

Traffic Measurement and Accounting

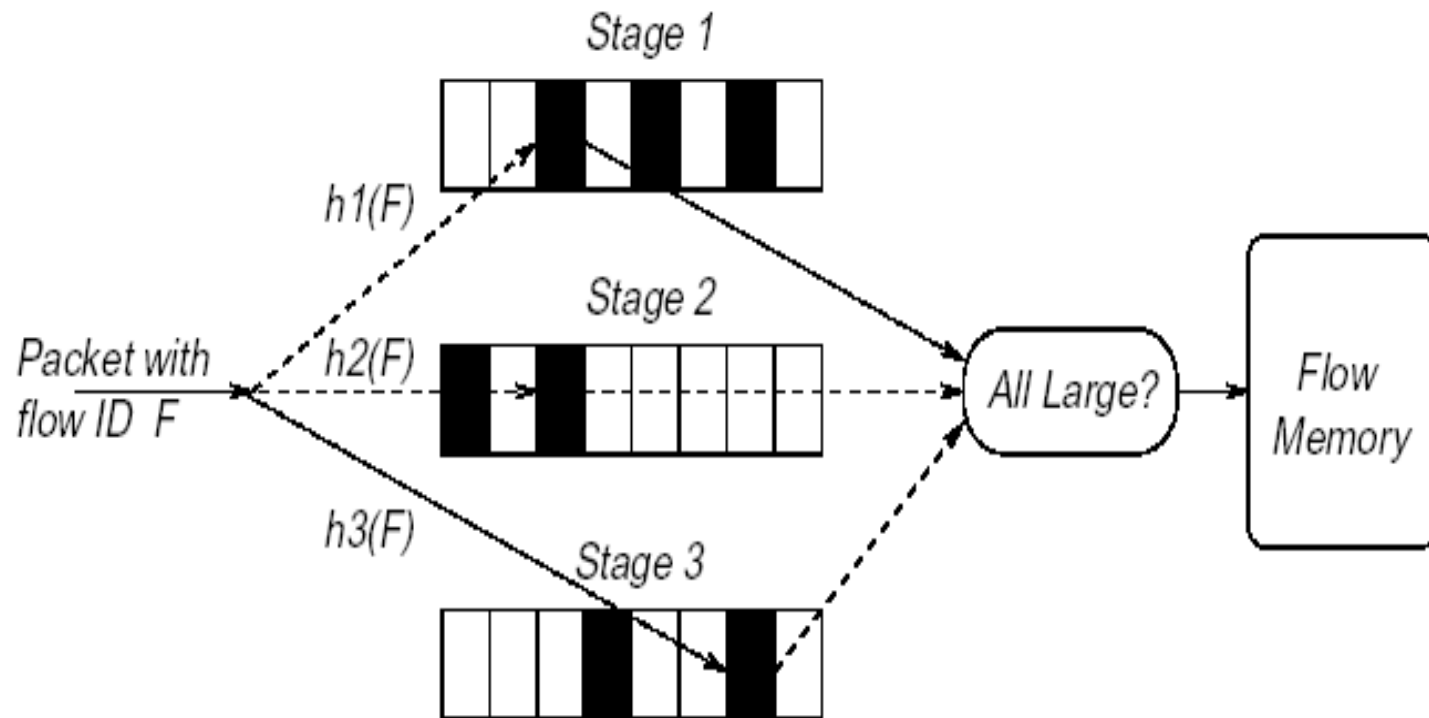
Denial Of Service & Accounting

- DOS attacks consume resources.
- If we can account for resource consumption, we can detect and identify attacks.
- Partial state -maintain limited amount of state -to track a few large flows
 - Rely on sampling and caching
- Today - more on accounting

Partial State

- Employs limited amount of state for fast processing
- Larger the SRAM, the larger number of flows monitored
- Other techniques?

Multistage Filters -- Identifying resource hogs



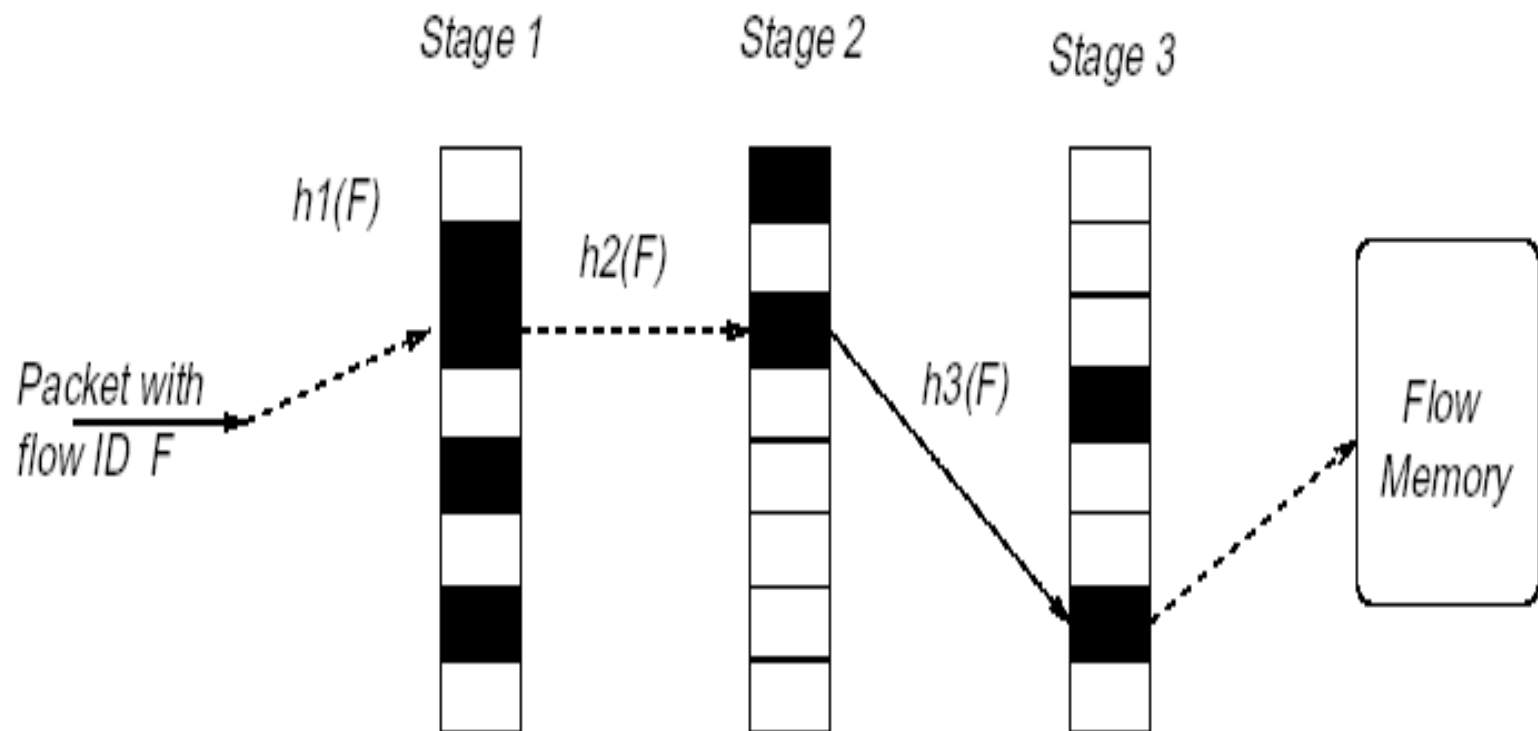
Multistage Filters

- Hash a flow into a limited number of bins
- Update the counter in that bin
- Declare a potential resource hog if the count exceeds a threshold?
- Hashing collisions, large number of flows, limited amount of memory

Multistage Filters

- Use multiple, independent hash functions
- Declare a flow resource hog only if threshold exceeded in all the bins
- Reduces the probability of false identification
- Can do this in parallel or serially

Serial multistage filter



Cisco's NetFlow

- Cisco's routers provide accounting
- Keep track of Flow's packet headers along with time and other information
- Takes a lot of memory
- Records stored in DRAM
- Sampling used at high speeds
- Records dumped to disk after a while

NetFlow

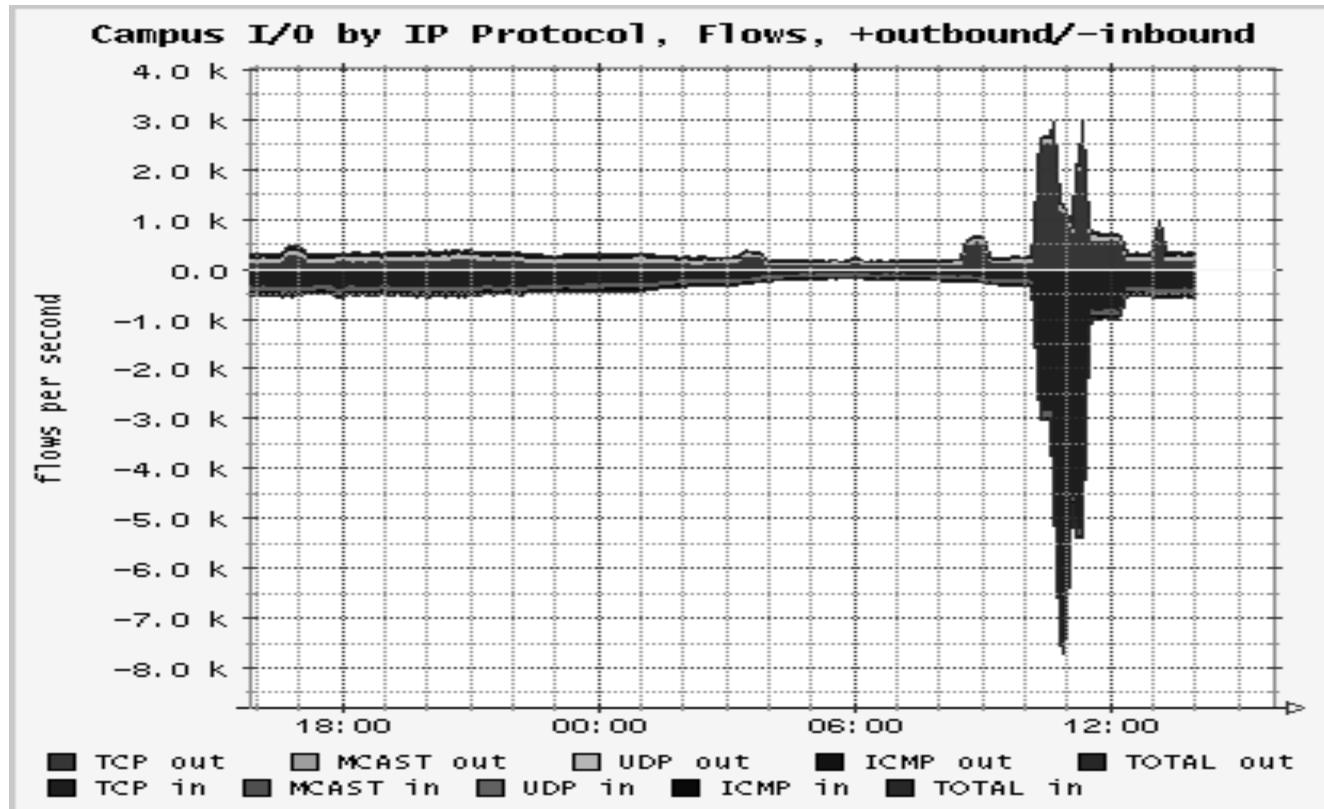
- Useful for post-mortem analysis
- Use time fields to group data into samples or buckets
- Can look at volume, number of flows, the number of bytes a flow sent etc.
- Provides raw data, needs to be analyzed

FlowScan

- Uses NetFlow data
- Counts the number of flows in a sample
- Large number of flows
 - Could be a signal of an attack

<http://www.caida.org/tools/utilities/flowscan/>

FlowScan



Counting Flows

- Look at packets with SYN
 - Don't need to look at all packets?
- Can count TCP flows with little overhead
- Does not take care of other flows
- Current traffic is 80-85% TCP
- Gives a reasonable estimate

tcpdump

- Can use tcpdump tool to look at all packets leaving a campus
- Again, memory intensive, slow
- Data needs to be analyzed

Snort

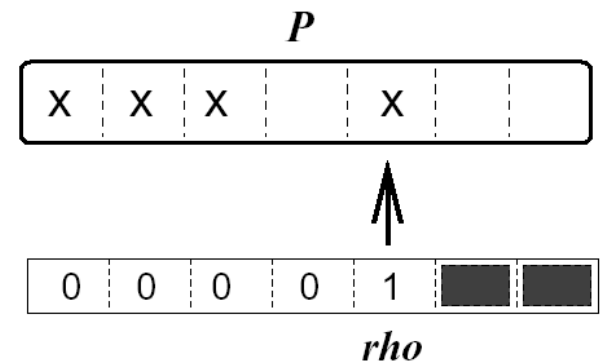
- Similar to tcpdump in data collection
- Provides a number of tools for analysis
- Specifically, used for testing “portscans”
 - Is anyone trying to access machines at every port number
 - Sign of an attacker looking for vulnerabilities?

Flow estimation

- Can we design cheaper estimation techniques than full counting?
- Probabilistic counting
 - Use a hash function that distributes the flow ids uniformly over n bits
 - Mark the LSB of sampled $\text{hash}(\text{flowid})$
 - At the end of sample, look at unmarked LSB –this gives an estimate of flows

Probabilistic counting

- <http://citeseer.nj.nec.com/flajolet85probabilistic.html>
- Estimate number of flows = $1.3 * 2^A$
- Intuition: hashing causes LSB = 1 with $p=0.5$, nextbit=1 with 0.25 etc..
- In other words, $p = 2^{-(k+1)}$ for a hash value with k zeros in LSBs



What about aggregates?

- Individual flows only tell so much
- A web server may advertise multiple addresses
- Traffic directed to all these addresses would attack a single site
- Will appear as multiple flows
- Flow accounting may not help enough

Aggregate accounting

- Could do this with partial state for known-length aggregates
 - Use destination address alone
 - Use prefix matching instead of exact matches
 - Look at top 24 bits for /24 prefixes...

Aggregate accounting

- RED-PD: look at drop history
- Merge the destination addresses into larger prefixes
- Requires off-line analysis
- Works only with RED routers
- Other algorithms: Estan, Varghese
 - Offline analysis of Netflow stats
 - Allows arbitrary-length prefixes

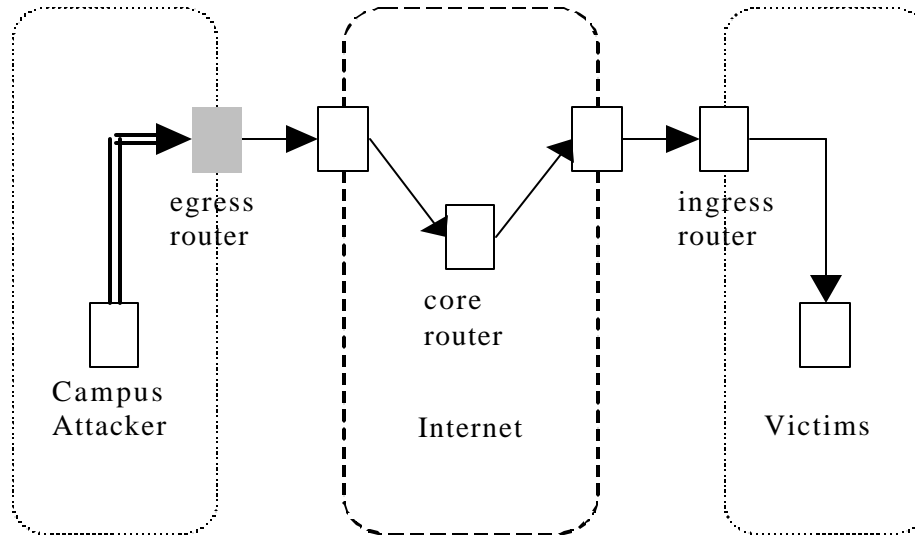
Destination vs. Source

- Source addresses are easy to spoof
- Source addresses are easy to filter at network egress
 - TAMU has two class B addresses
 - Packets leaving TAMU campus must have these as source addresses
 - Anything else is spoofed
 - Can still spoof within campus addresses

What else can we account?

- Volume of traffic
 - Fluctuates over a day over a campus
 - High resource hogs interesting at any given time
- Number of flows
 - Could give an indication of an attack
 - Flashcrowds (Recent SuperBowl incident)

Distribution of addresses?



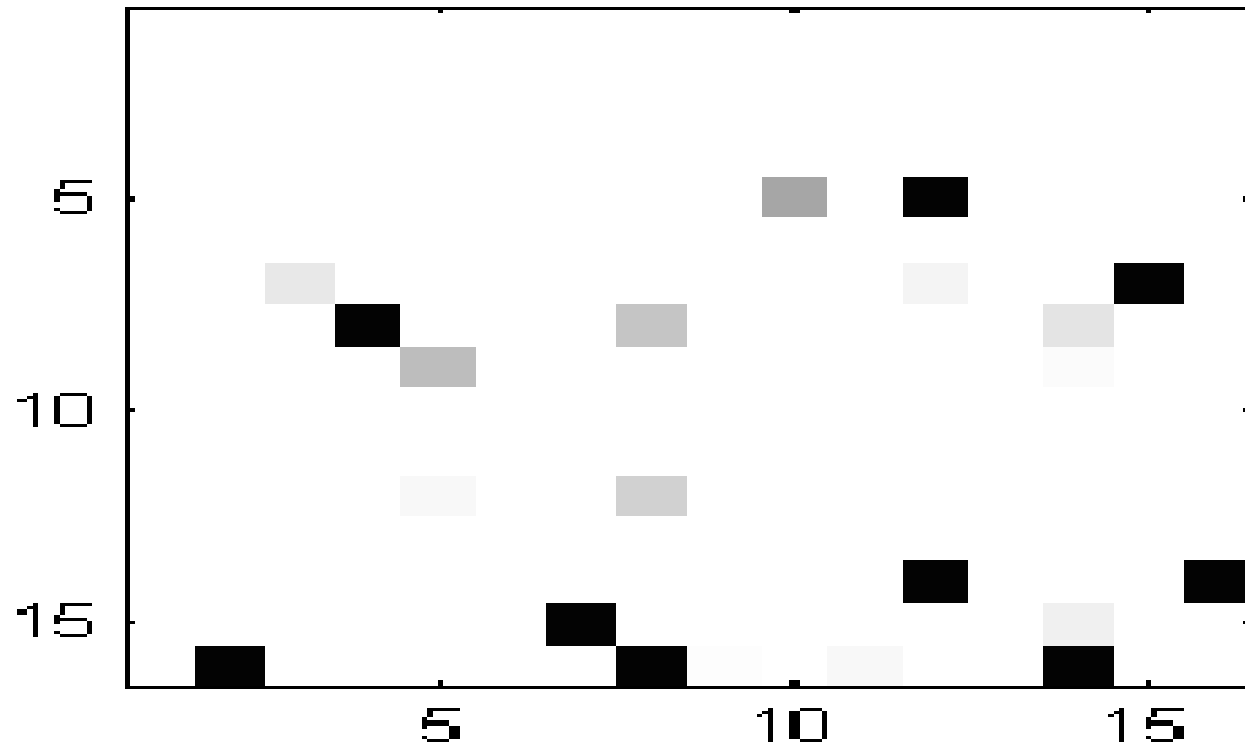
Address distribution

- 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

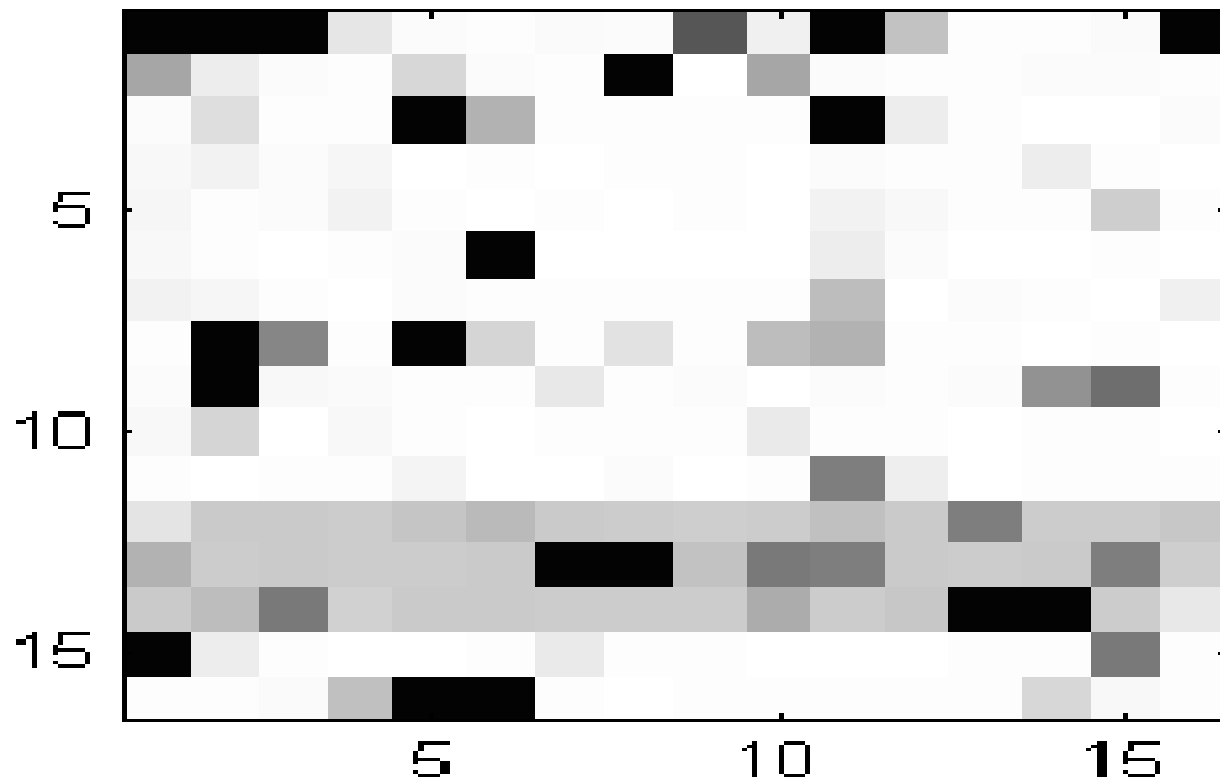
Data collection

- 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

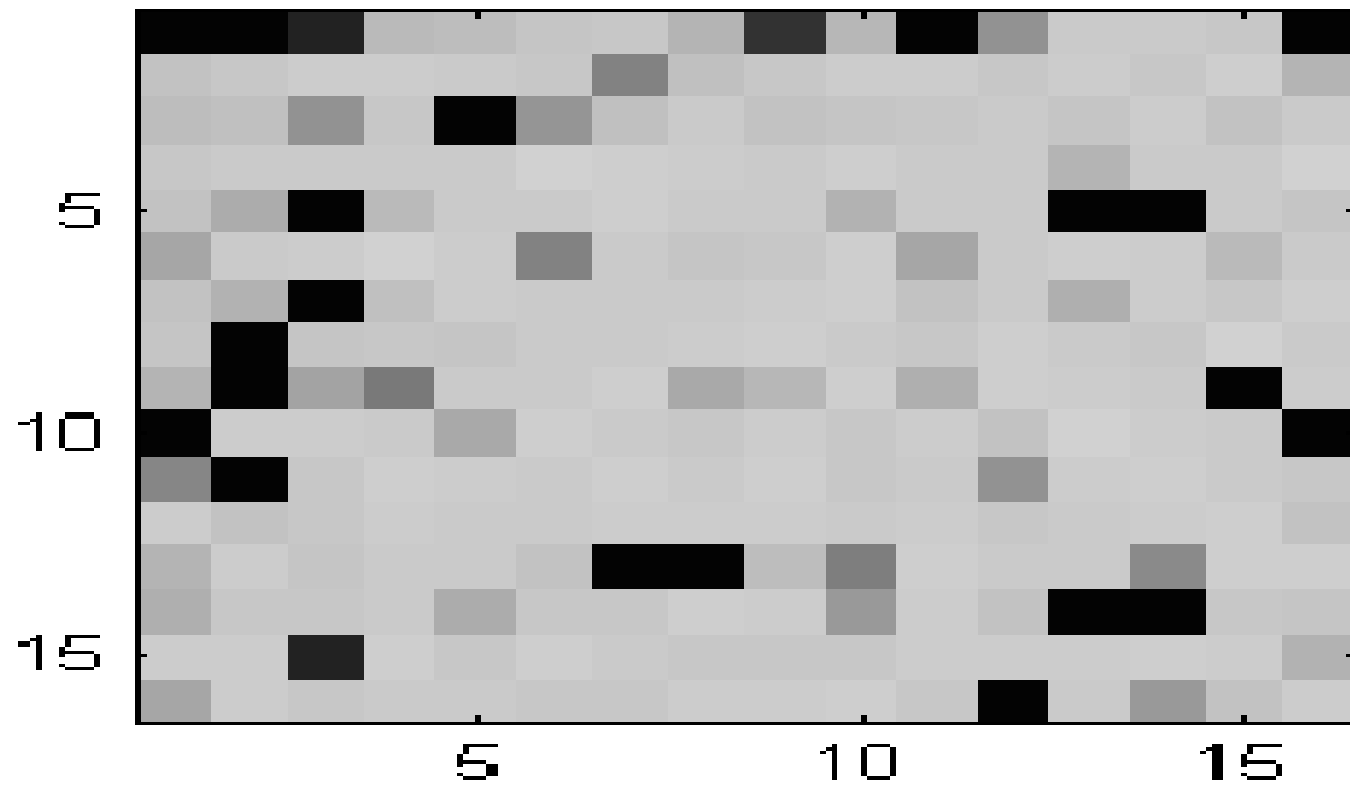
Data Collection



Address Distribution



Address Distribution



Address Distribution

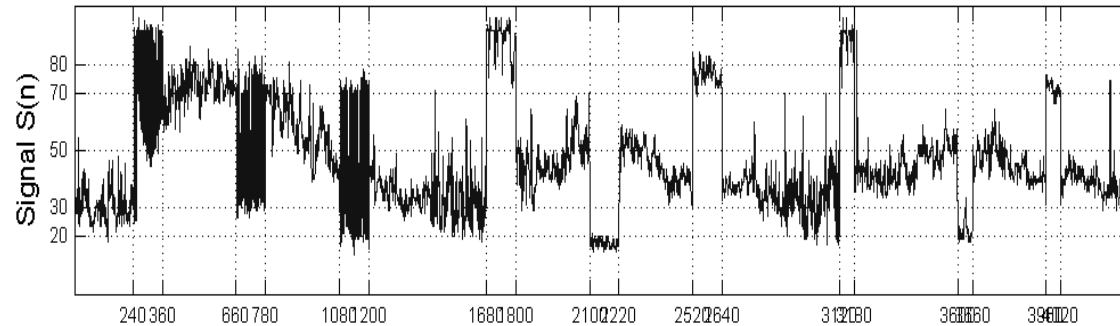
- Visually the distributions look different from normal, semi-random destination attacks, random attacks
- How to analyze, quantify to generate valid detection signals?
- Use Image processing ideas...
 - Address distribution sample as an image

Protocol Numbers?

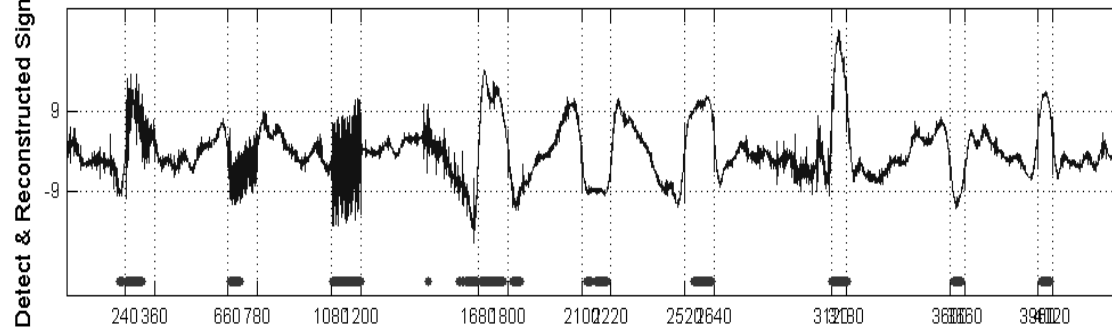
- Can we use protocol numbers as a signal?
- Recent worm attacks used distinct port numbers (1434 -Code Red)
- Port numbers discrete
 - How to generate a valid signal
- Apply auto-correlation over samples

Protocol 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], $-3.0\sigma < X < 3.0\sigma$



Measurement candidates

- Use every available piece of information for anomaly detection
 - Volume of traffic
 - Number of flows
 - Address distribution
 - Port numbers
 - Protocol distribution
- Encryption (VPN) throws a wrench

Classify, Measure, control

- A number of products allow you to classify packets along multiple dimensions
 - Packeteer, ...
- Put limits on individual bins
 - Rate limit kazaa traffic to 20% of link
- Work only on known types of traffic
- Worm attacks in email, web apps??

Packeteer

- Allows administrator's policy to guide resource usage
- Allows IP/TCP packet headers to be inspected for classification
- Uses TCP Ack-pacing to control rates of individual flows
 - Slow down the acks for aggressive flows
- Others: Allot's NetAccountant
 - Enterasys Dragon Sensor Appliance

TurnTide -Spam control

- Classify individual senders based on spam rankings
- Allow fewer SMTP connections for senders deemed high on spam ranking
- Easy to spoof email addresses, employ multiple source IP addresses
- Not clear how useful when good & bad reside at the same SMTP client
- Project ideas -anyone?

Summary of today's class

- Can measure traffic in many dimensions
- Measurements can help in identifying anomalies
- Classification and rate control and other measures can be applied