



Using Big Data for Good

Advanced Malware Protection as a Cloud Service

Gary Spiteri
Security Engineer
17 July 2012

Why Use Big Data for a Security Service?

- Because the traditional way is broken
 - Industry Average detection rate is ~40%
 - Ask yourself - when did this infection really occur?
- Security threats have moved on
 - 300 million new pieces of malware released in 2011
 - ~75% of attacks are seen only once
 - Have lifetimes measured in hours or days
- Static tools and Manual Responses are inadequate
- Malware Defence is a Data Analytics Problem





- The Vulnerability Research Team has been organising attack data for years
- There's no manual way to maintain over 20,000 individual rules
- The VRT built an automation system to stage OSeS and Threats and generate rules
- This is why Sourcefire has been the top of the NSS Labs IPS Protection Performance rankings for the last 4 years



3

SOURCEfire

Big Problems need Big Data

- Look at the problem from a data point of view
 - If you can see it, you can control it
- Design a collector
 - Lightweight Endpoint
 - Sensor for the Network
- Build a flexible back-end
 - Scalable, Elastic, Centralised, Adaptable
- Analyse the data on a massive scale
 - Mapreduce, Hadoop, Mongo and Riak
 - Look for patterns
 - Create Baselines



4

SOURCEfire

FireCLOUD and FireAMP

- What data is fed into the cloud
 - SHA256
 - Fuzzy SHA
 - Machine Learning Fingerprint
- Now that we have the data we can:
 - Determine malicious files even if we seen them once
 - Have complete file histories
 - See Patient Zero
 - Retroactive Protection

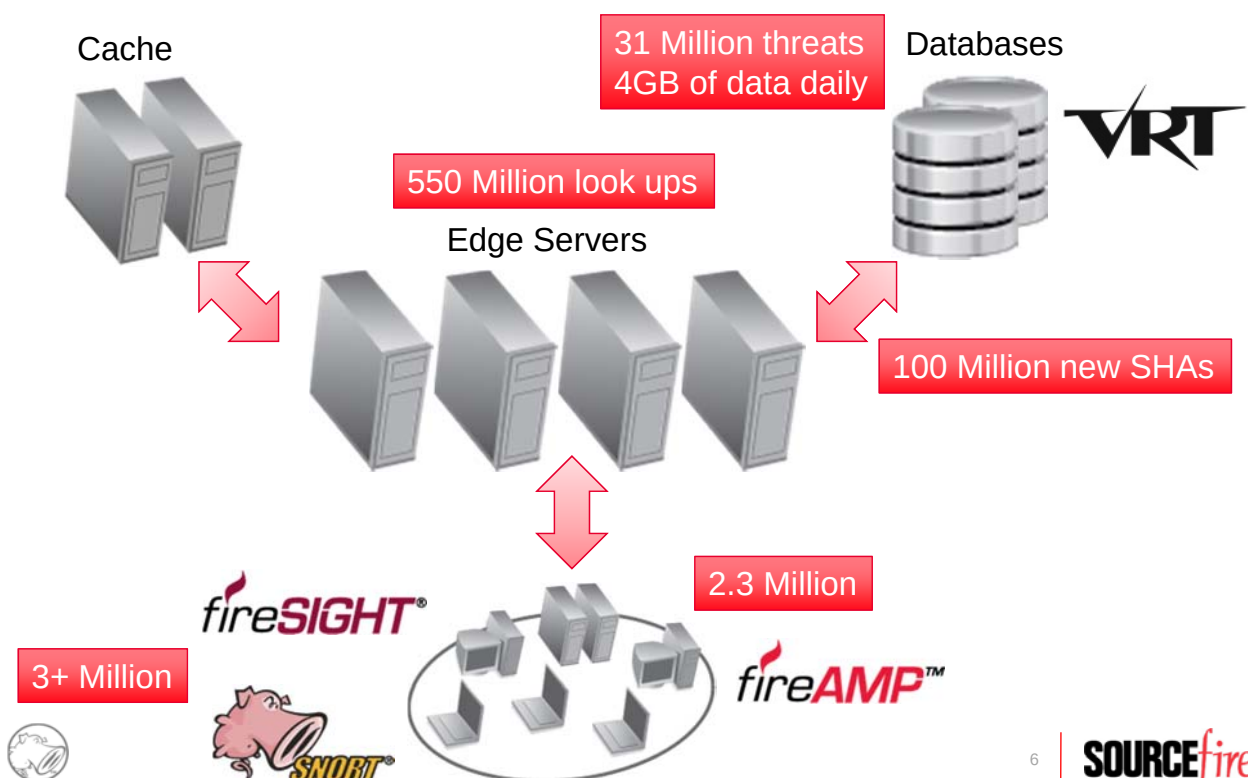
fireAMP™



5

SOURCEfire®

Inside **fireCLOUD™**

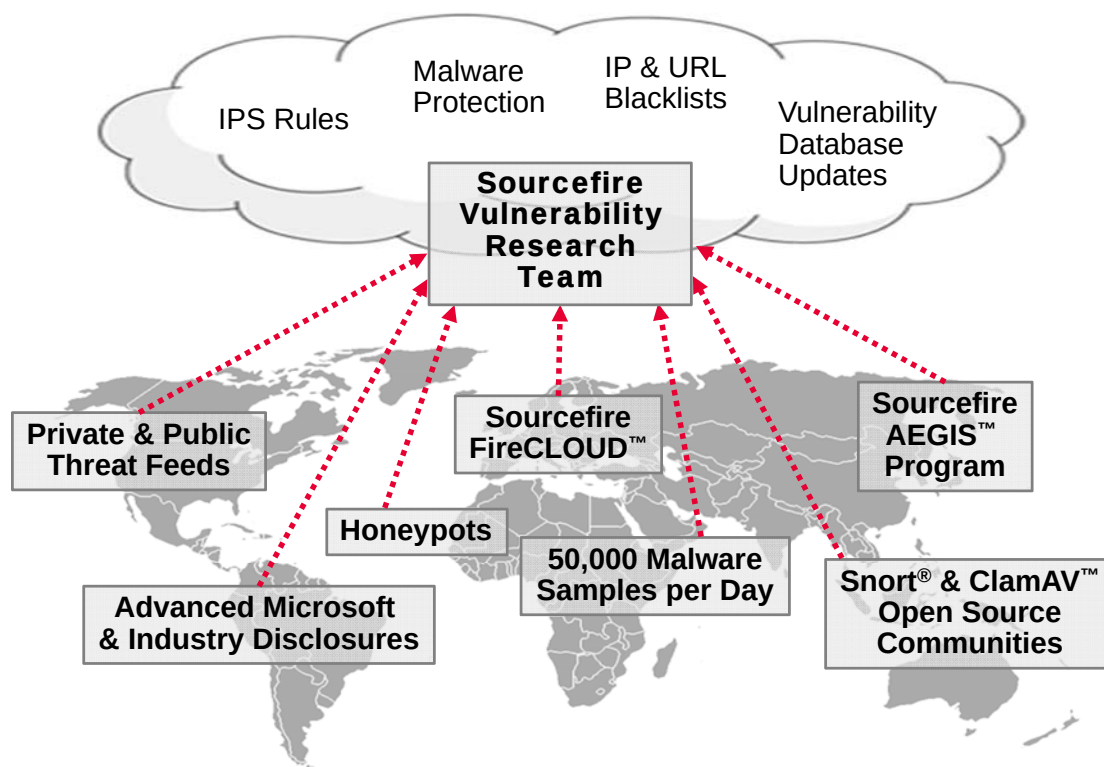


6

SOURCEfire®

Collective Security Intelligence

Global Visibility Through Open Community



Machine Learning and Decision Trees

- Engines
 - One-to-One
 - Generic Signature Engine
 - Machine Learning Engine
 - Advanced Analytics Engine
- Fuzzers
- Interchangeable
- Upgradeable
- Evolving



Example: The Drop Kick Engine

- New Engine introduced without connector update
- Uses Parent-Child Information to predict if a file is malicious
- Also applied to white-listing
- Dropkick identifies 1% additional Malware daily
 - Industry average for a new engine performance is 0.25%-0.50%
- It also whitelisted an additional 4,762,600 files in it's first 3 weeks



History and More History

- Massive data sets
 - Complete history yields Business Intelligence
- Trends
 - Bring structure to all this data
- Retrospective analysis
 - Actionable Data

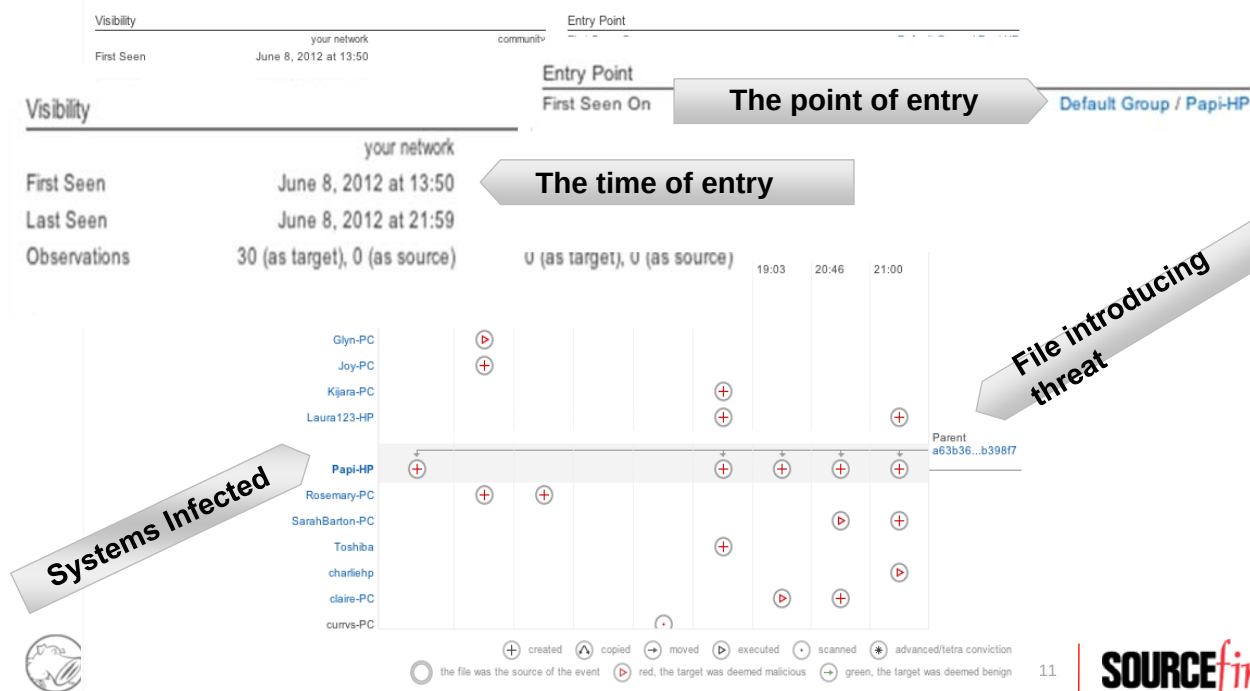


File Trajectory

The File Flight Recorder

- What systems were effected?
- What is the point and method of entry?
- What do I know about the threat?
- Can I stop the outbreak quickly?

File Trajectory for b0e1d4e24373cb7945c5630829c2c53954bed1617a636cadec2cab604705b4ac.



File Analysis

Sourcefire VRT Powered Insight

- What systems were effected?
- What is the point and method of entry?
- What do I know about the threat?
- Can I stop the outbreak quickly?

SourceFire Threat Analysis for e7337744b755cc039f32348a645ac4cdac01fec4f4725403e2b7d677e7c3c245

+ General information

- Summary

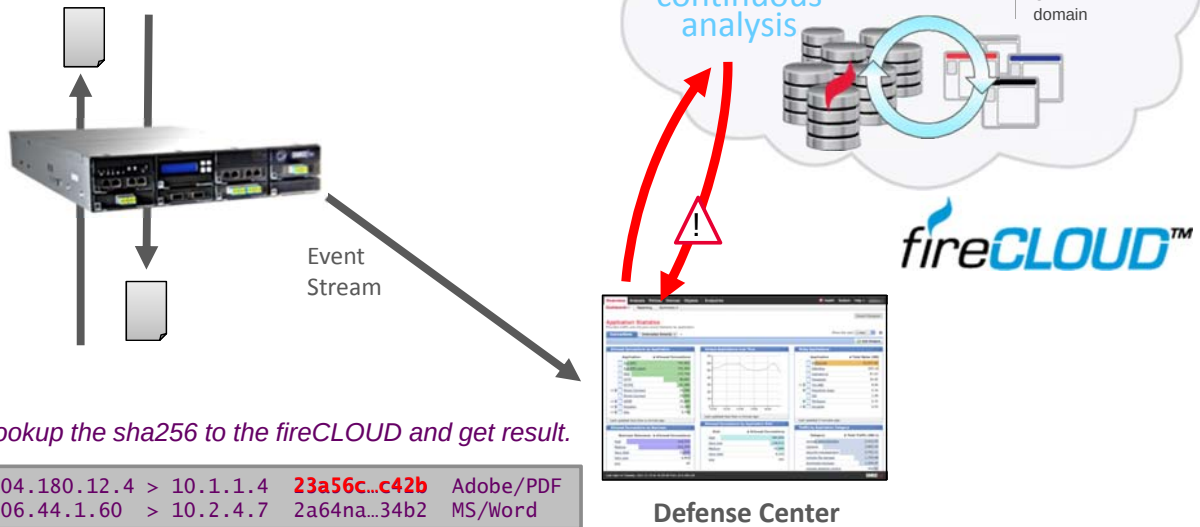
- Printf formatting strings found in memory and binary data
- Binary contains paths to debug symbols
- Queries a list of all running drivers
- Queries a list of all running processes
- Creates temporary files
- Performs DNS lookups
- Creates an software uninstall entry
- Creates files inside the program directory
- Creates files inside the user directory
- Creates a mutex to recognise infected hosts
- Downloads files from web servers via HTTP
- Creates an autostart registry key
- PE file contains sections with non-standard names
- Checks the online ip address of the machine
- Installs a browser helper object (BHO)
- Relocates Windows system DLLs to bypass security applications
- Hooks processes query functions(used to hide processes)
- Spawns drivers
- Modifies the prolog of usermode functions (usermode inline hooks)
- Binary may include packed or crypted data
- Modifies the interrupt descriptor table (IDT hooks)
- Modifies the prolog of kernelmode functions (kernelmode inline hooks)

Understand the file characteristics

Understand the file behavior

File Reputation

FireCLOUD lookup on files

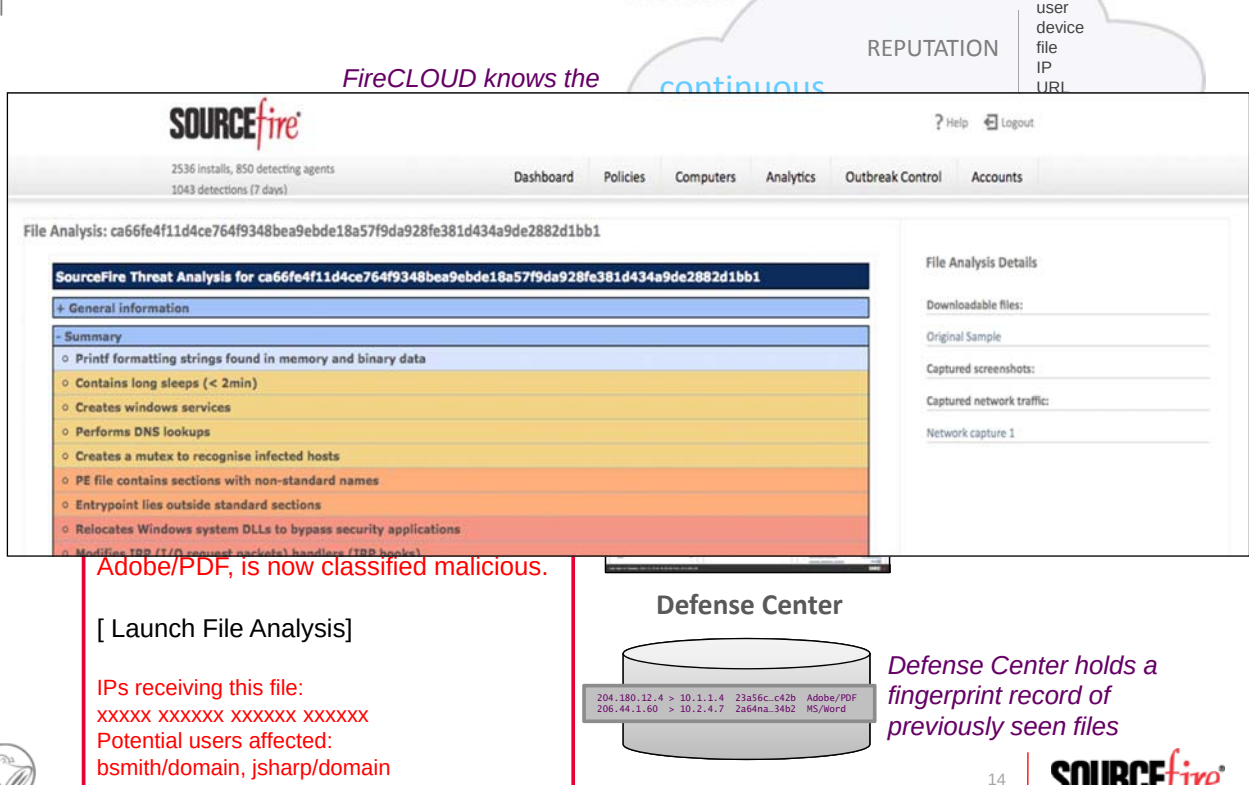


13

SOURCEfire

Retrospective Alerting

FireCLOUD convicts & alerts customers



14

SOURCEfire



Overview Analysis Policies Devices Objects FireAMP

Connection Events Intrusion Events Hosts Users Vulnerabilities Correlation

Impact to Destination

Drill Down of Impact and Destination ▶ Drill Down of Source and Destination IPs ▶ Table View of Events ▶

No Search Constraints (Edit Search)

Connection Events	Intrusion Events	Hosts	Applications	Application Details	Servers	Host Attr
1	1	10.1.8.230	SPECIFIC-THREAT			
1	1	10.11.8.230	SPECIFIC-THREAT			
1	1	10.21.8.230	SPECIFIC-THREAT			
2	2	50.11.8.114	BAD-TRAFFIC dns			
2	2	10.21.11.203	EXPLOIT Oracle Sc			
2	2	30.1.8.114	BAD-TRAFFIC dns			
2	2	30.11.8.114	BAD-TRAFFIC dns			
2	2	10.21.8.114	BAD-TRAFFIC dns			
2	2	10.21.8.114	BAD-TRAFFIC dns			
2	2	30.21.8.114	BAD-TRAFFIC dns			
2	2	50.21.8.114	BAD-TRAFFIC dns			
2	2	70.1.8.114	BAD-TRAFFIC dns			
2	2	70.11.8.114	BAD-TRAFFIC dns			
2	2	10.1.11.203	EXPLOIT Oracle Sc			
2	2	10.11.11.203	EXPLOIT Oracle Sc			
2	2	10.1.8.114	BAD-TRAFFIC dns			
2	2	10.11.8.114	BAD-TRAFFIC dns			
2	2	70.21.8.114	BAD-TRAFFIC dns			
2	2	50.1.8.114	BAD-TRAFFIC dns			
2	2	10.11.8.114	SPECIFIC-THREAT			

Last login on Tuesday, 2012-05-22 at 20:06:53 from unknown.tsg.sourcefire.com

Host: 10.1.8.230

Scan Host Generate White List Profile

Hostname
NetBIOS Name
Device (Hops)
MAC Addresses (TTL)
Host Type
Last Seen
Events
Intrusion Events
Current User

Operating System

Vendor Product Version

Microsoft Windows 2000, XP, Server 2003

Servers (3)

Protocol Port Application Protocol Vendor and Version

tcp 1189 pending
tcp 1077 pending
tcp 22 SSH OpenSSH 5.1p1 Debian-6ubuntu2

Applications (28)

Application Protocol Client Version

DNS DNS
Google Google
Google Safebrowsing Google Safebrowsing
IRC IRC
HTTP Internet Explorer 6.0
HTTP Internet Explorer 6.0
HTTP Internet Explorer 6.0
HTTP Internet Explorer 6.0
HTTP Chrome 15.0.874.120
HTTP Internet Explorer 6.0
HTTP Chrome 15.0.874.120
HTTP Internet Explorer 6.0
HTTP Chrome 15.0.874.120
HTTP Internet Explorer 6.0
HTTP Chrome 15.0.874.120

Client Applications
Client Version
Application

Who is at the host
OS & version Identified
Server applications and version

Only Sourcefire delivers complete network visibility in one appliance

Worm Propagation



Search

Nodes Links End Points

Legend

Nodes
Links
New process non white list
Destination becomes Source

Filters

User
System
Destination becomes Source
Impact 1 Non-Blocked
Legend
New services

Query Decorators

Property value

Destination...

Clear filter Save filter...

Data

name	ip address	type
NY_Eng_Host_285	172.22.0.71	workstation
NY_Eng_Host_318	172.22.0.96	workstation
NY_Fnn_Host_309	172.22.0.95	workstation

nodes(16) Links(0) EndPoints(0) Destination becomes Source(13) New process non white list(0)

Attributes

Property	value
datasource	TestData
DC_Controller	DC2
foreignSourceId	2299040
Impact 1 Non-blocked	true

Attributes

Property value

Destination...

Clear filter Save filter...

Search

Nodes Links End Points

Legend

Nodes
Links
New process non white list
Destination becomes Source

Filters

User
System
Destination becomes Source
Impact 1 Non-Blocked
Legend
New services

Query Decorators

Property value

Destination...

Clear filter Save filter...

Summing Up

- Big Data is a completely new way to tackle Malware
- Information Superiority creates actionable data to help with security incidents
- New ways to visualise this data gives new insights into the attack and how to recover



Thank You

Sourcefire Gartner
Newsletter

*Achieving Information
Superiority through the
Power of Big Data.*

