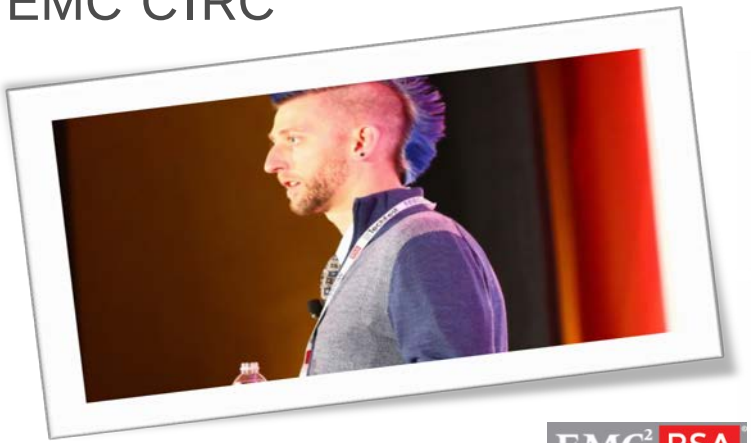
Blind Spot Analysis

Finding RAT Communications through Entropy and Analytics

Customer Profile

EMC CIRC



EMC² RSA

ANDREW RUTKIEWICZ

Principle IT Security Analyst

"Numerical weakness comes from having to prepare against possible attacks; numerical strength from compelling our adversary to make these preparations against us" – Sun Tzu

Challenge

- Develop Analytics
- Intel Data Management
- Behavioral Malware Categorization
- Reporting Data -> Actionable Items

Solution

- Quantitative Meta Creation
- Service Identification
- Defined Methodology
- Automated Processing

Applications

- SA
- ECAT
- Archer
- SAW

Results

- Fingerprinting Capabilities
- Analytical Disintermediation
- Lowered Analytical Transaction Costs

#RSAsummit

EMC²

RSA

Resources

- Phillip Evans – BCG – How Data Will Transform Business – Ted Talks
  - https://www.ted.com/talks/philip_evans_how_data_will_transform_business
- Data Streaming Algorithms for Estimating Entropy of Networks – 2006
  - http://www.cc.gatech.edu/~jx/reprints/Sigm06_entropy.pdf

What This Talk Is And Is Not

- As much theory as practice
  - Application of strategic business theory to information security
  - Tools are useless without a methodology
- Ideas to creatively solve issues and shortcomings
- No silver bullet
  - There is no magic to finding ALL malware

#RSAsummit

Porter, Strategy And Security

- "Blind spot" analysis
  - Methodology for decision makers to determine current practices/thoughts are antiquated
    - Perform Porter's Five Forces
    - Gather competitive intelligence
    - Compare the two
- Typically value chain analysis follows
  - How you add value via process or knowledge
- Easily transferred to ever changing security world

#RSAsummit

Five Forces

- Business
  - Threat Of New Entrants
  - Threat Of Substitutes
  - Barging Power of Suppliers
  - Barging Power of Buyers
  - Strength of Industry Rivalry
- Security
  - Threat of New Actors
  - Threat of New Malware
  - GA Tools\Code\Exploits
    - *OTS Sec Tools*
  - Actor's $ & Build to Suit Tools
    - *DIY Sec Tools*
  - Pitch of Battlefield

#RSAsummit

Competitive Intelligence

- Threat intelligence feeds & portals
- Internal intelligence
  - Level of expertise in security threats
- Counter intel operations
- Supply chain
  - Do threat actors target your upstream and downstream?
  - Does your supply chain communicate with you?

#RSAsummit

So… Where Do You Stand?

- Any major contradictions?
- Where are you out-matched?
- Where do have a competitive advantage?
- What you should you do next?
  - Data science/analytics – AI/ML – Panic/Freakout?
    - “Because Math” is not an acceptable answer
    - 95% have no staff or budget for analytics
    - Detection methods are unproven (in production environments)
    - It’s not science if its not repeatable
    - Can I make my own Skynet to help?

#RSAsummit

Blindspot Analysis Results (tool side)

- Few quantitative meta groups
- Packet meta is ugly in SAW
  - Lack of placeholders for HTTP dir,filename,action
  - Use of | in strings that are arrays
    - Nitty-gritty data manipulation
- Lack of building blocks
  - Summary tables
  - Profiling
- Data science free-for-all

#RSAsummit

A Common Blindspot – New RATs

- Difficult to detect on the wire
- Thousands of variants from many families
- All have the same basic functionality
- Many share sections of code
- Slight changes allow evasion of detection
- Many different C2 Comm channels

#RSAsummit

Typical APT RATs

Non-HTTP based

- 9002 aka Hydraq
- Gh0st aka Zegost aka LURK
- PlugX aka Sogu aka KorePlug
- Many others – Hupigon, Pirpi*, ZiYang, ZXShell, Poison Ivy

#RSAsummit

What Do They Have In Common?

9002, Gh0st, ZiYang and PlugX

- Packet length is incorporated into header
- Packet header is inside of first 16 Bytes
- Non-HTTP based
- Use compression or encoding, sometimes both
- Traditionally hard to detect and easy to modify

#RSAsummit

Packet Headers

RAT	01 02 03 04	05 06 07 08	09 10 11 12	13 14 15 16
Gh0st 5 Byte	G h 0 s	t L C M	P L U C	P
Gh0st 6 Byte	L O V E	T T L C	M P L U	C P
Gh0st Shifted	L C M P	L U C P	A Q Q K	L
9002	9 0 0 2	L C M P	L U C P	
ZiYang	XORed Cmd	L C M P		
PlugX	ENC Key	Comm Flag	C L U L (Encrypted)	

#RSAsummit

C2 Check-in Payloads

Uncompressed or decrypted all contain

- C2 Command
- CPU Speed and RAM
- Username and Computer Name
- OS Version
- Sys Info – IP, Webcam, Volume Info, etc.
- Typically Padded With 0x00
  - Compression is only effective on larger payloads

#RSAsummit

Value Chain: SA&SAW as a raw material

- Analytics work best with quantitative data
- OOB SA Packets has little available for DIY analytics
  - No client bytes vs server bytes
  - Can't parse entire stream (OOB first 128K [97.6K])
  - Can't parse all streams natively
    - Need token in parsers
    - Not always sure of what your looking for – "new rats"
- Limited avro data types – No INET type
- Need to write UDFs UDAFs to deal with |

#RSAsummit

DIY Analytics Cookbook

- Quantitative measures of sessions
  - Entropy, #of symbols, avg. packet length, byte frequency
- Repeatable (It's science!)
- Measures like entropy directly correlate to compression and encryption
  - Compression ratio is an expression of the entropy of information being compressed
  - Encryption = Close to maximum entropy
  - Obfuscation = High entropy

#RSAsummit

More Cookbook Fun

- Encoded data\compressed data have patterns
  - 39U 19! and \x4B63\x6060 → Gh0st - LZ Artifacts other than 789C
  - Bytes 26 and 33 tend to be identical
- Low entropy
  - Some PlugX variants are almost entirely 0x00
  - Hupigon uses 0x00 as padding (lots)
- High entropy
  - AES or other encryption without SSL
- Teach a computer why “traffic looks shady”

#RSAsummit

Analytics and Beyond...AIaaS...?

- Data enrichment -> ECAT DB, WWW
- Profiling of IP.DEST, Alias.host, Entropy
- Calculate entropy of HTTP directory or domain name
  - Base64 Identification
  - DGA Identification
- In-depth byte frequency analysis - NLP
- Is your dataset ready for AIaaS ?

#RSAsummit

Analytic Gotchas

- Entropy is computationally expensive (lots of LogN)
  - High CPU cost
    - 10% per 100Mb
- Boiling the ocean never works
  - Service identification must be >90% accurate
- Data scientists are not security professionals
- Standard analytic methods for network security carry high transaction costs and low yields
  - Outcome: Negative ROI – this is changing

#RSAsummit

New Analytical Meta

- Client.payload
- Client.entropy
- Client.entropy.m
- Client.mfb
- Client.mcb
- Client.ub
- Client.aps

- Server.payload
- Server.entropy
- Server.entropy.m
- Server.mfb
- Server.mcb
- Server.ub
- Server.aps

#RSAsummit

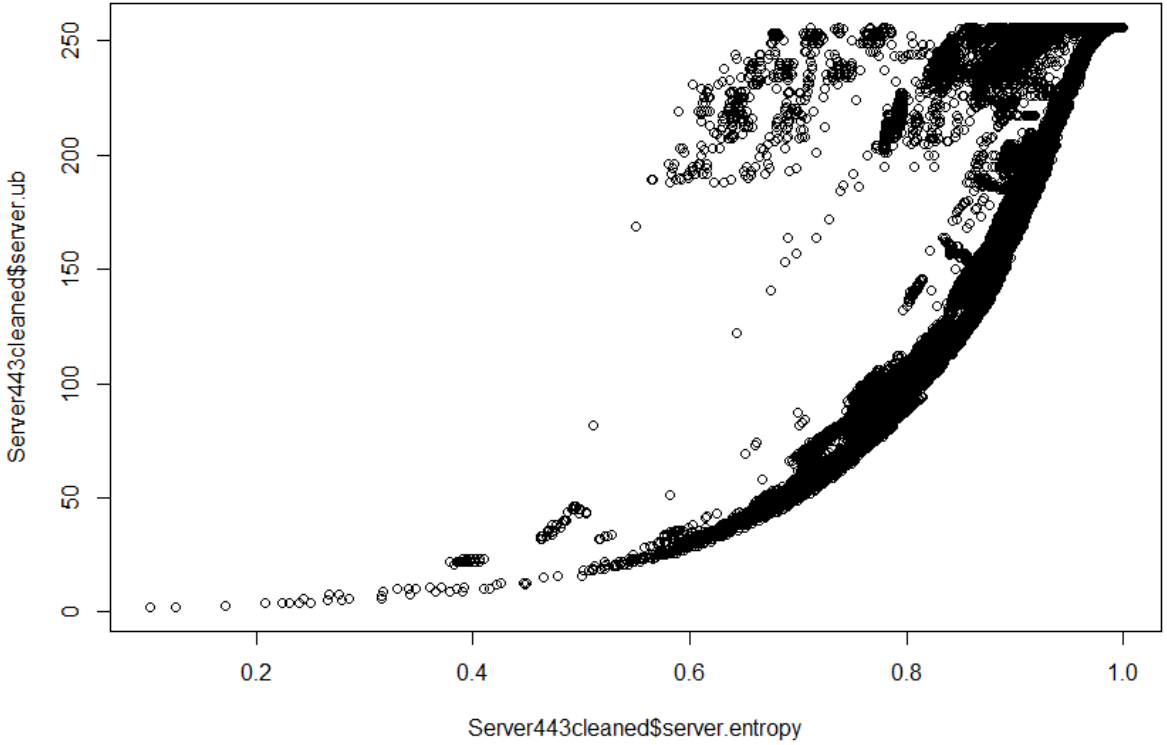
Derived Calculations

- *.ub /256 = % key space used
- *.mfb/*.payload = padding factor
- Server.payload/Client.payload= RxTx ratio
- Entropy / payload = metric entropy
- Client entropy / server entropy = needs a cool name

- Allows for finger printing of devices, websites, protocols and applications through network traffic

#RSAsummit

SSL Entropy VS Unique Bytes Profile



#RSAsummit

Summary

- Define the big problem, break problem into smaller, solvable questions
- Solutions to small problems should be reusable and optimally quantitative
- Meaningful patterns will not solve problems alone
- Must be repeatable, therefore scriptable ->AI
- Simple is better – keep transactions costs low!
- Follow a well defined methodology

#RSAsummit

Questions?

#RSAsummit

RSA® GLOBAL SUMMIT 2014

SECURITY REDEFINED

```
<key description="Client Entropy" format="Float32" level="IndexValues" name="client.entropy" defaultAction="Open"/>
<key description="Server Entropy " format="Float32" level="IndexValues" name="server.entropy" defaultAction="Open"/>
<key description="Client Metric Entropy" format="Float32" level="IndexValues" name="client.entropy.m" defaultAction="Open"/>
<key description="Server Metric Entropy" format="Float32" level="IndexValues" name="server.entropy.m" defaultAction="Open"/>
<key description="Client Payload" format="UInt32" level="IndexValues" name="client.payload" defaultAction="Open"/>
<key description="Server Payload" format="UInt32" level="IndexValues" name="server.payload" defaultAction="Open"/>
<key description="Client Most Common Byte" format="UInt8" level="IndexValues" name="client.mcb" defaultAction="Open"/>
<key description="Server Most Common Byte" format="UInt8" level="IndexValues" name="server.mcb" defaultAction="Open"/>
<key description="Client Unique Bytes" format="Int16" level="IndexValues" name="client.ub" defaultAction="Open"/>
<key description="Server Unique Bytes" format="Int16" level="IndexValues" name="server.ub" defaultAction="Open"/>
<key description="Client Occurrence of Most Common Byte" format="UInt32" level="IndexValues" name="client.mfb" defaultAction="Open"/>
<key description="Server Occurrence of Most Common Byte" format="UInt32" level="IndexValues" name="server.mfb" defaultAction="Open"/>
```

entro-lua

#RSAsummit