Privacy-Preserving Sharing and Analysis

Patrick Lincoln
SRI International
Outline

- **Context**
  - Information security, internet-scale defense

- **Requirements**
  - Preserve privacy
  - Enable aggregation and correlation
  - Practical performance

- **Approach**
  - Encryption, hashing, anonymization
Main Point

- Enable interesting interfaces to data
  - Protected by mathematics (cryptography)
  - Not just human trust (though that’s very useful too)
General Goal

Enable enterprises to share information for sharing, regulatory and business purposes while maintaining a high level of security and organizational privacy.
Privacy-Preserving Sharing and Correlation of Information Security Alerts

Patrick Lincoln
Phil Porras and Vitaly Shmatikov
SRI International
Context: Information Security

- Networked information assets
  - Huge and growing value of information on networks
    - JV2020 predicates national security on information superiority
- Defense against
  - Bots
  - Email viruses
  - Worms
- The scale of the problem is large and growing rapidly
  - Individual attacks taking over 100,000's machines
  - More and more successful attacks
Examples of Past Threats

- Attacks of growing sophistication
  - Control, speed, infection success %
- Already serious consequences
  - Though fortunately, generally benign payloads

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>OS</th>
<th>Service</th>
<th>Infected Machines</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lion</td>
<td>March 2001</td>
<td>Linux</td>
<td>BIND</td>
<td>10,000?</td>
<td>Days</td>
</tr>
<tr>
<td>Code Red</td>
<td>July 2001</td>
<td>Windows</td>
<td>IIS</td>
<td>200-400k</td>
<td>Days</td>
</tr>
<tr>
<td>Nimda</td>
<td>Oct 2001</td>
<td>Windows</td>
<td>IIS</td>
<td>100-200k</td>
<td>Hours</td>
</tr>
<tr>
<td>SQL Slammer</td>
<td>January 2003</td>
<td>Windows</td>
<td>IIS</td>
<td>100-200k</td>
<td>Minutes</td>
</tr>
<tr>
<td>MSBlaster</td>
<td>August 2003</td>
<td>Windows</td>
<td>IIS</td>
<td>300k?</td>
<td>Hours</td>
</tr>
</tbody>
</table>
Potential Future Threats

- **Worhol worms**
  - Famous for 15 minutes
- **Flash worms**
  - Even shorter timespans of visible attack

- These are too fast for human-oriented response mechanisms to address
Why Are Attackers More Successful?

- More systems being created
- More features being added
- More vulnerabilities
- More instances
- More networks

- Practice of sharing and building on others work
Out-Share or Lose

- Attackers and their machines share vulnerability information (rapidly, efficiently)
  - Virus writers, worm writers, crackers, script kiddies, etc.
  - Email, chat rooms, IRC, web sites, p2p, IM, etc.
  - Network effect works to their advantage
    - Build on others work

- Currently many defenders (network administrators) do not share detailed information automatically with others outside their organization
  - Precious little. Ponderously slow publication cycle
Attackers Share Openly

- Open-source code sharing
- Large email networks
- Chat rooms

Example: SDbot
  - Toolkit for building IRC-controlled bot armies
  - Rapid community uptake, extension of functionality, exploitation
Defenders Do Not Share As Openly

- Network defense viewed as local responsibility
- Individual sites defend themselves (only)
- Works OK against low or moderate levels of attack

- Internet-scale threats not well addressed in this mode
Defenders Do Not Share As Openly

- Sharing security alert information compromises organizational privacy and other business interests
  - Network topology
  - Network defenses
    - Which IDS, configured how, on which subnets?
    - Which Firewall, configured how?
  - Network monitoring capabilities
  - Customer information
    - Attacks come through ‘normal’ channels of communication
- Sharing takes time and energy
This is Complex Problem

- At least partially sociological
- Perhaps there are technological solutions that can play a role
- If it was easier and safer to share, more good guys would share more data more often
- Need to give defenders tools
  - Theories
  - Implementations
  - Accepted practices
that make their life easier
Starting to Get Technical

- What can the theoretical information security community do to help?
- How can we enable security alert sharing without compromising privacy interests?
Requirements

● Security alerts gathered from thousands of security sensors
  - Firewalls, Network IDS systems, Host-based IDS systems, Antivirus filters

● Network defenders (the good guys) need to perform coarse groupings of alerts
  - Geographic region, industry group, OS type, port

● Network defenders need to perform fine-grained correlation of certain alerts
  - Patterns of alerts highlights new attacks, vulnerabilities, and potential defenses
Example: DShield.org

Distributed Intrusion Detection System

DShield.org

Records Added

Last Month: 294,679,239
Last Week: 116,077,068
Today: 15,034,132

As of Tue May 25 10:36:56 2004 UTC

Internet Storm Center Status
green: Akamai Problems, New Angle(r) On An Old Phish
stop | start ticker
1027| 445| 80| 1026| 3128| 25| 113| 4672| 139

(ISC Daily Trends Page)

Top Attacker: 211.157.101.25
Most Attacked Port: 135

Geographic Distribution of attack sources. Last days

http://www.dshield.org
DShield.org
Distributed Intrusion Detection System

ZoneLog Analyser
Sorts the Threats from the Noise

http://www.dshield.org
The Internet Traffic Report monitors the flow of data around the world. It then displays a value between zero and 100. Higher values indicate faster and more reliable connections.

The Internet Traffic Report (ITF) wants to continue to provide useful information about networks from around the world. We want to make this information as accurate as possible.

View Graphs or Click a Continent to view more detailed information.

<table>
<thead>
<tr>
<th>Continent</th>
<th>Current Index</th>
<th>Avg. Response Time (ms)</th>
<th>Avg. Packet Loss (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>59</td>
<td>393</td>
<td>17 %</td>
</tr>
<tr>
<td>Australia</td>
<td>76</td>
<td>230</td>
<td>0 %</td>
</tr>
<tr>
<td>Europe</td>
<td>81</td>
<td>174</td>
<td>1 %</td>
</tr>
<tr>
<td>North America</td>
<td>90</td>
<td>84</td>
<td>1 %</td>
</tr>
</tbody>
</table>
The Symantec DeepSight Threat Management System is the industry’s most effective enterprise security threat management solution, providing early warning of attacks and bulletproof countermeasures to prevent attacks before they affect your enterprise.

Over 19,000 organizations in over 180 countries have registered to upload incident information to the Symantec Event Database. The expert team of Symantec Threat Analysts examines this global data, as well as hundreds of primary and secondary public and confidential sources, identifying imminent attacks and delivering comprehensive, detailed analysis based on your specific network configuration.

With the DeepSight Threat Management System, you can focus your security resources on proactively deploying critical countermeasures to mitigate the impact of attacks, rather than spending hours searching dozens of Web sites or hundreds of emails frantically trying to gather information on an attack and how to react and respond to it.
Deepsight (from their website)

- Monitors vulnerabilities in more than 18,000 technologies, operating systems, and application product versions from 2,200 vendors,
- Vulnerabilities monitored 24x7
- Enables secure, Web-based queries to an industry-leading vulnerability database
- Prioritized alerts
- Current and historical alerting and response reports
- Administrative user status gives control over subordinate users in order to share information, collaborate for early mitigation, and increase accountability
- Alert status tracking streamlines task assignment and reporting by providing status and documenting resolutions
Today's Diary

Handlers Diary May 23rd 2004

Updated May 24th 2004 15:21 UTC (Handler: Mike Poor)

Akamai problems. Quiet, well kinda quiet, day on the Internet

Update (Mon. May 24th 9 am EST, 13:00 UTC, 15:00 CEST)

It appears that websites that use Akamai's distribution system are currently not reachable. Security related web sites affected are symantec.com and trendmicro.com. Virus updates may fail as a result. Further details are currently not available and updates will be posted here as they become available. Thanks to Vidar Vålen for alerting us of this problem.

According to a post to NAS4U, the outage may be the result of a DDOS attack. At this point, Akamai has not ETA for a resolution.

Update 09:45 EST: Looks like some of the Akamai hosted sites start to come back.

Akamai posted this statement: " Due to a peering problem between AT&T and UUNet, a subset of UUNet users may have experienced problems accessing Akamai delivered sites between 8-10pm EDT on Saturday May 22, 2004. The problem has been fully resolved. "

http://sc.imd.nl/en/top10.php?src=5d5a2947462b0f9f9577145e5f4b2c
Is The Problem Solved?
Not Completely

● In order to use these services, one has to trust these organizations
  – Must trust them not to reveal private information
  – Must trust their sysadmins, janitorial staff, etc.
● Their privacy polices may not match yours
● They do not make data globally available for analysis by researchers and other services
  – Because their business model doesn’t allow that
  – Because that would compromise privacy of submitted alerts
Privacy Requirements

- No public release of information about vulnerabilities
  - May invite and enable more attacks
- No public release of information about defenses
  - May enable attackers to adapt and avoid defenses
- No public admission of successful attack
  - May compromise public confidence
- No public release of information even about network topology
In Addition

- Don’t require (too much) trust in a repository
- Don’t require unreasonably large computational or communications overhead
  - Shared computation
  - Secure multiparty computation
  - Threshold cryptography
Are These Requirements in Conflict?

- At least partially...

- Any information gleaned from exposed reports may be viewed as a privacy breech

- Need to agree on acceptable disclosures
Specific Question

Can we partially resolve the tension between sharing information security alert information and the maintenance of organizational privacy?
Proposed Approach

- Cleanse alerts
- Share anonymously
- Protect alert repositories
Proposed Approach

- Protect privacy
  - Encrypt certain fields of alerts
  - Anonymize reports where appropriate
  - Route alerts anonymously

- Enable correlation of alerts
  - Exploit “weaknesses” in cryptographic and hashing algorithms
    - Malleability, etc
  - Make certain properties of alerts checkable
## Example Security Alert Content

<table>
<thead>
<tr>
<th>Field</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source ipAddress</td>
<td>FW, ID</td>
<td>Typically refers to the source IP address of the machine that initiated the session or transferred the transaction that caused the alert to fire. In IDS alerts, this field may represent the victim, not the attacker, since some systems alert upon an attack reply rather than request.</td>
</tr>
<tr>
<td>Source Port</td>
<td>FW, ID</td>
<td>Source TCP or UDP port of the machine that initiated the session or transferred the transaction that caused the alert to fire.</td>
</tr>
<tr>
<td>Destination IP</td>
<td>FW, ID, AV</td>
<td>Typically refers to the destination IP address of the machine that initiated the session or transferred the transaction that caused the alert to fire. In AV systems, Dest_IP can identify the machine in which the infection is discovered.</td>
</tr>
<tr>
<td>Destination Port</td>
<td>FW, ID</td>
<td>Destination TCP or UDP port of the machine that initiated the session or transferred the transaction that caused the alert to fire.</td>
</tr>
<tr>
<td>Protocol</td>
<td>FW, ID</td>
<td>Protocol type (e.g., UDP, TCP, ICMP).</td>
</tr>
<tr>
<td>Timestamp</td>
<td>FW, ID, AV</td>
<td>May incorporate incident start time, end time, incident report time.</td>
</tr>
<tr>
<td>SensorID</td>
<td>FW, ID, AV</td>
<td>May incorporate the brand and model of the sensor and a unique identifier for the individual instantiation of the sensor.</td>
</tr>
<tr>
<td>Count</td>
<td>FW, ID, AV</td>
<td>Often used to represent some notion of repeated activity, either at the alert or event (e.g., packet) level.</td>
</tr>
<tr>
<td>EventID</td>
<td>FW, ID, AV</td>
<td>Uniquely defines the alert type for the given sensor.</td>
</tr>
<tr>
<td>Outcome</td>
<td>FW, ID, AV</td>
<td>Reports the status or disposition of the reported activity. For firewalls, it may report whether the log entry was associated with an allow or deny rule. For AV, it may indicate infection disposition (e.g., Symantec’s AV indicates whether the infected file is cleaned or quarantined). Outcome fields for IDS tools are highly vendor-specific.</td>
</tr>
<tr>
<td>Captured Data</td>
<td>ID</td>
<td>Some IDS sensors have the ability to report part or all of the data content in which the alert was applied.</td>
</tr>
<tr>
<td>Infected File</td>
<td>AV</td>
<td>Antivirus logs include the identity of the file that was infected.</td>
</tr>
</tbody>
</table>
Basic Privacy Protection

● Scrubbing sensitive fields
  - Infected file
  - Outcome
  - Captured data
    ▪ In future, hope to enable analysis, for now, scrub

● Hiding IP addresses
  - Source_IP
  - Dest_IP
Hiding Information

- Encrypting IP addresses under private key is unacceptable
  - Prevents correlation of attacks
- Hashing under SHA-1 or MD5 has issues
  - Enables dictionary attacks
  - Attacker could precompute hash values of subset of network, monitor repository for hits
  - Only 65536 or 256 addresses (or smaller)
- Solution: balance privacy and utility
  - Own network use keyed hash
  - External networks use standard hash function
Abstractly

● Want to enable (approximate) equality tests under hash or encryption, but not enable simple dictionary attacks, etc

● Decreased functionality is acceptable
  - Matching Source_IP in one attack to Dest_IP address in another may not be simple
Other Privacy Protection

- **Re-keying by repository**
- **Randomized hot list thresholds**
  - For collaborative detection of high-volume alerts, it is sufficient for the repository to publish only the hot list of reported alerts that have something in common where the number exceeds some threshold
    - Vulnerable to flooding attacks
    - Defense includes random thresholds that the attacker cannot predict
- **Controlled delay alert publication**
  - For reports on historical (hours) trends, delayed alert publication provides feasible defense against probe-response attacks
Example Firewall Alert Sanitization

<table>
<thead>
<tr>
<th>Field ID</th>
<th>Raw firewall alert</th>
<th>Sanitized firewall alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source_IP</td>
<td>172.16.30.2</td>
<td>0x16e9368f</td>
</tr>
<tr>
<td>Source_Port</td>
<td>1147</td>
<td>1147</td>
</tr>
<tr>
<td>Dest_IP</td>
<td>173.19.33.1</td>
<td>0x78a65237</td>
</tr>
<tr>
<td>Dest_Port</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td>Protocol</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Timestamp</td>
<td>09032003:01:03:10</td>
<td>09032003:01:03:00</td>
</tr>
<tr>
<td>Sensor</td>
<td>PIX-4-10060231</td>
<td>PIX</td>
</tr>
<tr>
<td>Count</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Event_ID</td>
<td>Deny</td>
<td>Deny</td>
</tr>
<tr>
<td>Outcome</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Capture_Data</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Infected_File</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>
Example IDS Alert Sanitization

<table>
<thead>
<tr>
<th>Field ID</th>
<th>Raw IDS alert</th>
<th>Sanitized IDS alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source-IP</td>
<td>172.16.30.49</td>
<td>0xb09956c2</td>
</tr>
<tr>
<td>Source-Port</td>
<td>1299</td>
<td>1299</td>
</tr>
<tr>
<td>Dest-IP</td>
<td>176.20.22.43</td>
<td>0xd6e79b79</td>
</tr>
<tr>
<td>Dest-Port</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Protocol</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Timestamp</td>
<td>10132003:11:41:09</td>
<td>10132003:11:41:00</td>
</tr>
<tr>
<td>Sensor</td>
<td>EM-HTTP-90209321</td>
<td>EM-HTTP</td>
</tr>
<tr>
<td>Count</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Event-ID</td>
<td>CGI-ATTACK</td>
<td>CGI-ATTACK</td>
</tr>
<tr>
<td>Outcome</td>
<td>NO-REPLY</td>
<td>none</td>
</tr>
<tr>
<td>Capture-Data</td>
<td>/scripts/.%255c%255c./winnt/system 32/cmd.exe?/c+dir</td>
<td>none</td>
</tr>
<tr>
<td>Infected-File</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>
## Example Antivirus Alert Sanitization

<table>
<thead>
<tr>
<th>Field ID</th>
<th>Raw AV Alert</th>
<th>Sanitized AV alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source_IP</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Source_Port</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Dest_IP</td>
<td>176.30.22.11</td>
<td>0x84ddc807</td>
</tr>
<tr>
<td>Dest_Port</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Protocol</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Timestamp</td>
<td>11172003:09:39:00</td>
<td>11172003:09:39:00</td>
</tr>
<tr>
<td>Sensor</td>
<td>NORTON-AV-02209302</td>
<td>NORTON-AV</td>
</tr>
<tr>
<td>Count</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Event-ID</td>
<td>W32.Sobig.F.Dam</td>
<td>W32.Sobig.F.Dam</td>
</tr>
<tr>
<td>Outcome</td>
<td>Left alone</td>
<td>none</td>
</tr>
<tr>
<td>Capture_Data</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Infected_File</td>
<td>A0014566.pdf</td>
<td>none</td>
</tr>
</tbody>
</table>
Further Privacy Protection

- **Multiple alert repositories**
  - Sensor information can be shared directly with multiple repositories
  - Spreading alerts reduces opportunities for collaborative pattern finding, but still enables real-time monitoring

- **Randomized alert routing**
  - Overlay network for p2p routing of alerts hides sources
  - Mix networks, Crowds, or Onion routing
Alert Sharing Infrastructure
Anonymity Estimates

- With $n$ routers, if $c$ are controlled by attacker, the probability a route contains a node controlled by the attacker is $z$.

- This is close to $\frac{c}{n}$.

- So $1-z$ that many alerts are not observed by the attacker at all.

- For each of the alerts that is observed by the attacker, the probability that its apparent source is the actual source is $\frac{n-p(n-c-1)}{n}$. 
Attacks on Anonymous Sources

Attacker workload to determine origin of observed alerts

Percentage of routers controlled by attacker:
- 20%
- 40%
- 60%
- 80%

average length of routing path for each alert
A Few Loud Sensors

Thus we have to be careful how we anonymize sensor IDs
Implementation

- Reasonable-seeming performance costs
  - CPU at the sensor
  - Outbound networking
- 1.5GHz P3, Mark Shellor’s SHA and HMAC
- SRI firewall during Kuang2 outbreak
  - 4M alerts
- 1 day of DShield records
  - 19M alerts
Implementation

- Costs are modest

<table>
<thead>
<tr>
<th></th>
<th>baseline</th>
<th>hashed</th>
<th>delta</th>
<th>cached-8</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>DShield.org</td>
<td>29.81</td>
<td>64.16</td>
<td>34.35</td>
<td>56.84</td>
<td>27.02</td>
</tr>
<tr>
<td>Laboratory</td>
<td>75.80</td>
<td>110.34</td>
<td>34.54</td>
<td>106.20</td>
<td>30.40</td>
</tr>
</tbody>
</table>

Table 5: CPU Impact of IP Hashing (seconds per 1 million alerts).
Analysis Mode

- After anonymous routing, analysis is needed
- Enables key EMERALD alert aggregation and correlation
  - EMERALD is an intrusion-detection and hierarchical correlation infrastructure built at SRI
  - Similar tools using similar APIs into alert repositories could also use this data
Next Steps

- Expand types of analysis can we enable on alert records?
- Connect with Cornell work on multiparty computation
- Handle rare alert events better
- Implement Snort plugin for wide deployment
- Connect to actual onion routing infrastructure
- Convince people to use it