Breach Detection -- in 2013
Gunter Ollmann, CTO
About Me

• Gunter Ollmann
  ▪ CTO - IOActive
    ▪ University of Georgia – CS Advisory board
  ▪ Formerly:
    ▪ Damballa – CTO & VP Research
    ▪ IBM – Chief Security Strategist
    ▪ ISS – Director of X-Force & EMEA SAS
    ▪ NGS – Professional Services Director
  ▪ Can be found/followed/located at:
    ▪ Email – gunter.ollmann@ioactive.com
    ▪ Twitter - @gollmann
    ▪ Blog – http://blog.ioactive.com
    ▪ Blog - http://technicalinfodotnet.blogspot.com
Breach Detection

• So what?
• The Ecosystem
• Incident Response
• Machine learning 101
• Early breach detection
Breached again and again...

In the first 2 1/2 years, they gathered more than 13,000 internal passwords and raided servers that could give them detailed information about the company and how it was organized...
Breach Attribution

• Does attribution really matter?
  – China has your secrets
  – Does it matter precisely who in China?

• Differentiate between attacker types
  – Insider vs. State vs. Criminal
The Malware Author/Attacker of Old
The Malware Business of Today

Create  Distribute  Infect  Manage  Monetize
The Malware Author of Today

Create

Distribute

Infect

Manage

Monetize
The Malware Author of Today

Create → Distribute → Infect → Manage → Monetize
The Malware Author of Today

Specialists of the trade

- Typically post-graduate degree holders
- 5+ years of software development experience
- Often try to operate within the “letter of the law”
The Malware Author of Today

Kit-based Solutions

- Botnet construction kits for premium rates
- Enterprise management capabilities
- Include infection vectors and phishing components
- Servicing the “average” cyber-criminal
The Malware Ecosystem

- Create
  - Armoring
  - QA Testing
  - Anti Sandbox
- Distribute
  - PPI
  - Spam
- Infect
  - Drive-by
  - Bullet proof Hosting
- Manage
  - Hosting
  - Rental
  - PI Sales
  - IP Sales
- Monetize
  - PII Sales
  - DDoS

0-day Sales
Herd Sales
Attack Construction

• All attack components and phases can be purchased
  – Directly or through middlemen

• Target Agnostic
  – Sellers don’t care who purchases

• Commission based projects too
  – E.g. “0-day” exploits
Malware Construction

• “Average” capability malware is sufficiently advanced and fully featured

• Capabilities can be extended by specialist tools & services
  – Armored to guarantee detection at time of infection
  – Inserted in to processes to permanently evade local inspection
  – Survive a disk format and reimaging of the victim device
Reconnaissance

• “Barrage” via range of vectors
  – Which work, which don’t...
  – Timing and type of response noted

• Any host will do
  – Monitor all traffic & usage

• Map network and relationships
  – Slow and steady wins the race
Target Selection

• Objectives are important
  – Criminal (financial)
  – Criminal (espionage)
  – State (intelligence gathering)
  – State (cyberwar)

• Determines injection vector & persistence requirements
Breach Detection

• If it walks like a duck, flies like a duck, and quacks like a duck...

• Feature selection:
  – Walking
  – Flying
  – Quacking
The IR Timeline

Reality

- A
- B
- C
- D
- E
- F
- G
- H

Breach Occurred
Compromise Detected

Window of compromise
Incident Response

PANIC!!!
The IR Timeline

Hackers

- Knock knock
- Knock knock
- Breach occurs
- Attack propagation
- Assets traded
- Extraction

- Attacker Recon
- Window of compromise
- Incident Response
The IR Timeline

Ideal Scenario

Breach Occurred

Breach Detected

Breach Remediated

Window of compromise
The IR Timeline

Network

Breach Occurred

Network Anomalies

Destination IP address & country,
Domain name & Authoritative DNS,
Frequency & automation,
Protocol attributes, etc.
Breach Detection

• Network-based detection taking precedence
  – Easiest point to monitor the enterprise
  – Much more difficult to lie & subvert
  – Don’t need pesky desktop agents
  – Passive observations
  – If the attacker isn’t communicating, it isn’t extricating data

• Rich but voluminous data set...
  – Big Data
If it walks like a duck, flies like a duck, and quacks like a duck...
Signature Matching

• Signature systems match specific features

Walks

Flies

Talks

• Specific matches to listed attributes
Detection/Classification

• Threat classification dependencies
  – Parameter selection and classification
  – Seen before and/or categorized

• Attackers knowledge of parameters
  – False flag operations
  – Evasion
Machine Learning Primer

- Normal Network Traffic
- Breach Traffic
- Malicious Traffic
Machine Learning Primer

Feature Selection
- Adds new dimensions for grouping and clustering

Breach Traffic
Malicious Traffic
Normal Network Traffic
Machine Learning Primer

• Automatic learning and feature selection
  – Supply training data (+ve and –ve classes)

Ducks

Not Ducks
Machine Learning Primer

• Attributes
  – Walks
  – Flies
  – Talks

• Attributes
  – Feet (width, color, nails, etc.)
  – Gait (length, pace, etc.)
  – Wings (length, color, width)
  – Beak (color, length, tongue)
  – Feathers (shape, color, etc.)
  – Quack (sound, volume, etc.)
  – Texture (plastic, wood, etc.)
Big Data Clustering

- Clustering
  - Green: Good (Duck)
  - Red: Bad (Not a Duck)
  - Light Gray: Unknown
Big Data Clustering

- Clustering
  - Good (Duck)
  - Bad (Not a Duck)
  - Unknown

0.9784 Bad

0.8631 Bad

0.0021 Bad
Big Data Clustering

- Clustering
  - Good (Duck)
  - Bad (Not a Duck)
  - Unknown

0.9784 Bad

0.8631 Bad

0.0021 Bad
Big Data Clustering

- Clustering
  - Good (Allowed)
  - Bad (Attack Traffic)

China APT

Romanian Cybercrime

Microsoft Update
Big Data Analytics

• Three sets of data:
  – Labeled “good” (small)
  – Labeled “bad” (small)
  – “Live” data (big)

• “Live” data can encompass many sources and features
  – Streaming network traffic
  – Event logs
  – Etc.
Blacklists/Whitelists

• Legacy Blacklists/Whitelists
  – Not rich enough for good ML training sets

• Fallacy of blacklists

  – $1 + 1 = 2$ (wrong)
  – $0.02 \times 0.03 = 0.006$ (closer to the truth)
Lessons Learned

• Big data, machine learning and clustering
  – Changing the way threat detection is done
  – Increased demand for labeled data sets

• Communication between victim & attacker
  – Changes in behavior when control is transferred between attackers
  – Attribution based off network traffic & labeled clusters
You’re the Victim

• Paradigm change
  – You will be breached...
    ... prompt detection is critical

• Attribution is possible...
  ... for phases of the breach
  ... but may not be useful
Lessons Learned

• Machine Learning
  – Higher confidence systems

• Trust the machine...
  – Allows for Automatic remediation
  – Reimage & reboot within 15 minutes
Thank you!

Gunter Ollmann
CTO, IOActive Inc.
Email: gunter.ollmann@ioactive.com
Twitter: @gollmann