

# Mapping Internet Sensors with Probe Response Attacks

John Bethencourt, Jason Franklin, and Mary Vernon

`{bethenco, jfrankli, vernon}@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
- ▶ collaborative intrusion detection systems
- ▶ honeypots and honeynets
- ▶ Internet sinks and network telescopes

## 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
- ▶ noticing attacks before the majority of vulnerable systems are compromised
- ▶ classifying the pervasiveness of threats like port scans, DoS attacks, and botnet activity

## A Few Example Systems

### Example Internet Sensor Networks

- ▶ SANS Internet Storm Center
- ▶ Symantec's DeepSight network
- ▶ myNetWatchman
- ▶ University of Michigan's Internet Motion Sensor
- ▶ Distributed Overlay for Monitoring InterNet Outbreaks (DOMINO) system
- ▶ Cooperative Association for Internet Data Analysis (CAIDA)
- ▶ University of Wisconsin's iSink

# 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

We introduce a new class of attacks called probe response attacks which are 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.

## 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

Date and Time	Source IP	Source Port	Dest. IP	Dest. Port
1/04/05 10:32:15	209.237.231.200	1956	64.15.205.183	132
1/04/05 10:30:41	216.187.103.168	4659	169.229.60.105	80
1/04/05 10:30:02	24.177.122.32	3728	216.187.103.169	194
1/04/05 10:28:24	24.168.152.10	518	209.112.228.200	1027

## 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

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

## Procedure to Discover Monitored Addresses

### Core Idea

- ▶ probe an IP address with activity that will be reported if the address is monitored
- ▶ wait for next report to be published, check for the activity, and decide whether the address was monitored
- ▶ repeat for every IP address

### 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

## Procedure to Discover Monitored Addresses

### Problem

There are too many addresses to check one after another.

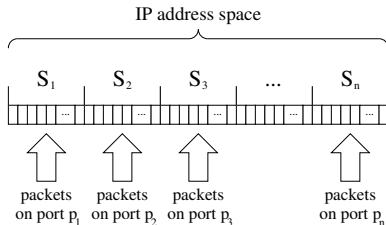
- ▶ most participants only submit logs to the ISC every hour
- ▶ there are about 2.1 billion valid, routable IP addresses

### Solution

Check many in parallel. This is possible for several reasons.

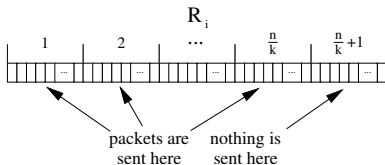
- ▶ only a very small portion of addresses are monitored, so send same probe to many addresses
  - ▶ if no activity is reported they can all be ruled out
  - ▶ otherwise report reveals the number of monitored addresses
- ▶ since activity reported by port, send probes with different ports to run many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

## Detailed Procedure: First Stage



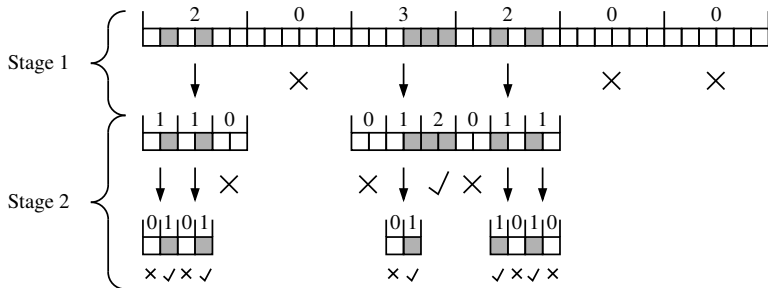
- ▶ begin with list of 2.1 billion valid IP addresses to check
- ▶ divide up into  $n$  search intervals  $S_1, S_2, \dots, S_n$
- ▶ send SYN packet on port  $p_i$  to each address in  $S_i$
- ▶ wait two hours and retrieve port report
- ▶ rule out intervals corresponding to ports with no activity

## Detailed Procedure: Second Stage



- ▶ distribute ports among  $k$  remaining intervals  $R_1, R_2, \dots, R_k$
- ▶ 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



## 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

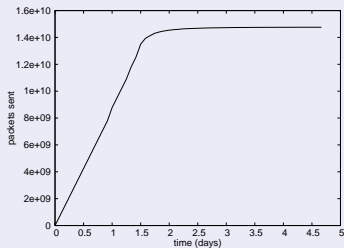
### Adversarial Models for Simulation

- ▶ **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

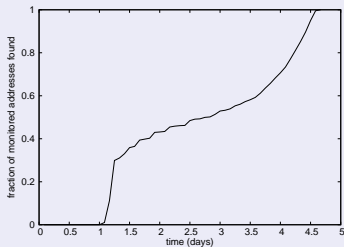
## Attack Details

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

### Probes Sent



### Attack Progress



## Random Sensor Sets

### Question

The previous simulations showed that probe response attacks could map the ISC in practice. But what about other sets of monitored IP addresses? The feasibility may depend on how clustered they are.

### Simulate Mapping Other Sets

We generate random sets of monitored IP addresses and determine how quickly they may be mapped. By varying the degree to which the addresses are clustered, we may determine if probe response attacks may also effectively map other Internet sensor networks.

## 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 also simulate mapping a set of addresses with no special clustering (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

bandwidth	set of addresses	data sent	time to map
OC6	ISC	1,300GB	2 days, 22 hours
T3	ISC	687GB	4 days, 16 hours
T1	ISC	440GB	33 days, 17 hours
T3	average cluster size $\geq 10$	$\sim 600$ GB	$\sim 2$ days
T3	average cluster size $\sim 1.6$	$\sim 1,100$ GB	$\sim 8$ days
T3	totally random	$\sim 860$ GB	$\sim 9$ days

### 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

It is very difficult to recover from a successful mapping attack. If the list of monitored addresses was published publicly, data from those addresses could never again be considered an accurate picture of Internet activity.

## 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

## 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

<b>network</b>	<b>bandwidth</b>	<b>probes sent</b>	<b>time to map</b>
DeepSight	-	2.1 billion	single pass of probes
myNetWatchman	-	2.1 billion	single pass of probes
SANS ISC	T3	14 billion	4 days 16 hours

## 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 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.

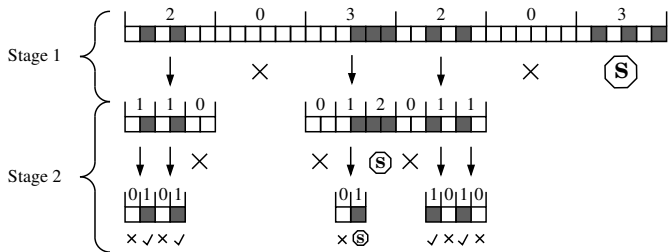
For example:

- ▶ 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

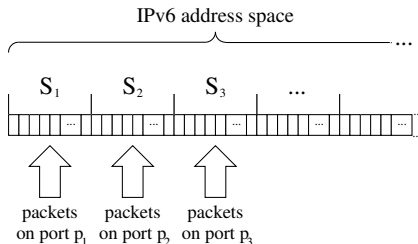
- ▶ to reduce the probability of a false negative, the attacker would need to send multiple probes
- ▶ for instance, to reduce the false negative rate of  $\frac{4}{5}$  to 1%, an attacker would need a twenty-fold increase in bandwidth



## Scan Prevention

### IPv6

- ▶ increases IP addresses from 32 bits to 128 bits
- ▶ greatly reduces the feasibility of TCP/UDP scanning
- ▶ effective countermeasure if widely adopted
- ▶ 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

## 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

Organization	Size
Regional ISP	/24, /24
Large Enterprise	/18
Academic Network	/22, /23
National ISP	/8
Broadband Provider	/17, /22, /23

## 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

- ▶ “Privacy-Preserving Sharing and Correlation of Security Alerts” by Lincoln, Porras, and Shmatikov. *Proceedings of the 13th USENIX Security Symposium*, 2004.
- ▶ “A High-level Programming Environment for Packet Trace Anonymization and Transformation” by Pang and Paxson. *Proceedings of SIGCOMM 2003*, August 2003.

### 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>