SOURCEfire

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

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
  - Scaleable, Elastic, Centralised, Adaptable
- Analyse the data on a massive scale
  - Mapreduce, Hadoop, Mongo and Riak
  - Look for patterns
  - Create Baselines

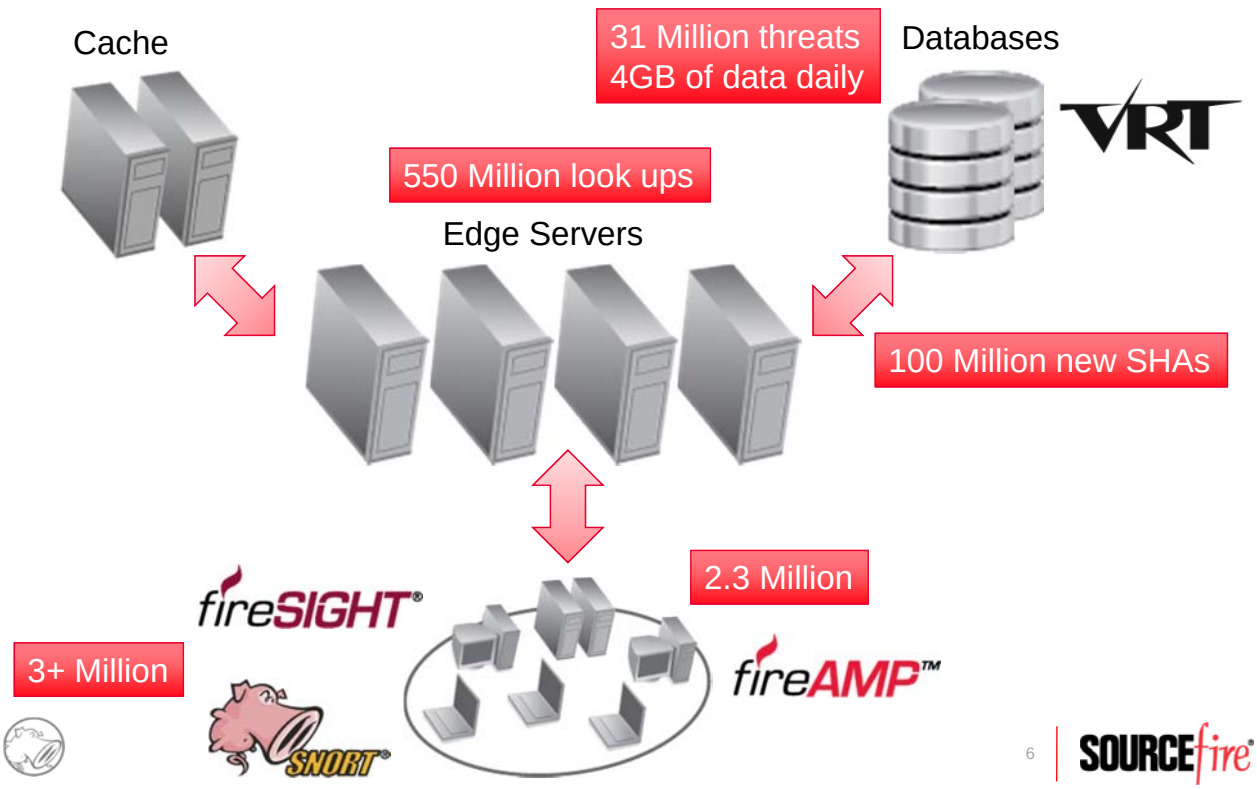
fireCLOUD™

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

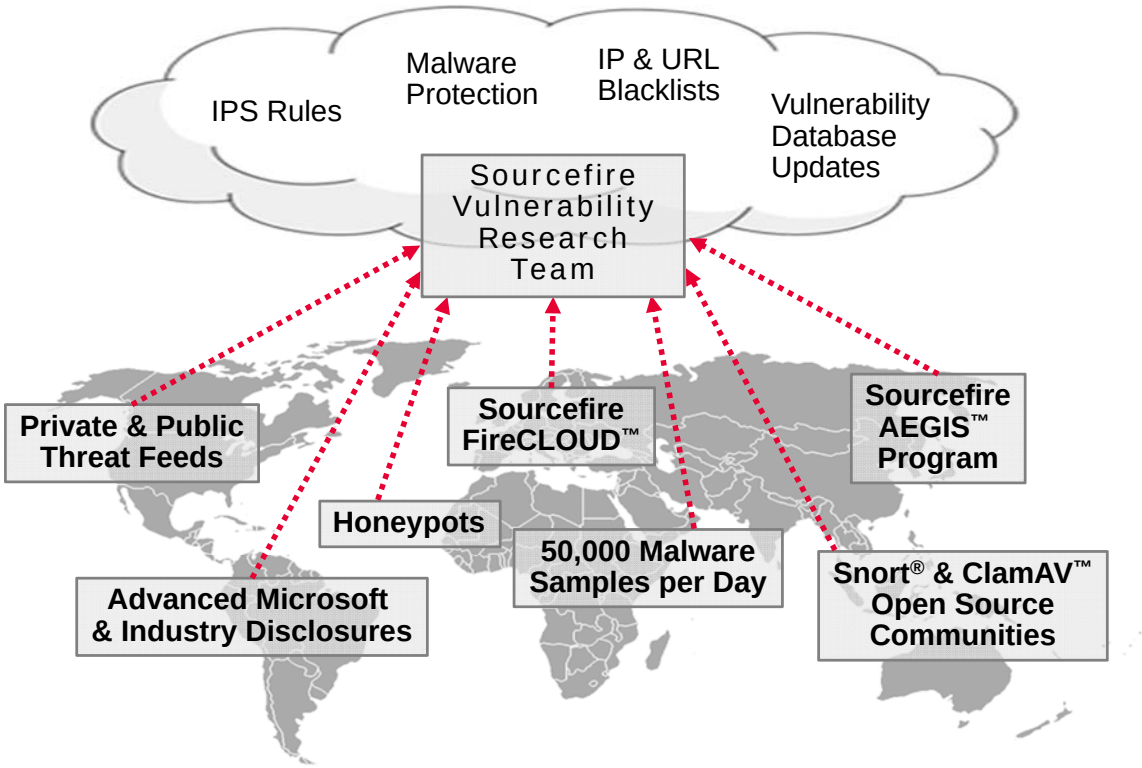
fireAMP™

Inside fireCLOUD™



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

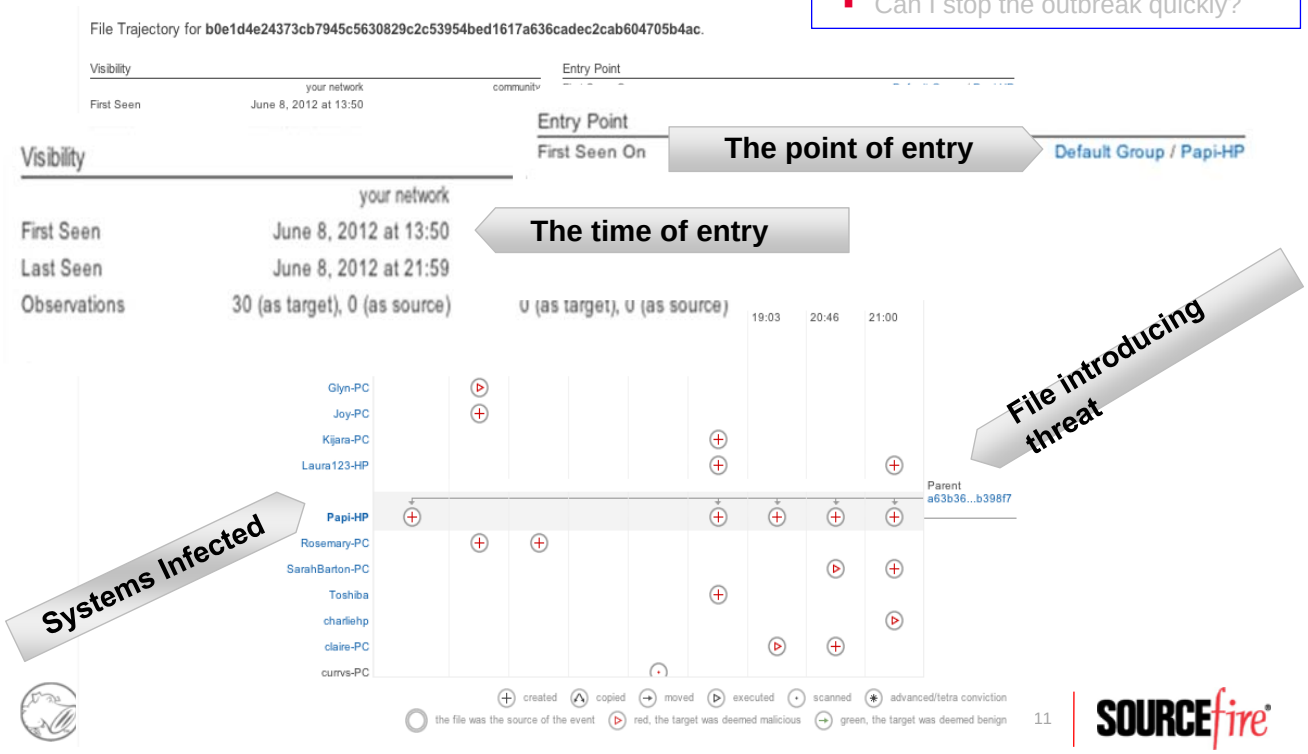
History and More History

- 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?

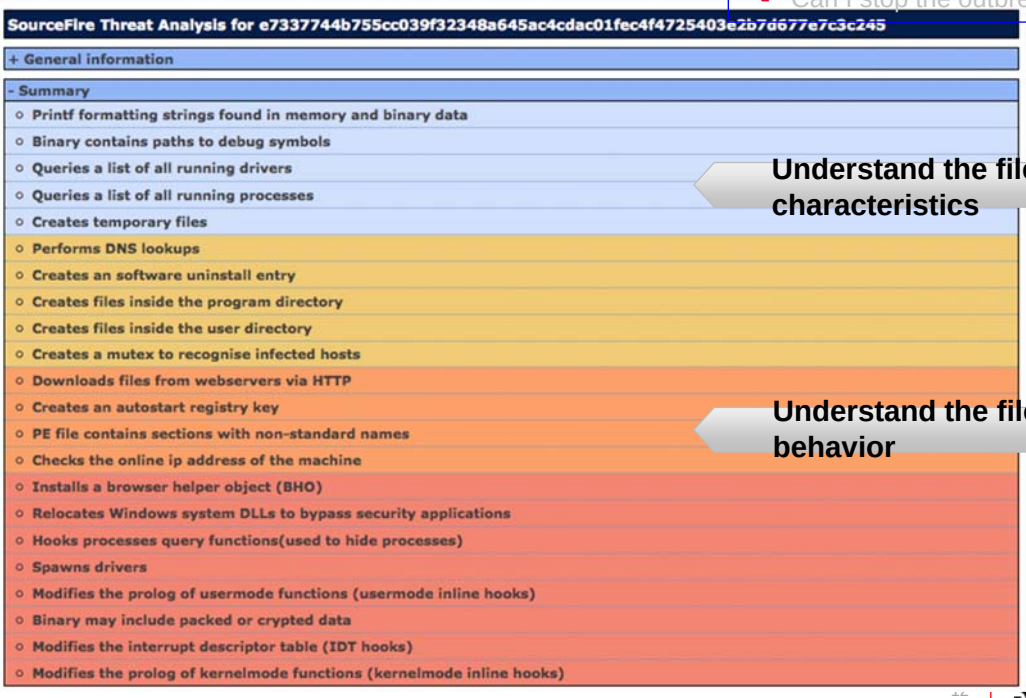


SOURCEfire

File Analysis

Sourcefire VRT Powered Insight

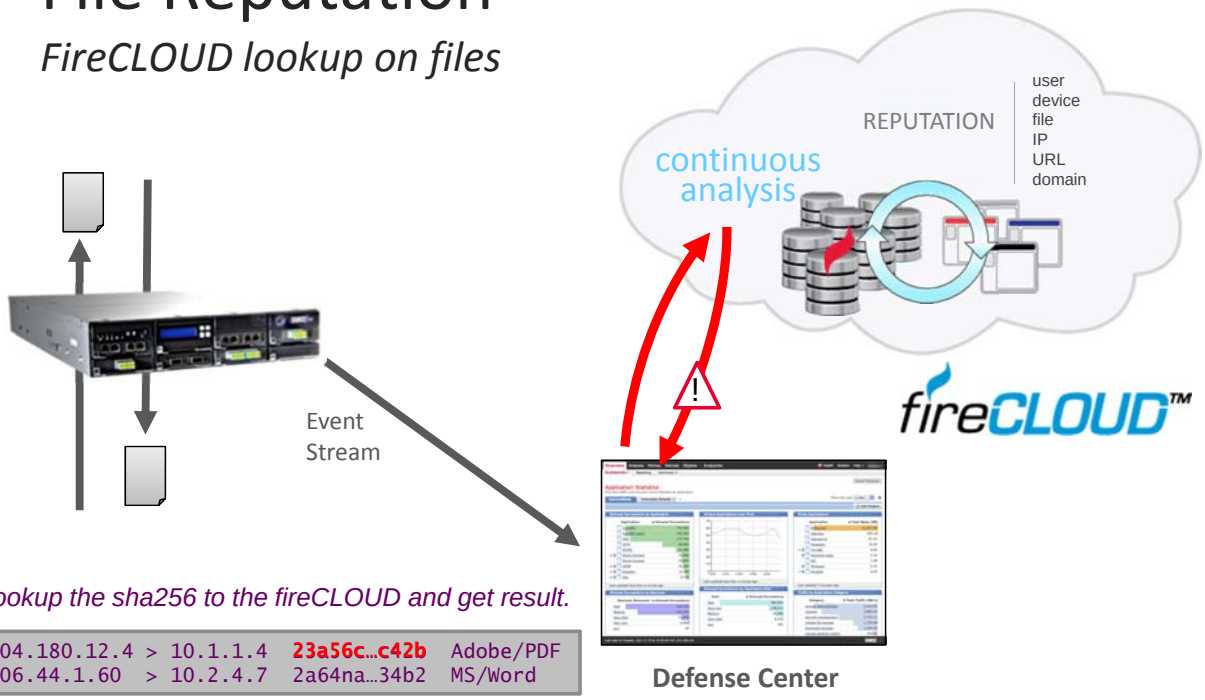
- What systems were effected?
- What is the point and method of entry?
- What do I know about the threat?
- Can I stop the outbreak quickly?



SOURCEfire

File Reputation

FireCLOUD lookup on files

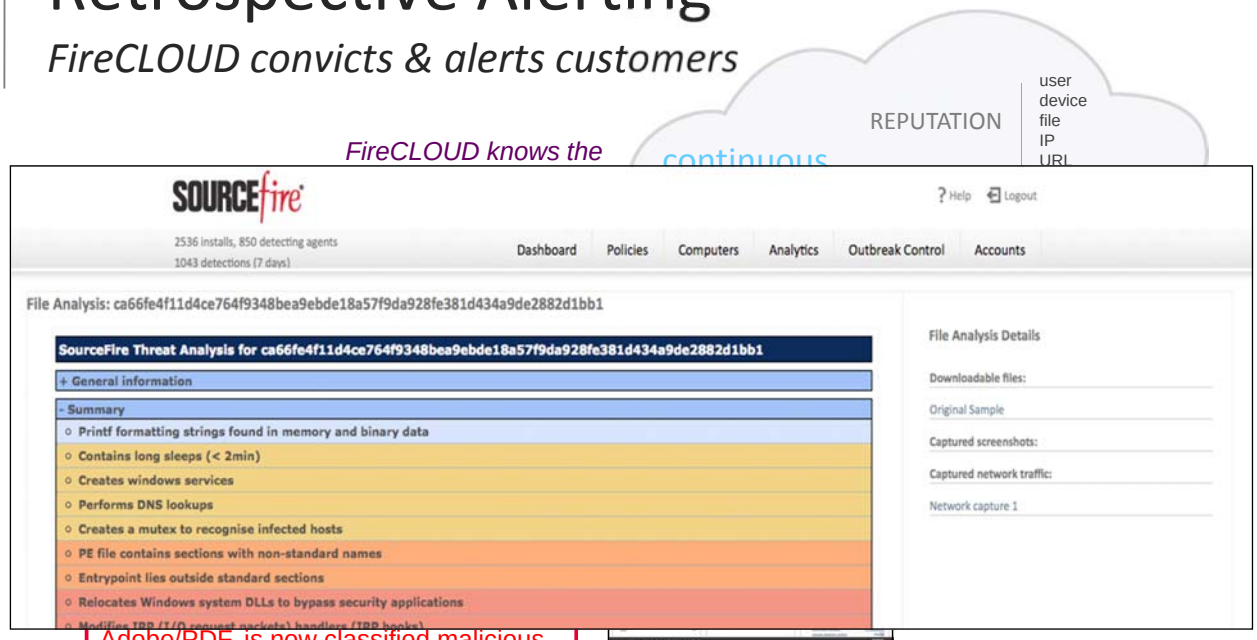


Lookup the sha256 to the fireCLOUD and get result.

```
204.180.12.4 > 10.1.1.4   23a56c…c42b   Adobe/PDF
206.44.1.60  > 10.2.4.7   2a64na…34b2   MS/Word
```


Retrospective Alerting

FireCLOUD convicts & alerts customers



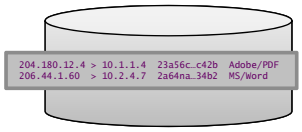
FireCLOUD knows the

Adobe/PDF, is now classified malicious.

[Launch File Analysis]

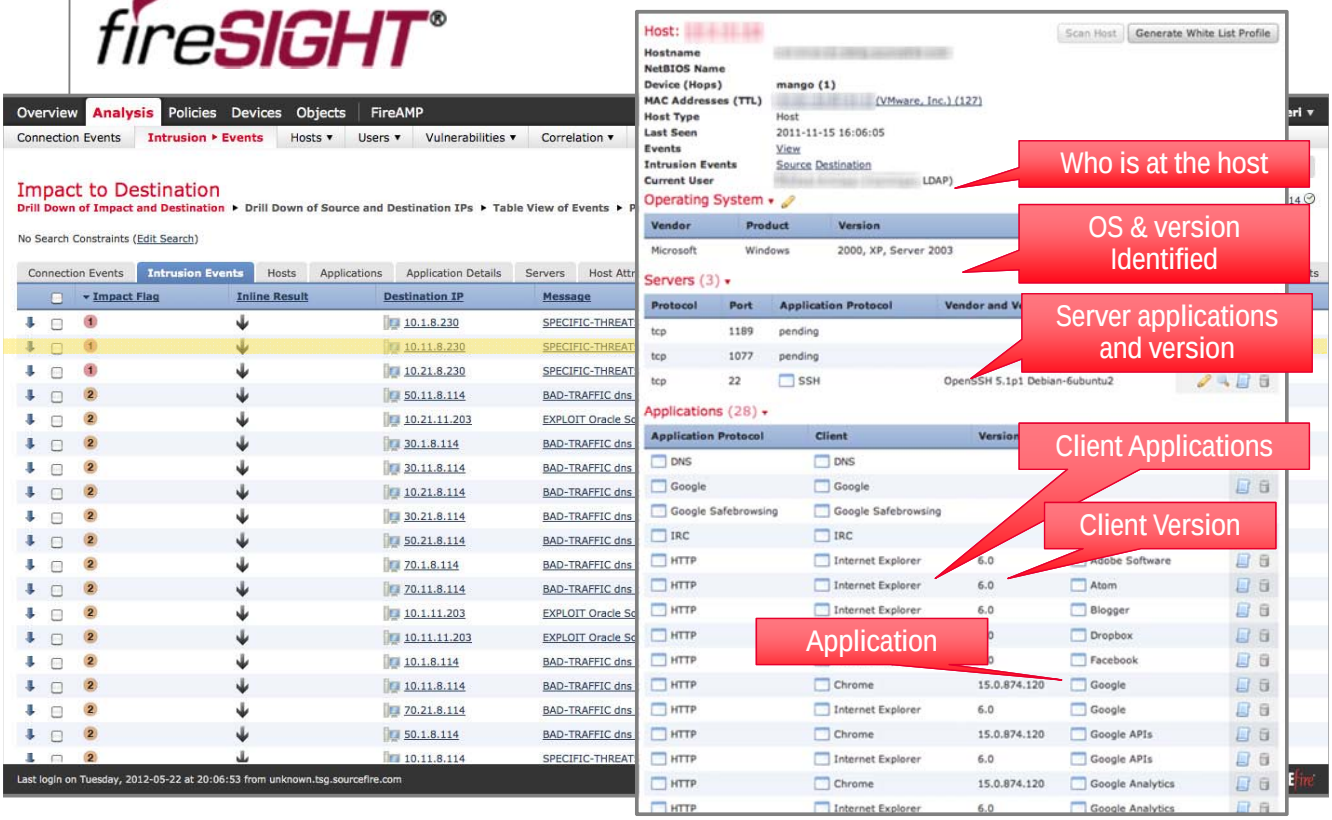
IPs receiving this file:
xxxxx xxxxxx xxxxxx xxxxxx
Potential users affected:
bsmith/domain, jsharp/domain

Defense Center



Defense Center holds a fingerprint record of previously seen files

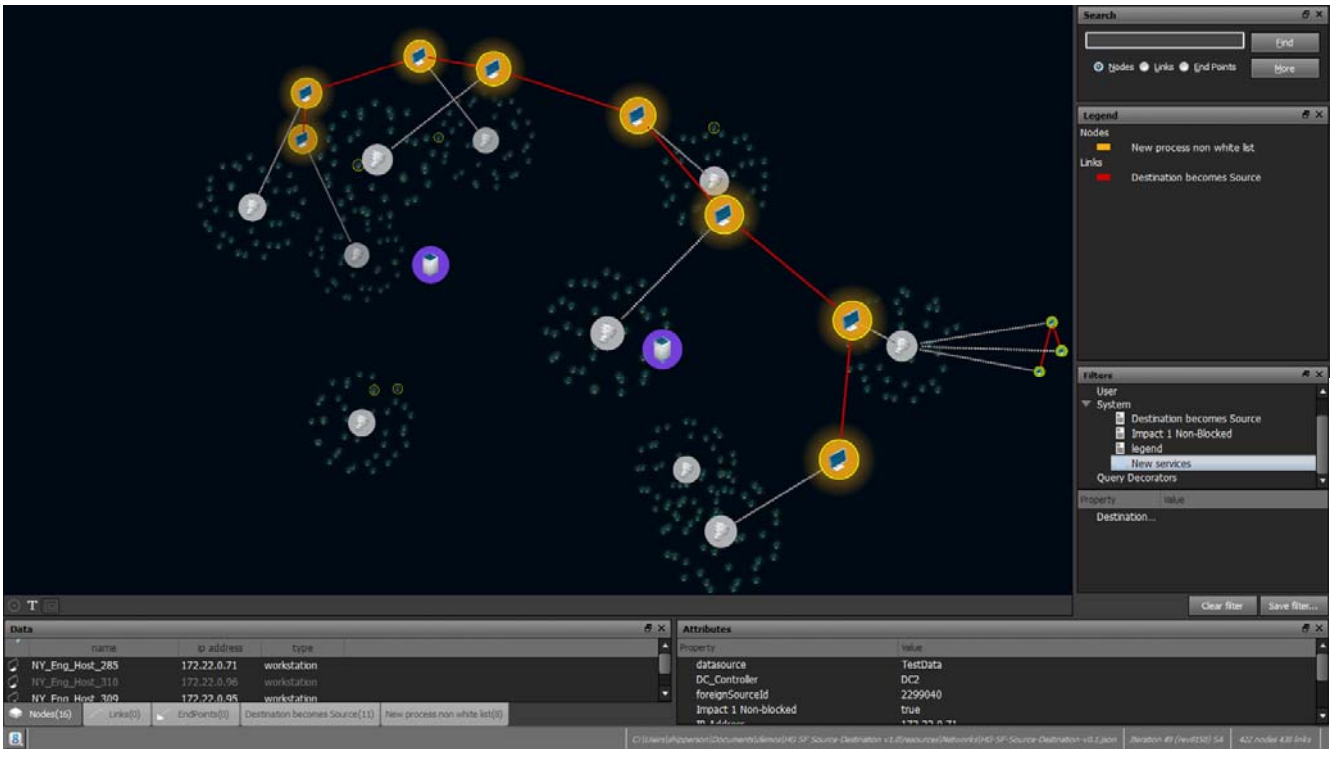
fireSIGHT®



Only Sourcefire delivers complete network visibility in one appliance

Worm Propagation

realstatus
IT insight



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.