Framework for Classifying DoS Attacks [Hussain02b]: Hussain, Heidemann, Papadopoulos

CSci551: Computer Networks
SP2006 Thursday Section

John Heidemann

Preview: Security Problems in the Internet

- virus
- worms
- denial-of-service attacks
- phishing attacks
- eavesdropping
- imposters / authorization
- defenses:
  - anti-virus (at a host)
  - firewalls: try to keep bad stuff out
    - typically look at packet headers
  - intrusion detection systems (IDS):
    - look at signatures in traffic
    - look look for anomolous traffic patterns

Key ideas

- way to classify DoS attacks
  - single source vs. multisource
  - header analysis
  - ramp-up behavior (new)
  - spectral analysis (new)
- applications of approaches
- looks at *why* attack traffic looks this way
  - wrt ramp-up and spectral

Approach and Motivation

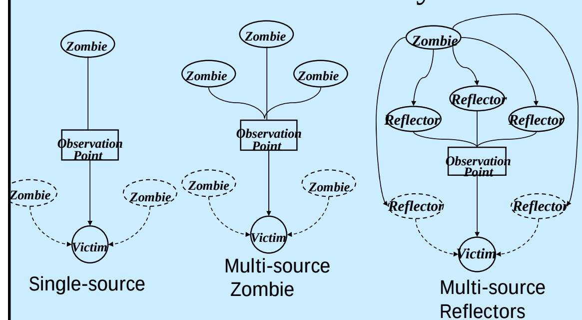
- develop methods to classify DDoS attacks
  - headers, ramp-up, spectral analysis
- applications
  - determine single- vs. multi-source to select response
  - use to validate accuracy of simulation models
  - (but applications are not completely compelling)
- side benefit: explore spectral analysis

Related Work: Intrusion Detection Systems

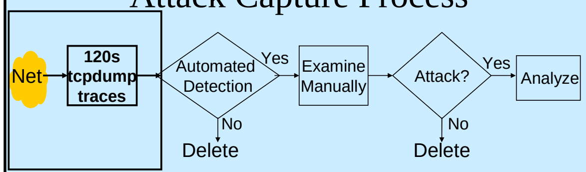
- idea: look in packet stream for known patterns
- strengths?
  - 100% detection of known attacks
  - can be fast (just byte matching)
- weaknesses?
  - have to look at packet contents
  - 0% detection of unknown attacks
- idea: characterize normal traffic, detect anomalies
  - define "normal" traffic, look for things outside normal
- strengths?
  - can detect previously unknown attacks
- weaknesses?
  - probably has higher false positives
  - defining normal is hard

Attack Taxonomy

[Hussain03b, Figure 1]

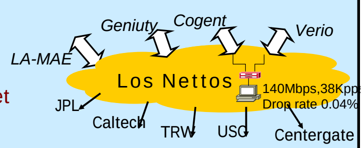


Attack Capture Process

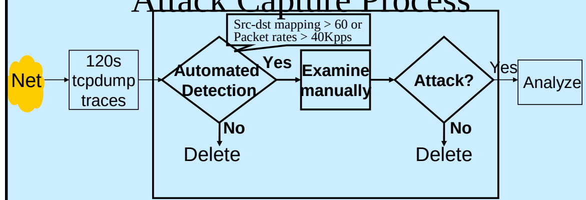


Trace Collection

- Off-the-shelf PC, FreeBSD
- Modified driver to support partial packet transfer



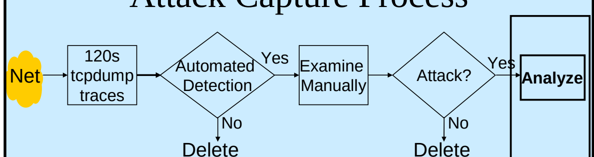
Attack Capture Process



Attack Detection

- Mapping of source IP to destination IP or traffic rates
- Empirically derived thresholds
- 80 attacks from July–Nov 2002

Attack Capture Process



Attack Analysis

- Header content
- Packet stream characteristics
  - Ramp-up behavior
  - Spectral analysis

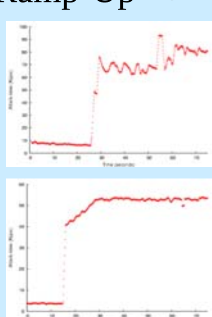
Analysis: Header Content

- two approaches:
  - ID field
    - linear change => single source
    - multiple concurrent linear changes => multi-source
    - randomized => can't tell
  - TTL
    - all the same => not multi-source
    - different => can't tell
- spoofability? yes, easily--just randomize the fields
- why bother with other approaches?
  - other approaches needed because this is spoofable
  - provides ground truth to test other approaches

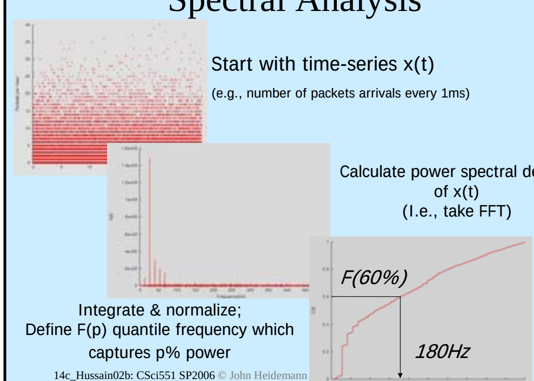
Analysis: Ramp-Up

[Hussain03b, Figure 5]

- ramp-up: change in rate at the beginning of an attack
- ramp up => single start right away, multi-source has larger ramp-up
- why?
  - multiple sources start at different times
- spoofability?
  - hard to make multi-source get rid of ramp up
  - easy to make single source like multiple sources



Spectral Analysis



Start with time-series x(t)

(e.g., number of packets arrivals every 1ms)

Calculate power spectral density of x(t)
(I.e., take FFT)

Integrate & normalize;
Define F(p) quantile frequency which captures p% power

Spectral Analysis: Math

$$c(k) = 1/N \sum_{t=0}^{N-k} (x(t)-\bar{x})(x(t+k)-\bar{x});$$
$$r(k) = c(k)/c(0)$$

ACF at lag k

$$S(f) = \sum_{k=0}^{M} r(k) e^{-\imath 2\pi f k}$$

Power spectrum

$$P(f) = \sum_{i=0}^{f-1} \frac{(S(i)+S(i+1))}{2};$$
$$C(f) = \frac{P(f)}{P(f_{max})};$$

Integrate and normalize S(f)

$$F(p) = \min_{0 \le f \le f_{max}} f \text{ such that } C(f) \ge p$$

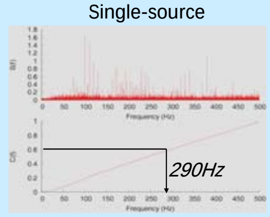
Determine *p* quantile

14c_Hussain02b: CSci551 SP2006 © John Heidemann 26

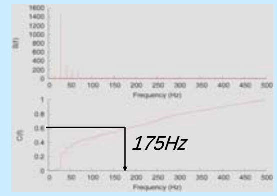
Single vs. Multi-source Attacks

[Hussain03b, Figure 6]

Single-source



Multi-source



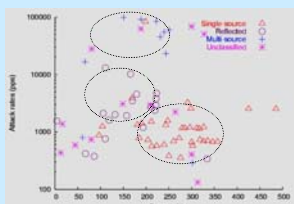
- Single src attack produces linear cumulative spectrum
- Multi-src attacks produce localization of power in low frequencies

14c_Hussain02b: CSci551 SP2006 © John Heidemann 27

Classifying Attacks

Steps:

- Compare F(60%) to identify single-/multi-source attacks
- Single-source:
  F(60%) mean 268Hz (240-295Hz)
- Multi-source:
  F(60%) mean 172Hz (142-210Hz)
- Robustly categorize Unclassified attacks



[Hussain03b, Figure 7]

14c_Hussain02b: CSci551 SP2006 © John Heidemann 28

DDoS Attacks: *Why* Does Spectra Change?

intuition:

- single flow has characteristic signature
  - determined by sending process, bottleneck interface, etc.
  - results in high-frequency components
- multiple flows *loose* this signature because they are not synchronized
  - instead their interactions produce low frequencies

14c_Hussain02b: CSci551 SP2006 © John Heidemann 29

Implications of Why

why care about why? need to figure out about tomorrow: protocol changes, or attack countermeasures

- single source wants to appear like multiple
  - possible, but reduces attack effecitivness
- multiple sources wanted to be like single

=> complex interaction in spectrum
  - very hard: would have to have close, distributed synchronziation
- what about countermeasures?
  - find things to observe that are inherrent
    - i.e., to conceal what's happening must slow the attack

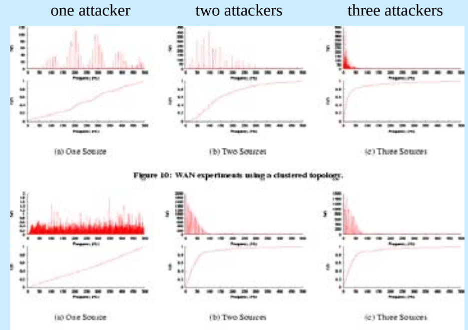
14c_Hussain02b: CSci551 SP2006 © John Heidemann 32

Validating Why

one attacker | two attackers | three attackers

clustered

distributed



14c_Hussain02b: CSci551 SP2006 © John Heidemann 33

Why Validate Why?

- compares things in several ways
  - real traces
  - real traces from another site (too small)
  - testbed experiments
  - simulations
- focusing on carefully explaining and proving phenomena is important
  - ex: compare "in Africa, lots of people have anemia"
  - vs. "in Africa, people have anemia, *and* they tend to have sickle-cell blood cells, *and* people who don't tend not to have anemia, *and* that's correlated with a feature on Gene #X, *and* it's plausible that the sickle cell actually helps protect against malaria"
    - you know a *lot* more and can actually make informed decisions
- like with [Aguayo04a], methodology and depth are important

Future Directions

- active area of work at USC
- lot of open questions
  - trade-offs in representation of network traffic as signal
  - comparing on new attacks
  - countermeasures and counter-countermeasures
  - applying spectral analysis to other networking problems? (like…)
  - automating procedure

Other questions/observations?

- xxx