Le point de vue d'un WOMBAT sur les attaques Internet

Marc Dacier
Directeur, Symantec Research Labs Europe
Overview

• Introduction
  – Data Acquisition
  – Data Enrichment
  – Threats Analysis

• Conclusions
Foreword

• What is presented here is the result of a joint collaboration between all WOMBAT partners over the last 14 months

(see www.wombat-project.eu for the list of publications and deliverables)
Take-away Message

• WOMBAT is collecting and offering data for collaboration with other organizations

• A lot remains to be done to efficiently identify, analyze and counter the modus operandi of the malicious actors on the Internet

• Understanding these strategies is key to enable cyber security situational awareness.

• Looking at raw material, eg malware, is not enough. We must enrich it with metadata and contextual information.
The WOMBAT approach

Data acquisition (WP3)

Data enrichment (WP4)

Threat analysis (WP5)

New collection practices

Storage Analysis

Meta-data Analysis

Malware analysis

Context analysis

Honeypots

Crawlers

External feeds

New security technologies

New security practices

Knowledge
Overview

• Introduction

  – Data Acquisition
  – Data Enrichment
  – Threats Analysis

• Conclusions
• Wombat builds upon two complementary approaches:
  – A WAPI API
  – A federated Database (proxy) for non persistent datasets

• New sensors are developed
  – SGNET
  – Honey clients
  – Bluetooth, WIFI
  – …
Example of a new sensor: SGNET

- Code Injection information
- Malware

Internet

SGNET

AV identification statistics

VIRUS TOTAL

Behavioral information

Anubis

SGNET dataset
Normal operation
New exploit encountered
Global update of the FSM knowledge
Submission of a shellcode sample
Analyze new malware sample
Overview

• Introduction
  – Data Acquisition
  – Data Enrichment
  – Threats Analysis

• Conclusions
SGNET data enrichment framework

- Code Injection information
- Malware

Internet

- SGNET dataset

Clustering techniques

- AV identification statistics
- Behavioral information
- Generated alerts

Models

SGNET dataset

Anubis

Virus Total

Snort

WOMBAT
1. Activity classification

- The interaction with the FSM model can be used to characterize the network interaction
- What is taken into account
  - Features of the protocol (e.g. “HELO”)
  - Features of the specific exploit tool (e.g. same username)
1. Activity classification

The Conficker example

- November 2008: raise of the Conficker worm
  - SGNET generates a new path for the anonymous NetBIOS authentication used by the worm
How difficult is it?

- Total number of traversals generated by the deployment
- Process of death and birth of traversals
• An increase in load handled by the system is “absorbed” by the learning process
A host connects to another peer on port 9988 and pushes data.
Multi-headed propagation

- Known propagation strategies for Rbot.bni
  - The same malware type uses very different propagation strategies
  - IDS such as Snort expect the ASN.1 exploit only on port 445, instead we are witnessing it also on port 139!
3. Automated threat response

The problem

- Each different activity type is plotted according to the number of involved attackers and victims (its “size”)
Overview

• Introduction
  – Data Acquisition
  – Data Enrichment
  – Threats Analysis

• Conclusions
From raw events to meta data
Raw tcpdump traces
Clusters
Attack Events

Attack Time Series
(2006-12-13 -> 2007-04-02)
• Evolution of attack events may reveal information about the “keep alive” strategies of the attackers.

• Multi dimensional analysis enables us to derive hidden links between attack events.

• Contextual information regarding the malware gives insight on code evolution, transformation.
Attack Events split by ...
Empirical CDF

- $T_{\text{Platform}}$
- $T_{\text{Country}}$

CDF vs. common source ratio

Not a single explanation
2 distinct survival strategies

- 2 distinct groups of attack events highlighting different evolutions
  - Left side: attack events have common IPs only with their direct predecessor and successor event
  - Right side: attack events have common IPs with all other events over a 700 days period of time
Can zombie armies exist for that long?
Multidimensional analysis

- If long term phenomena exist, how can they be explained?

- Are they simply due to some coincidence or experimental errors?

- If these events have not been grouped randomly together, they should have some other characteristic(s) in common.
Example 1+: Adding the platforms viewpoint

- Those 4 botnet waves have hit the same group of platforms

- **But:** not all botnet waves came from the same groups of IP Netblocks
  - Dynamic evolution of the botnet population
  - Still, certain “stable” clusters of IP blocks (see ipmaps on next slides)
Example 1: Botnet wave 1
**Example 1:**

*Botnet wave 2*

![IPv4 Map](image)

<table>
<thead>
<tr>
<th>Local</th>
<th>2</th>
<th>13</th>
<th>12</th>
<th>17</th>
<th>18</th>
<th>23</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>31</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>31</td>
<td>28</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>56</td>
<td>57</td>
<td>54</td>
<td>53</td>
<td>32</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>62</td>
<td>61</td>
<td>60</td>
<td>59</td>
<td>58</td>
<td>34</td>
<td>39</td>
<td>38</td>
</tr>
</tbody>
</table>

**Intensity (rel. to max)**
- > 0.66
- 0.33 < x < 0.66
- < 0.33

IPv4 Map
- IANA reserved
Example 1: Botnet wave 3
Example 1:
Botnet wave 4

IPv4 Map
- IANA reserved

Intensity (rel. to max)
- > 0.66
- 0.33 < x < 0.66
- < 0.33
Example 2: Dimension-4 viewpoints

- "Multi-headed" attack tool
  - Nov. 2006
  - Port. Seq.: 1433T – 5900T
  - 7.3K sources
  - Dimension-4 concept:
    - Same group of countries
    - Same group of subnets
    - Same time series
    - Same group of 3 sensors hit, all in the same /8!
Ongoing work

- Many features can be used to find relationships between groups of events.

- Not all features are relevant all the time

- There is work in progress on building an automated framework that includes the expert knowledge in order to extract meaningful sets to reason about the modus operandi of the malicious actors.
Overview

• Introduction
  – Data Acquisition
  – Data Enrichment
  – Threats Analysis

• Conclusions
Conclusions

• The WOMBAT likes to have new friends.
  – Join the team!

• The WOMBAT has plenty of toys and is eager to share them with his partners.
  – Benefit from the datasets and tools developed so far

• The WOMBAT is always hungry for new datafeeds.
  – Install a sensor at your place.

• CONTACT POINT: marc_dacier@symantec.com
Thanks!

Questions?