Mapping Internet Sensors with Probe Response Attacks

Jason Franklin jfrankli@cs.wisc.edu

Computer Sciences Department
University of Wisconsin, Madison

Outline

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

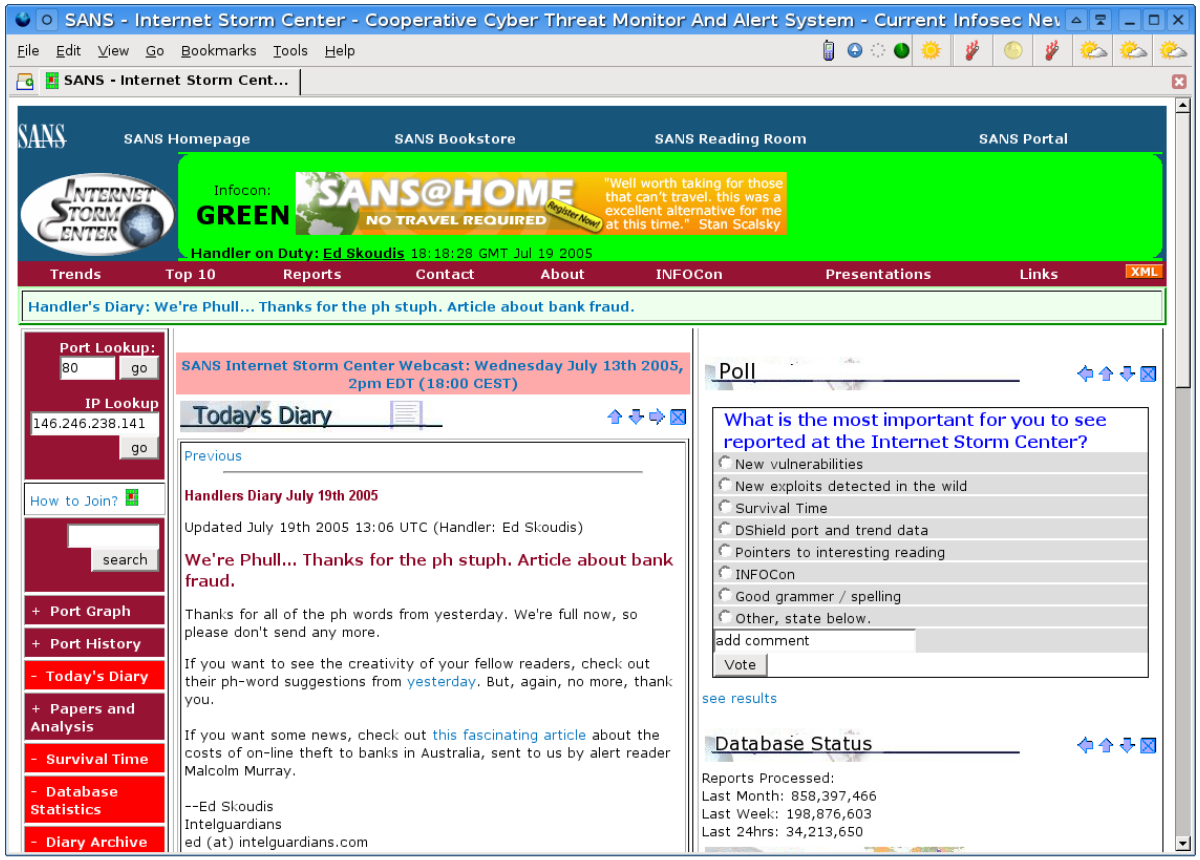
Probe Response Attacks

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

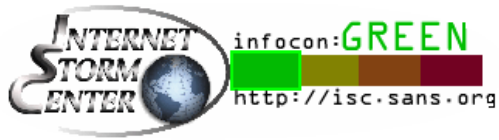
Countermeasures

We also provide countermeasures which are effective in preventing probe response attacks.

SANS Internet Storm Center



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

for each IP address *i* **do**
  probe *i* with reportable activity *a*
  wait for next report to be published
  check for activity *a* in report
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

Procedure to Discover Monitored Addresses

Problem

There are too many addresses to check one after another.

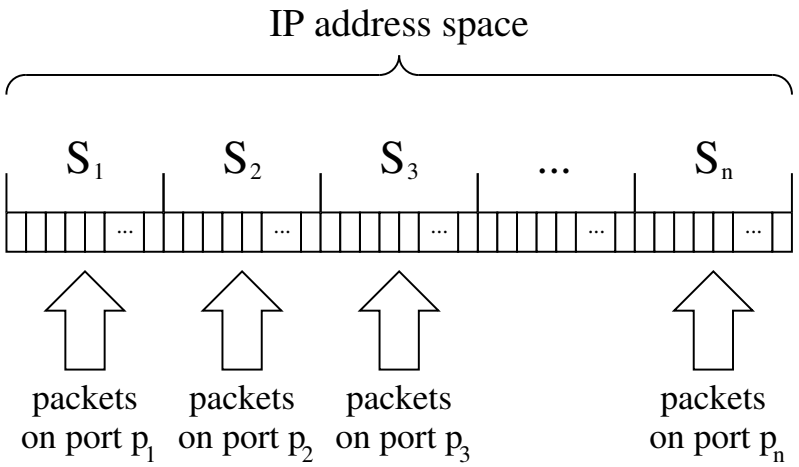
- most participants only submit logs to the ISC every hour
- there are about 2.1 billion valid, routable IP addresses

Solution

Check many addresses in parallel.

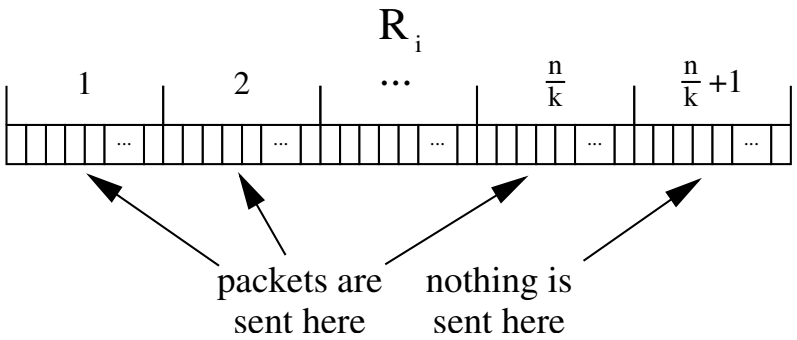
- only a small portion of addresses are monitored, so send same probe to many addresses
  - if no activity is reported they can all be ruled out
  - otherwise report reveals the number of monitored addresses
- since activity reported by port, send probes with different ports to run many independent tests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

Detailed Procedure: First Stage



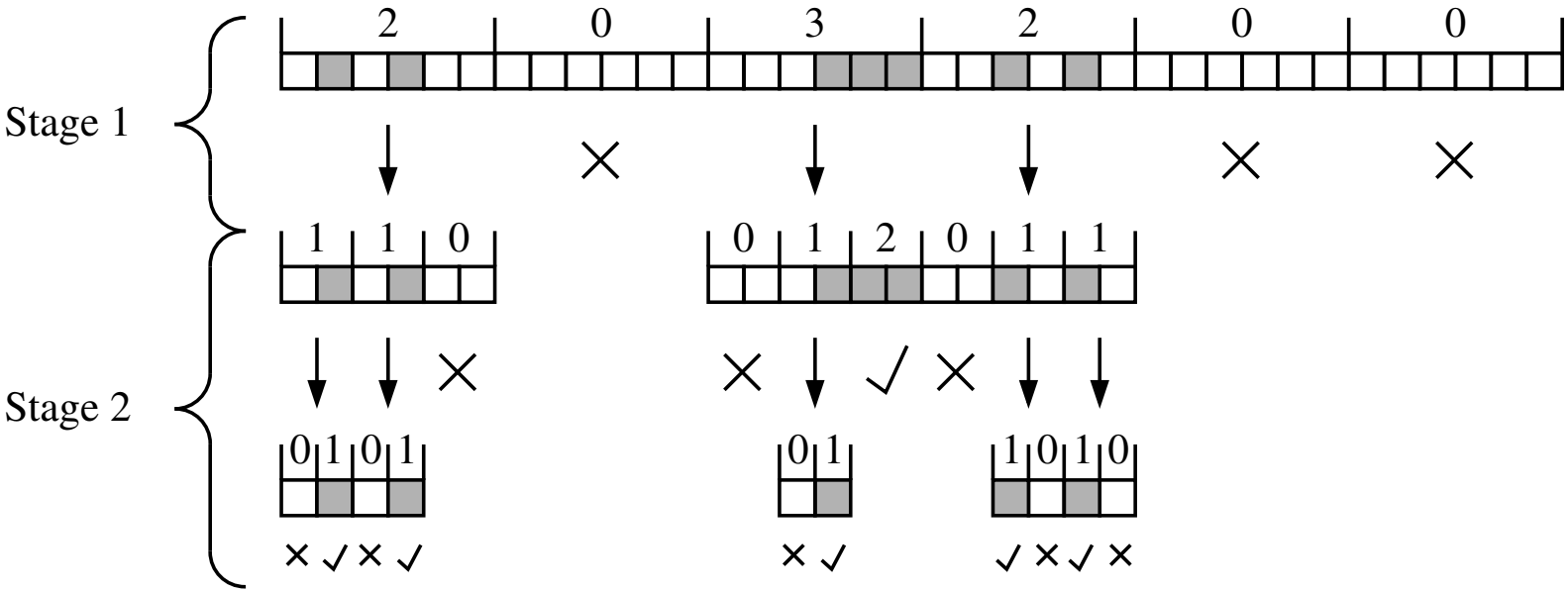
- begin with list of 2.1 billion valid IP addresses to check
- divide up into n search intervals $S_1, S_2, \ldots S_n$
- send SYN packet on port p_i to each address in S_i
- wait two hours and retrieve port report
- rule out intervals corresponding to ports with no activity

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 jth 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

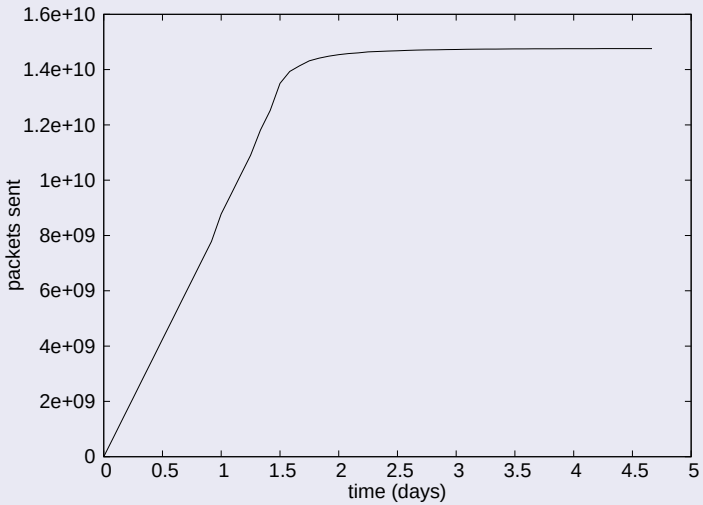
- **T1 attacker** 1.544 Mbps of upload bandwidth
- **Fractional T3 attacker** 38.4 Mbps of upload bandwidth
- **OC6 attacker** 384 Mbps of upload bandwidth

- 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

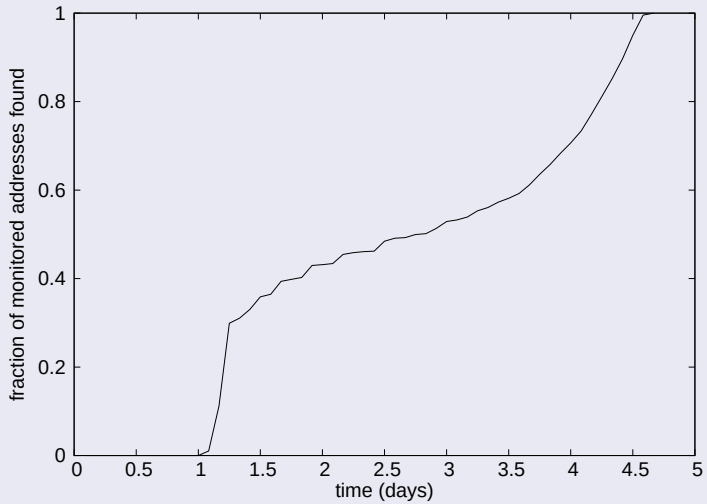
Attack Details

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

Probes Sent



Attack Progress



Random Sensor Sets

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 choosen uniformally 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

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 ≥ 10	$\sim$ 600GB	$\sim$ 2 days
T3	average cluster size ~ 1.6	$\sim 1,100$GB	$\sim$ 8 days
T3	totally random	$\sim$ 860GB	$\sim$ 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

- 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.

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

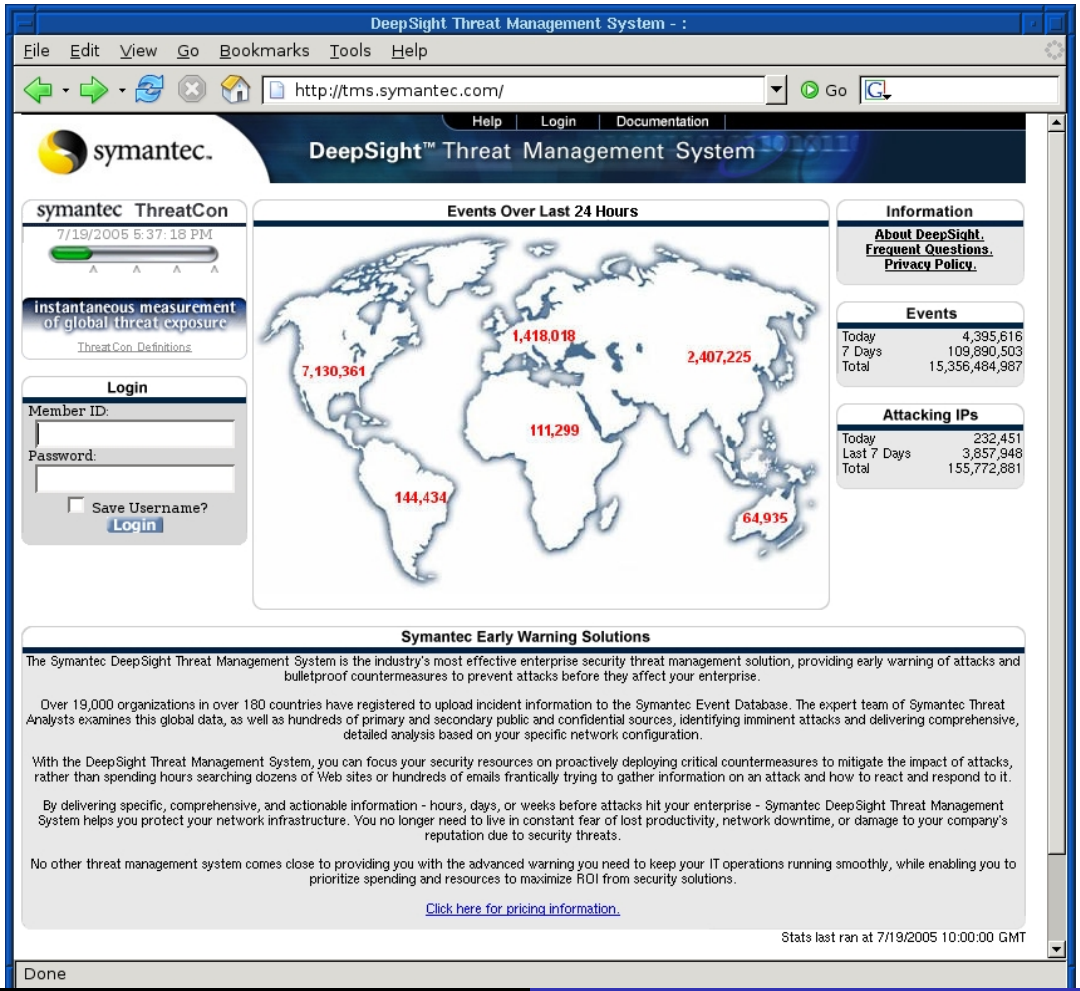
Covert Channels

- 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

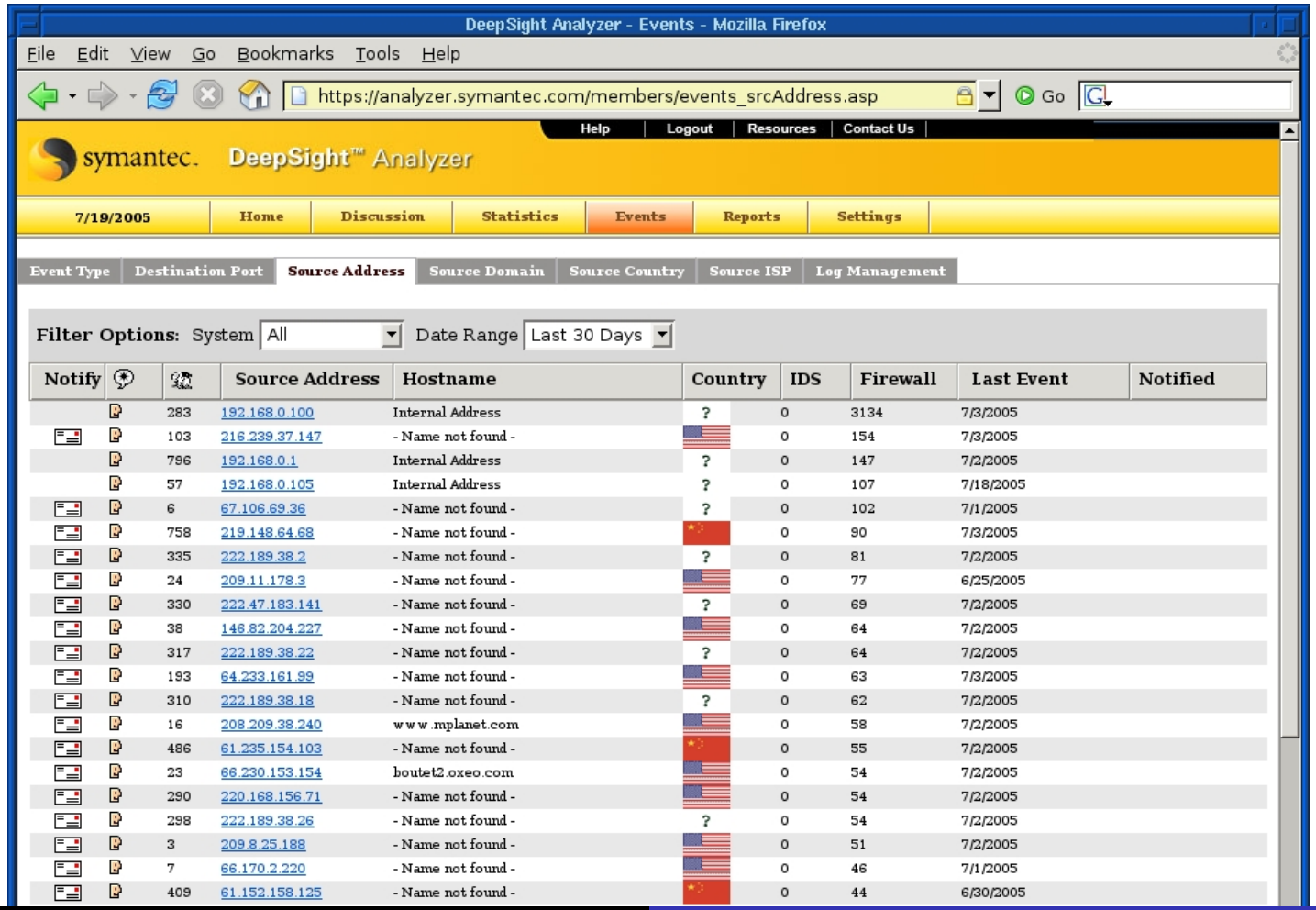
Symantec's DeepSight

- 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

network	bandwidth	probes sent	time to map
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

Symantec's DeepSight Report



Notify		Source Address	Hostname	Country	IDS	Firewall	Last Event	Notified
	283	192.168.0.100	Internal Address	?	0	3134	7/3/2005	
	103	216.239.37.147	- Name not found -		0	154	7/3/2005	
	796	192.168.0.1	Internal Address	?	0	147	7/2/2005	
	57	192.168.0.105	Internal Address	?	0	107	7/18/2005	
	6	67.106.69.36	- Name not found -	?	0	102	7/1/2005	
	758	219.148.64.68	- Name not found -		0	90	7/3/2005	
	335	222.189.38.2	- Name not found -	?	0	81	7/2/2005	
	24	209.11.178.3	- Name not found -		0	77	6/25/2005	
	330	222.47.183.141	- Name not found -	?	0	69	7/2/2005	
	38	146.82.204.227	- Name not found -		0	64	7/2/2005	
	317	222.189.38.22	- Name not found -	?	0	64	7/2/2005	
	193	64.233.161.99	- Name not found -		0	63	7/3/2005	
	310	222.189.38.18	- Name not found -	?	0	62	7/2/2005	
	16	208.209.38.240	www.mplanet.com		0	58	7/2/2005	
	486	61.235.154.103	- Name not found -		0	55	7/2/2005	
	23	66.230.153.154	boutet2.oxeo.com		0	54	7/2/2005	
	290	220.168.156.71	- Name not found -		0	54	7/2/2005	
	298	222.189.38.26	- Name not found -	?	0	54	7/2/2005	
	3	209.8.25.188	- Name not found -		0	51	7/2/2005	
	7	66.170.2.220	- Name not found -		0	46	7/1/2005	
	409	61.152.158.125	- Name not found -		0	44	6/30/2005	

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 resonse 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.

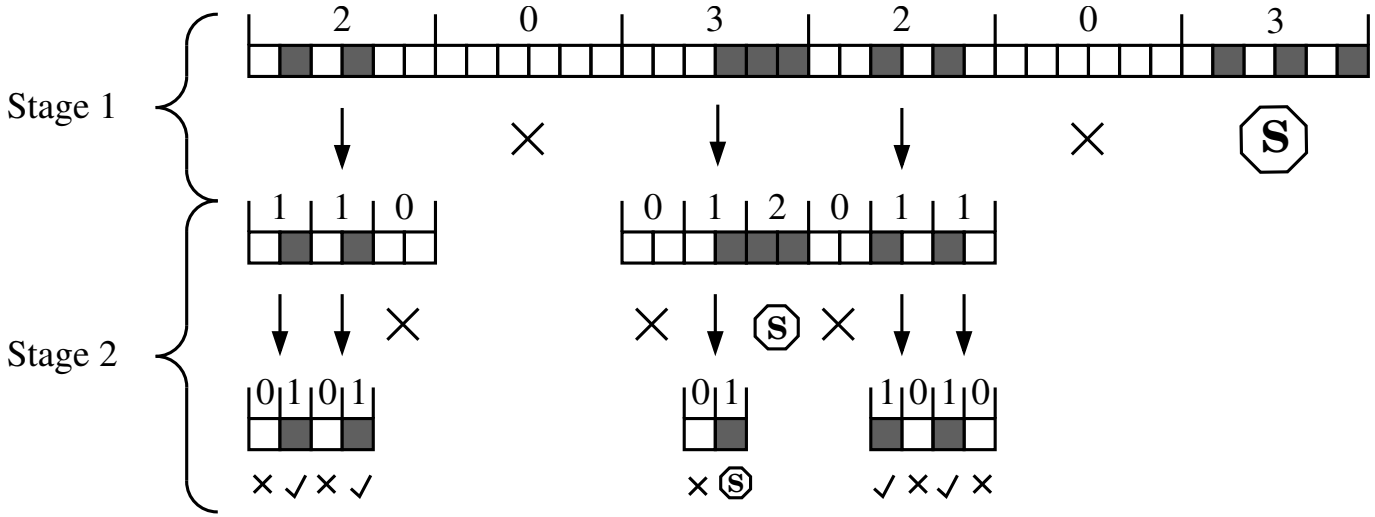
For example:

- 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

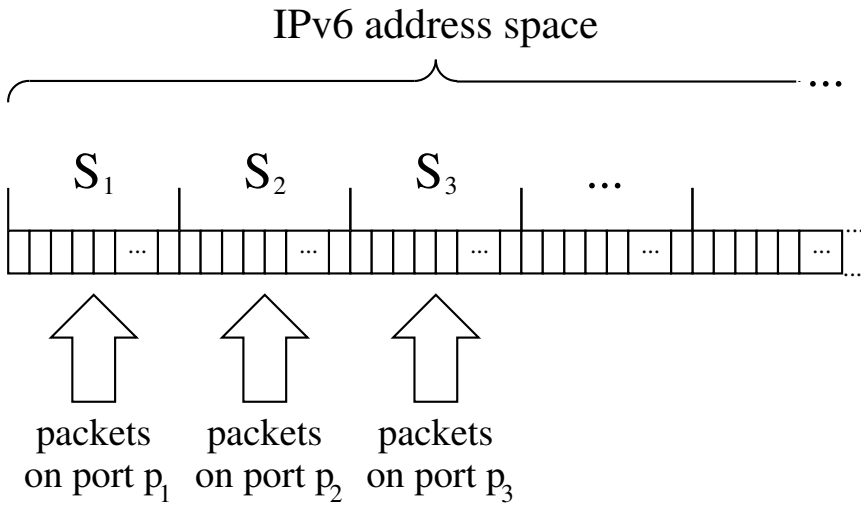
- to reduce the probability of a false negative, the attacker would need to send multiple probes
- for instance, to reduce the false negative rate of $\frac{4}{5}$ to 1%, an attacker would need to a twenty-fold increase in bandwidth



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

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.
- "Vulnerabilities of Passive Internet Threat Monitors" by Yoichi Shinoda, Ko Ikai, Motomu Itoh. *Proceedings of the 14th USENIX Security Symposium*, August 2005.

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