Fast fingerprinting of OLE2 files:
heuristics for detection of exploited OLE2 files based on specification non-conformance
Fast fingerprinting of OLE2 files

Authors

- Stephen Edwards, SophosLabs UK
- Paul Baccas, SophosLabs UK
- {stephen.edwards, paul.baccas}@sophos.com
Fast fingerprinting of OLE2 files

• What were we trying to achieve?
• How we went about it?
• What we found?
• What happened next?
Fast fingerprinting of OLE2 files

What were we trying to achieve?

• Troj/DocDrop-S
• Binary specification
• Prototyped a non-conformance scanner
Fast fingerprinting of OLE2 files

What were we trying to achieve?

• Fast method to hash on files that we wanted to spend time on.
• Differentiate different threats
Fast fingerprinting of OLE2 files

How we went about it?

- Read the Microsoft Specifications
- Initially we implemented the detection as an internal detection
- Implemented it in python
Fast fingerprinting of OLE2 files

How we went about it?

**BOF record** – 128 bits

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>reserved1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>vers</td>
<td>dt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rupBuild</td>
<td>rupYear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fast fingerprinting of OLE2 files

How we went about it?

**BOF record** – 128 bits: 104 can violate the specification

|   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 3 | 0 | 1 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| vers |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| dt |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| rupBuild |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| rupYear |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | reserved1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| verLowestBiff |   | O |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| reserved2 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Fast fingerprinting of OLE2 files

How we went about it?

Fast fingerprinting of OLE2 files

How we went about it?

Fast fingerprinting of OLE2 files

What we found? Clustering
Fast fingerprinting of OLE2 files

What we found?

- Troj/DocDrop-S grouped well
- XM97/Hidemod-A also grouped
Fast fingerprinting of OLE2 files

What we found? Troj/DocDrop-S

- CVE-2009-3129
- MS09-067
- Interesting sample names
- Some more on my blog
Fast fingerprinting of OLE2 files

- CVE-2009-3129
- MS09-067
- Interesting sample names
- Some more on my blog

What we found?

Troj/DocDrop

VB Conference Barcelona 2011
Fast fingerprinting of OLE2 files

• CVE-2009-3129
• MS09-067
• Interesting sample names

What we found?

Troj/DocDrop-S

VB Conference Barcelona 2011
# Fast fingerprinting of OLE2 files

What we found? Troj/DocDrop-S

<table>
<thead>
<tr>
<th>Bit index</th>
<th>Property name</th>
<th>Spec notes</th>
<th>Probable meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>rupYear</td>
<td>“The value MUST be 0x07cc or 0x07cd”</td>
<td>Maps to Excel spec version. Is set to 0x0700 in this instance; 1792!</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Excel97 writes 0x07cc for rupYear” (1996)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>fRisc</td>
<td>“MUST be 0”</td>
<td>Unknown. This bit is so commonly set that the given meaning in the spec seems unlikely.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“last edited on RISC platform”</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>fWinAny</td>
<td>“SHOULD be 1”</td>
<td>Flag denoting if the file has ever been edited on Windows.</td>
</tr>
<tr>
<td>8</td>
<td>fMacAny</td>
<td>“MUST be 0”</td>
<td>Flag denoting if the file has ever been edited on Mac. Again, this violation is very common, and it would make little sense to have this bit, have Mac products, yet never set it!</td>
</tr>
<tr>
<td>11</td>
<td>2(^{nd}) bit of unused1 field</td>
<td>“Undefined and MUST be ignored”</td>
<td>Unknown.</td>
</tr>
<tr>
<td>35</td>
<td>reserved2</td>
<td>20 bits, “MUST be zero”</td>
<td>Unknown. The 20 following bits often have a consistent pattern within a group it is likely these bits are in use and have an undocumented meaning.</td>
</tr>
<tr>
<td>38-39</td>
<td>Bits 2-3 of the reserved2 bits</td>
<td>As above</td>
<td>As above</td>
</tr>
</tbody>
</table>
Fast fingerprinting of OLE2 files

What we did next?

- Looked at a more robust directory chain parser
- Extended to WordDocument
- …
Fast fingerprinting of OLE2 files

What we did next? Extended to WordDocument

• File Information Block (Fib)
• Starts with FibBase
• 32 bytes 12 possible Must and 2 possible Should
• Going further in the docs looks promising
Fast fingerprinting of OLE2 files

<p>| | | | | | | | | | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>wIdent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>unused</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pnNext</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>nFibBack</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>reserved3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>reserved4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>reserved5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>reserved6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fast fingerprinting of OLE2 files

File Information Block (Fib)

- Starts with `FibBase`
- 32 bytes, 12 possible Must and 2 possible Should

Going further in the docs looks promising

What we did next? Extended to WordDocument

[Diagram showing file information block structure with columns for wIdent, nFib, unused, lid, pnNext, nFibBack, lKey, envr, reserved3, reserved4, reserved5, reserved6]
Fast fingerprinting of OLE2 files

What we did next? WordDocument
Fast fingerprinting of OLE2 files

What we did next? WordDocument

• Group 7
• 36 files all same violation
• …
Fast fingerprinting of OLE2 files

Conclusion

• Spec violations can provide good grouping.
• Quick/cheap less than 4 loads
• The fingerprints are common over diverse campaigns
Fast fingerprinting of OLE2 files

Questions?

• ...

VB Conference Barcelona 2011