The Reconnaissance Phase

Detecting the Enemy *Before* the Attack

Carrie Gates
PhD Candidate, Dalhousie University
Visiting Scientist, CERT, Carnegie Mellon University
Outline

- Indicate a gap in our defences
- Talk about how we’re addressing it now
- Talk about how it can be addressed
- Give examples to indicate why we should address it

- Concluding statement
Timeline

Defence

Protection (e.g. firewalls)
Prevention (e.g. patches)

GAP

Detection (e.g. IDS)

Gaining Access
Maintaining Access

Intrusion

Covering tracks and hiding

Reconnaissance
Port Scanning

(Skoudis, 2002)
How to cover the gap?

- Detection of reconnaissance
- Detection of port scanning
- Correlation of information from different sources (technical)
- Correlation of information from different sources (non-technical)
Detection of Reconnaissance

This is hard! 😊

- E.g. who-is databases, newsgroup browsing
- We don’t have access to many of these logs (and we would be swamped if we did!)
- BUT, can track web browsing (but how to tell benign from malicious?)
- AND, can track social engineering attacks
Detection of Port Scanning

Here is where we have concentrated the most effort.
Why is this hard?

- How can you determine if a packet is legitimate?
- What is suspicious?
  - Too many destinations?
    - May still all be legitimate. What is too many?
  - Malformed packets?
    - Yes, but not very common – usually SYN scans
  - Too many SYN packets with no connections?
    - People are now faking connection-like traffic
Some Solutions...

- Snort (Roesch, 1999) – malformed packets, $x$ destinations in $y$ seconds
- Bro (Paxson, 1999) – uses threshold on number of destinations, plus some payload analysis

- Require (unidirectional) packet-level information
- Thresholds prone to false positives (if too low) or false negatives (if too high)
Spice (Staniford et al., 2000) – examines “anomalous” (as determined by Spade) incoming packets, grouping them using a simulated annealing procedure

- Still unidirectional packet level
- Groups represent more than just scans
- (Robertson et al., 2003) – examines return traffic and thresholds on number of missing/rejected responses
- (Jung et al., 2004) – examines return traffic and builds hypothesis based on number of hits (SYN-ACKs) versus misses (no response/RSTs)

- Require packet-level information
- Require packets in *both* directions
Flow-dscan (Fullmer and Romig, 2000) – Uses thresholds on destinations/source with suppress lists, ports < 1024 only

MISSILE (work in progress at CERT) – Uses combination of various metrics to indicate likelihood of a scan

- Uses unidirectional flow data
MISSILE

- Examines characteristics of all TCP flows from each source, looking for activity that indicates a scan
- Also looks at “event level” for scans (e.g. majority of flows just SYN, malformed packets)
Sample output

- One class B for one week:
  - 24,114,559 flows
  - 3 hours to process
  - 7481 unique sources identified as scanning
  - 1436 unique sources identified as attempting exploit during scan
  - 5667 sources identified as SYN scanning
  - Average: 452 destinations/source
  - Maximum: 196073 destinations – 3 ports (1080, 3128 and 10080) on 65469 IPs in ~ 8 hours
How is this scan information used?

- To proactively block scanning IP addresses to prevent information gain
  - Can be used as a denial of service
  - Some network admins don’t want the performance hit from extra routers
- To send complaint letters – RARE!

Largely ignored 😞
How *could* this information be used?

- What was targeted?
- Who answered?
- Who (destinations) should I watch?
- Is someone about to attack?

Tells you:

- Who to patch!
- Who might have been compromised!
Scans can include exploit
- Who responded? Was there a conversation? What machines might have been compromised?
- The Honeynet Project has noted that there is an increase in attackers performing scan bundled with exploit
  - Nearly 20% of attackers identified in previous example (1436/7481)
Scans can be for pure reconnaissance – so attacker might return later

Who responded? What is likely to be targeted?
Are patches up to date on that service on the responding machines?

We don’t know how common this is
What if someone comes back from a new IP?
TCP SYN scan of port 80 (web)

Example Scan

Internal Addresses Scanned

- days 01 - 06
  21,234,579
  53%
- days 14 - 21
  19,124,423
  47%

Internal Addresses Replied

- days 01 - 06
  299,511
  90%
- days 14 - 21
  34,454
  10%
Distribution of Replies to this Scan

- 1-pkt, TCP: rst 51%
- 1-pkt, TCP: syn ack 48%
- Others 1%
- >1-pkt, TCP
- ICMP
- other
This same activity occurred over the following timeline:

- Days 1 – 6: scanning (scattered over the days)
- Days 8 – 9: (6 hours) apparent attacks on selected subnets; seems to have targeted only hosts that had replied to the earlier scan with
  - 1-pkt, TCP: syn ack, or
  - multiple TCP packets in a flow
- Days 14 – 21: scanning of additional subnets (scattered over the days)
- ...

Correlation of Information

Technical

- Correlation between logs is being researched, but concentration is on correlation between different IDSs (e.g. (Cuppens and Miège, 2002) and (Ning et al., 2002))

- We need to add in other forms of information, e.g. IDS, NetFlow, web logs
Correlation of Information

Non-technical

- Need to add into correlation of information all non-technical information, such as if a social engineering attack has been attempted
Network Intelligence Analysis

- Trying to bridge the gap between Protection/Prevention and Detection has been likened to intelligence gathering (e.g. SIGINT, HUMINT), e.g. (Shimeall and Dunlevy, 2001)

- However, network intelligence includes an even broader perspective:
  - Political events (e.g. hactivism)
  - Social events (e.g. holidays)
  - Technical events (e.g. vulnerability releases)
Concluding Remarks

“Without a solid network intelligence, defenders are required to respond equally to all intrusions. This is untenable in the long run ....”

(Shimeall and Dunlevy, 2001)
Thanks! 😊

- CERIAS
- CERT Analysis Center
- IBM Centers for Advanced Studies
- Dalhousie University