Anomaly Detection on User-agents

Peter van Bolhuis
Overview

- Introduction
- Research Question
- User-agents
- Scoring host anomalies
- Verification
- Conclusion
Introduction

• Methods
  – Statistical
  – Knowledge based
  – Machine learning
Research Question

What is the effectiveness of statistical anomaly detection when applied to user-agent strings?
User-agents

- Programs that “act on behalf of a user”
- Identify themselves with a string
  - Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; WOW64; Trident/7.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; eSobiSubscriber 2.0.4.16; BRI/1; MAAR; .NET4.0C; AskTbORJ/5.15.9.29495; .NET4.0E; BRI/2) Funshion/1.0.0.1
User-agents (2)

• Problem:
  – Mozilla/4.0 (compatible; version 1.33.7)
  – Mozilla/4.0 (compatible; version 1.33.8)
  – Dalvik/1.4.2 (AskTbORJ/5.15.9.29495)
User-agents (3)

- Splitting on elements
  - Mozilla/4.0
  - Dalvik/1.4.2
  - Compatible
  - Version
Anomaly Detection on User-agents
Scoring host anomalies

- Elements with the lowest \( n \) occurrences give a host a +1

<table>
<thead>
<tr>
<th>User-agent element</th>
<th>#Occurrences</th>
<th>Increases score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozilla/4.0</td>
<td>100</td>
<td>No</td>
</tr>
<tr>
<td>AppleWebKit/537.36</td>
<td>20</td>
<td>No</td>
</tr>
<tr>
<td>Dalvik/1.4.0</td>
<td>10</td>
<td>Yes</td>
</tr>
<tr>
<td>AppWorld/5.0</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Q10/10.2.1.3175</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>zh-cn</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>4012FREE</td>
<td>1</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table for \( n = 3 \)
Scoring host anomalies (2)

![Graph showing hosts with scores greater than 1](image)
## Verification

<table>
<thead>
<tr>
<th>Host</th>
<th>Score</th>
<th>Result of verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>299</td>
<td>Host was a phone: Compliance incident</td>
</tr>
<tr>
<td>B</td>
<td>64</td>
<td>Host infected with Conduit Browser Hijacker</td>
</tr>
<tr>
<td>C</td>
<td>157</td>
<td>Host was a proxy</td>
</tr>
<tr>
<td>D</td>
<td>353</td>
<td>Host was a proxy</td>
</tr>
</tbody>
</table>

![Anomaly Detection Graph]
Conclusion

• User-agent strings can be used for anomaly detection
  – Best results on uniform networks
  – Anomalies are not necessarily infections, but rather installed software packages
Demo
Thank you