Mapping Internet Sensors with Probe Response Attacks

Jason Franklin jfrankli@cs.wisc.edu

Computer Sciences Department
University of Wisconsin, Madison
Outline

Background

Example Attack
  Introduction to the Attack
  Basic Probe Response Algorithm

Attack Simulation
  Internet Storm Center Distribution
  Other Internet Sensor Network Distributions

Generalizing the Attack
  Covert Channels
  Other Networks

Countermeasures

Conclusion
Internet Sensor Networks

Definition

An Internet sensor network is a collection of systems which monitor the Internet and produce statistics related to Internet traffic patterns and anomalies.

Example categories of Internet sensors include:

- security log collection and analysis centers such as the SANS Internet Storm Center, myNetWatchman, and Symantec’s DeepSight network
- collaborative intrusion detection systems
- Internet sinks and network telescopes such as the University of Michigan’s Internet Motion Sensor and CAIDA
Usage of Internet Sensor Networks

Internet sensors are useful for distributed intrusion detection and monitoring such as:

- quickly detecting worm outbreaks
- enabling a wide area perspective of the Internet
- aggregating rare events from globally distributed monitors
- classifying the pervasiveness of threats like port scans, DoS attacks, and botnet activity
Data Integrity, Sensor Anonymity, and Privacy

Critical Assumption

The integrity of an Internet sensor network is based upon the critical assumption that the IP addresses of systems that serve as sensors are secret.

The results of violating this assumption include:

- integrity of the data produced by network is greatly reduced
- potential loss of anonymity and privacy of sensors
Maintaining Privacy

Current attempts to maintain the privacy of organizations submitting logs to Internet sensor networks include the following:

**Techniques**

- **black marker approach**: eliminating sensitive fields from published reports
- **hashing**: using a hash function on fields of a report
- **bloom filters**: encoding data in an efficient data structure for set membership tests and set unions
- **permutations**: applying a prefix-preserving permutation to IP addresses
Attacks and Countermeasures

Probe Response Attacks

- new class of attacks called **probe response attacks**
  - capable of compromising the **anonymity and privacy** of individual sensors in an Internet sensor network.

Countermeasures

We also provide countermeasures which are effective in preventing probe response attacks.
SANS Internet Storm Center

Background

Mapping Internet Sensors with Probe Response Attacks

SANS Internet Storm Center

Computing Environment

SANS Internet Storm Center

Survey

Today's Diary

We're Phull... Thanks for the ph stulty. Article about bank fraud.

SANS Internet Storm Center Webcast: Wednesday July 13th 2005

Poll

What is the most important for you to see reported at the Internet Storm Center?

New vulnerabilities
New exploits detected in the wild
Survival Time
Stirrings port and trend data
Pointers to interesting reading
INFOcan
Good grammar / spelling
Other, state below.

Add comment
Add comment

Database Status

Reports Processed:
Last Month: 558,397,466
Last Week: 1,268,876,603
Last 24hrs: 34,019,660

Jason Franklin jfrankli@cs.wisc.edu

Mapping Internet Sensors with Probe Response Attacks
Case Study: the ISC

SANS Internet Storm Center

To evaluate the threat of probe response attacks in greater detail, we analyzed the feasibility of mapping a real-life Internet sensor network, the ISC.

▶ one of the most important existing systems which collects and analyzes data from Internet sensors
▶ challenging to map
  ▶ large number of sensors (over 680,000 IP addresses monitored)
  ▶ IP addresses broadly scattered in address space
ISC Sensors

Currently, ISC collects packet filter (firewall) logs.

- logs primarily contain failed connection attempts
- over 2,000 organizations and individuals participate
- logs typically uploaded hourly

Sample Packet Filter Log

<table>
<thead>
<tr>
<th>Date and Time</th>
<th>Source IP</th>
<th>Source Port</th>
<th>Dest. IP</th>
<th>Dest. Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/04/05 10:32:15</td>
<td>209.237.231.200</td>
<td>1956</td>
<td>64.15.205.183</td>
<td>132</td>
</tr>
<tr>
<td>1/04/05 10:30:41</td>
<td>216.187.103.168</td>
<td>4659</td>
<td>169.229.60.105</td>
<td>80</td>
</tr>
<tr>
<td>1/04/05 10:30:02</td>
<td>24.177.122.32</td>
<td>3728</td>
<td>216.187.103.169</td>
<td>194</td>
</tr>
<tr>
<td>1/04/05 10:28:24</td>
<td>24.168.152.10</td>
<td>518</td>
<td>209.112.228.200</td>
<td>1027</td>
</tr>
</tbody>
</table>
ISC Analysis and Reports

The ISC publishes several types of reports and statistics - we focus on the “port reports.”

Port Reports

- port reports list the amount of activity on each destination port
- this type of report is typical of the reports published by Internet sensor networks in general

Sample Port Report

<table>
<thead>
<tr>
<th>Port</th>
<th>Reports</th>
<th>Sources</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>325</td>
<td>99321</td>
<td>65722</td>
<td>39</td>
</tr>
<tr>
<td>1025</td>
<td>269526</td>
<td>51710</td>
<td>47358</td>
</tr>
<tr>
<td>139</td>
<td>875993</td>
<td>42595</td>
<td>180544</td>
</tr>
<tr>
<td>3026</td>
<td>395320</td>
<td>35683</td>
<td>40808</td>
</tr>
<tr>
<td>135</td>
<td>3530330</td>
<td>155705</td>
<td>270303</td>
</tr>
<tr>
<td>225</td>
<td>8657692</td>
<td>366825</td>
<td>268953</td>
</tr>
<tr>
<td>5000</td>
<td>202542</td>
<td>36207</td>
<td>37689</td>
</tr>
<tr>
<td>6346</td>
<td>2523129</td>
<td>271789</td>
<td>2558</td>
</tr>
</tbody>
</table>

Jason Franklin jfrankli@cs.wisc.edu
Procedure to Discover Monitored Addresses

Core Idea

```latex
\textbf{for} each IP address $i \ \textbf{do}$
\begin{itemize}
  \item probe $i$ with reportable activity $a$
  \item wait for next report to be published
  \item check for activity $a$ in report
\end{itemize}
\textbf{end for}
```

Details

- only one TCP packet necessary for each probe
- bandwidth requirements of sending a packet to every possible address will be addressed in discussion of simulations
Mapping Internet Sensors with Probe Response Attacks

Example Attack

Introduction to the Attack

Procedure to Discover Monitored Addresses

**Problem**

There are too many addresses to check one after another.
- most participants only submit logs to the ISC every hour
- there are about 2.1 billion valid, routable IP addresses

**Solution**

Check many addresses in parallel.
- only a small portion of addresses are monitored, so send same probe to many addresses
  - if no activity is reported they can all be ruled out
  - otherwise report reveals the number of monitored addresses
- since activity reported by port, send probes with different ports to run many independent tests at the same time

Jason Franklin jfrankli@cs.wisc.edu
Detailed Procedure: First Stage

- begin with list of 2.1 billion valid IP addresses to check
- divide up into $n$ search intervals $S_1, S_2, \ldots, S_n$
- send SYN packet on port $p_i$ to each address in $S_i$
- wait two hours and retrieve port report
- rule out intervals corresponding to ports with no activity
Detailed Procedure: Second Stage

- distribute ports among $k$ remaining intervals $R_1, R_2, \ldots R_k$
- for each $R_i$
  - divide into $\frac{n}{k} + 1$ subintervals
  - send a probe on port $p_j$ to each address in the $j$th subinterval
  - not necessary to probe last subinterval (instead infer number of monitored addresses from total for interval)
  - if subinterval full, add to list and discard
- repeat second stage with non-empty subintervals until all addresses are marked as monitored or unmonitored
Example Run With Six Ports

Stage 1

Stage 2
External Activity

Problem

What if other activity is present in port reports? External activity may be considered noise which obscures the signal in the port reports.

Solution

Use a noise cancellation technique.

- use ports that consistently have less than $k$ reports per time interval
- send $k$ SYN packets in each probe
- use the “reports” field of the port report
- divide number of reports by $k$ and round down
Attack Simulation Overview

We provide detailed results of a simulated probe response attack on the ISC including:

- time required to complete
- number of packets sent
- attack progress (percentage of monitored addresses discovered)

Additional Simulation Results

- mapping distributions of addresses other than the ISC distribution
- consequences of a successful mapping attack
Adversarial Models

Adversarial Models for Simulation

- **T1 attacker** 1.544 Mbps of upload bandwidth
- **Fractional T3 attacker** 38.4 Mbps of upload bandwidth
- **OC6 attacker** 384 Mbps of upload bandwidth

- our algorithm is not dependent upon a particular Internet connection or attacker configuration
  - can be executed on a single machine or a distributed collection of machines (botnet)
  - time to complete is dependent only on upload bandwidth
  - does not require significant state or complete TCP connections
Attack Details

Details of fractional T3 attacker mapping the addresses monitored by the ISC.

### Probes Sent

![Graph showing the number of packets sent over time](image1)

### Attack Progress

![Graph showing the fraction of monitored addresses found over time](image2)

Jason Franklin  
jfrankli@cs.wisc.edu
Random Sensor Sets

Simulation Results

- previous simulations show that probe response attacks can map the ISC
- but what about other sets of monitored addresses

Generalized Sets of Addresses

- feasibility of mapping other sets of monitored addresses depends in part on how they are clustered
- to extend our results we work with generalized sets of address
  - generate random sets of monitored IP addresses
  - vary the degree to which the addresses are clustered
Random Sensor Sets

Clustering Model

- a “cluster” is set of sensors with sequential IP addresses
- model cluster size with Pareto distribution
- model sizes of gaps between clusters with exponential distribution

Results

- with parameters set to match actual ISC addresses, time to map is roughly the same
- with larger average cluster sizes mapping becomes easier
- with smaller average cluster sizes mapping takes longer, but remains feasible
Random Sensor Sets

Totally Random Addresses

- as an extreme case, we map a set of addresses chosen uniformly at random
  - (i.e., each address is monitored with equal probability)
- this may be considered a worst case for the attacker

Results

- attack remains feasible
- under the T3 attacker model, about 9 days necessary to map 680,000 addresses
Simulation Summary

<table>
<thead>
<tr>
<th>bandwidth</th>
<th>set of addresses</th>
<th>data sent</th>
<th>time to map</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC6</td>
<td>ISC</td>
<td>1,300GB</td>
<td>2 days, 22 hours</td>
</tr>
<tr>
<td>T3</td>
<td>ISC</td>
<td>687GB</td>
<td>4 days, 16 hours</td>
</tr>
<tr>
<td>T1</td>
<td>ISC</td>
<td>440GB</td>
<td>33 days, 17 hours</td>
</tr>
<tr>
<td>T3</td>
<td>average cluster size ≥ 10</td>
<td>~ 600GB</td>
<td>~ 2 days</td>
</tr>
<tr>
<td>T3</td>
<td>average cluster size ~ 1.6</td>
<td>~ 1,100GB</td>
<td>~ 8 days</td>
</tr>
<tr>
<td>T3</td>
<td>totally random</td>
<td>~ 860GB</td>
<td>~ 9 days</td>
</tr>
</tbody>
</table>

Key Simulation Results

Probe response attacks are a serious threat.
- both a real set of monitored IP addresses and various synthetic sets can be mapped in reasonable time
- attacker capabilities determine efficiency, but mapping is possible even with very limited resources
## Results of Successful Attack

### Consequences

The consequences of an attacker successfully mapping the addresses monitored are severe.

- attacker may avoid monitored addresses in malicious activities (e.g., port scanning)
- worms may avoid monitored addresses and go undetected
- sensors may be flooded with errant data

### Recovery

- very difficult to recover from a successful mapping attack
- data from publicly published list of monitored addresses can not be considered an accurate picture of Internet activity.

---

*Jason Franklin*  
jfrankli@cs.wisc.edu
Covert Channels in Reports

In our attack, an attacker gains information by:

- sending probes with different destination ports to different IP addresses
- considering which ports have activity reported
- using activity reported to determine the set of IP addresses that could have possibly received probes

Probe Response Attack Covert Channel

In this way, the destination port appearing in the packet sent out and later in the port reports is used by the attacker as a **covert channel in a message to themselves**.
## Example Covert Channels

### Covert Channels

- many possible fields of information appearing in reports are suitable for use as covert channels
- characteristics of attacks or probes may be reported in almost any field which an attacker can influence
- using covert channels an attacker can encode partial information about a destination IP address in a packet

### Example Fields

- Time / date
- Source IP
- Source port
- Destination subnet
- Destination port
- Captured payload data
Symantec’s DeepSight
Other Networks

Symantec’s DeepSight

- reports include time, source IP and port, destination port, and number of other sensors affected by attack
- requires attacker to submit a log containing each unique probe
- easily mapped by encoding destination IP address in source IP address of probe

Simulation Results

<table>
<thead>
<tr>
<th>network</th>
<th>bandwidth</th>
<th>probes sent</th>
<th>time to map</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepSight</td>
<td>-</td>
<td>2.1 billion</td>
<td>single pass of probes</td>
</tr>
<tr>
<td>myNetWatchman</td>
<td>-</td>
<td>2.1 billion</td>
<td>single pass of probes</td>
</tr>
<tr>
<td>SANS ISC</td>
<td>T3</td>
<td>14 billion</td>
<td>4 days 16 hours</td>
</tr>
</tbody>
</table>

Jason Franklin jfrankli@cs.wisc.edu
Symantec’s DeepSight Report

Jason Franklin  jfrankli@cs.wisc.edu

Mapping Internet Sensors with Probe Response Attacks

Generalizing the Attack

Other Networks
Current Countermeasures

- Hashing, Encryption, and Permutations
  - simply hashing report fields is vulnerable to dictionary attack
  - encrypting a field with a key not publicly available is effective, but reduces utility of fields
  - prefix-preserving permutations obscure IP addresses while still allowing useful analysis

- Bloom Filters
  - allow for space efficient set membership tests
  - configurable false positive rate
  - vulnerable to iterative probe response attacks as a result of the exponentially decreasing number of false positives

These current methods of anonymization do not prevent probe response attacks.
Information Limiting

One approach to prevent probe response attacks is to limit the information provided in public reports in some way.

- private reports
  - eliminate public reports entirely
  - effective, but severely limits utility of network
- top lists
  - only publish most significant events
  - provides some useful information, but not complete picture of Internet phenomena
  - may allow attackers to consistently avoid detection by keeping their activity below thresholds
- query limiting
  - slow queries against public reports
  - may require monetary payment, computational puzzle, or CAPTCHA to perform query
  - will only slow down mapping attacks
Sampling Countermeasure

Random Input Sampling Technique

Randomly sample the logs coming into the analysis center before generating reports to increase the probability of false negatives.

For example:

- suppose an analysis center discards every log it receives with probability $\frac{4}{5}$
- large scale phenomena such as worm outbreaks and port scanning should remain visible in the reports
- however, a probe response attack becomes more difficult because the probability of a single probe resulting in a false negative for the attacker would be $\frac{4}{5}$
Sampling Countermeasure

Overcoming Random Input Sampling

- to reduce the probability of a false negative, the attacker would need to send multiple probes.
- for instance, to reduce the false negative rate of \( \frac{4}{5} \) to 1%, an attacker would need to a twenty-fold increase in bandwidth.

Jason Franklin  
jfrankli@cs.wisc.edu
Scan Prevention

IPv6

- increases IP addresses from 32 bits to 128 bits
- greatly reduces the feasibility of TCP/UDP scanning
- effective countermeasure if deployed correctly
- widespread adoption is out of our control
Delayed Reporting

Another strategy in preventing mapping is delaying the publication of public reports.

- publish reports reflecting old data (e.g., last week’s data)
- forces attacker to either wait a long period between iterations of attack or use non-adaptive algorithm
- a sufficiently long delay will make an adaptive attack infeasible
- non-adaptive (or offline) algorithms do not base the probes of the current rounds on previous rounds
  - much larger search space
  - likely to use many more probes and take much longer
  - more detailed investigation remains as future work
- delaying reports greatly reduces effectiveness of Internet sensor network in providing real-time notification of new phenomena
Eliminating Inadvertent Exposure

- publishing information about the specific distribution of addresses monitored by an Internet sensor network
- aids attacker by reducing the number of probes necessary
- if a sensor network publishes the fact that they monitor a /8, the number of probes required for an attack drop from around 8 billion to 256 probes

Sample Distribution

<table>
<thead>
<tr>
<th>Organization</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional ISP</td>
<td>/24, /24</td>
</tr>
<tr>
<td>Large Enterprise</td>
<td>/18</td>
</tr>
<tr>
<td>Academic Network</td>
<td>/22, /23</td>
</tr>
<tr>
<td>National ISP</td>
<td>/8</td>
</tr>
<tr>
<td>Broadband Provider</td>
<td>/17, /22, /23</td>
</tr>
</tbody>
</table>
Conclusion

- Internet sensor networks monitor the health of the Internet.
- Secrecy of the monitored addresses is essential to the effectiveness of the sensor network.
- Probe response attacks can be used to quickly and efficiently locate Internet sensors.
- Scan prevention, sampling, and limited and delayed reporting can be effective countermeasures against probe response attacks.

Final Advice

Internet sensor networks should be designed to resist probe response attacks.
Questions?

Related Work


Resources for Further Information

**USENIX Security ’05**  
“Mapping Internet Sensors with Probe Response Attacks” by John Bethencourt, Jason Franklin, and Mary Vernon.

**CIPART Project**  
http://www.cs.wisc.edu/~vernon/cipart.html

**Presentation Slides**  
http://www.cs.wisc.edu/~jfrankli/

Coauthor Information

- **John Bethencourt**  
  Affiliation: University of Wisconsin, Madison  
  Email: bethenco@cs.wisc.edu

- **Professor Mary Vernon**  
  Affiliation: University of Wisconsin, Madison  
  Email: vernon@cs.wisc.edu