Advanced Network Security

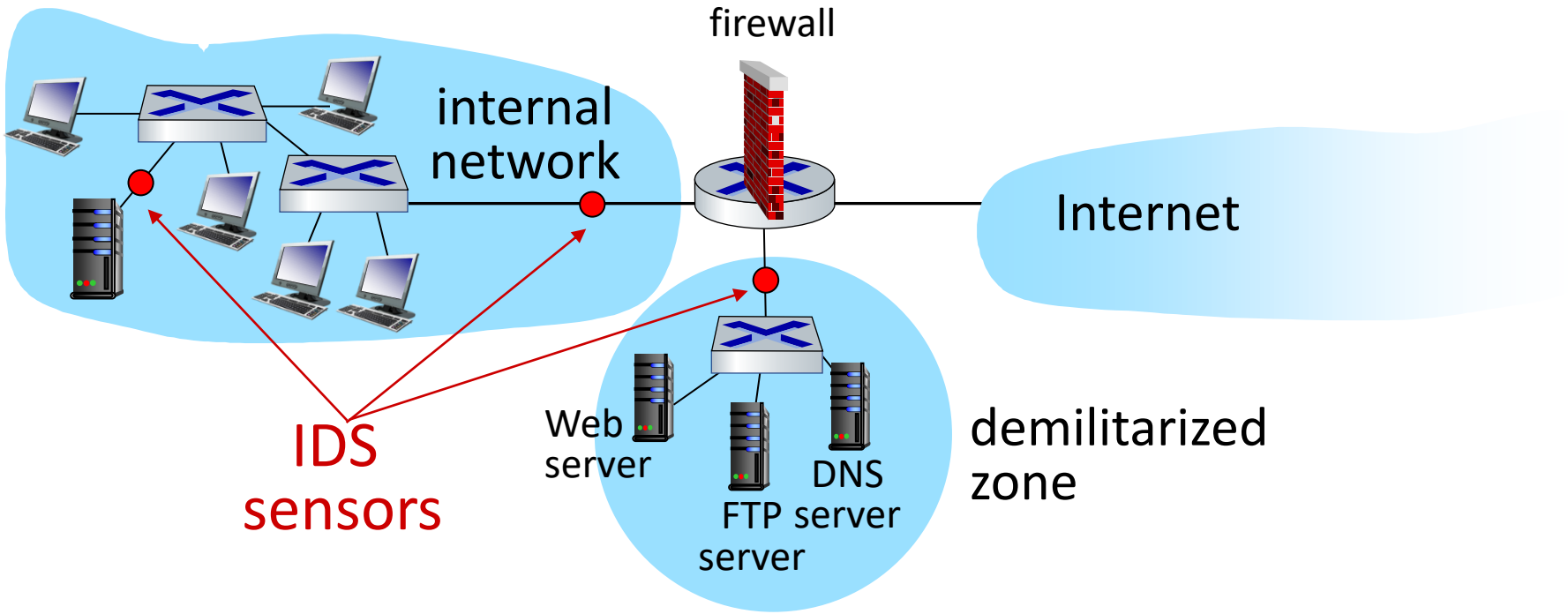
Firewalls and IDS

Amir Mahdi Sadeghzadeh, Ph.D.

Most slides have been adapted from Steven Bellovin, COMS W4180, Columbia University, 2006, and Mehdi Kharrazi, CE40-817, Sharif University of Technology, 2015

Intrusion detection systems

multiple IDSs: different types of checking at different locations



What is IDS?

- An Intrusion Detection System (IDS) is a system that attempts to identify intrusions.
- Intrusion detection is the process of identifying and responding to malicious activity targeted at computing and networking resources.

- The goal of IDS is to detect fingerprints of malicious activity.

Anomaly Detection

- Using a model of normal system behavior, try to detect deviations and abnormalities
- Any large deviation from the model is thought as anomaly.
  - E.g., raise an alarm when a statistically rare event(s) occurs

- Pro: can detect previous unseen attacks
- Con: have higher false positives, and hard to train a system for a very dynamic environment.
- Approaches: statistical methods, Machine Learning

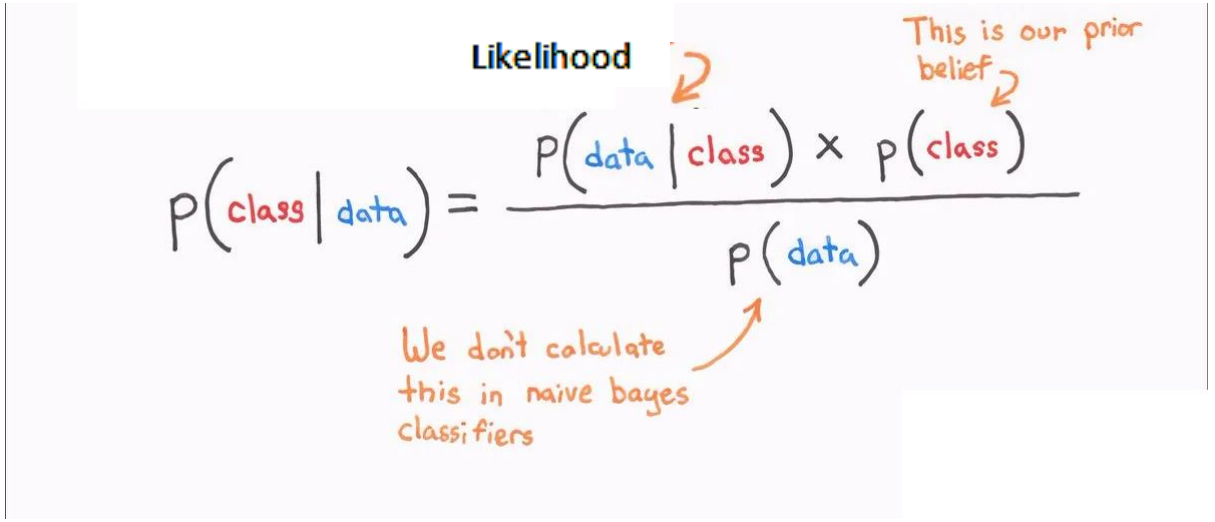
Anomaly Detection with NIDS

- High false positive rate
  - False identifications are very costly because sys admin will spend many hours examining evidence
- Training is difficult
  - Lack of training data with real attacks
  - Network traffic is very diverse, the definition of “normal” is constantly evolving
    - What is the difference between **a flash crowd** and a **denial of service** attack?

Naïve Bayes Classifier (supervised learning)

- A Bayes Classifier is a probabilistic model that uses Bayes' theorem to classify data into different categories or classes.

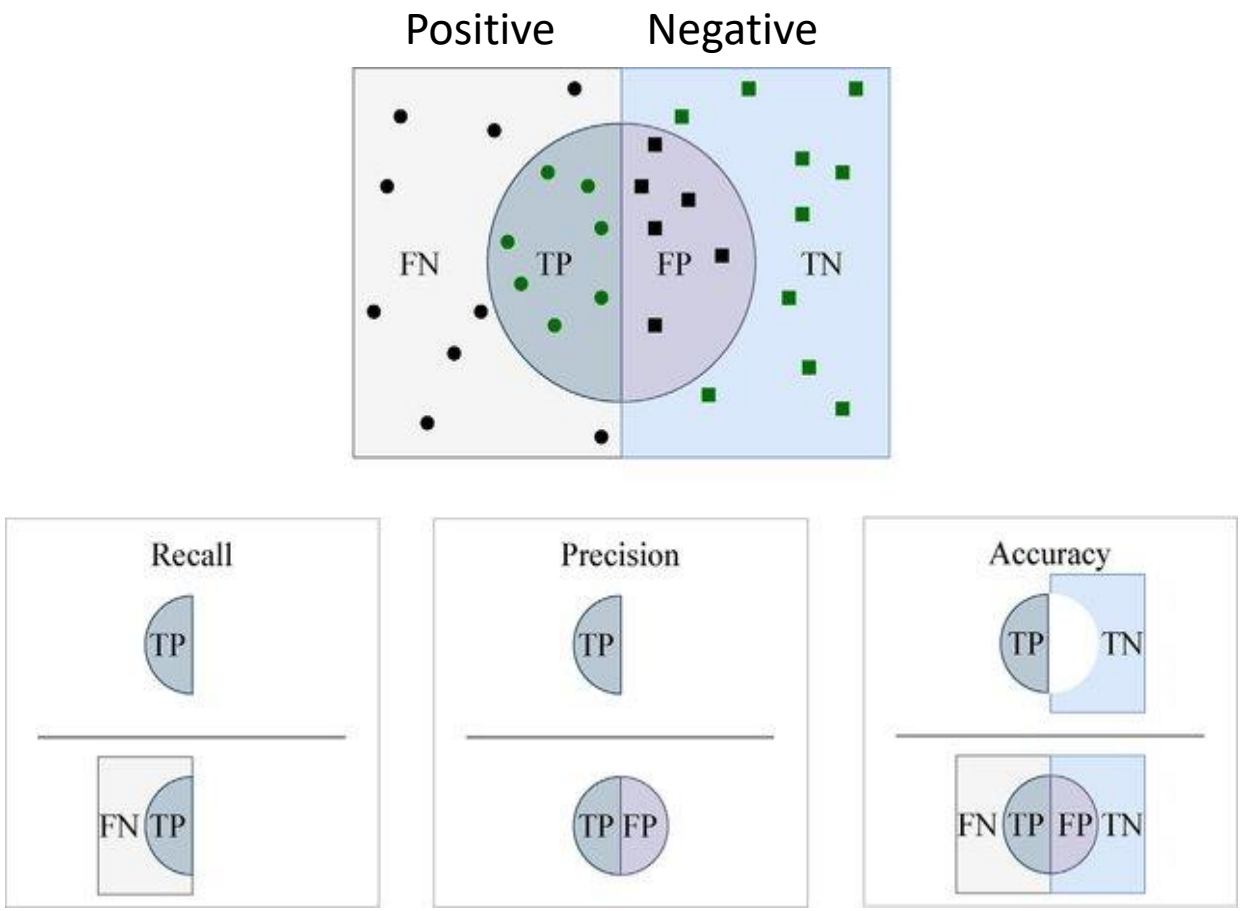
- **P(Class | Data):** Probability of the data belonging to a specific class.
- **P(Data | Class):** Probability of observing the data given the class.
- **P(Class):** Prior probability of the class.
- **P(Data):** Probability of the data.



IDS Evaluation

- Accuracy: false positives and false negatives should be minimized.
- Performance: the rate at which audit events are processed.
- Completeness: to detect all attacks.
- Fault tolerance: resistance to attacks.
- Timeliness: time elapsed between intrusion and detection.

Precision and Recall



Base-Rate Fallacy

- 1% of traffic is SYN floods; IDS accuracy is 90%
  - IDS classifies a SYN flood as attack with prob. 90%, classifies a valid connection as attack with prob. 10%
- What is the probability that the connection flagged as a SYN flood by IDS is actually valid?

$$\Pr(\text{valid} \mid \text{alarm}) = \frac{\Pr(\text{alarm} \mid \text{valid}) \bullet \Pr(\text{valid})}{\Pr(\text{alarm})}$$

$$= \frac{\Pr(\text{alarm} \mid \text{valid}) \bullet \Pr(\text{valid})}{\Pr(\text{alarm} \mid \text{valid}) \bullet \Pr(\text{valid}) + \Pr(\text{alarm} \mid \text{SYN flood}) \bullet \Pr(\text{SYN flood})}$$

$$= \frac{0.10 \bullet 0.99}{0.10 \bullet 0.99 + 0.90 \bullet 0.01}$$

= 92% chance raised alarm is false!!!

slide

Sensor Locations

- Outside the firewall?
  - We know there are bad guys there; what's the point?
- Just inside? What's the threat model?
- On sensitive internal nets?
- In front of each sensitive host?
- In "dark space"?

What's the Purpose?

- Inside the firewall? Detect data exfiltration
- Sensitive internal nets: detect threats aimed at them
- Watching each host? Detect attacks on inside hosts from other hosts on the same LAN
- Dark space? Detect scanning worms (and attackers)

What's Dark Space?

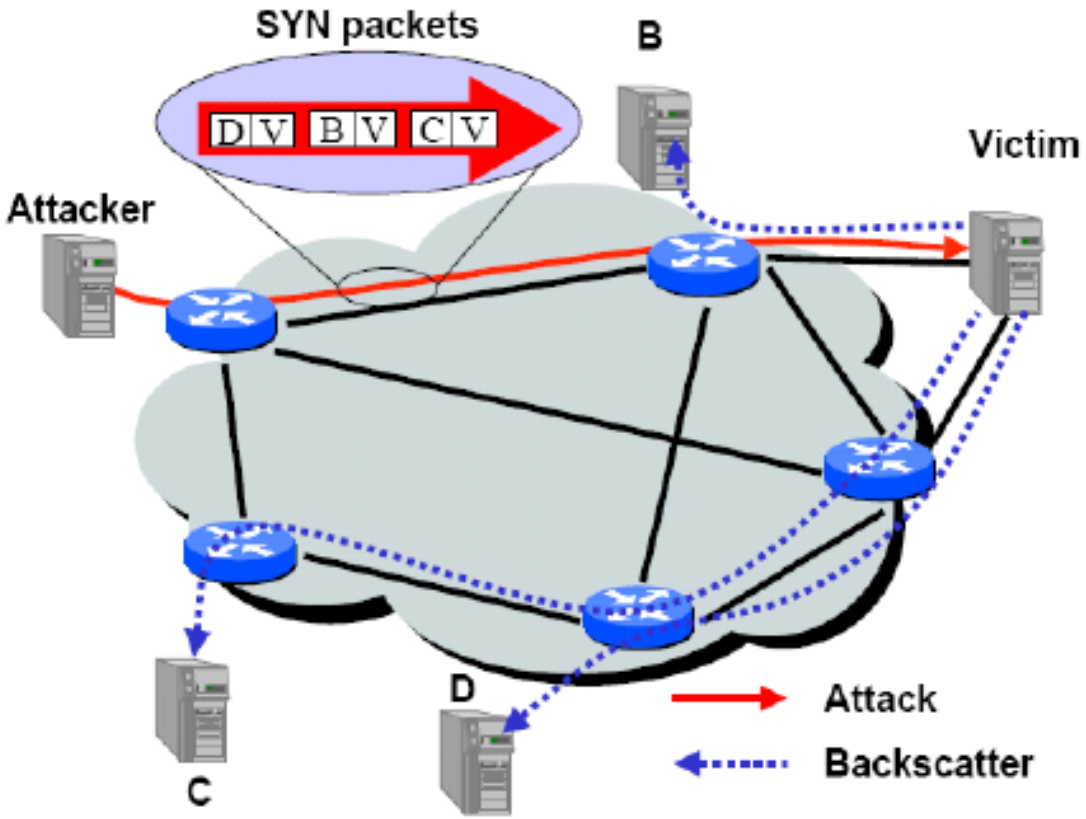
- A block of address space not used by real machines and not pointed to by DNS entries
- There is no legitimate reason to send packets to such addresses
  - Therefore, any host sending to such addresses is up to no good
- Commonly used to detect scanning worms

Network Telescopes and Honeypots

- Monitor a cross-section of Internet address space
  - Especially useful if includes unused “dark space”
- Attacks in far corners of the Internet may produce traffic directed at your addresses
  - “Backscatter”: responses of DoS victims to SYN packets from randomly spoofed IP addresses
  - Random scanning by worms
- Can combine with “honeypots”
  - Any outbound connection from a honeypot behind an otherwise unused IP address means infection
  - Can use this to analyze worm code

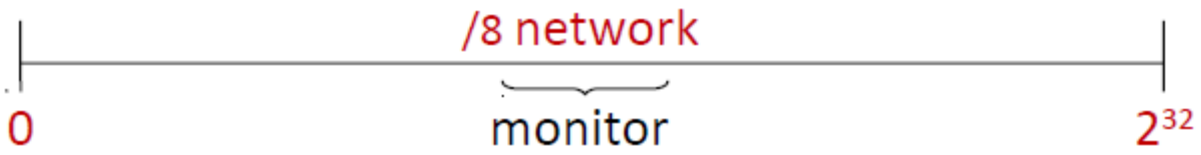
Backscatter of SYN Floods

- SYN with forged, random source IP address ⇒ SYN/ACK to random host



Measuring Backscatter

- Listen to unused IP addresses space (dark space)



- A lonely SYN/ACK packet is likely to be the result of a SYN attack
- 2001: 400 SYN attacks/week
- 2013: 773 SYN attacks/24 hours
- 2016: 1654 SYN attacks/24 hours

Honeypots and Honeynets

- Special-purpose host or network designed to be attacked
  - Lure the attacker in deeper
- Equipped with many monitoring options
- Waste the attacker's time; study the attacker's technique
- Note well: keeping honeypot (and dark space) addresses secret is vital

Auto-Quarantine

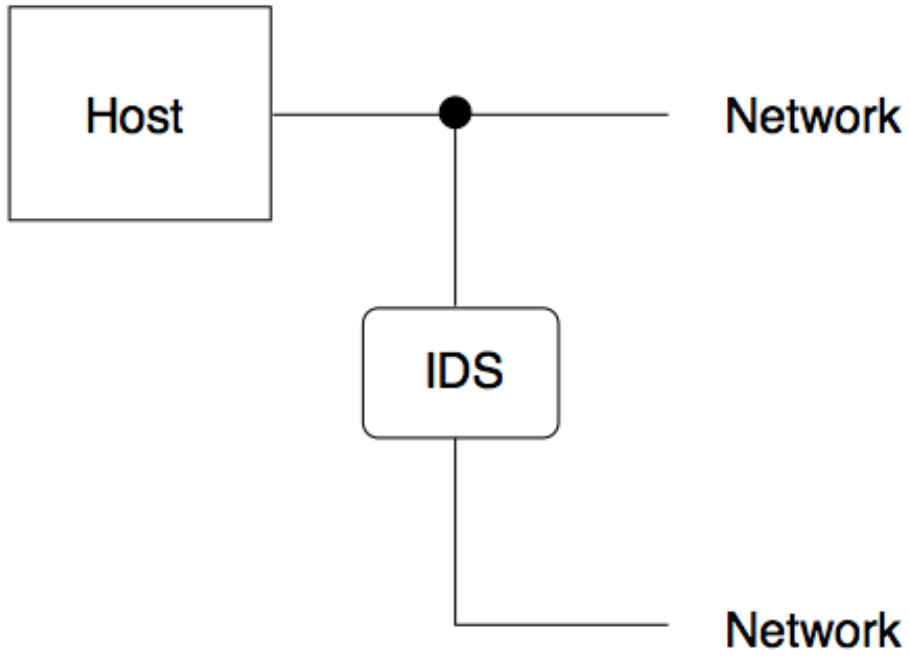
- Many organizations implement “auto-quarantine”
- This is especially common for university residence hall networks
- Machines that do too much scanning (and in particular attempt to probe dark space) are assumed to be virus-infected
- They’re moved to a separate net; the only sites they can contact are Windows Update, anti-virus companies, and the like

Host- or Net-Resident?

- Suppose you want to monitor each host. Where does the monitor live?
- Dedicated in-line hardware: good, but expensive
- On the host: cheap, but subvertible

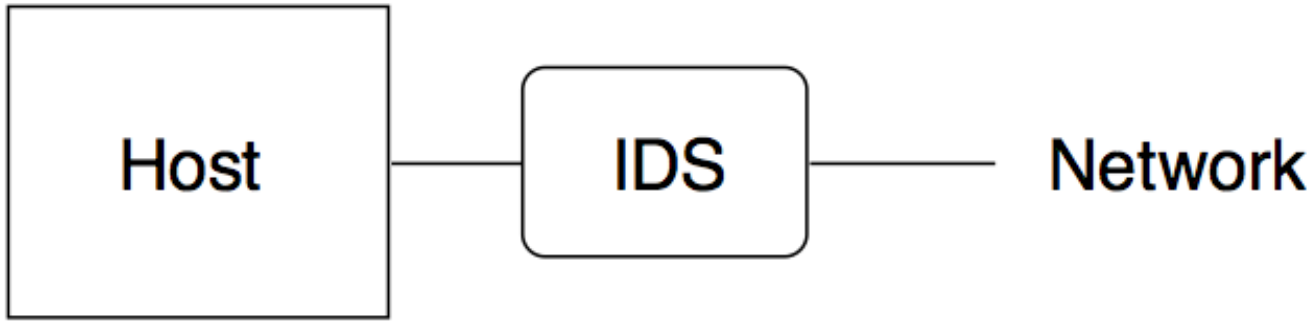
Net-Resident: Parallel

- Very unobtrusive
- But — need special hardware to tap an Ethernet
- Need some network connection to the IDS



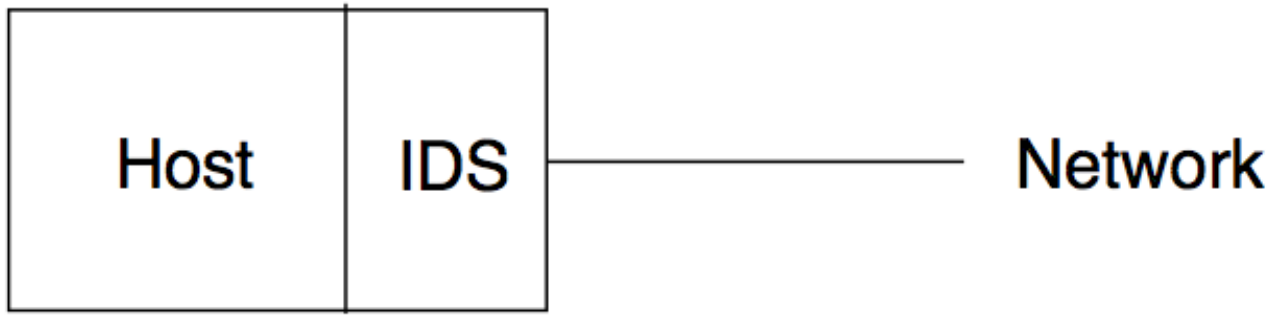
Net-Resident: Serial

- Can’t miss packets
- But — if it crashes, the host is unreachable
- Can the IDS box be hacked?



Host-Resident Monitor

- No special hardware needed
- IDS sees exactly what host sees
- But — subvertible
- Useful precaution: immediately transmit IDS data elsewhere



The Big Advantages of Host IDS

- More time
- More context
- Everything is reassembled
- Look at entire item, not streams

Extrusion Detection

- Detect bad things leaving your network
- Detect sensitive things leaving your network
- Finds theft of inside information, either by attacker or by rogue insider
- Can be done in the network or in application gateways

Simple Logging

Simple Logging

- I (Steven Bellovin) ran this command for a while, on two hosts:
  - tcpdump -p -l "tcp[13] == 0x2 and dst $us"
- What does it do?
- Logs all TCP SYN-only packets addressed to us (tcp[13] is the flags byte in the TCP header; 0x2 is SYN)

TCP Header

Offsets		0								1								2								3							
Octet	Bit	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
0	0	Source port																Destination port															
4	32	Sequence number																															
8	64	Acknowledgment number (if ACK set)																															
12	96	Data offset				Reserved 0 0 0 0				CWR	ECE	URG	ACK	PSH	RST	SYN	FIN	Window Size															
16	128	Checksum																Urgent pointer (if URG set)															
20	160	Options (if *data offset* > 5. Padded at the end with "0" bits if necessary.)																															
⋮	⋮																																
56	448																																

Some Results

- About 85 probes apiece, during a 30-hour run
- 63 different ports scanned
- Some obvious: http, ssh, Windows file-sharing, SMTP, web proxy
- Some strange: 49400–49402, 8081–8090, 81–86
- Some threatening: terabase, radmin-port
- Most probers looked at one port; one looked at 46 ports

The Most Probed Ports

Scans	*Port*
3	ms-wbt-server
3	ssh
5	8000
5	http-alt
6	ms-sql-s
6	radmin-port
7	BackupExec
8	smtp
9	WebProxy
9	http

What Did The Probers Want?

- WebProxy and SMTP are probably for spam email and connection-laundering
- The others look like probes for known vulnerabilities

Bad Neighborhoods

- I see more probes here than elsewhere. Why?
- There are different “neighborhoods” — ranges of IP addresses — in cyberspace
- University networks are good hunting grounds — few firewalls, good bandwidth, many poorly-administered machines
- Newly-allocated network blocks have few hosts, and aren’t scanned as much

Finding Compromised Hosts

Finding Compromised Hosts

- Suppose you've identified a compromised host. Now what?
- Get data: IP address and (when feasible) MAC address
- Find it

Databases

- Must be able to map IP address to location
- Must be able to map IP address to person
- Difficult on campus — wide-open nets
- Primary reason for host registration in many places

Layer 2 Data

- Enterprise-grade switches are “managed”
- They can map an IP address or a MAC address to a physical port
- Especially useful if the attacker is forging addresses. . .

Switch Data

- Note that a single MAC address has shown up on two different switch ports, in different buildings. This is reasonable for a laptop, but not for a server!

Home + Switch View + Port View + Jacks View + Search Jacks + Search Host

MAC Address:	0003BA1077F7
Submit	Reset

0003BA1077F7 is not staticly registered

Location	First Seen	Last Seen
cs-4-1.net:5/15	02-aug-2004 16:03:27	13-nov-2006 18:08:29
cepsr-7-1.net:6/9	09-may-2006 21:39:18	31-oct-2006 14:52:13

ARP cache

IP	MAC	Last Seen
128.59.16.72	0003BA1077F7	13-nov-2006 22:17:50

Problems with (Commercial) IDS

- Cost of update and keeping current is growing
  - Organizations lack internal expertise
- Knowledge based IDS systems suffer from False Negative Problem
  - New augmented IDS with Anomaly Detectors are appearing in the commercial market
- IDS are inherently noisy and chatty and suffer from the False Positive problem
  - Volumes of alerts are crushing
  - Zooming in on most serious threats is hard
- NIDS positioned at the perimeter
  - The most serious/predominant threat is the insider
  - Host and LAN-based IDS now more crucial

References

- Kurose, James F., and Keith W. Ross. "Computer networking: A top-down approach edition." Addision Wesley (2007), chapter 8.
- Steven Bellovin, COMS W4180, Columbia University, 2006
- Mehdi Kharrazi, CE40-817, Sharif University of Technology, 2015
- Vitaly Shmatikov, CS 361S, UT Austin, 2014