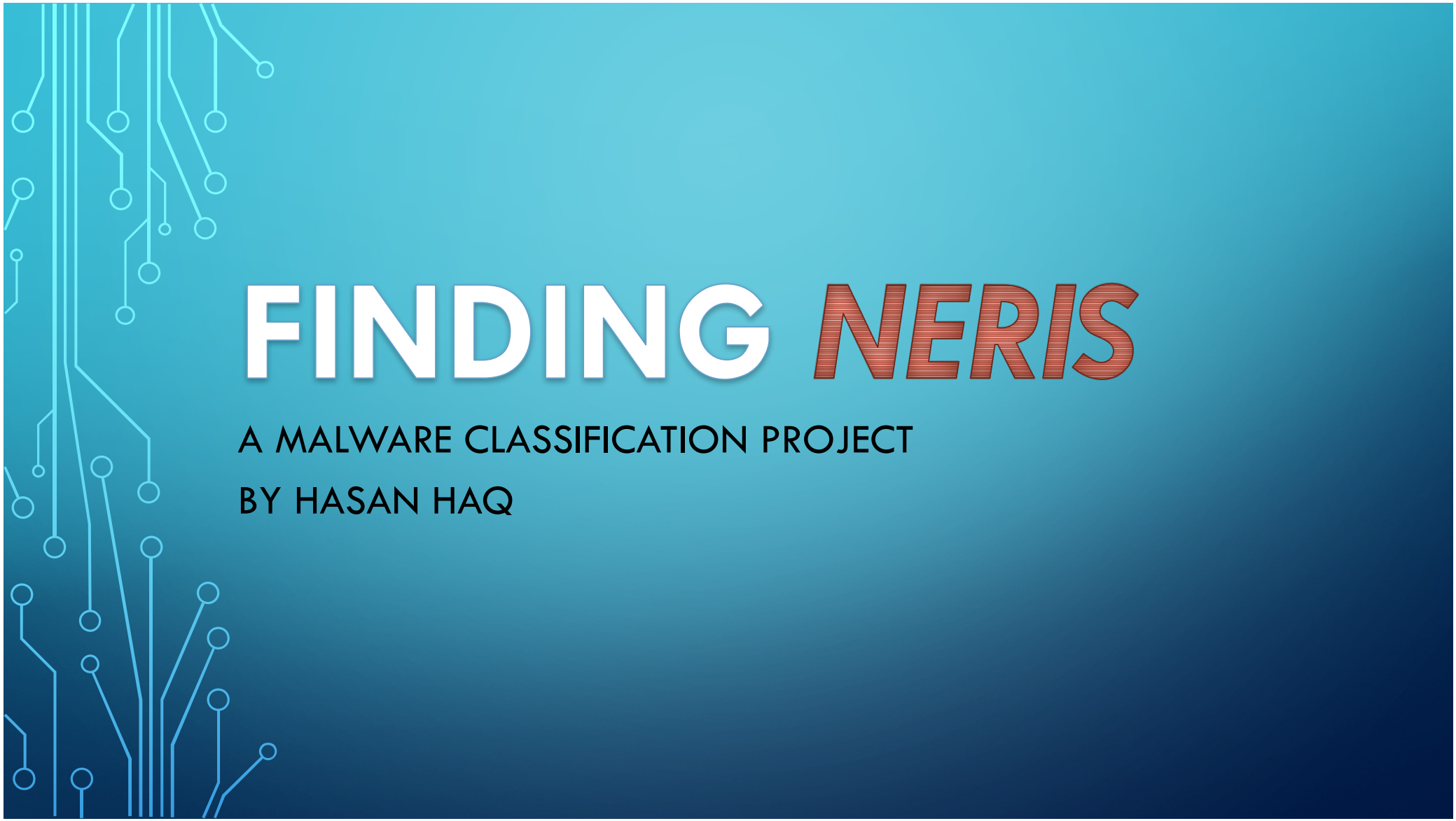

FINDING NERIS
A MALWARE CLASSIFICATION PROJECT
BY HASAN HAQ

NERIS.EXE

- Botnet ran on University network for 6.15 hours
- HTTP-based C&C channel
- Send SPAM; perform click-fraud
- Md5: bf08e6b02e00d2bc6dd493e93e69872f

Source: http://mcfp.weebly.com/ctu-malware-capture-botnet-42.html

THE HORROR!!

POST/?c799959d9582d49995979194948 2d19995939782d29997909691 82c699959c949c92959c82c0999582d79995969c959d9d9482c199e79ef8f3edeae0ebf3f7f8f0e1e9f4f893ccd

dddcccad3c28ac1dcc182c399cdcacdd0a4 HTTP/1.1

HTTP/1.1 200 OK

Date: Wed, 10 Aug 2011 09:41:53 GMT

Server: Apache/2.2.8 (Fedora) DAV/2 PHP/5.2.6 mod_ssl/2.2.8 OpenSSL/0.9.8g

X-Powered-By: PHP/5.2.6

Content-Length: 26

Connection: close

Content-Type: text/html; charset=UTF-8

CB2=212.117.171.138:65500

The Tools

Argus - The All Seeing
System and Network Monitoring Software

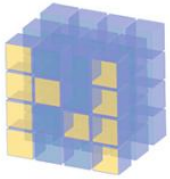
pandas
$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

python™

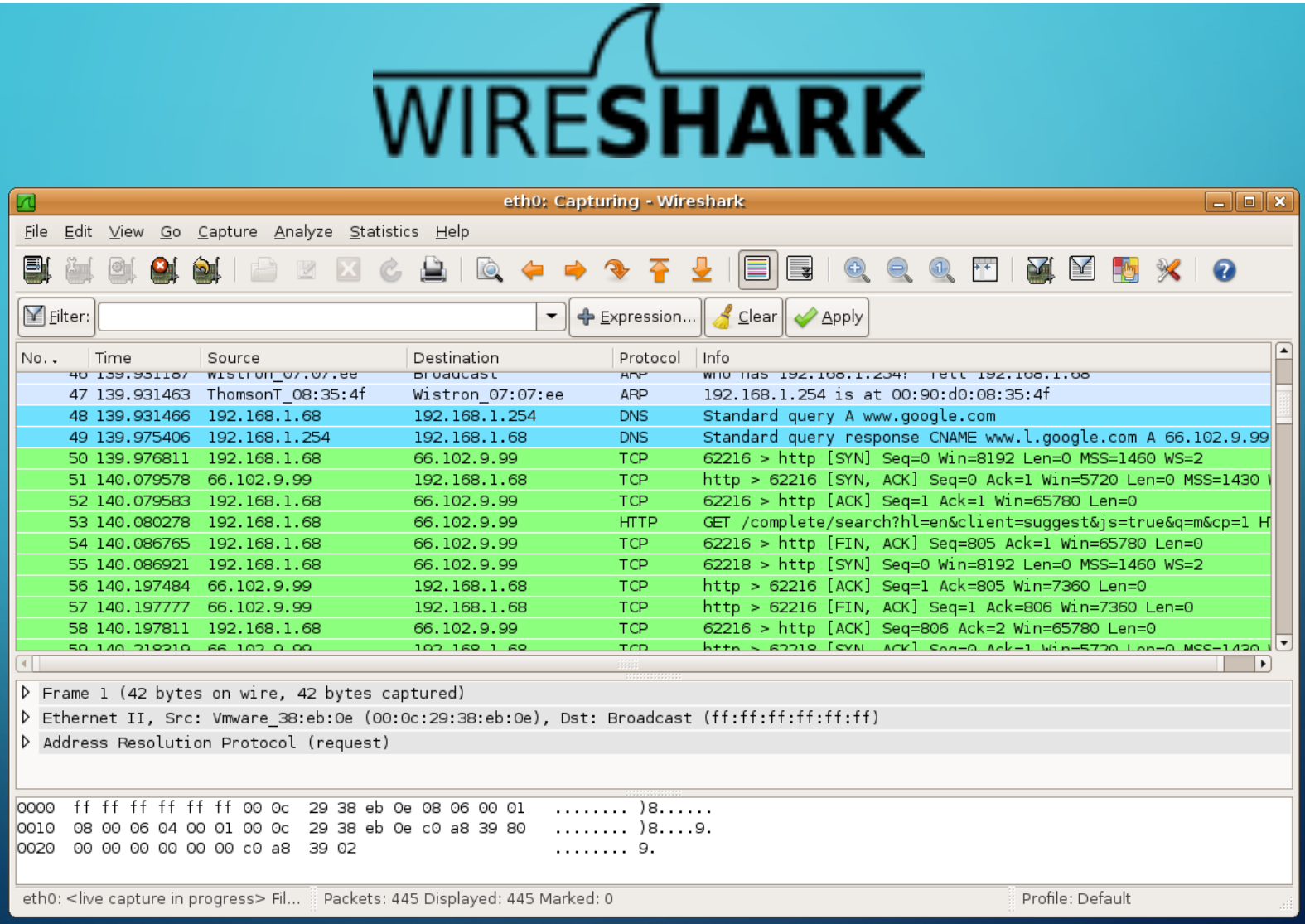
WIRESHARK©

TCPDUMP

scikit learn



NumPy


WIRESHARK
eth0: Capturing - Wireshark
File Edit View Go Capture Analyze Statistics Help
Filter:
Expression... Clear Apply
No. Time Source Destination Protocol Info
47 139.931463 ThomsonT_08:35:4f Wistron_07:07:ee ARP 192.168.1.254 is at 00:90:d0:08:35:4f
48 139.931466 192.168.1.68 192.168.1.254 DNS Standard query A www.google.com
49 139.975406 192.168.1.254 192.168.1.68 DNS Standard query response CNAME www.l.google.com A 66.102.9.99
50 139.976811 192.168.1.68 66.102.9.99 TCP 62216 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
51 140.079578 66.102.9.99 192.168.1.68 TCP http > 62216 [SYN, ACK] Seq=0 Ack=1 Win=5720 Len=0 MSS=1430
52 140.079583 192.168.1.68 66.102.9.99 TCP 62216 > http [ACK] Seq=1 Ack=1 Win=65780 Len=0
53 140.080278 192.168.1.68 66.102.9.99 HTTP GET /complete/search?hl=en&client=suggest&js=true&q=m&cp=1
54 140.086765 192.168.1.68 66.102.9.99 TCP 62216 > http [FIN, ACK] Seq=805 Ack=1 Win=65780 Len=0
55 140.086921 192.168.1.68 66.102.9.99 TCP 62218 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
56 140.197484 66.102.9.99 192.168.1.68 TCP http > 62216 [ACK] Seq=1 Ack=805 Win=7360 Len=0
57 140.197777 66.102.9.99 192.168.1.68 TCP http > 62216 [FIN, ACK] Seq=1 Ack=806 Win=7360 Len=0
58 140.197811 192.168.1.68 66.102.9.99 TCP 62216 > http [ACK] Seq=806 Ack=2 Win=65780 Len=0
Frame 1 (42 bytes on wire, 42 bytes captured)
Ethernet II, Src: Vmware_38:eb:0e (00:0c:29:38:eb:0e), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
Address Resolution Protocol (request)
0000 ff ff ff ff ff ff 00 0c 29 38 eb 0e 08 06 00 01 ....)8......
0010 08 00 06 04 00 01 00 0c 29 38 eb 0e c0 a8 39 80 ....)8....9.
0020 00 00 00 00 00 00 c0 a8 39 02 .... 9.
eth0: <live capture in progress> Fil... Packets: 445 Displayed: 445 Marked: 0 Profile: Default

Argus - The All Seeing

System and Network Monitoring Software

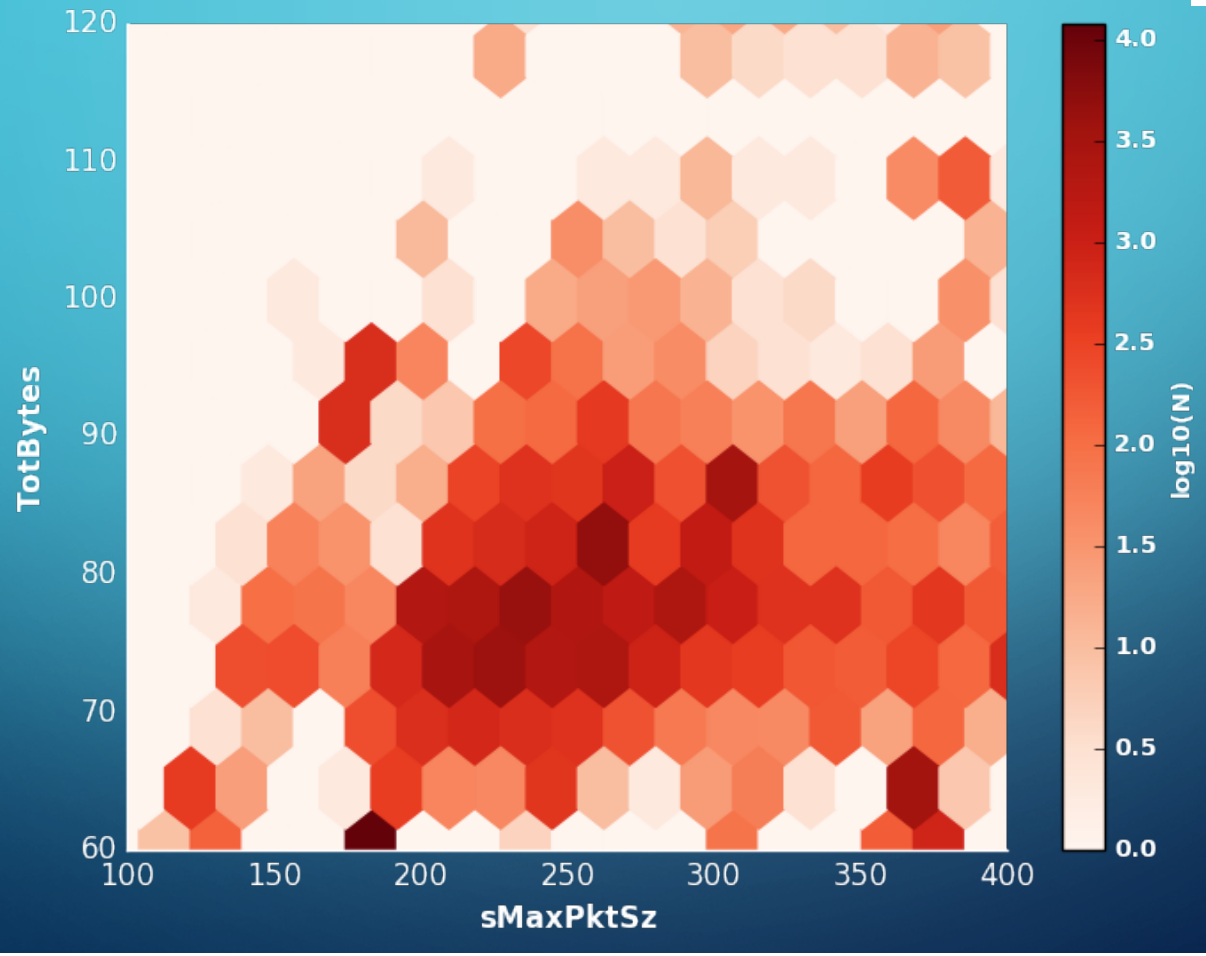
```
StartTime,Dur,Proto,SrcAddr,Sport,Dir,DstAddr,Dport,State,sTos,dTos,TotPkts,TotBytes,SrcBytes,Label
2011/08/17 05:01:12.984851,0.976560,tcp,93.45.141.223,seispoc,    ->,147.32.84.118,6881,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:01:15.616709,0.950668,tcp,84.16.60.37,64136,    ->,147.32.84.118,6881,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:01:23.209772,1.006908,tcp,94.44.197.227,rnm,    ->,147.32.84.118,6881,RST,0,0,4,276,156,flow=Background-TCP-Attempt
2011/08/17 05:01:24.216680,1.000400,tcp,94.44.197.227,rnm,    ->,147.32.84.118,6881,RST,0,0,4,276,156,flow=Background-TCP-Attempt
2011/08/17 05:03:27.216922,1.093291,tcp,147.32.3.51,dec-mbadmin-h,    ->,147.32.87.22,10010,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:03:50.442721,0.982337,tcp,93.45.211.241,4332,    ->,147.32.84.118,6881,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:04:26.344286,1.636698,tcp,85.114.48.118,35566,    ->,147.32.84.118,6881,RST,0,0,4,276,156,flow=Background-TCP-Attempt
2011/08/17 05:06:55.248668,0.993942,tcp,85.114.48.118,41173,    ->,147.32.84.118,6881,RST,0,0,4,276,156,flow=Background-TCP-Attempt
2011/08/17 05:08:42.994651,8.159175,tcp,95.143.208.138,19797,    ->,147.32.86.165,http,RST,0,0,5,364,216,flow=Background-TCP-Established
2011/08/17 05:08:50.682989,1.140232,tcp,93.47.141.213,49669,    ->,147.32.84.118,6881,RST,0,0,4,252,132,flow=Background-TCP-Attempt
2011/08/17 05:08:45.115361,9.299190,tcp,95.143.208.138,19801,    ->,147.32.86.165,https,RST,0,0,5,364,216,flow=Background-TCP-Established
2011/08/17 05:08:51.153826,11.275245,tcp,95.143.208.138,19797,    ->,147.32.86.165,http,RST,0,0,6,442,294,flow=Background-TCP-Established
2011/08/17 05:09:05.710077,0.999752,tcp,147.32.3.51,prolink,    ->,147.32.84.46,10010,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:10:49.007927,1.219781,tcp,2.158.166.41,60860,    ->,147.32.84.118,6881,RST,0,0,4,268,148,flow=Background-TCP-Attempt
2011/08/17 05:10:57.724292,0.287592,tcp,67.210.234.164,ofsd,    ->,147.32.87.191,http,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:10:58.011884,0.287865,tcp,67.210.234.164,ofsd,    ->,147.32.87.191,http,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:10:58.299749,0.287522,tcp,67.210.234.164,ofsd,    ->,147.32.87.191,http,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:10:58.587271,0.287716,tcp,67.210.234.164,ofsd,    ->,147.32.87.191,http,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:10:58.874987,0.289964,tcp,67.210.234.164,ofsd,    ->,147.32.87.191,http,RST,0,0,4,244,124,flow=Background-TCP-Attempt
2011/08/17 05:11:49.887787,1.727371,tcp,12.130.124.18,41802,    ->,147.32.87.5,urd,RST,0,0,4,252,132,flow=Background-TCP-Attempt
2011/08/17 05:11:58.264951,1.138943,tcp,93.47.141.213,49772,    ->,147.32.84.118,6881,RST,0,0,4,252,132,flow=Background-TCP-Attempt
```

STRATEGY

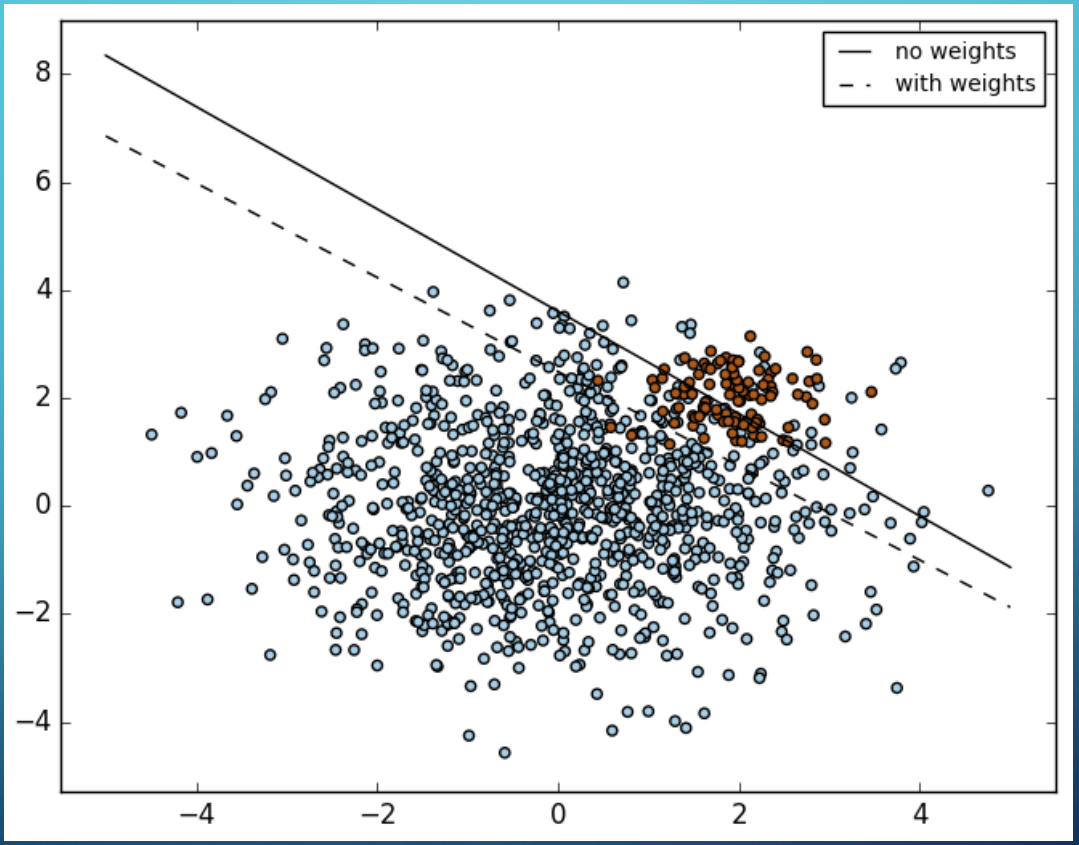
- Feature Selection
- “Masa” Cross-Validation (F1 Score)
- Model / Parameter Selection
- Test Totally New Data
- Deploy to Web (…)

TRAINING DATA

- 360 MB CSV
- 1.8mm rows raw
- 902k rows clean
- 2% Malware



SAMPLE WEIGHT (Example)



Source: http://scikit-learn.org/stable/auto_examples/svm/plot_separating_hyperplane_unbalanced.html

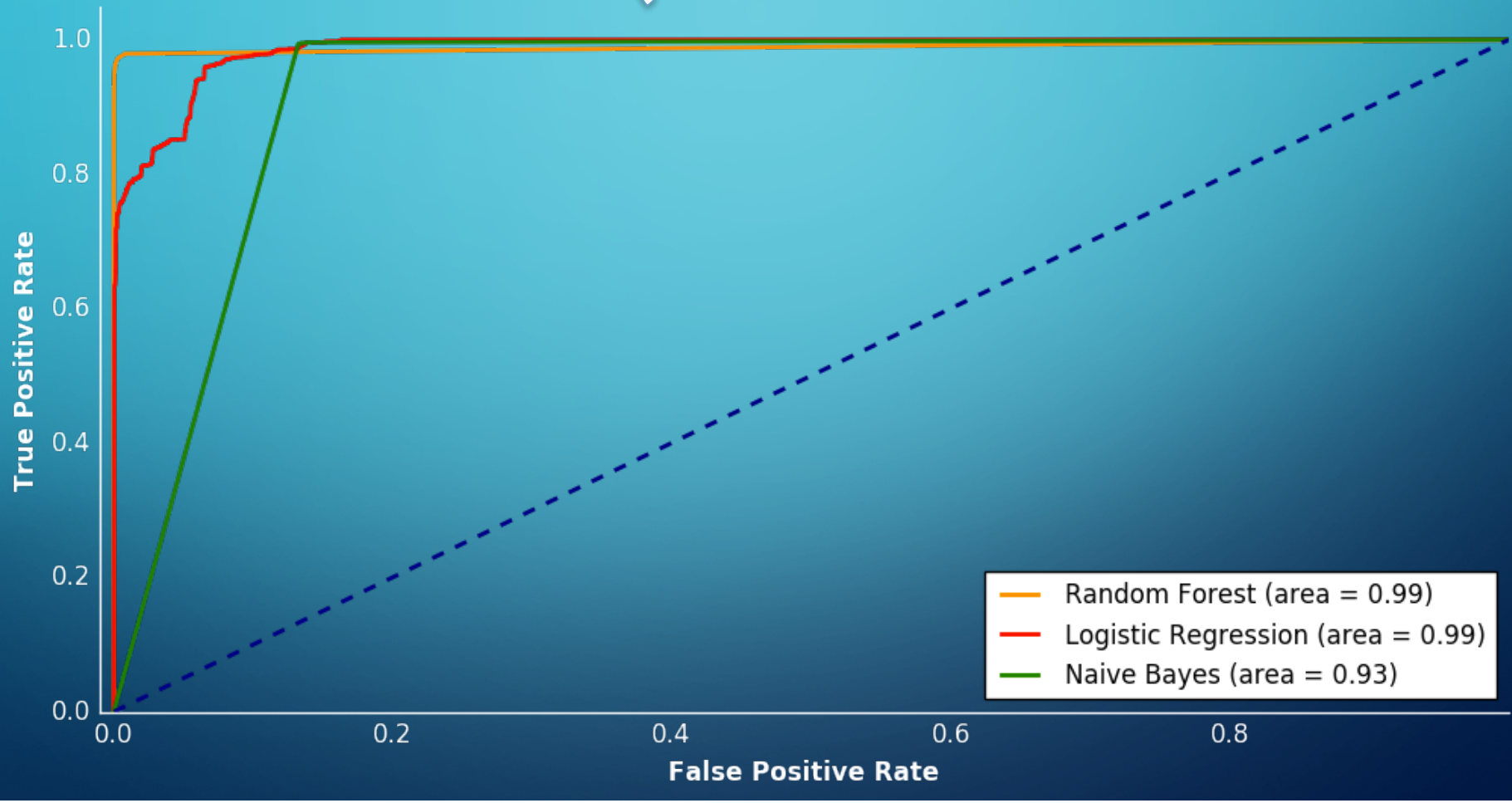
Packet Flow FEATURES

START TIME	SOURCE ADDRESS	MIN PACKET SIZE
DURATION	DEST ADDRESS	MEAN PACKET SIZE
TOS	SOURCE PORT	MAX PACKET SIZE
TTL	DES PORT	PACKET LOSS
HOPS	PROTOCOL	TOTAL BYTES
FLOW DIRECTION	STATE	APP BYTES

RANDOM FORESTS

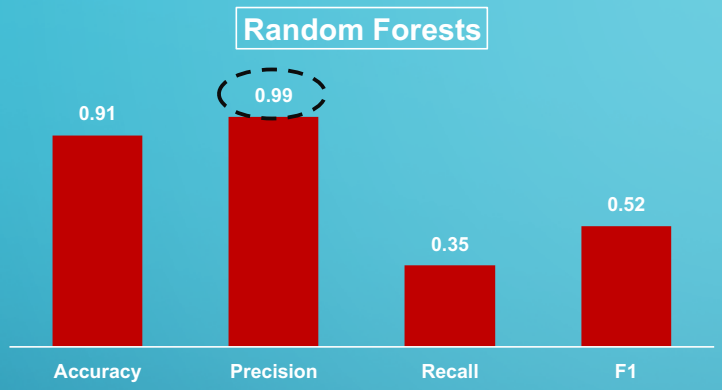
Feature Selection

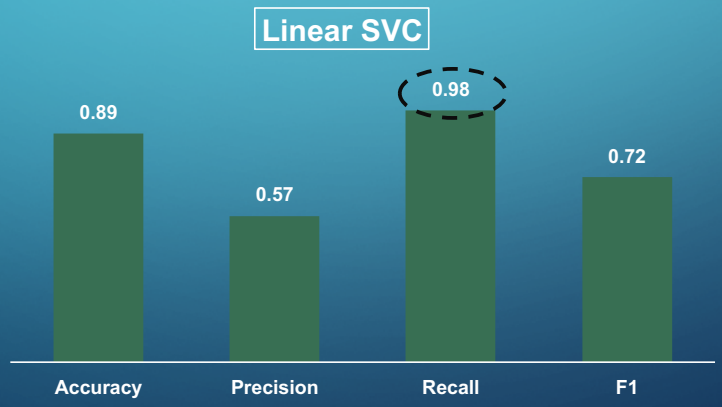
ROC CURVE (OUT OF SAMPLE)

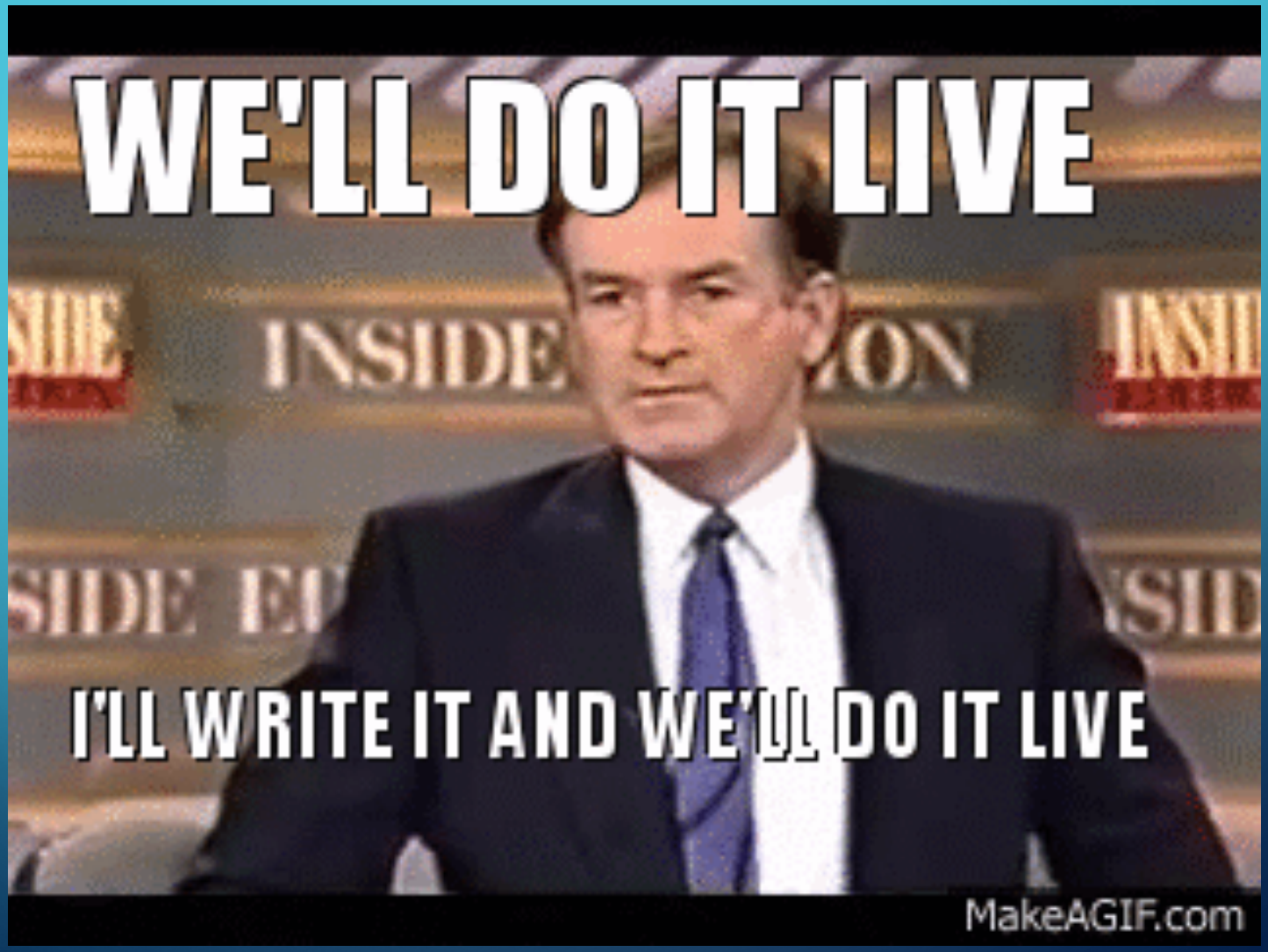


NEW TEST SET

- 400 MB CSV
- 2.1mm rows raw
- 1.3mm rows clean






WE'LL DO IT LIVE
I'LL WRITE IT AND WE'LL DO IT LIVE
MakeAGIF.com