Automated Worm Fingerprinting

Awais Aslam   Attique dawood

FAST, Islamabad

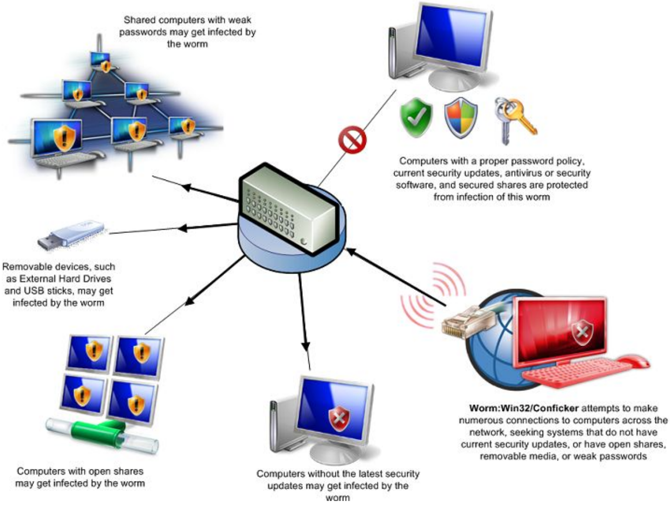
March 15, 2012

Presentation Organization

- What are computer worms?
- Need for efficient worm detection techniques
- Solution approach: Dynamic signature generation
- Experimental results of Earlybird
- Related work and assessment

Problem: Worms

- Standalone computer program
- Designated as a *malware*
- Replicates itself in order to spread
- Can use network to spread to connected hosts
- Can be malicious or useful (very rare)


Shared computers with weak passwords may get infected by the worm
Computers with a proper password policy, current security updates, antivirus or security software, and secured shares are protected from infection of this worm
Removable devices, such as External Hard Drives and USB sticks, may get infected by the worm
Worm:Win32/Conficker attempts to make numerous connections to computers across the network, seeking systems that do not have current security updates, or have open shares, removable media, or weak passwords
Computers with open shares may get infected by the worm
Computers without the latest security updates may get infected by the worm

Malicious Activities

- Increase in network traffic
- Can open backdoors and install rootkits
- Zombie for botnets
- Delete files
- Reboot systems

Worm Out-breaks in Recent Times

- Software homogeneity and unrestricted internet access
- Code Red worm took fourteen hours to infect vulnerable population (360,000 hosts)
- Slammer worm infected same number of hosts in under 10 minutes

Need for Efficient Detection Techniques

- Signature based
- Anomaly based
- Challenges in detection: Need to quickly generate signatures

Detection Techniques

Three methods for worm detection:

- Scan detection (telescopes)
- Honeypots
- Behavioural techniques at end hosts

Solution Approach: Earlybird

- Detects anomalies to generate signatures
- Invariant content in worms used as signatures
- Spreading mechanism of worms atypical of internet applications
- Frequently repeated and widely dispersed content treated as signature

A Priori Signature Creation

- Characterization is the process of analysing and identifying a new worm
- Using a priori vulnerability signatures from already know worms
- New worms exploiting known vulnerabilities
- Traffic content compared with known database of attack signatures
- Can only be used for well–known vulnerabilities

Signature Extraction

- Use decoys programs/systems in a controlled environment to get infected
- Extract infected regions
- Apply heuristics and techniques to identify invariant code strings
- Refine set of signatures by comparing them against known infections

Defining Worm Behaviour

- Behaviour different from normal applications
- Content invariance
- Content prevalence
- Address dispersion
- Extensive traffic generation

Content Invariance and Content Prevalence

- Existing worm signatures invariant across all copies
- Can make use of polymorphism methods
- Encrypting the actual worm code
- Decryption routines are still invariant
- Invariant portion of the worm will appear frequently on network

Address Dispersion

- Number of infected hosts will grow over time
- Packets containing worms will reflect varied source and destination addresses

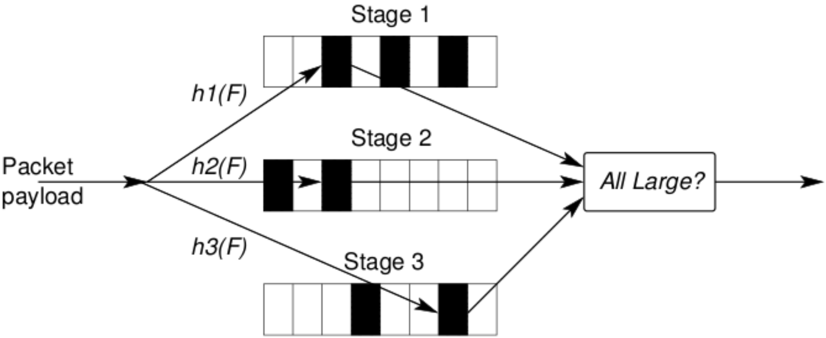
Experimental Testing and Results

- Algorithm must have small processing requirements
- Ideally can be parallelized
- Should not depend on a symmetric vantage point in the network

Estimating Content Prevalence

- Find packet payloads appearing at least x times among N packets
- Instead of storing packet payload, hash of payload is stored
- Collision probability in hash table can be reduced by using a suitably large range
-

Multi-Stage Filter



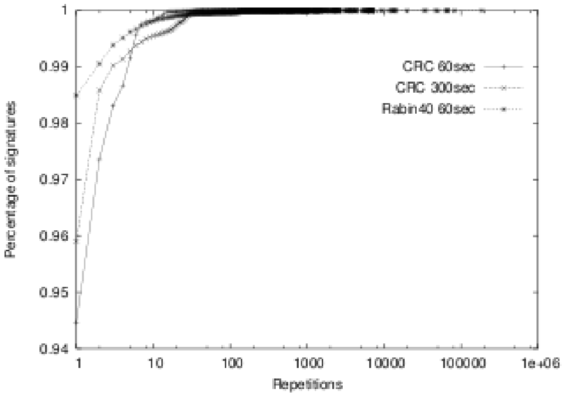
Estimating Address Dispersion

- Count distinct source IP and destination IP addresses
- Different from estimating only repetitions of a distinct string
- Trading between accuracy and memory requirements
- Using bitmaps

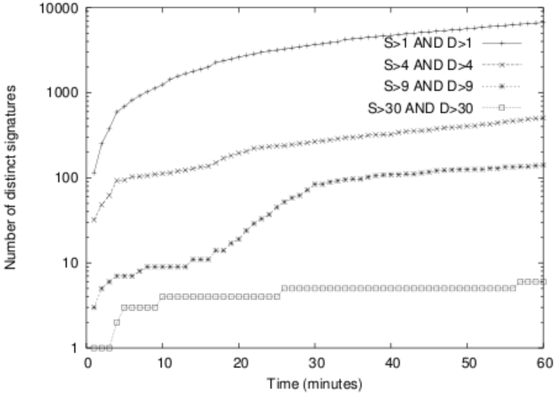
Parameter Tuning

- Content prevalence threshold set to 3
- Address dispersion threshold set to 30 source and 30 destination addresses

CDF of Signatures



Distinct Signatures Detected Over Time



Earliest Automation System

- By Kephart and Arnold
- Used decoy programs in controlled environment
- Allows decoys to get infected
- Extraction of infected regions and used as signatures

Autograph

- By Kim and Karp
- Uses network–level data to infer worm signatures
- Rabin fingerprints to index counters of content sub–strings
- Compared to Earlybird cannot detect a large number of worms

Assessment

- Depends on invariant code or strings
- Cannot detect truly polymorphic worms
- Works well for older worms but may will not be able to adapt to polymorphic behaviour

References

- S. Singh, C. Estan, G. Varghese, and S. Savage, "Automated worm fingerprinting," in *OSDI*, pp. 45–60, 2004.
- http://en.wikipedia.org/wiki/Intrusion_detection_system.
- http://en.wikipedia.org/wiki/Computer_worm.
- http://vx.netlux.org/lib/ajm01.html.