27th ANNUAL FIRST BERLIN CONFERENCE 14-19 JUNE 2015

UNIFIED SECURITY: IMPROVING THE FUTURE
Rod Rasmussen

IID founder, CTO

Co-chair Anti-Phishing Working Group’s Internet Policy Committee

Member of:
- ICANN's Security and Stability Advisory Committee
- Online Trust Alliance’s Steering Committee
- FCC Communications Security, Reliability and Interoperability Council
- Messaging Malware Mobile Anti-Abuse Working Group
- Forum of Incident Response and Security Teams (FIRST Representative)
- DNS-OARC

MBA from Haas School of Business UC-Berkeley; bachelor's degrees in Economics and Computer Science from University of Rochester
Cutting through cyberthreat intel noise

- Threat intel source overload
- How to cut out noise
- Threat intel plug and play with security appliance
Problem

- Over 90% of data breaches in 1H 2014 could have been avoided with simple controls and best practices*
- Security controls and best practices are valuable but only you have the right threat intelligence
- How to choose data from thousands of threat intelligence sources
Threat intel source overload

INDUSTRY ORGS
OPEN SOURCE
Shadowserver
ZeusTracker

VENDORS

SECRET SQUIRREL

INTERNAL ALERTS
From appliances and analysis

CUSTOMERS

ISACS

CERTS
Vendor Sponsored
OPEN SOURCE
SURBL
Spamhaus
Phishtank

PEERS

GOVERNMENT
FBI
Interpol
Sources cover many important things

- IPs, hosts, malware hashes, TTP’s, e-mail, account info, etc.
- Recent CERTCC study showed amazing lack of overlap amongst most popular “open source” data feeds
  - Over 96% of hostnames were unique to one feed only (sample size >30 million hosts on 13 lists)
  - Over 82% of IP addresses were unique to one feed only (sample size >120 million IPs on 38 lists)
What do you want?

Totally stolen from Tom Millar’s lightning talk yesterday!
Even less capacity to use data

<table>
<thead>
<tr>
<th>Product</th>
<th>Rule/protection type</th>
<th>Max Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firewall vendor 1</td>
<td>Security rules (IPs/ASNs)</td>
<td>40,000</td>
</tr>
<tr>
<td>Firewall vendor 2</td>
<td>Maximum Firewall Policies</td>
<td>100,000</td>
</tr>
<tr>
<td>Firewall vendor 3</td>
<td>Maximum Firewall Policies</td>
<td>40,000</td>
</tr>
</tbody>
</table>

- IDS systems have rule or practical performance limits that kick in depending on hardware
- RBL’s, DNSBL’s, web proxies, and other in-line products that match against known threats all run into capacity and performance problems eventually
- Flexible analysis tools (e.g. Splunk) have cost considerations for large sets
All intel is useful for something—use case matters most!

• Life is shades of gray, not black and white
• Reputation and context are key for use
• Block | Alert | Inform scoring | “Fits a pattern” | “Kill Chain” point
• For example, google.com
  In an ISP blacklist = disaster.
  In a malware analysis tool doing wireshark on
  a bare-metal honeypot = sign of malware activity
• Fit the data to your purpose
Dangers of threat intel that’s just noise

- False positives
- Incomplete or missing context
- No concept of TTL or useful life
- Lack of understanding good applications for data
- Not relevant to your risk situation (one size fits all data)
Noiseworthy vs. noise

Determine trustworthiness of source

Use internal threat intel and reputation to determine false positives

Analyze metrics across all data

Increase confidence with correlation, frequency and source reputation

Expand context by linking related data points to previous unknowns
Example: Conficker

- Nasty malware, but never really exploited by bad guys
- 50,000+ rendezvous domains per day (DGA)
- Most never resolve, those that do are legit sites and security sinkhole operators
- Your infected hosts will still ping a random set every day
- High volume, extremely accurate indicator, but also a false-positive generator – what to do?
- Dozens more DGA’s out there…
Um, that’s a lot of data…

- Appliances can only handle so much data
- Prioritize based on problem you’re solving and implementation ease
- Refresh rates
  - Performance
  - Timeliness
  - Cost/bandwidth
So what do you need?

- Research tools/repositories – give me everything!
- RBL’s/RPZ’s/DNBL’s, Firewall rules
  - Just the facts get delivered (hosts/IPs/TTLs)
- Still need context for making decisions about what to do
  - Give me the necessary context and I’ll decide
  - Provide your data in a “pure” form that is well described
  - I’ll give you my rules & criteria – you give just that data
Machine to machine delivery

- With your game plan set, how to get data into security appliances and analysis tools
- Scale is key—attacks are ubiquitous
- Hub and spoke vs. peer to peer
- Correlation, analysis, prioritization
- Feedback loops
Architecture may get complicated
Probably going to end up here some day…
You still need manual data in production

• Translating a research project or buddy’s email into network protection
• That latest security report (PDF!) from vendor X
• Inventory how you do (or wish you did) things today
• Automating a bunch of manual processes
Choose the right security appliances

- Next Gen Firewall
- SIEM
- IDS/IPS
- Web Proxy
- DNS Server
- Email Filter
- Internal TI Repository
- Research Tool
- Log Analysis
- Advanced Threat Detection
Choose the right data format

- CSV
- JSON
- XML
- Open IOC
- STIX
- IODEF
- NMSG
- XLS
- CEF
Working with various formats

**Battle plan:** format that delivers for the given use case

- The right tools to translate
- Push through repositories or services to normalize
- Hook up to the right “pipes” to put in play with your security systems
So how easy is this?

- Bad news – it isn’t easy
  - Lots of custom, manual processes for each environment
  - Spot solutions – no comprehensive tools
  - Lots of standardization problems still beyond data format (taxonomy, reputation, BCP’s for handling)
- Good news – lots of people working on the problems
  - Open source and commercial efforts
  - Governments and industry organizations driving standards forward
Some further thoughts
What tools and staff do you have?

- People to build scripts to automate
- Investigators to look at things in-depth
- Big commercial repository, homegrown system or spreadsheets?
- Lots of bandwidth?
- Can you put things in the cloud?
Why do you need ALL the data attributes from one source?

- Many data sources don’t provide “feeds”
- Reputation and context for the indicators you see can be found in many places
- Utilize API’s, scraping (if ok) and subscriptions
- “Fractional Access” – i.e. get just what you need, just in time
  - Saves time/money/equipment/
  - Flesh out your stored data with attributes for further work
How about your own alerts/data?

Data Source 1

Data Source 2

Data Source X

Data Feed Manager

Threat Data Management System

Repository

SIEM

Firewalls

Anti-Spam

IDS
Ubiquity of “baseline” threat intelligence data

• Not all of your valuable assets are “on network” anymore
• Different device profiles and access methods
• Different security vendors have different TI incorporated in-device or on-path
• How can you guarantee protection against “threat X” across all?
• You will HAVE to insert data across all environments - eventually
Questions

KEEP CALM AND ASK QUESTIONS
Rod Rasmussen, President & CTO
rod.rasmussen <at> internetidentity.com