Statistical Traffic Measurement
Denial Of Service & Accounting

- DOS attacks consume resources.
- Looked at many approaches to account for resources.
- Use every available piece of information for anomaly detection
  - Volume of traffic
  - Number of flows
  - Address distribution
  - Port numbers
  - Protocol distribution
Measurement

• How do you detect an anomaly?
• What is anomalous behavior?
• Use statistics and past distributions
  – So far, we looked at static thresholds
  – Flows above 1% of link bandwidth
  – Top 100 flows etc.
• How to get data for statistical measurements?
FlowScan
False Alarms & False negatives

• Accurate anomaly detector requires
  - Few false alarms
  - Few false negatives

• Setting of thresholds requires careful analysis of data

• Today, a few case studies and a platform for statistics collection
General Problem

• Collect information about a signal over the signal space each sample
  - Address space of $2^{32}$
  - Port numbers of $2^{16}$
• Collect information over many samples
• Does current sample look different?
• How to collect, analyze and conclude?
• Real-time versus Post-mortem
Distribution of addresses?
Destination Addresses

- Address space is large \((2^{32})\)
- Address space is discrete
- Hard to analyze
- A few things make it feasible
  - Our habits - access same sites again
  - Popular sites - get lots of hits
- On an aggregate - likely to have high correlation over time
Destination addresses

- Instead of looking at individual addresses, look at a summary signal
- Each sample, keep track of individual addresses to generate a summary
- Keep track of summary values over several samples
- See if current sample is anomalous
- Data reduction approach
Destination addresses

• Use a simpler data structure
• Count each byte of IP address
  - 4 arrays of 256 locations
  - Count frequency of incidence
• Volume of traffic fluctuates over time
• Use Relative frequency
  - Normalize with total packet count
Address correlation signal

- Generate a correlation signal by multiplying two successive samples
  - \( a_{in} \times a_{i(n-1)} \) over all addresses \( i \)

- Hypothesis: at the time of attack
  - Correlation will increase if few victims
  - Correlation will decrease if random victims (as in worms)
Address correlation

• Data requirements: 1024 words per sample, n past summary values
• Computation requirements?
Analysis of address correlation

• What timescales should we consider?
• Sample to sample, 2/4/... samples?
• Wavelets allow analysis at multiple timescales
• Look at differences at multiple timescales
Real-time analysis

Real-time Detector in Auckland IV traces, 1m sampling period, 2-hour DWT window [4201:4320], -3.0σ < X < 3.0σ

Signal Set

(cD1) Time range: 0 hours 2 minutes
(cD2) Time range: 0 hours 4 minutes
(cD3) Time range: 0 hours 8 minutes
(cD4) Time range: 0 hours 16 minutes
(cD5) Time range: 0 hours 32 minutes
(cD6) Time range: 1 hour 4 minutes
(cD7) Time range: 2 hours 8 minutes

Real-time Indication

Detect

sampling points
Fig. 6. A Multilevel two-band wavelet decomposition and reconstruction
Fig. 2. The block diagram of our detector.

- Network Traffic
- Signal Generation & Data Filtering
- Statistical Analysis & Wavelet Transform
- Anomaly Detection & Thresholding
Data Collection
Address Distribution
Address Distribution
Addresses as Images

• We can employ many image processing techniques

• DCT - Discrete Cosine Transform
  - Few coefficients are normally enough
  - Allows data to be compressed

• Delta encoding of two samples
  - Reduce the data requirement
Visualization

- Allows visualization tools to be built
- Allows playback of events
- Work in progress
Address data structure

- Compact
- Collisions
- Provides a first hint of where to focus on
  - Identification requires more work
Summary of today’s class

• Many measurement techniques
• Signal/image/time-series based
• Different data structures
• Wavelets, statistical techniques
• Real-time vs. post-mortem
• General approaches feasible
• Need more work
Port Numbers

Signal of PORT Numbers Correlation in 3-day Auckland-IV traces, 1m sampling period

Postmortem Detector, 3-day DWT window [1 : 4320], 20-minute DET window [4301 : 4320], -3.0σ < X < 3.0σ
How do we do this in general?

• Different signals, different analyses, different thresholding techniques
• Can we design a building block to enable different detectors?
• Time series analysis
• What if we wanted to identify anomalous flows?
• Different statistical measures?
Sketches

- Employ a HxK matrix
- H is the number of hash functions
- K is the range of each hash function
- Similar to earlier parallel hashing
  - Used for identifying resource hogs
- Internet Measurement Conf. 2003
  - Krishnamurthy et al
Operations on Sketches

\textbf{UPDATE}(S, a, u): For } \forall i \in [H], T_S[i][h_i(a)] + = u.

2. \textbf{ESTIMATE}(S, a): Let } \text{sum}(S) = \sum_{j \in [K]} T_S[0][j] \text{ be the sum of all values in the sketch, which only needs to be computed once before any } \text{ESTIMATE}(S, a) \text{ is called. Return an estimate of } v_a

\[ v_a^{est} = \text{median}_{i \in [H]} \{ v_a^{h_i} \} \]

where

\[ v_a^{h_i} = \frac{T[i][h_i(a)] - \text{sum}(S)/K}{1 - 1/K} \]
Operations on Sketches

3. \textsc{EstimateF2}(S): Return an estimate of the second moment

\[ F_2^{\text{est}} = \text{median}_{i \in [H]} \{ F_2^{h_i} \} \]

where

\[ F_2^{h_i} = \frac{K}{K - 1} \sum_{j \in [K]} (Ts[i][j])^2 - \frac{1}{K - 1} (\text{sum}(S))^2 \]
Operations on Sketches

4. \textsc{Combine}(c_1, S_1, \cdots, c_\ell, S_\ell): The linearity of the sketch data structure allows us to linearly combine multiple sketches

\[ S = \sum_{k=1}^{\ell} c_k \cdot S_k \]

by combining every entry in the table:

\[ T_S[i][j] = \sum_{k=1}^{\ell} c_k \cdot T_{S_k}[i][j] \]
Basic modules

• Sketch module - keep track of current data
• Forecast module - forecast what current data should be based on past samples
• Change detection module - compare error between forecast and observed values - see if above some threshold
Forecast bases

Moving Average (MA)  This forecasting model assigns equal weights to all past samples, and has a single integer parameter $W \geq 1$ which specifies the number of past time intervals used for computing the forecast for time $t$.

$$S_f(t) = \frac{\sum_{i=1}^{W} S_f(t-i)}{W}, \quad W \geq 1$$

S-shaped Moving Average (SMA)  This is a class of weighted moving average models that give higher weights to more recent samples.

$$S_f(t) = \frac{\sum_{i=1}^{W} w_i \cdot S_f(t-i)}{\sum_{i=1}^{W} w_i}, \quad W \geq 1$$
Forecast bases

- Exponentially weighted moving average
- ARIMA models
- Non-seasonal Holt-winters

\[
S_s(t) = \begin{cases} 
\alpha \cdot S_o(t-1) + (1 - \alpha) \cdot S_f(t-1), & t > 2 \\
S_o(1), & t = 2 
\end{cases}
\]

\[
S_t(t) = \begin{cases} 
\beta \cdot (S_s(t) - S_s(t-1)) + (1 - \beta) \cdot S_t(t-1), & t > 2 \\
S_o(2) - S_o(1), & t = 2 
\end{cases}
\]

The forecast is then \( S_f(t) = S_s(t) + S_t(t) \).
Sketches validation

• Looked at traces, post-mortem
• To reduce false alarms, $H = 5, K = 8K$
• Conclusion: Per-flow state not needed
• Possible to implement many change detectors on different forecasts
• Scope for data structure research!!
• Only allows linear estimators
More on sketches

- Not sure if feasible in real-time
- Special hash functions allow lookup based computation – interesting read
- $5 \times 8K \times \text{number of samples} = \text{getting large}!!$
- Delta encoding across sketches?
- Any other architectures?