A TAXONOMY OF JAVASCRIPT REDIRECTION SPAM

Kumar Chellapilla, Alexey Maykov
Microsoft Live Labs
AIRWeb 2007

CSE 450 Web Mining Seminar
Presented by Liangjie Hong
Feb 20th, 2008
BACKGROUND & INTRODUCTION

What is Spam?
- Any deliberate human action that is meant to trigger an unjustifiably favorable relevance or importance for some web page, considering the page’s true value.

- Get higher rank in Search Engines’ index

Categories
- Boosting
  - seeks to achieve high relevance and/or importance
- Hiding
  - hide the adopted boosting techniques from human web users.
BACKGROUND & INTRODUCTION

Hiding Techniques

- Content Hiding
- Cloaking
- Redirection

Search Engine or Crawler

Intermediates (proxies, doorways)

Human Users

Destination Page
**JavaScript Redirection**

- **Why JavaScript?**
  - Obfuscation
  - Dynamic code generation
  - Self modifying

- **Hard to detect!**
  - Script-agnostic
  - Only use static analysis
  - Almost impossible to predict the final behavior
Basic Redirection Techniques

- Directly change Location

  ```html
  <script type="text/JavaScript">
    window.location = "http://www2007.org/"
  </script>
  <script>
    document.location = "http://www2007.org/"
    location.href = "http://www2007.org/"
    location.replace("http://www2007.org/")
  </script>
  ```

- Using Timeout

  ```html
  <script>
    function delayed_redirect()
      { window.location = "http://www2007.org/" }
  </script>
  <body onLoad="setTimeout('delayed_redirect()', 5000)">
  ```

- DOM Event

  ```html
  Document: onload, onunload,
  onchange, onsubmit, onreset, onselect, onblur, onfocus,
  Keyboard: onkeydown, onkeypress, onkeyup,
  Mouse: onclick, ondblclick, onmousemove,
  onmousedown, onmouseover, onmouseout, onmouseup
  ```
JavaScript Redirection Techniques I

- **String Manipulation and Eval()**

```javascript
var a1="win", a2="dow.", a3="loca", a4="tion.", a5="replace", a6="('http://www.partypoker.com/index.htm?wm=2501068')";
var i,str="";
for(i=1;i<=6;i++){
    str += eval("a"+i);
}
eval(str);

http://party-poker-bonus.cjb.net/
```

- Avoid presenting the whole redirect action as a single string
- Using string concatenation and just `eval()`
- **Throw off parsers** without the ability to aggregate constants!
**JavaScript Redirection Techniques II**

- **Unescape and Decode (Obfuscation)**

```javascript
var s = '\%5CBEOD%5C%05GDU_BDE\%6C%0C__%5B%11%04%04%5C%5C%5C%05
SMYNFD%5DBNX%05HDF%04%0C';
var e = '', i;
 eval(unescape('s\%3Dunescape%28s%29%38for%28s%3D0%3B%3Cs.length%3B%2B%2B%2B%7Be%2B%3DString.fromCharCode%28s.charCodeAt%28s%29%5E43%29%3B%7D%3Beval%28e%29%3B'));
http://freegaypornfortodays.blogspot.com/2006_10_01_freegaypornfortodays_archive.html
```

- Avoid discovery of direct checks
  - **Encoding characters**
  - **Custom decoding schemes**

```javascript
var tt, kk="", mm;
  tt="w\nd\w\s!\c#|\n;\{[\*]!!*r\l|\n$s\f>dl$s#rc($*(\")q;\c(\">+c\|g\r>\{s\"");
  for (i=0; i<tt.length+1; i++)
  {
    mm=tt.substring (i,i+1);
    if (mm=="(") mm="h^"; if (mm=="*\") mm="p"; if (mm=="l\") mm="/";
    if (mm==":" ) mm="e"; if (mm=="$\") mm=";"; if (mm=="l\") mm="t";
    if (mm=="f\") mm="a"; if (mm=="^\") mm="o"; if (mm=="l\") mm="?";
    if (mm=="@\") mm="k"; if (mm=="(" ) mm="&"; if (mm=="l") mm=".";
    if (mm=="\") mm="s"; if (mm=="|\") mm="l"; if (mm=="+" ) mm="+");
    kk=kk+mm;
  }
  eval (kk);
```
Injection (Script, Form, Event ...)

- Dynamic code generation and self modifying
  - Script on other sites
  - Post to redirect
  - Simulate click event!
PREVALENCE OF JAVASCRIPT REDIR-SPAM

Table 1. Percentage occurrence of JavaScript redirection among popular and blogspot.com pages

<table>
<thead>
<tr>
<th>URL Type</th>
<th>Count / Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popular</td>
<td>2,712 / 782,937 = 1 in 288</td>
<td>0.35%</td>
</tr>
<tr>
<td>Blogspot</td>
<td>7,196 / 934,876 = 1 in 130</td>
<td>0.77%</td>
</tr>
</tbody>
</table>

---

![Graph showing prevalence of JavaScript redirection spam](graph1.png)

![Graph showing JavaScript redirection spam by technique](graph2.png)
# Prevalence of JavaScript Redir-Spam

Table 1. Percentage occurrence of JavaScript redirection among popular and blogspot.com pages

<table>
<thead>
<tr>
<th>URL Type</th>
<th>Count / Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Popular</td>
<td>2,712 / 782,937 = 1 in 288</td>
<td>0.35%</td>
</tr>
<tr>
<td>Blogspot</td>
<td>7,196 / 934,876 = 1 in 130</td>
<td>0.77%</td>
</tr>
</tbody>
</table>

**Bar Chart:**
- Percentage of JavaScript redirection spam pages using the technique.

**Chart:**
- Percentage of JavaScript redirection spam by number of techniques used.

**Legend:**
- **POPULAR**
- **BLOGSPOT**
CONCLUSION

- Crawlers and search engines cannot be script-agnostic.

- Obfuscation techniques limit the effectiveness of static analysis and static feature based systems.

- Machine-learning based systems may only be partially effective.

- Dynamic analysis needs!