordinated anomaly detection at multiple measurements

Singular Value Decomposition

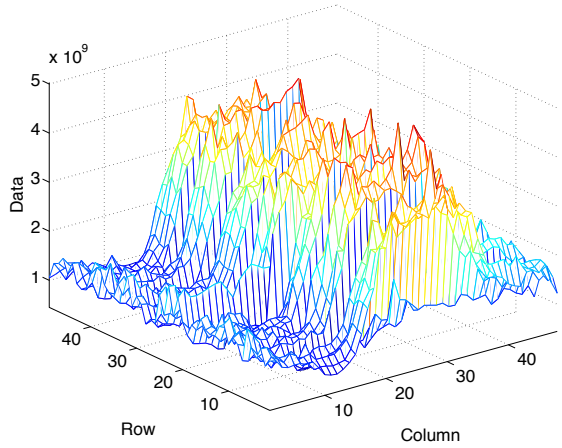
- This is my own research. Not intend to detect anomalies, but do provide some information related to that.
- Main Target: try to find patterns of Internet traffic trace.
- UNC campus data, Monday June 9, 2003 - Sunday July 27, 2003.
  49 days packet counts data.
  Bin size: 1 minute



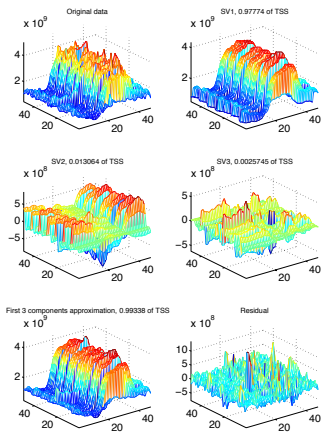
Time series plot for the 49 days data, 1 minutes bin size

Time Series Plot of the trace

- From the plot
  - Approximately 49 spikes, daily pattern?
  - 7 groups of spikes, weekly pattern?
  - weekday-weekend effect?
- How to treat the data?
  - Assume the daily pattern and form a data matrix instead of a time series
    The data matrix will be 49×1440 (Our data is 49×1436)



Mesh plots for half hour bin data, for better visualization.



Analysis Result

- Two clusters
  - Weekday-Weekend have different network usage
  - Weekday-weekend have different daily shapes
- Outliers
  - Outlying date (rows)
  - Outlying time (columns, or cell)

Definition of Singular Value Decomposition

Singular Value Decomposition of a data matrix $X = (X_{ij})_{m\times n}$ with $\text{rank}(X) = r$ is defined as

$$X = USV^T \quad (4)$$

$$= s_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + s_r\mathbf{u}_r\mathbf{v}_r^T \quad (5)$$

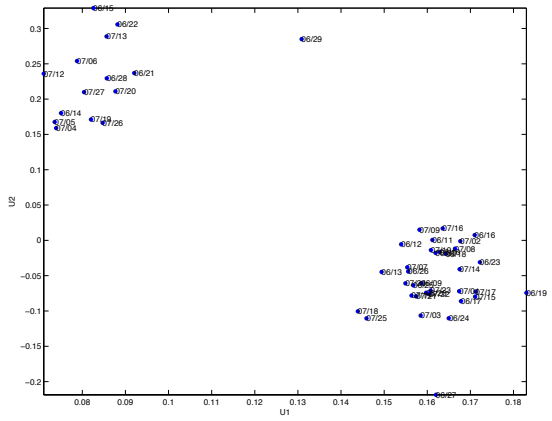
where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r)$, $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r)$, $S = \text{diag}\{s_1, s_2, \cdots, s_r\}$ with $s_1 \geq s_2 \geq \cdots \geq s_r$. $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ are called *singular columns* and *singular rows* respectively; $\{s_i\}$ are called *singular values*; and matrices $\{s_i\mathbf{u}_i\mathbf{v}_i^T\}(i = 1, \cdots, r)$ are referred to as *SVD components*.

Properties of Singular Columns and Rows

- Singular Columns ($\mathbf{u}$'s)
  - $\mathbf{u}_i$ forms orthonormal basis for columns spaces spanned by the columns of the data matrix X.
  - $\mathbf{u}_i$ also gives relative scores of original data X projects on the corresponding $\mathbf{v}_i$.
- Singular Rows ($\mathbf{v}$'s)
  - $\mathbf{v}_i$ forms orthonormal basis for rows spaces spanned by the rows of the data matrix X.
  - $\mathbf{v}_i$ also gives relative scores of original data X projects on the corresponding $\mathbf{u}_i$.

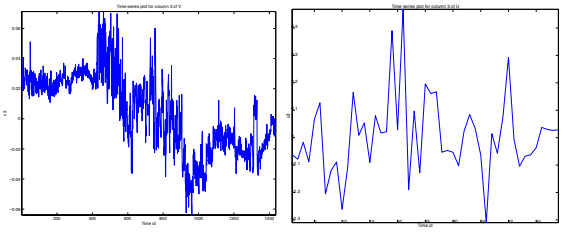
Scatterplots and TimeSeriesPlot of Singular Columns or Rows

- This gives the projection on a rotated plan, might help to detect outliers in the corresponding (column or row) space.
- Scatter plots of $\mathbf{u}$ help to detect outlying day
- Time Series plots of $\mathbf{v}$ help to find special pattern of specified time within a day.
- combining together will help to detect cell outliers.



Scatter plot between $\mathbf{u}_1$ and $\mathbf{u}_2$

- Two clusters, Weekday-weekend
- Friday July 4, is among the weekend data
- Sunday June 29, is isolated between the two clusters
  - The first Sunday of the second Summer Session
- Friday June 27, is a little isolated from the two clusters
  - The last registration day for second Summer Session



Time Series of $\mathbf{u}_3$ and $\mathbf{v}_3$

- $\mathbf{v}_3$
  - 420-600 of $\mathbf{v}_3$ is with higher variability.
- $\mathbf{u}_3$
  - Row 19, 21 is with large projection. Which indicates that Row 19 and Row 21 contribute a lot on $\mathbf{v}_3$.
- Last registration day

Multiple Time Series and Multivariate Methods

- Anomaly detection base on modeling and analysis of traffic on all links simultaneously. Jeff presented several papers already. This might induced Multiple Time Series Methods.
- Anomaly might not be detected from a single measurement, but possibly be detected from analysis of several measurements together.
- The correlation between time series, or measurements might mask some anomalies (outlier).

Measurements

- Measurements at single link
  - packet counts, bytes, etc.
  - simple statistics of network information, entropy of packet, etc.
- Measurement of whole network(?)
  - flow packet, bytes etc.

Multivariate Outlier Detection

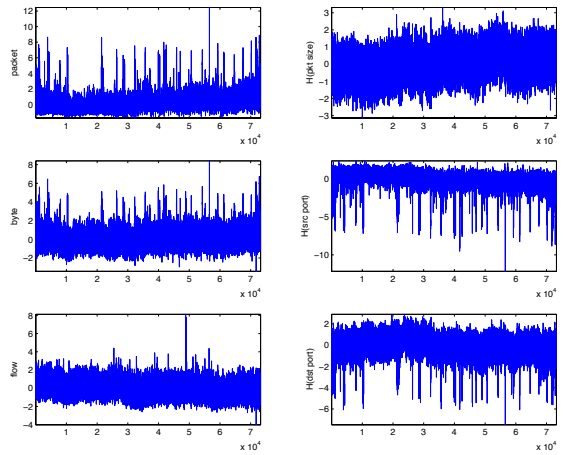
- Outlier in multivariate sense is much harder to detect.
- Univariate outlier detection method is not enough.
- Outlier might be masked for the multivariate correlation.
- Resent research topic in Statistics

An example

- To show one multivariate outlier detection method
- To illustrate the outlier might not be detected from single measurement outlier detection method
- Problem of the example: do not know the false positive and false negative rate

Idea of the detection

- The multivariate data forms point cloud in high-dimensional space. The point cloud might consist of one main cluster, and other points or small clusters out of it.
- Calculate the distance of each point to the center point of the main cluster, and define those with large value as (multivariate) outliers.

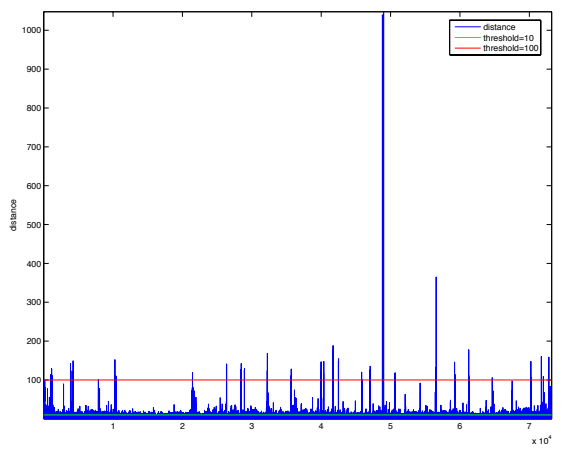


Multivariate Distance

- Assume we have $X = (x_{ij})_{n\times p}$, n observations with p variables. $\mu = (u_1, \cdots, u_p)$ is the center point (fox example, mean of each column or median of each column) in $\mathbb{R}^p$, Σ is the (robust) covariance matrix of the p variables, then the multivariate distance (d_i) of the ith observation will be

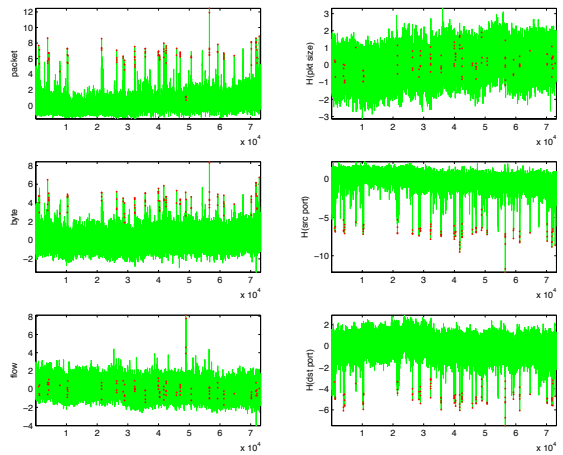
$$d_i^2 = [(x_{i1}, \cdots, x_{ip}) - \mu]\Sigma^{-1}[(x_{i1}, \cdots, x_{ip}) - \mu]^T$$

- Under multivariate normal distribution, if Σ is known, then d_i are approximately normal distribution. if Σ is unknown, then d_i are approximately t distribution.



Multivariate Distance (continued)

- Set threshold
  - We might use the critical value of t or normal distribution as the threshold, not discussed here.
  - set threshold as 10, there are 7871 (10.75%) observations greater than it.
  - set threshold as 100, there are 105 (0.14%) observations greater than it.
- Problem of this method
  - data are not necessarily gaussian
  - robust location and robust covariance matrix needed
  - if there does not exist the variance?



Results

- Most outlier appears in the spikes
- For some measurements, the outliers appear in the center.
- Multivariate analysis is appropriate, but need to go deep.
- if the location and covariance matrix exist and the robust estimation are provided, this method can be used to do online detection.

PCA(SVD) Method

- Another popular method to detect multivariate outliers
- Basically project the observations onto the principal direction, and find extreme values with respect to those projections.
- Statisticians also form some statistics of the former projections to detect outliers
- Flow data already talked by Jeff.

Part III – Further Work

Statistical Properties of Normal traffic and Anomalies

- Which measurements best fit for analysis
- Correlation between measurements
- Signals of existing anomalies

New Statistical Methods

- Anomalies might affect the normal traffic decomposition. robust method might help, decrease false negative.
- Statistical Theory under non-Gaussian, Long range dependence or heavy tailed sample

Multi-Scale Detection

- Anomalies vary from different scales. One scale statistical method is not appropriate?
- Multivariate Wavelet Method?
- Multi-Scale data?

Online Detection

- Current work mostly done at offline sense
- Online detection is essential
- Update algorithm every fix time interval
- Processing time should be short, response time should be quick, false negative and false positive should be as small as possible

References

[1] P. Barford, J. Kline, D. Plonka and A. Ron, (2002) A Signal Analysis of Network Traffic Anomalies, *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, Marseille, France, pp 71-82.

[2] Chen-Mou Cheng, H. T. Kung and Koan-Sin Tan, (2002) Use of Spectral Analysis in Defense Against DoS Attacks *Proceedings of IEEE GLOBECOM 2002*, Taipei, Taiwan.

[3] Laura Feinstein, Dan Schnackenberg, Ravindra Balupari and Darrell Kindred, (2003) Statistical Approaches to DDoS Attack Detection and Response, *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX03)*, Washington, DC.