Comprehensive Understanding of Malicious Overlay Networks

Georgia Institute of Technology

Wenke Lee

9/17/2013
Team Profile

- **Georgia Tech and GTRI**
  - Wenke Lee, David Dagon, and Chris Smoak
- **University of Georgia**
  - Roberto Perdisci
- **Dissect Cyber**
  - April Lorenzen
- **Global Cyber Risk LLC**
  - Jody Westby
- **Open Information Security Foundation**
  - Matt Jonkman
- **Farsight Security**
  - Paul Vixie
Customer Need

Security in 1990s

Security in 2013
Customer Need

• Classify/cluster malware samples: known/family, new/capabilities
  – what really matters

• Attribution analysis: malware related in evolution and shared network infrastructures
  – how it happens

• Takedown
Approach: Federated Malware Execution
Approach: Scaling Execution

- Analyze “bootstrap” malware dataset
  - Run each sample for a relatively long time (e.g., few hours)
  - Group samples that behave similarly into *malware families*
  - Extract *family behavior profiles* for each malware family
Approach: Scaling Execution

• Running new samples (post-bootstrap phase)
  – Frequently vet network/system behavior against family behavior profiles
  – If a profile matches a known family:
    • do malware in the family exhibit new behaviors if run for longer?
  – Stop/continue execution accordingly
**Approach: Identify Criminal Network**

- RHIP on all domains pulled each day
- Vertices are /24s
- Initial concerns about scaling
- Updated version has multiple levels of granularity
- Whitelisting:
  - Very aggressive
  - Alexa top 10k
  - Remove IPs w/ at least one RHDN in whitelist
  - e.g., doubleclick

Edges denote historical overlap in domain name resolutions to both vertices. Edges are weighted by the Jaccard index of domain overlap:

\[ J(v_i, v_j) = \frac{|D(v_i) \cap D(v_j)|}{|D(v_i) \cup D(v_j)|} \]
Example: Rustock Criminal Network
Approach: Takedown Recommendation

Input: \{D_s\}

Enumerate Infrastructure

Input: \{D_e \cup D_i\}

Interrogate Malware

Classify Malware Behavior

Finite Domains/ IPs

No Behavioral Changes

DGA

P2P

1.) Revoke D

1.) Counter P2P
2.) Revoke D

1.) Reverse engineer DGA
2.) TLD cooperation
3.) Revoke D
Benefits

• Improved malware intelligence and situation awareness
  – Define legal boundaries for “live” malware analysis
  – Collection of malware intelligence at scale

• Improved remediation efforts
  – Recommendations for more effective takedowns
Current Status

- Legal issues associated with FMAS
  - Identified applicable laws, developing operational documents, agreements, and policies

- Developed malware clustering algorithms based on output of static and dynamic analysis

- Developed preliminary versions of attribution and takedown recommendation algorithms

- Papers accepted to the 2013 ACM CCS and RAID
Next Steps

• Complete legal framework for FMAS within the next six months

• Start scaling and improving malware clustering, attribution, and takedown recommendation systems

• PI and Co-PI are (co)founders of security start-ups
  – Data from real-world environments
  – Direct evaluation and transition technologies
Contact Information

- Wenke Lee: wenke@cc.gatech.edu
- David Dagon: dagon@sudo.sh
- Roberto Perdisci: perdisci@cs.uga.edu