(Im)possibility of Enumerating Zombies

Yongdae Kim (U of Minnesota - Twin Cities)
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bot Bin/Sources + Bots</td>
<td>Sell Bots - HTTP/IRC etc here...</td>
</tr>
<tr>
<td>Stealers / Keyloggers / Rats</td>
<td>Sell Firefox/Steam etc Stealers here...</td>
</tr>
<tr>
<td>Accounts</td>
<td>Sell Cpanels/WHM's etc here...</td>
</tr>
<tr>
<td>Crypters/Downloader</td>
<td>Sell Packers/Crypters/Binders here...</td>
</tr>
<tr>
<td>Servers and Hosting</td>
<td>Sell Servers/Roots/VPS's/Hosting/Shells etc here...</td>
</tr>
<tr>
<td>Other</td>
<td>Sell Other stuff here, which doesn't fit in other categories, eg. Droped sites...</td>
</tr>
<tr>
<td>Exploits</td>
<td>Sell 0day Exploits here...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC's</td>
<td>Sell CC's, Specify Country, Price, Minimum Amount</td>
</tr>
<tr>
<td>Gift Cards</td>
<td>Sell Any Gift Cards in here</td>
</tr>
<tr>
<td>Cardable</td>
<td>Post Sites you've carded here &amp; Chat...</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Gunter Ollmann at Damballa's blog
Botnet and DDoS

Botnets becoming the major tool for DDoS

5 million nodes Botnet

- 5 PHz CPU (1 GHz CPU/bot)
- 5 PB RAM (1 GB RAM/bot)
- 5 TB upload bandwidth (1 MB/bot)

When we detect DDoS, we might be too late!

Either kill the botnet or at least, find zombies!
Botnet Architectures

万博 Centralized
  ‣ IRC
  ‣ Central server is a critical weak point
  ‣ If disabled, the botnet fails

万博 Decentralized
  ‣ More robust
  ‣ Often P2P architecture
  ‣ Each peer performs server functions
Decentralized Botnet Architecture

Bot master

Master servers

Proxy bots

Worker bots

Overnet
P2P Systems

How to find the desired information?
- Centralized structured: Napster
- Decentralized unstructured: Gnutella
- Decentralized structured: Distributed Hash Table (DHT)

A DHT provides a hash table's simple put/get interface
- Insert a data object, i.e., key-value pair \((k,v)\)
- Retrieve the value \(v\) using key \(k\)

O: offerer of the file
P: node looking for a file
X: a node searching for a file
Query: match
Download: retrieve \((K_1)\)
P2P Routing Type

Recursive Routing

Iterative Routing
DHT Protocol Message Types

♀ Connect
  ▸ To start a node, it needs other contacts for its routing table.
  ▸ Ask other nodes about their contacts.

♀ Publicize
  ▸ Ping message to check/verify liveness

♀ Routing
  ▸ Returns k contacts

♀ Publish
  ▸ Store information

♀ Search
  ▸ Find published information
Kademlia Protocol

- \( d(X, Y) = X \odot Y \)
- An entry in k-bucket shares at least k-bit prefix with the nodeID
  - k=20 in overnet
- Add new contact if
  - k-bucket is not full

- Parallel, iterative, prefix-matching routing
- Replica roots: k closest nodes
Storm Worm operation

- Usually infected by clicking links in spam mail, malicious binaries, and everything else
- Installs rootkit
- Disable windows firewall
- Overnet routing table bootstrapping
- Connect to overnet
- Put and get a lot of hashes
- Download and decrypt secondary injection URL
- Execute secondary injection
Finding Nodes in a P2P Network
Take 1: Confirmation Attack

火烧 If handshake algorithm is known, crawl the whole Internet!

Example: Conficker C

Expensive

Yelling from admins ;-(

University of Minnesota
Take 2: Global Observer

If network signature is known, each ISP checks if its client is infected!
Sharing information

No incentive for ISP
Politics!
Take 3: Targeted Enumeration

If we know what they are looking for

Conficker A and B C&C channel blocked by Microsoft and Cabal group
Take 4: Crawler

A node relies on other nodes to publish/search information

Two possible cases
- Iterative routing: information about other nodes have to be sent to help routing
- Bootstrap: Need to know information about other nodes to start a node

Algorithm

Input: $IP = \{\text{known IPs having bots}\}$

While (1) {
    Send connect or search;
    Receive and store IP;
    If no new IPs are found, break;
}

Output $IP$
Take 4: Crawler (cnt)

♀Pros
  › Quickly find nodes reachable from outside
    ♀ 11 minutes to crawl 2M Kad Network [Steiner 07]

♀Cons
  › Nodes behind a firewall/NAT box cannot be found
    ♀ Typically, worker bots...
Take 5: Passive P2P Monitoring

Input
- $IP = \{\text{known IPs having bots}\}$
- $PPM\ nodes = \{n_1, n_2, ..., n_k\}$

Algorithm
PPM nodes join Storm overnet
While (1){
    Receive packets from Storm and store IP;
}

Output $IP$ periodically
P2P Network Monitoring (cnt.)

Pros
- Continuous monitoring
- Sufficient backpointers by running it long time
  - Eclipse Attack

Cons
- Passive...
- Spoofed communication?
P2P Network Monitoring Result

Aug 30, 2007
- Collect 24G of logs from 256 nodes
- Initial IP: Results of one targeted attack (180 IPs)
- Detect 230k (probable) bots

Jan 28, 2008

```
PPM (224K)

161K  41K  16K

63K  22K

Crawler, Reply (57K)
Crawler, No reply (85K)
```

- Why are they different?
Firewall/NAT Checker

_possible_reason could be because of NAT boxes and firewalls.
  - Not reachable by crawler
  - But, they can still send queries to PPM.

_How do we verify that a node is under firewall/NAT?
Firewall Checker Design

Message 4 means bot IP is not spoofed.

Message 6 means bot is under firewall/NAT box.
Result (PPM vs. FWC)
Crawler vs. PPM: # of IPs found

Number of IP Adresses found

Day

0 2 4 6 8 10 12 14 16 18 20

PPM
Crawler

0 100000 200000 300000 400000 500000 600000 700000 800000 900000 1000000 1100000 1200000
Lifetime of Ips found by Crawler, PPM
Analysis of Coverage of PPM

When

- $p$ is the probability of PPM receiving a message from a bot for a particular hash
- $k$ is the number of nodes a bot sends a message with that hash to

Then probability of PPM receiving a message from a bot is calculated as

$$L = 1 - (1 - p)^k$$

How do we obtain $p$ and $k$?

- Experimentally
In-degree comparison

- Search
- Get
- Put
Node distribution: search/publish
256 Node PPM Coverage (k message)
Future Botnets

♀ Current botnet design is terrible!
♀ Does unenumerable botnet exist?
Questions?
Send e-mail to kyd@cs.umn.edu