Detecting Network Anomalies using Traffic Modeling

Anant Shah
Anomaly Detection

- Anomalies are deviations from established behavior
- In most cases anomalies are indications of problems
- The science of extracting deviations (and potentially uncovering problems) is called Anomaly Detection
Focus: Network Anomalies

- Network operators often face challenges such as DoS, Scans, Worms, Outages, etc.
- They manifest themselves as deviations in network traffic
- Sometimes these deviations are obvious to observe, sometimes they are hidden
Obvious Network Anomaly

Burma DDoS Attack

November 1 -2, 2010
All Times EDT

Gbps

12:00 PM 1:20 PM 2:40 PM 4:00 PM 5:20 PM 6:40 PM 8:00 PM 9:20 PM 10:40 PM 12:00 PM 1:20 AM 2:40 AM 4:00 AM 5:20 AM 6:40 AM 8:00 AM 9:20 AM 10:40 AM 12:00 AM

Monday Nov 1 Tuesday Nov 2

Credit: Arbor Networks

Detecting Network Anomalies using Traffic Modeling
Sometimes Network Anomalies are Hidden

- How many Anomalies can you point out?
- Let’s model the degree of randomness: Entropy of destination IP and Port

What factors contributed to the successful detection of this anomaly?
How was this Anomaly Detected?

- Count based metrics #bytes, #packets
- Simple plot of time series didn’t reveal the anomaly
- Data used is header fields destination IP and Port
- Scope of analysis is single incoming link
- Model of Entropy made the anomaly stand out
Network Anomaly Detection Dimensions

• Data
  • Network Traffic and Statistics
    – Bytes/sec, Packets/sec, Header fields, etc.

• Scope
  • Per Link (Temporal analysis)
  • All links at a time (Spatio-Temporal analysis)

• Model
  • Model network traffic characteristics
    – Exploit an intrinsic characteristic of data that the anomaly impacted
    – Frequency, Entropy, Principal Components, etc.
Interesting Questions to Answer

What data to use?
- Variety in available data is limited to data collection in operational networks

What scope to focus on?
- Choices range from one link to all the links in the network
  - How they function
  - Intuitive understanding of why they work
  - Use of compound models

Which modeling technique to use?
- Challenge to match model to anomaly
- Most interesting aspect of anomaly detection
Taxonomy

Data

Sketches, Sampling

PCA

Data Reduction

Correlation

Probability

Temporal Dynamics

Entropy

Frequency, Distribution

Detection

Detecting Network Anomalies using Traffic Modeling
Example

Roadmap of the method

Subspace Method by Lakhina et al:

- Detection
- Correlation
- Data Reduction
- Data

EWMA
Principal Components
Sampling
SNMP

Detecting Network Anomalies using Traffic Modeling
Single Link Methods
Feature Vector Method

- Detection
- Data Reduction
- Data

- Variance
- Sampling
- Count Metrics

{Variance, Count Metrics, Single Link}
Generating Envelopes

How to capture dynamic nature of traffic?
- Use envelopes

Detecting Network Anomalies using Traffic Modeling
Expected behavior as Envelopes

For each metric find which envelope the observed value falls and output corresponding integer

- Some other metrics: Number of broadcast packets, Load on adjacent network, percentage bandwidth used, etc
Matching Algorithm

• Anomaly Vectors
  – State of network

• Fault Feature Vector
  – Expected deviation of anomaly

• If distance is < 3 then there is a match

Distance:
(1,1) => 0
(1,1) => 0
(2,2) => 0
(0,0) => 0
(0,-1) => 1

Sum = 1
Anomaly Detected
Results

- Authors analyze their department LAN data
- Best detection for broadcast storm
- Success for hardware faults is low
  - Authors attribute this to lack of more metrics
  - Claim addition of more metrics may help detecting hardware faults
Comments: Feature Vector Method

• Generating a good feature vector for an anomaly is very difficult and there are large number of anomalies

• Will not work for anomalies that do not cause significant volume changes

• Predictions are based on previous days data

• Good for enterprise wide LAN but not a solution for ISPs
Signal Analysis Method

Detection

Temporal Dynamics

Data Reduction

Data

Frequency

Sampling

SNMP, Count Metrics

Deviation Score
Why Signal Analysis?

• Some anomalies are instantaneous with sudden changes in the count metrics

• Simple variance may not be able to capture these in given time window

• Need: Segregate instantaneous changes with slow varying trends
  - Analyzing time series as a signal can help
Time Series Decomposition

- Each band captures different traffic variations
- We can now focus on just high and mid band to find instantaneous deviations
Signal Analysis Method Overview

- Step 1: Sliding window method
- Step 2: Weighted average
- Step 3: Deviation score

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

Detecting Network Anomalies using Traffic Modeling
Results

- Wavelet method shows good results for sharp instantaneous anomalies
  - E.g., DoS, Link Outage

- Anomalies that impact a specific set of flows, could not be detected
  - E.g., Port Scan

<table>
<thead>
<tr>
<th>Anomaly type</th>
<th>Total</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoS attack</td>
<td>38</td>
<td>30</td>
</tr>
<tr>
<td>Port scan</td>
<td>198</td>
<td>2</td>
</tr>
<tr>
<td>Large file transfer</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Prefix outage</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Link outage</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Measurement gap</td>
<td>136</td>
<td>6</td>
</tr>
<tr>
<td>Unknown</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>395</td>
<td>56</td>
</tr>
</tbody>
</table>
Comments: Signal Analysis Method

• Large anomalies may mask other subsequent small anomalies

• Selecting appropriate deviation threshold is challenging, since it will be different for different networks

• Need: Method to exploit flow level characteristics
Equilibrium Method

Detection

Temporal Dynamics

Data Reduction

Data

Error Bounds

Fitting Distribution

Sampling

IP Flows

Detecting Network Anomalies using Traffic Modeling
Why Equilibrium Property?

A Short Timescale Uncorrelated Traffic Equilibrium: ASTUTE

• What we have so far:
  – Anomalies that impact traffic dispersion: Use Variance
  – Anomalies that are instantaneous: Use Frequency
  – These are just count metrics, how to find anomalies in flows?

• Need: A way to detect anomalies that impacts volume of flows w.r.t each other

• Observation: Over a short duration of time (<15mins) IP Flows are independent and stationary
  – Intuitively: Independent flows cancel each other out
ASTUTE Method Example

- CLT: for large |F|, the AAV* has a standard Gaussian distribution

A toy example:

\[ K' = 2 \]

3 flows

\[
\begin{bmatrix}
0 \\
+2 \\
-1
\end{bmatrix}
\]

\[
\hat{\delta} = 1/3 \\
\hat{\sigma}^2 = 7/3
\]

\[ K' = \frac{\hat{\delta}_i}{\hat{\sigma}_i} \sqrt{F} \]

AAV: K(F) ≈ 0.378

No Alarm

*AAV: Astute Assessment Value
ASTUTE v/s Wavelet Method

- Unlike wavelet method detection is based on a flow’s correlation with other flows
- Best performance for port scan and prefix outage
- Poor performance for DoS detection
**Single Link Methods Summary**

<table>
<thead>
<tr>
<th>Model</th>
<th>Target anomaly class</th>
<th>Best detected anomalies</th>
<th>(+)Pros</th>
<th>(-)Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance envelopes</td>
<td>Affecting dispersion</td>
<td>Broadcast Storm</td>
<td>Accounts for daily patterns</td>
<td>Needs significant volume changes</td>
</tr>
<tr>
<td>Frequency</td>
<td>Instantaneous</td>
<td>DoS</td>
<td>Segregation of highly variable data</td>
<td>Small anomalies can be masked</td>
</tr>
<tr>
<td>Equilibrium</td>
<td>Affecting flow correlations</td>
<td>Prefix outage</td>
<td>No a priori knowledge needed</td>
<td>Keeping per flow state not feasible</td>
</tr>
</tbody>
</table>

- **Use Variance envelopes of multiple parameters**
- **Use Equilibrium property**
  1. How to detect anomalies that span multiple links?
  2. Can we separate out anomalous space?

**Need better ways to capture instantaneous spikes**
Network Wide Methods
Subspace (PCA) Method

Detection

Correlation

Data Reduction

Data

EWMA

PCA

Sampling

SNMP

Detecting Network Anomalies using Traffic Modeling
Why Subspace Method?

• Data from multiple links

• Allows analysis of variance across all links

• Extracts principal components in decreasing order of variance, which allows you to define subspaces
Basic Understanding of PCA

- PC-1 captured most variance
- If there were $n$ dimensions we would have $n$ PCs
  - In decreasing order of variance captured
Dimensionality Reduction: PCA

- Most variance is captured by first 3 or 4 PC
- Separate into normal and anomalous subspace
Detecting Anomalies in Subspace

First few components represent Normal Subspace

Later components represent Anomalous Subspace

Detecting Network Anomalies using Traffic Modeling
Subspace Entropy Compound Method

Detection
- EWMA

Correlation
- PCA

Probability
- Entropy

Data Reduction
- OD Flows

Data
- IP Flows

Detecting Network Anomalies using Traffic Modeling
Origin Destination Flows

- A relation between links and

\[ A_{ij} \] where flow \( j \) traverses link \( i \)

Detecting Network Anomalies using Traffic Modeling
Multi-way Subspace Method

- For all OD Flows calculate entropy of header fields
- Tile all 4 matrices and apply subspace method
  - This reduces dimensionality
- Analyze residual vectors for anomalies
Results

Detected by Subspace method
Input: Byte count

Detected by Entropy+Subspace method
Input: Entropy of OD Flows
(Aggregated IP Flows)

<table>
<thead>
<tr>
<th>Network</th>
<th># Found in Volume Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Géant</td>
<td>464</td>
</tr>
<tr>
<td>Abilene</td>
<td>152</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anomaly Label</th>
<th># Found in Volume</th>
<th># Additional in Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha Flows</td>
<td>84</td>
<td>137</td>
</tr>
<tr>
<td>DOS</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Flash Crowd</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Port Scan</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Network Scan</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Outage Events</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Point to Multipoint</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Unknown</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>False Alarm</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>152</td>
<td>292</td>
</tr>
</tbody>
</table>
Comments: Multi-way Subspace

• Feature metrics capture different anomalies than volume ones

• Manual inspection is needed to derive anomalous IP flows from OD flows

• PCA is computationally heavy, need reduce data before using subspace
  – Solution: Use Sketches!
Subspace Entropy Compound Method

Detecting Network Anomalies using Traffic Modeling
What is a Sketch, Why use it?

- Keeping per flow state is not feasible
- A method to detect anomalies in massive data sets is needed
  - Should work below any general detection method
- Sketch is a projection of IP Flows on a smaller space
- Sketch preserves the characteristics of original data
Sketches + Entropy + Subspace

• Build multiple sketches of feature entropies

• Applying this compound model:
  1. Compute Local Sketches of entropies at each router for 4 header fields
  2. Compute Global Sketches and find entropies
     • Combine entries belonging to same hash
     • Intuitively: We have aggregated IP flows at all routers that hashed to same value
  3. Detect anomalies by subspace analysis
Results

- This method detects different anomalies than detected by just Entropy method

**Detected by Entropy+Subspace method**
Input: Entropy of OD Flows
(Aggregated IP Flows)

**Detected by Sketches + Entropy+ method**
Input: IP Flow Sketches
Network Wide Methods Summary

<table>
<thead>
<tr>
<th>Subspace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not find anomalous flows</td>
</tr>
<tr>
<td>Use OD Flows with Entropy Model</td>
</tr>
<tr>
<td>Need IP Flow level details with reduction in search space</td>
</tr>
<tr>
<td>Use Sketches before applying PCA and Entropy</td>
</tr>
</tbody>
</table>

Detecting Network Anomalies using Traffic Modeling
## Network Wide Methods Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Target anomaly class</th>
<th>Best detected anomalies</th>
<th>(+)Pros</th>
<th>(-)Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subspace</td>
<td>Network wide volume changes</td>
<td>Alpha flows</td>
<td>Exploit correlations among network links</td>
<td>Does not drill down to anomalous flows</td>
</tr>
<tr>
<td>OD flows + Entropy + Subspace</td>
<td>Behavioral distribution changes</td>
<td>Port/Network Scans</td>
<td>Detect anomalous flows</td>
<td>Needs topology information</td>
</tr>
<tr>
<td>Sketches + Entropy + Subspace</td>
<td>Long lasting behavioral changes in IP flows</td>
<td>Port/Network Scans</td>
<td>Detect anomalous flows with less false positives</td>
<td>Sensitive to choice of sketch keys</td>
</tr>
</tbody>
</table>
General Limitations

• Is sampled data sufficient?
  – Sampling reduces the effectiveness of modeling methods
  – For change in 10% sampling rate more than half anomalies were not detected

• Common assumptions may not be valid
  – Availability of clean data
  – False positives are okay
  – Synthetic anomalies are representative
  – Interpretation of anomalies is left up to the operator
Conclusion

- The problem space comprises of 3 dimensions
  - Data, Scope and Model

- Each method can be broken down in multiple layers

- Multiple layers can used together to achieve multiple goals such as:
  - Targeting different set of anomalies
  - Reduction in dimensionality
  - Achieve better granularity
Thank You

Any Questions?
Backup Slides
Anomaly Detection Basic Example

- Use Exponential moving average (EMA) to model
- Anomaly if observed value is more/less than 40% of expected value
Why is Anomaly Detection Important

- Detect attacks and outages
- Maintenance of networks

Model network traffic over a given period of time
## Signature Based v/s Anomaly Based Detection

<table>
<thead>
<tr>
<th>Signature Based</th>
<th>Anomaly Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known signatures</td>
<td>No signature required</td>
</tr>
<tr>
<td>Difficult to detect unknown anomalies</td>
<td>Can detect unknown anomalies</td>
</tr>
<tr>
<td>No further identification is required</td>
<td>Anomaly identification or classification is needed</td>
</tr>
<tr>
<td>No data modeling is needed</td>
<td>Data modeling is the prime step</td>
</tr>
<tr>
<td>Difficult to understand network behavior</td>
<td>Data modeling gives idea about network behavior</td>
</tr>
<tr>
<td>Works great for attacks where signature is known</td>
<td>Works for attacks, outages, traffic engineering</td>
</tr>
</tbody>
</table>
Types of Network Anomalies

- **Attack**
  - DoS
  - Port Scan

- **Abnormal usage**
  - Flash Crowd
  - P2P Node

- **Measurement error**
  - Buggy – Software Config Error

- **Network Change**
  - New ingress and egress routers
Anomaly Impact

- Volume
  - DoS, Flash Crowd, Alpha Flows, Outages

- Traffic Profile
  - Metric Counts
  - Interactive Feature
    - Correlations
      - Scans, Worms
    - Probabilistic
      - Prefix outage, Measurement gap
    - Distribution
All Layers Together

<table>
<thead>
<tr>
<th>Detection</th>
<th>Deviation Score, Holt-Winters, Variance, Error Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Dynamics</td>
<td>Frequency, Distribution</td>
</tr>
<tr>
<td>Correlations</td>
<td>PCA, Kalman Filter</td>
</tr>
<tr>
<td>Probability</td>
<td>Entropy, Mutual Information</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>Sketches, OD Flows, Sampling</td>
</tr>
<tr>
<td>Data</td>
<td>SNMP, Count Metrics, IP Flows</td>
</tr>
</tbody>
</table>
Feature Vector Method Overview

- **Step 1:** Track variance of different metrics compared to their variance the day before

- **Step 2:** Determine network state as a vector representing magnitude of above metrics called anomaly vector

- **Step 3:** Network operator predefines anomaly signature called fault feature vector

- **Step 4:** To flag an anomaly compare anomaly vector to fault feature vector
ASTUTE Overview

• Step 1: For each flow present in consecutive time bins keep track of flow volume

• Step 2: Calculate ASTUTE assessment value (AAV) using change in flow volumes

• Step 3: Flag anomaly if AAV falls outside the error bounds
Assessment Value

• For each flow measure its volume change between two time bins

• Compute assessment value

\[
\hat{\delta}_i = \sum_{f=1}^{F} \frac{\delta_{f,i}}{F} \quad \therefore \quad \hat{\sigma}_i = \left[ \sum_{f=1}^{F} \frac{(\delta_{f,i} - \hat{\delta}_i)^2}{F - 1} \right]^{\frac{1}{2}}
\]

\[
K' = \frac{\hat{\delta}_i}{\hat{\sigma}_i} \sqrt{F}
\]
Comments: ASTUTE Method

• Immune to large volume changes in one flow

• Time bins should be less than 15 minutes
  – Or the assumptions do not hold

• Not a replacement for wavelet method but should be used in complement with it

• Keeping per flow state is difficult
Sketch Method

Data
Scope
Model

{Aggregation, IP Flows, Single Link}

Detection
Data Reduction
Data

Holt-Winters
Sketches
IP Flows

Detecting Network Anomalies using Traffic Modeling
Sketches give similar detection

- Forecasting models can be used on top of sketches
- Similarity of detection is more than 90%
Comments: Sketch Method

• A very promising way to reduce search space
  – Without compromising data characteristics

• But does not separate normal and anomalous space
Comments: Subspace Method

• Good method to reduce dimensionality

• Detection is based on relationship between multiple links
  – Provides a new correlative knowledge of the network

• Works on link counts and then derives underlying anomalous flow
  – Does not directly find anomalous flows
Non-header metrics for Entropy

- Some feature metrics are highly correlated
- Multiway subspace method uses four header fields and then apply PCA to reduce dimensionality
  - PCA is computationally heavy
- There are 3 new possible metrics:
  - Flow size distribution
  - In Degree (Ratio of times an IP address appeared as destination)
  - Out Degree (Ratio of times an IP address appeared as source)

<table>
<thead>
<tr>
<th></th>
<th>Out Deg</th>
<th>Src Addr</th>
<th>Dst Addr</th>
<th>Src Port</th>
<th>Dst Port</th>
<th>FSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>InDeg</td>
<td>0.102</td>
<td>0.100</td>
<td>0.097</td>
<td>0.000</td>
<td>0.007</td>
<td>0.414</td>
</tr>
<tr>
<td>OutDeg</td>
<td>-</td>
<td>-0.034</td>
<td>-0.033</td>
<td>-0.054</td>
<td>-0.015</td>
<td>-0.018</td>
</tr>
<tr>
<td>SrcAddr</td>
<td>-</td>
<td>-</td>
<td>0.994</td>
<td>0.962</td>
<td>0.956</td>
<td>0.307</td>
</tr>
<tr>
<td>DstAddr</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.966</td>
<td>0.969</td>
<td>0.286</td>
</tr>
<tr>
<td>SrcPort</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.989</td>
<td>0.171</td>
</tr>
<tr>
<td>DstPort</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.181</td>
</tr>
</tbody>
</table>

Correlation among different feature metrics
Kalman Filter Method

Data
Scope
Model

Detection
Variance
Correlation
Kalman Filter
Data Reduction
Sampling
Data
IP Flows

Detecting Network Anomalies using Traffic Modeling
Why Kalman Filter?

• Kalman Filter is a way of predicting system state using current state
  – Difference in estimated and observed value is error

• Better than PCA when data is noisy

• Can be used for both multiple and single link approaches
Applying Kalman Filter

• Step 1: Aggregate data into OD Flows

• Steps 2: Generate matrix of OD Flows v/s byte counts

• Step 3.1: Prediction:
  – Predict expected value of OD flows
  – Predict variance of the above estimation

• Step 3.2: Estimation:
  – Using estimated value, estimated variance and observed value calculate correction

• Step 4: Keep track of correction values, if a large correction is required, flag anomaly
Results

- Kalman filter performs very similar to Wavelet method
- For better detection Kalman filter should be used in combination with other methods like ASTUTE
Network Wide Methods Summary

Subspace
- Does not find anomalous flows
  - Use OD Flows with Entropy Model
    - Need IP Flow level details with reduction in search space

Use Sketches before applying PCA and Entropy
- Which header fields to use for Entropy
  - Use FSD with Entropy
    - Need a predictive approach which is similar to PCA
      - Kalman Filter
# Conclusive Summary

<table>
<thead>
<tr>
<th>Single link</th>
<th>Volume metric</th>
<th>Flow metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Feature Vector Method (Variance)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Wavelets (Frequency)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi link</td>
<td>• Subspace (Principal Components)</td>
<td></td>
</tr>
<tr>
<td>• Multi-way Subspace (Entropy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Kalman Filter (Predictive)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Equilibrium (Independence)
- Sketches (Aggregation)