security is not an island
HILTON MALTA

24th Annual FIRST Conference
MALTA
17 - 22 June 2012
Incident Response in Large Complex Business Environments

Ramses Martinez
Ismail Guneydas
Yahoo!

FIRST Conference
24th Annual
MALTA
17 - 22 June 2012
Agenda

1. Definitions
2. Challenges
3. Solutions
4. Case Studies
Definition of ‘Large & Complex’

1. Scale:
   – >100k Production Systems.
   – > 1 Petabyte of data generated per week

2. Diversity:
   – > 4 Major Business Lines.
   – Business lines must ‘interact’ with each other.
   – Business lines must be have internal/external dependencies.
   – Heterogeneous technology environment.

3. Geographical Distribution:
   – Employee base in at least 10 different countries.
   – Providing services globally.
Challenges: Scalability, Cost & Resources

1. System Forensic Tools
   – Per-node approach is not cost effective.
   – Speed of traditional acquisition not adequate.
   – Resources required may not be available.

2. Network Forensic Tools
   - Bandwidth/cost limitations.
   - Geographical distribution constraints.

3. Detection, Alerting, and Correlation
   - Per byte log analysis model not cost effective.
   - High false positive rate.
   - Linear searches simply break down at this scale.

4. Resources

   However, what ever alternatives to traditional methods we decide to use must always preserve the integrity of the investigative process, comply with the law and obviously yield good results.
Solutions: Scalability & Cost

System Forensic Tools

GRR Live Forensic Framework:
- Lightweight and very fast
- Accessible anywhere
- Open source
- Secure communications channel
- Memory and disk forensics
- Multiple platform support
- Supports multiple system sequential analysis
- Great detective control, can be configured to do ongoing analysis of processes, registry keys and other system artifacts to detect infections at early stages.

http://code.google.com/p/grr/downloads/detail?name=GRRArchitecture.png&can=2&q=
Solutions: Scalability & Cost

Network Forensic Tools
Hadoop cluster & Machine Learning:
- Average of 900% gain in speed vs. linear searches
- Open source.
- Multiple platform support.
- Supports multiple system parallel queries.
- Highly customizable.
- Can be configured to do ongoing analysis.

http://hadoop.apache.org/
Case Study: Fraudulent Ticket Sales

- Set of 1962 potentially fraudulent yahoo e-mails with passwords along with other information was reported to us by an external resources to us on December 02, 2011.
- Extracted account ID’s and possible passwords from the file
- Run a grid script to match e-mail addresses to user ids.
- Run a grid script to check if the reported passwords were real
- Run a grid script to check for associations to unreported accounts

Before:
- 2358|Maria|Surrova|mariasurrova@yahoo.com|c0deb4910|GWTG56

After
- One file with all e-mails, one file with all passwords
- mariasurrova@yahoo.com
- c0deb4910
Case Study: Initial Data Analysis

• All of them has a unique characteristic:
  – 9 characters with all lower cases and numbers: c0deb4910

• Accounts have same verification questions
  – What is the first name of your favorite uncle?
  – What was your favorite food as a child?

• All of the answers were 33 character lower case combined with numbers.
  – Ahsdufkdoplsjkd3jd7j8ks8d6hr64jks

  **100% Match**

• Not compromised users but machine registrations. But for what? What were attackers’ goals?
Case Study: Account Analysis

- Accounts created in last two months and registration IP’s geographically distributed across the US.
- Moreover, IP addresses are from both residential (right pic.) businesses (i.e. hosting companies) as well as proxy servers (Left pic.).
- There was no failed login activity on those accounts.
- There was no e-mails sent from those e-mail boxes.
- All accounts used to registered with a particular VOIP company.

The other commonality between those accounts, there were tickets were purchased using those e-mails from a company ticket sales and distribution company based.
Case Study: IP → Geolocation Correlation

Direct correlation between the registration IP of each account and the state where sporting events tickets were being purchased.
Case Study: How the Attack Works

**Step 1**
Attacker uses Botnet to create Yahoo! accounts

**Step 2**
Attacker uses Yahoo! accounts to register on Skype service

**Step 3**
Attacker uses Botnet to subvert TicketMaster online sales system using stolen credit cards

**Step 4**
Yahoo! accounts used for verification of fraudulent transactions with TicketMaster

**Step 5**
Attacker uses Yahoo! accounts to conduct fraudulent sale of tickets across several US jurisdictions.

* Red arrows indicate points where Yahoo! infrastructure is abused

The FBI investigated a similar case of defrauding ticket vendors last year:
Case Study: FBI Wiseguy Operation

*In case of Wiseguy's attack, human labor was used to solve the audio CAPTCHAs.*
Case Study: Conclusions

- Attackers use some clever techniques to beat CAPTCHA mechanisms.
- Attackers had access to a botnet or compromised systems across the USA (literally every single state in the US).
- They focused on high-end expensive seats at events.
- They purchase a ticket in a state where they had compromised systems.
- They have enough people to go through all email accounts to respond to verification mechanisms.
- After initial attack, the accounts were used for other fraudulent schemes like targeting jewelry stores and online banking.
- The attack involved a strong physical (human) component and was likely conducted by an organized criminal group.