Detection of Botnets Using Combined Host- and Network-Level Information

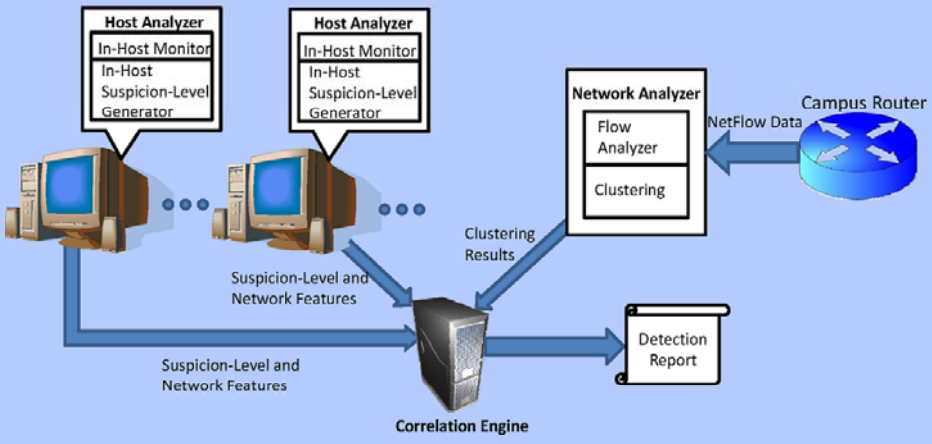
Yuanyuan Zeng, Xin Hu, Kang G. Shin
The University of Michigan
{gracez, huxin, kgshin} @eecs.umich.edu

Motivation

- Botnets are one of the most serious security threats to Internet services and applications.
- Traditional IRC- or HTTP-based botnets are vulnerable to a central-point-of-failure; attackers have recently shifted towards a decentralized Command & Control (C&C) protocol such as P2P.
- Most existing botnet-detection approaches only look for traffic signatures or flow patterns, and are thus unlikely to have a complete view of botnets' behavior.
- Host-based approaches are useful to identify each bot's in-host behavior, and hence are susceptible to the host-resident malware, if used alone.

System Architecture

- Host Analyzer
  - -- In-Host Monitor
  - -- In-Host Suspicion-Level Generator
- Network Analyzer
  - -- Flow Analyzer
  - -- Clustering
- Correlation Engine



Methodology

In-Host Monitor: Captures system-wide behavior in real time at different locations and transforms it into a behavior vector

Index	Behavior Features
1	DLL or EXE Creation into System Directory
2	Modification of Files in System Directory
3	Creation of AutoRun Key in Registry
4	Creation of Process Injection Key in Registry
5	Modification of Critical Registry Key (Disabling taskmgr; Overriding antivirus,firewall keys, etc.)
6	Number of Ports Opened
7	Number of Suspicious Ports
8	Number of Unique IPs Contacted
9	Number of SMTP Flows

In-Host Suspicion-Level Generator: Generates a suspicion level for each behavior vector

Flow Analyzer: Processes flow records to extract trigger-action patterns and form feature vectors; coordinated behaviors are invariant to all types of botnets

Index	Flow Features
1 to 4	Duration Mean, Variance, Skewness and Kurtosis
5 to 8	Totalbytes Mean, Variance, Skewness and Kurtosis
9 to 12	Number of Packets Mean, Variance, Skewness and Kurtosis
13	Number of TCP Flows
14	Number of UDP Flows
15	Number of SMTP Flows
16	Number of Unique IPs Contacted
17	Number of Suspicious Ports

Clustering: Uses hierarchical clustering to group similarly-behaving hosts

Correlation Engine: Considers both the suspicion level and the quality of clustering to produce a detection result

Evaluation Results

- Data Collection
  - -- 5-day NetFlow data from a core router in our campus network
  - -- Botnet traces from virtual machines

Trace	Duration	Number of Bots
IRC-rbot	24h	4
IRC-spybot	32m	4
HTTP-BobaxA	4h	4
HTTP-BobaxB	20h	4
Storm	48h	4
Waledac	24h	4

- Detection Results
  - -- False-positive (FP): a benign host classified as bot-infected
  - -- False-negative (FN): a bot-infected host fails to be detected

Trace	Average FP Hosts	Average FP	Average FN Hosts	Average FN	Duration
IRC-rbot	3.208	0.0016	0.125	0.0313	24h
IRC-spybot	2.833	0.0014	0	0.0000	24h
HTTP-BobaxA	1.000	0.0005	0	0.0000	24h
HTTP-BobaxB	1.083	0.0005	0	0.0000	24h
Storm	2.563	0.0013	0	0.0000	48h
Waledac	0.9167	0.0005	0	0.0000	24h