Detecting Malicious Websites using Machine Learning

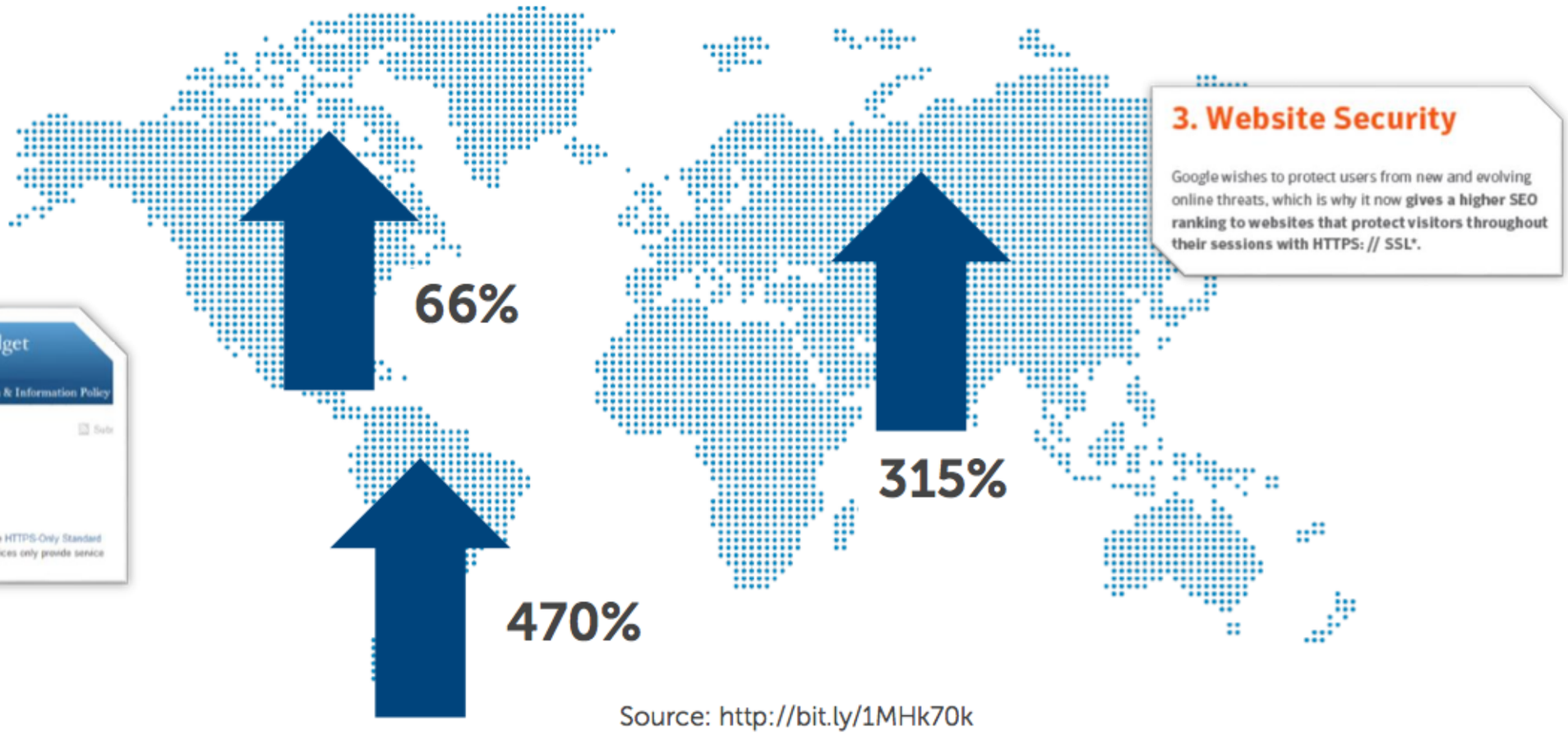
Andrew Beard
Ajit Thyagarajan

Encrypted web traffic growth

SSL comprises **1/3** of typical enterprise traffic

SSL traffic is growing **20%** per year

50% of all attacks are predicted to use SSL by 2017



3. Website Security

Google wishes to protect users from new and evolving online threats, which is why it now **gives a higher SEO ranking to websites that protect visitors throughout their sessions with HTTPS: // SSL*.**


Office of Management and Budget

About | OMBlog | The Budget | Management | Regulation & Information Policy

HTTPS-Everywhere for Government

Posted by Tony Scott on June 08, 2015 at 03:57 PM EDT

E-Mail Tweet Share +

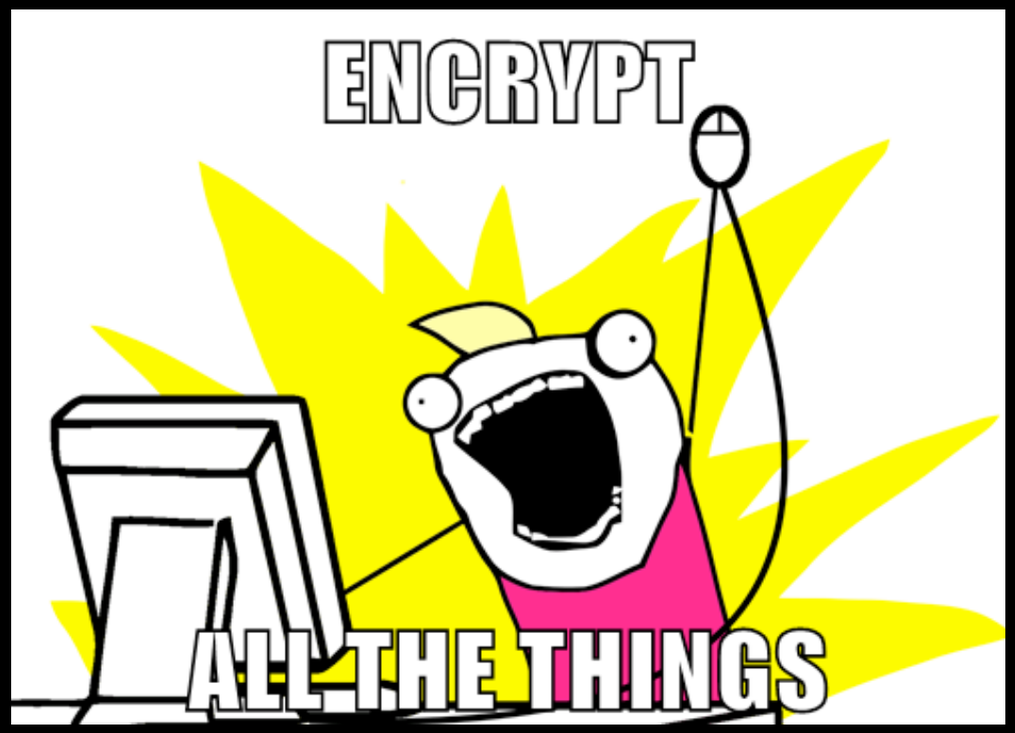
day, the White House Office of Management and Budget (OMB) issued the HTTPS-Only Standard ve, requiring that all publicly accessible Federal websites and web services only provide service

DellWorld'15

Source: http://bit.ly/1MHk70k

http://www.slideshare.net/DellWorld/key-security-insights-examining-2014-to-predict-emerging-threats

Users, Websites



NOC

```
MIICRzCCAbACCQDFYxWoDWqG5TANBgkqhkiG9w0BAQsFADBoMQswCQYDVQQ
GEwJBVTEPMA0GA1UECAwGZjJ0ZWU0MRQwEgYDVQQHDAtnZjIzZXQ2NWFkdD
ERMA8GA1UECgwIdGc0cjZ0ZHMxDDAKBgNVBAsMA3JzdDERMA8GA1UEAwwIc
nZndnRmZGYwHhcNMTYwNjA4MTc1MTU2WhcNMTcwNjA4MTc1MTU2WjBoMQsw
CQYDVQQGEwJBVTEPMA0GA1UECAwGZjJ0ZWU0MRQwEgYDVQQHDAtnZjIzZXQ
2NWFkdDERMA8GA1UECgwIdGc0cjZ0ZHMxDDAKBgNVBAsMA3JzdDERMA8GA1
UEAwwIcnZndnRmZGYwgZ8wDQYJKoZIhvcNAQEBBQADgY0AMIGJAoGBALiaY
8j+pPVAhmR5FpLaigtv02bhO7pmx+dAipjnsbhGapSpvZ2E5iKi1J6rwNAN
5kzqDYRWBwLYlN+NpUNbJj8icTrYBfZmbbqXI/
0zv66GB+g2fqvdaCNwtJv/E/+GF4w9DVZWGhiSlDcLLQubGwutgS3rHJj/
daReaMsE4PmtAgMBAAEwDQYJKoZIhvcNAQELBQADgYEAMtxqHvDHlNB65Jm
VfjSFlpx/YMVJol61AmEEu1noLMWhp20nwwE/
rZ1Guf7qF1XK6sEqP8YIoAGQsl1IW0hz0XHb24+foCAbboPCHM3b1Xt0Bof
Ry1PX+gcy84SjbeTpOZ4iuKeeZjcUH2oINdz7nNGnz+2Dmyn2buLe2PoYza
c=
```

```
abeard@pcap:~/certs$ base64 -d trickbot_cert.der | hexdump -C
00000000  30 82 02 47 30 82 01 b0  02 09 00 c5 63 15 a8 0d  |0..G0.......c...|
00000010  6a 86 e5 30 0d 06 09 2a  86 48 86 f7 0d 01 01 0b  |j..0...*.H......|
00000020  05 00 30 68 31 0b 30 09  06 03 55 04 06 13 02 41  |..0h1.0...U....A|
00000030  55 31 0f 30 0d 06 03 55  04 08 0c 06 66 32 74 65  |U1.0...U....f2te|
00000040  65 34 31 14 30 12 06 03  55 04 07 0c 0b 67 66 32  |e41.0...U....gf2|
00000050  33 65 74 36 35 61 64 74  31 11 30 0f 06 03 55 04  |3et65adt1.0...U.|
00000060  0a 0c 08 74 67 34 72 36  74 64 73 31 0c 30 0a 06  |...tg4r6tds1.0..|
00000070  03 55 04 0b 0c 03 72 73  74 31 11 30 0f 06 03 55  |.U....rst1.0...U|
00000080  04 03 0c 08 72 76 67 76  74 66 64 66 30 1e 17 0d  |....rvgvtfdf0...|
00000090  31 36 30 36 30 38 31 37  35 31 35 36 5a 17 0d 31  |160608175156Z..1|
000000a0  37 30 36 30 38 31 37 35  31 35 36 5a 30 68 31 0b  |706081751 56Z0h1.|
000000b0  30 09 06 03 55 04 06 13  02 41 55 31 0f 30 0d 06  |0...U....AU1.0..|
000000c0  03 55 04 08 0c 06 66 32  74 65 65 34 31 14 30 12  |.U....f2tee41.0.|
000000d0  06 03 55 04 07 0c 0b 67  66 32 33 65 74 36 35 61  |..U....gf23et65a|
000000e0  64 74 31 11 30 0f 06 03  55 04 0a 0c 08 74 67 34  |dt1.0...U....tg4|
000000f0  72 36 74 64 73 31 0c 30  0a 06 03 55 04 0b 0c 03  |r6tds1.0...U....|
00000100  72 73 74 31 11 30 0f 06  03 55 04 03 0c 08 72 76  |rst1.0...U....rv|
00000110  67 76 74 66 64 66 30 81  9f 30 0d 06 09 2a 86 48  |gvtfdf0..0...*.H|
00000120  86 f7 0d 01 01 01 05 00  03 81 8d 00 30 81 89 02  |............0...|
00000130  81 81 00 b8 9a 63 c8 fe  a4 f5 40 86 64 79 16 92  |.....c....@.dy..|
00000140  da 8a 0b 6f d3 66 e1 3b  ba 66 c7 e7 40 8a 98 e7  |...o.f.;.f..@...|
00000150  b1 b8 46 6a 94 a9 bd 9d  84 e6 22 a2 d4 9e ab c0  |..Fj......".....|
00000160  d0 0d e6 4c ea 0d 84 56  07 02 d8 94 df 8d a5 43  |...L...V.......C|
00000170  5b 26 3f 22 71 3a d8 05  f6 66 6d ba 97 23 fd 33  |[&?"q:...fm..#.3|
00000180  bf ae 86 07 e8 36 7e ab  dd 68 23 70 b4 9b ff 13  |.....6~..h#p....|
```

```
abeard@pcap:~/certs$ base64 -d trickbot_cert.der | openssl x509 -inform der -text -noout
Certificate:
    Data:
        Version: 1 (0x0)
        Serial Number:
            c5:63:15:a8:0d:6a:86:e5
    Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=AU, ST=f2tee4, L=gf23et65adt, O=tg4r6tds, OU=rst, CN=rvgvtfdf
        Validity
            Not Before: Jun  8 17:51:56 2016 GMT
            Not After : Jun  8 17:51:56 2017 GMT
        Subject: C=AU, ST=f2tee4, L=gf23et65adt, O=tg4r6tds, OU=rst, CN=rvgvtfdf
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:b8:9a:63:c8:fe:a4:f5:40:86:64:79:16:92:da:
                    8a:0b:6f:d3:66:e1:3b:ba:66:c7:e7:40:8a:98:e7:
                    b1:b8:46:6a:94:a9:bd:9d:84:e6:22:a2:d4:9e:ab:
                    c0:d0:0d:e6:4c:ea:0d:84:56:07:02:d8:94:df:8d:
                    a5:43:5b:26:3f:22:71:3a:d8:05:f6:66:6d:ba:97:
                    23:fd:33:bf:ae:86:07:e8:36:7e:ab:dd:68:23:70:
                    b4:9b:ff:13:ff:86:17:8c:3d:0d:56:56:1a:18:92:
                    94:37:0b:2d:0b:9b:1b:0b:ad:81:2d:eb:1c:98:ff:
                    75:a4:5e:68:cb:04:e0:f9:ad
                Exponent: 65537 (0x10001)
    Signature Algorithm: sha256WithRSAEncryption
         32:dc:6a:1e:f0:c7:94:d0:7a:e4:99:95:7e:34:85:96:9c:7f:
         60:c5:49:a2:5e:b5:02:61:04:bb:59:e8:2c:c5:a1:a7:6d:27:
         c3:01:3f:ad:9d:46:b9:fe:ea:17:55:ca:ea:c1:2a:3f:c6:08:
         a0:01:90:b2:5d:48:5b:48:73:d1:71:db:db:8f:9f:a0:20:1b:
         6e:83:c2:1c:cd:db:d5:7b:74:06:87:d1:cb:53:d7:fa:07:32:
         f3:84:a3:6d:e4:e9:39:9e:22:b8:a7:9e:66:37:14:1f:6a:08:
         35:dc:fb:9c:d1:a7:cf:ed:83:9b:29:f6:6e:e2:de:d8:fa:18:
         cd:a7
```

```
abeard@pcap:~/certs$ base64 -d trickbot_cert.der | openssl x509 -inform der -text -noout
Certificate:
    Data:
        Version: 1 (0x0)
        Serial Number:
            c5:63:15:a8:0d:6a:86:e5
    Signature Algorithm: sha256WithRSAEncryption
        Issuer: C=AU, ST=f2tee4, L=gf23et65adt, O=tg4r6tds, OU=rst, CN=rvgvtfdf
        Validity
            Not Before: Jun  8 17:51:56 2016 GMT
            Not After : Jun  8 17:51:56 2017 GMT
        Subject: C=AU, ST=f2tee4, L=gf23et65adt, O=tg4r6tds, OU=rst, CN=rvgvtfdf
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
                Public-Key: (1024 bit)
                Modulus:
                    00:b8:9a:63:c8:fe:a4:f5:40:86:64:79:16:92:da:
                    8a:0b:6f:d3:66:e1:3b:ba:66:c7:e7:40:8a:98:e7:
                    b1:b8:46:6a:94:a9:bd:9d:84:e6:22:a2:d4:9e:ab:
                    c0:d0:0d:e6:4c:ea:0d:84:56:07:02:d8:94:df:8d:
                    a5:43:5b:26:3f:22:71:3a:d8:05:f6:66:6d:ba:97:
                    23:fd:33:bf:ae:86:07:e8:36:7e:ab:dd:68:23:70:
                    b4:9b:ff:13:ff:86:17:8c:3d:0d:56:56:1a:18:92:
                    94:37:0b:2d:0b:9b:1b:0b:ad:81:2d:eb:1c:98:ff:
                    75:a4:5e:68:cb:04:e0:f9:ad
                Exponent: 65537 (0x10001)
    Signature Algorithm: sha256WithRSAEncryption
         32:dc:6a:1e:f0:c7:94:d0:7a:e4:99:95:7e:34:85:96:9c:7f:
         60:c5:49:a2:5e:b5:02:61:04:bb:59:e8:2c:c5:a1:a7:6d:27:
         c3:01:3f:ad:9d:46:b9:fe:ea:17:55:ca:ea:c1:2a:3f:c6:08:
         a0:01:90:b2:5d:48:5b:48:73:d1:71:db:db:8f:9f:a0:20:1b:
         6e:83:c2:1c:cd:db:d5:7b:74:06:87:d1:cb:53:d7:fa:07:32:
         f3:84:a3:6d:e4:e9:39:9e:22:b8:a7:9e:66:37:14:1f:6a:08:
         35:dc:fb:9c:d1:a7:cf:ed:83:9b:29:f6:6e:e2:de:d8:fa:18:
         cd:a7
```

Issuer and Subject Fields

- Subject and Issuer fields are structured by CONVENTION
- Essentially free-form text
- Free form text fields make awesome fingerprints for tracking adversaries

Project Sonar

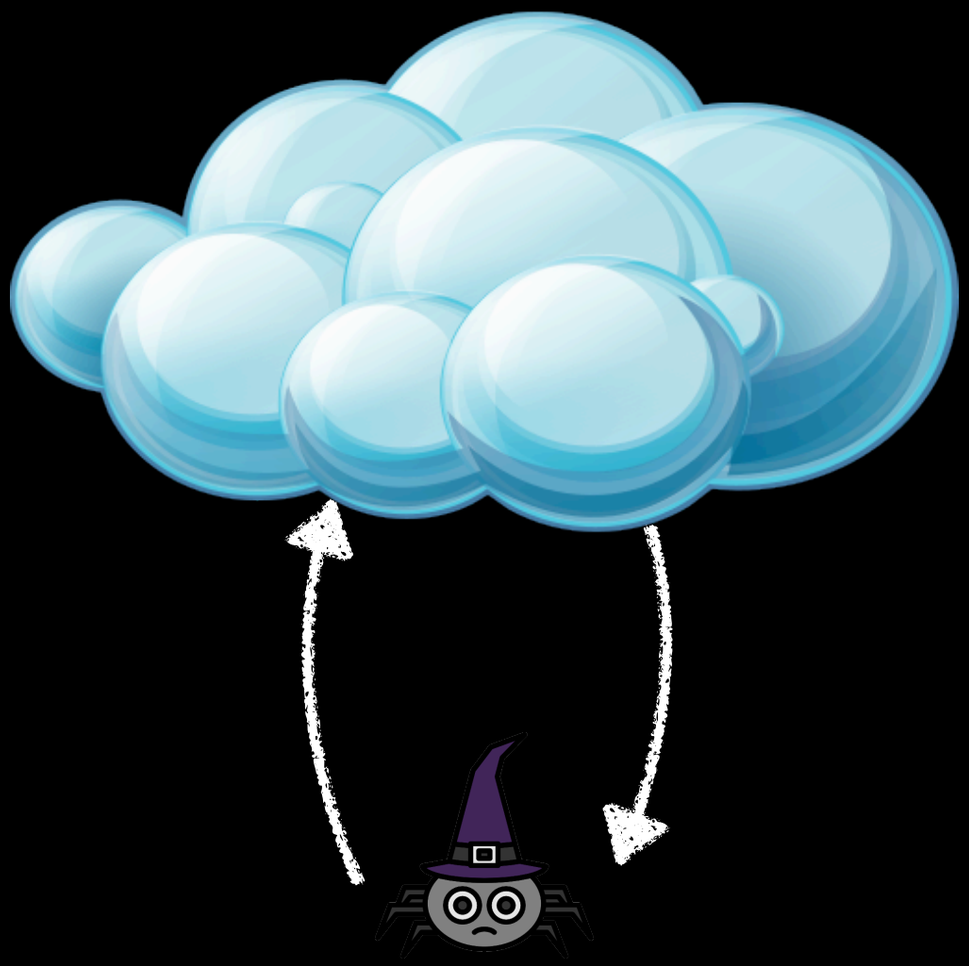
- Scans complete IPv4 space every 7 days
- Only looks at port 443
- Data published every week starting in May 2013
- ~82 million certificates observed to date
- Complete x.509 certificates are available in base64 DER format

SSLBL

- Blacklist of SSL certificates used by malware
- Published by abuse.ch
- Mostly commodity C2 servers
- Certificate fingerprints (SHA1 hashes)

A Couple Assumptions

- The Internet is big, and malicious sites are (relatively) rare
- Certificate blacklists do not provide anywhere near full coverage (even of a particular family)
- We can trust that the certificates that are identified are bad
- Some actors generate certificates for malicious sites in batches that are “similar”



Project
Sonar

abuse.ch
SSLBL

python +
openssl

Amazon
Redshift

date	sha1	version	subject_cn	subject_l	subject_c	subject_o	subject_ou	subject_st	subject_emailaddress	subject_unstructuredname	subject_serialnum...
2015-07-27	8172ab15fecc9af2866...	3	79.192.163.185								
2015-07-27	80cf042056e19496c7...	3	www.dlink.com	Taipie	TW	D-LINK	DHPD Dept.	Taiwan			
2015-07-27	590ec0d54f277470d8...	1	192.168.1.1		US						
2015-07-27	8a416ea83845705d41...	3	217.234.153.102	Wuerselen	DE	LANCOM Systems GmbH	General	NRW	info@lancom-systems.de		
2015-07-27	b273ab4b68b37d90d6...	3	OpenWrt	Leipzig	DE			Saxony			
2015-07-27	098642d2b6aee7fcaa0...	3	93.220.244.227	Wuerselen	DE	LANCOM Systems	Engineering	NRW	info@lancom-systems.de		
2015-07-27	95404cfe74ede49832...	3	79.199.0.92	Wuerselen	DE	LANCOM Systems GmbH	General	NRW	info@lancom-systems.de		
2015-07-27	6c2b35d71bcbe72252...	3	79.211.14.8	Wuerselen	DE	LANCOM Systems GmbH	General	NRW	info@lancom-systems.de		
2015-07-27	50cbfd012e3d65cba78...	3	92.192.37.182								
2015-07-27	a8cd4cc5835eaad4ee6...	3	hagelin.remotewebacce...				Domain Control Validat...				
2015-07-27	39b4794e0c4c7f1a5c...	3	212.41.107.41								
2015-07-27	01a3750f24fb085cd84...	3	62.225.222.252								
2015-07-27	7a1191cb889e38439e...	3	www.dlink.com	Taipie	TW	D-LINK	DHPD Dept.	Taiwan			
2015-07-27	6323084f8f6f0bb3336...	3	semi-kindergarten.noel....		AT	Amt der NOe Landesreg...			helpdesk@noel.gv.at		393719710864
2015-07-27	674b67bf7096ed7d9a...	3	10.0.25.254								
2015-07-27	e444c304e9242613a7...	3	79.232.252.83								
2015-07-27	143fb0c4a0f572b635d...	3	92.77.137.28	Wuerselen	DE	LANCOM Systems	Engineering	NRW	info@lancom-systems.de		
2015-07-27	4ed16397cc613277a9...	3	Netgear VPN Firewall		US	Netgear Inc.	Netgear Prosafe				
2015-07-27	5fdc5935d653581de7...	3	92.73.176.29								
2015-07-27	c06dfef287248ae5c2d...	3	84.39.86.82								
2015-07-27	9ad372fa2d9f884611a...	3	62.226.116.49								
2015-07-27	e2df44a162cbde9a254...	3	bieberstefan.anydns.info								
2015-07-27	a58c6614d58d5ab228...	3	OpenWrt	Leipzig	DE			Saxony			
2015-07-27	9240d848a49272dda6...	3	83.218.54.210								
2015-07-27	34a9b6fc5826875748...	3	Gateway Authentication		US	2Wire	Gateway Device				48141N149129
2015-07-27	a8f6dd2584316ee27ff...	1	TRIL-NYC-INET01							TRIL-NYC-INET01	FOC1750V2HG
2015-07-27	402ebbaff21a7e39fbe...	1	Fshop-Peraia							Fshop-Peraia	FGL154421JV
2015-07-27	51b4c522a30de622a9...	3	h66a.24-7.ro	H	DE	SRiT	INT	NDS			
2015-07-27	1e14efcdaeac5d46216...	3	657245.vps-10.com						ssl@657245.vps-10.com		
2015-07-27	fe468ddfdbd50e9ee1fb...	3	50.76.126.109								
2015-07-27	2e5b485be0388ff1471...	3	213.177.85.139								
2015-07-27	b652c68f59153eac96...	3	84.136.3.181								
2015-07-27	ddbff3d9263a67321dd...	3	193.83.7.72								
2015-07-27	b9232a13bd45a45873...	3	52.10.57.105	Reston	US	ScienceLogic		Virginia	support@sciencelogic.c...		
2015-07-27	f4fdbe898ab8d21abf0...	3	80.53.98.178								
2015-07-27	96c800c2fbdef253922...	3	87.122.129.142	Wuerselen	DE	LANCOM Systems	Engineering	NRW	info@lancom-systems.de		
2015-07-27	63c44480a40831e662...	1	inetgwsw01.							inetgwsw01.	
2015-07-27	6857f28f0e6c81bcf57...	3	93.198.192.63								
2015-07-27	86d6aca2c2708ea0a3...	3	HPD6324E	Vancouver	US	HP	HP-IPG	Washington			
2015-07-27	c3170e6e942c599e45...	1	PlayBook: 94:eb:cd:36:...			Research In Motion Limi...					
2015-07-27	aba23a701ef9325177...	3	americanroller.com				Domain Control Validat...				
2015-07-27	c18fd5e8488a1f253a7...	1	192.168.1.1		US						
2015-07-27	91c53066751eb665dd...	3	Ruckus Wireless	Sunnyvale	US	Ruckus Wireless		California			
2015-07-27	3e75db46432a678fffcf...	3	60.191.209.254			Product Root CA	IP Camera				
2015-07-27	a93a5ea360bf90f4203...	3	91.15.220.84								
2015-07-27	bf938e1d252a2ec92ef...	3	84.183.197.249								
2015-07-27	46cae286caa8229395...	3	ram-maurer.dyndns.org								
2015-07-27	ac5cc2b2e338131aee...	3	DNS-320L48ee0c23cb...	US	US	D-LINK		D-LINK			
2015-07-27	052cfa7145d885199a...	3	91.15.195.36								
2015-07-27	4212be212d801b6001...	1	PlayBook: a4:e4:b8:21:...			Research In Motion Limi...					
2015-07-27	bae850a7609ddf68ce9...	3	*.ongravity.com	Unites States	US	ongraviti	ongraviti	Non-US/O...	jo1.magat@gmail.com		

What We Can’t Find

- Hijacked or compromised certificates
- Targeted Attacks
- Initial wave of new campaigns

What We Can Find

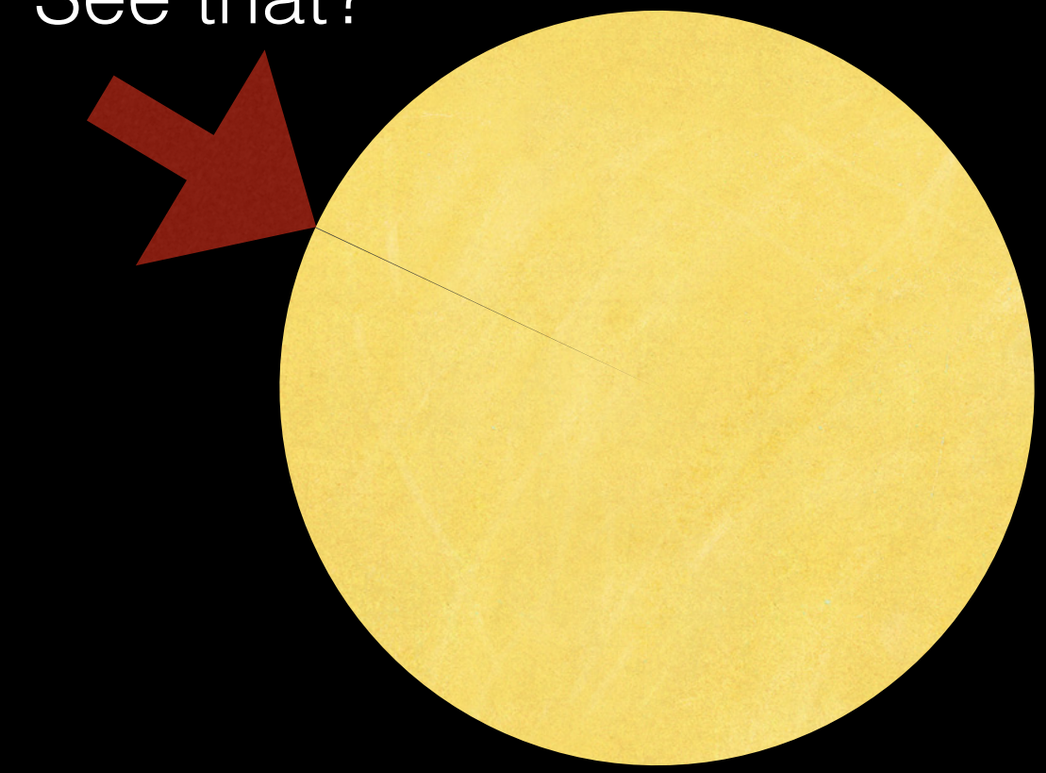
- Certificates that were issued with malicious intent
- Commodity Malware C2
- A lot of dumb

Logistic Regression

- Certificates are either malicious or not -> binary classifier
- Very common method, lots of implementations
- Supervised learning model (data is labeled)
- First try, let’s just feed it all the labeled data…

I have 5 million labeled “good” records, and 1335 “bad” records

See that?



5,000,000 / 5,001,335 = .99973

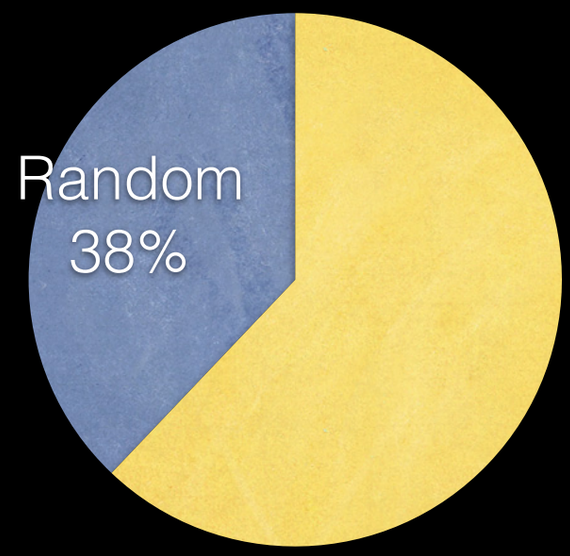
99.97% accuracy by always saying it's not malicious

EPIC FAIL

FOR WHEN NORMAL FAIL JUST DOESN'T CUT IT

More specific case, random looking values

- Possible to recognize via transformations (Shannon entropy, unique character counts, etc)
- We want to try without creating more problem-specific attributes
- Easy for a human to label our 1335 bad certificates as random or not



sha1	subject_c	subject_cn	subject_l	subject_o	subject_st	looks_random
0bf637618e1994571fbb5425ea8aa700a7dc96be		secureinishman.com				FALSE
0cbdafc2494b2c2072c9a0964bb0fdc338428157	GM	km0ss69CDS8LOgxXblVxs5c4ZlvOyOw	khrFxmrvs02CM0RjwUnoZRUp3	xy0vNynspRynpuoXn7EFQs	VP	TRUE
0d2274585486af6aacc8afe0dd7660b90a17717b	GB	localhost	York	MyCompany Ltd.	Yorks	FALSE
0d2f7ba6bde70603ea7f2a88e2eae1e132f09842	FP	mnlsw7gd9vA	vo1sjobMRrJLkxMhOpn5LYq46j...	vG0DvoxRB9MhZsVALo	IX	TRUE
0d8e5b2cb22c0c3b1bb4cc9eef3fa74788351165	XX		Default City	Default Company Ltd		FALSE
0dee94cca8a85c6b6de16d889b90584b056cff74		marinova.am				FALSE
0e034408346e2c66faeca8f89724ea1ff6c75a5e		svr2				FALSE
0e225cf941fd1e7be0f5e9fe21dd13c373100e44	AU			Internet Widgits Pty Ltd	Some-St...	FALSE
0e2e20c6f25cc88757d81dc492dc4afe1eacd8ce	OS	LPjtpA06yybGesp	njkdGx1AdDIRTH5gQwlycjZGsC...	VS4elaKMMOrz4aAluW9E7O7c6...	HR	TRUE
0e44c7aaadd1186c17f4f1364e3722c172a7ce2e		docknetworks.com				FALSE
0e63759575633cb2626cfe23f0ac5f67506cff1b	ZA	nyctradersacademy.com	Bnz	Hosl	FX	FALSE
0e66f3d415a6e83a25ce595e409a8c8be4e87af8	XX		Default City	Default Company Ltd		FALSE
0e9405e67cbd7e1ef295bfbf5d8b0ef34582e84a	CN	DcNJ0tZ3LSrvQT0iok685067	tp9V0DykxMVHBhouzay7t7Vl	AVy17f7kfvMglrglicqS4Wie	ST	TRUE
0ea588347268749ad831a7188acc6df00e443863	US		Springfield	Dis	Denial	FALSE
0eb47680ff6e869ea0a6a73358bbdfba617ea218	FM	8cAbHmGiPGu4	YCP6bb2xaAjcPthdVOpYa4ws2u...	bOMFiGqjUvC9du4IwyXzpG1piDW	DZ	TRUE
0ec788ce6e3146fca8e91179106d734097617a25	QI	h9nCYUulExL6	8qWB59pYCfFnY	54FA7GBmjcnq0dfU	NI	TRUE
0ed0d70dafffa4ea7249176333f7fa76b2d2c4a3	US	ubkbtnttefm tvazwlslyyzc	Ridgeville	aovgrmymqhitrvq gdvqwiulmfyyd...	South Ca...	FALSE
0eda9c1bf5f4ac1f5bd44eb93d8e7bf0acf3a691	US	pisppsthmhe.sa	Lauderdale	Monsanto	FL	FALSE
0edd722452c12c686f16a7ee7be74b56e89a6db5		kpai7ycr7jxqkilp.torexplorer.com				FALSE
0ee1e19b755a24f497e45be308608413ab34ca90	IR	cyberwise.biz		Boo	Some-St...	FALSE
0f206c9a9522ca50c6df640e3f7dbebfaab817fc	US		Springfield	Dis	Denial	FALSE
0f2dae7905a578d96530ab2c9f55dd5dcf8db862		legallyjumps.com				FALSE
10530dc91451b4775017c95fdd11763011f95004	US	whaovxeynxctdrvzn.com	Tampa	Realtek Semiconductor Corp.	FL	FALSE
10541e30879e48f459d657ed67fad0300af3eb90	CN	1apsIIT6VRuj8FIAFR4I7bie	llvBvWY4NkSdGA3vKhPu3LAZ	gvIBA6JftP2DMpSn8CNGqVRv	ST	TRUE
10d2f538a6930aaa53665bfaeb786a421ced13d3	YX	fP77y3zUQ54JWD	DN9J2SCURMAsxYuBOR7Lg	xA0CKUId90xR	BP	TRUE
10d8ee8221dfeaf28cf1700a7ec1a34650244b8b	US	nryvmrxygenzmdb gpksaaohxhmkohsbutr	Spring Grove	cxsurhnmqkdou lqsitubdrtk	Illinois	FALSE
10ee62a1a2d369e3a5b55f7c71273afd424f08d0	GB	localhost	York	MyCompany Ltd.	Yorks	FALSE
10fa0d1ddfe71212feff953d6ea275c62c2055a2	AU			Internet Widgits Pty Ltd	Some-St...	FALSE

Again with the logistic regression

- Good news: Pretty accurate in this case
- Bad news: Entirely dependent on state and country code. Pretty easy to figure out which are uncommon
- Easier way: Use a Bloom Filter

Clustering

- All about counts of occurrences
- Start with most common values for a given attribute
- Look for other common attributes that occur in the same set of results
- Harder with categorical values since there’s no real distance, but scripts can help

L = Taipei

L = London

L = York

L = Springfield

L = <EMPTY>

L = Taipei

L = London

O = Company

O = MyCompany rk Ltd.

L = <EMPTY>

L = Springfield
O = Dis

O = Internet Widgits Pty Ltd

- Patterns should be more common in the bad set than the overall set

- Find a ratio of matches in overall set: matches bad set, normalize to overall matches per bad match

  - Exactly 1 - Only found in bad set, not useful

  - Greater than ~1000 - Appears often in overall set. Probably not a pattern marker.

  - Between 1 < x < ~250 : The sweet spot. *Potentially* useful

```
Bad:   1335
Total: 82331027
Ratio: 61671.1812734
…
subject_o:      189 -> 52724965 (278968.068783)
subject_o: MyCompany Ltd.      110 -> 870 (7.90909090909)
subject_o: Dis      106 -> 4770 (45.0)
subject_o: Default Company Ltd      76 -> 38476 (506.263157895)
subject_o: Internet Widgits Pty Ltd      75 -> 186885 (2491.8)
subject_o: Global Security      7 -> 275 (39.2857142857)
subject_o: Private      5 -> 4549 (909.8)
subject_o: My Company Ltd      4 -> 10448 (2612.0)
subject_o: YouPorn Ltd      4 -> 5 (1.25)
subject_o: XX      4 -> 57114 (14278.5)
subject_o: SomeOrganization      3 -> 927436 (309145.333333)
subject_o: Companyname      3 -> 339 (113.0)
subject_o: none      3 -> 115216 (38405.3333333)
subject_o: koalabride      2 -> 4 (2.0)
subject_o: protectthegays      2 -> 13 (6.5)
subject_o: US Aid      2 -> 4 (2.0)
subject_o: Ubiquiti Networks Inc.      2 -> 642654 (321327.0)
subject_o: Domain      2 -> 31 (15.5)
subject_o: Monsanto      2 -> 26 (13.0)
subject_o: democracy      2 -> 11 (5.5)
subject_o: International Security Depart      2 -> 3 (1.5)
subject_o: Joel Fisker      2 -> 4 (2.0)
subject_o: Hosl      2 -> 12 (6.0)
subject_o: Microsoft      2 -> 4033 (2016.5)
subject_o: Realtek Semiconductor Corp.      2 -> 17 (8.5)
```

```
Bad:   1335
Total: 82331027
Ratio: 61671.1812734
…
subject_o:      189 -> 52724965 (278968.068783)
subject_o: MyCompany Ltd.      110 -> 870 (7.90909090909)
subject_o: Dis      106 -> 4770 (45.0)
subject_o: Default Company Ltd      76 -> 38476 (506.263157895)
subject_o: Internet Widgits Pty Ltd      75 -> 186885 (2491.8)
subject_o: Global Security      7 -> 275 (39.2857142857)
subject_o: Private      5 -> 4549 (909.8)
subject_o: My Company Ltd      4 -> 10448 (2612.0)
subject_o: YouPorn Ltd      4 -> 5 (1.25)
subject_o: XX      4 -> 57114 (14278.5)
subject_o: SomeOrganization      3 -> 927436 (309145.333333)
subject_o: Companyname      3 -> 339 (113.0)
subject_o: none      3 -> 115216 (38405.3333333)
subject_o: koalabride      2 -> 4 (2.0)
subject_o: protectthegays      2 -> 13 (6.5)
subject_o: US Aid      2 -> 4 (2.0)
subject_o: Ubiquiti Networks Inc.      2 -> 642654 (321327.0)
subject_o: Domain      2 -> 31 (15.5)
subject_o: Monsanto      2 -> 26 (13.0)
subject_o: democracy      2 -> 11 (5.5)
subject_o: International Security Depart      2 -> 3 (1.5)
subject_o: Joel Fisker      2 -> 4 (2.0)
subject_o: Hosl      2 -> 12 (6.0)
subject_o: Microsoft      2 -> 4033 (2016.5)
subject_o: Realtek Semiconductor Corp.      2 -> 17 (8.5)
```

```
subject_o: protectthegays + subject_l: Quebeck  2
subject_o: protectthegays + subject_ou: gay team  2
subject_o: protectthegays + subject_serialNumber:  2
subject_o: protectthegays + subject_c: CA  2
subject_o: protectthegays + subject_emailAddress:  2
subject_o: protectthegays + subject_unstructuredName:  2
subject_o: protectthegays + subject_st: Monreal  2
```

Project Sonar Redshift › projectsonar › SQL Query — Connected. — Redshift 1.0.1101 TLS 1.2

```
1 SELECT sha1, serial_number, subject_c, subject_cn, subject_l, subject_o, subject_ou, subject_st, feed_match FROM cert_metadata WHERE
subject_o = 'protectthegays'
```

Load Query... | Save Query... | Cancel | Execute Statement

sha1	serial_number	subject_c	subject_cn	subject_l	subject_o	subject_ou	subject_st	feed_match
0bf417e3aaee820bc63a92b3d756aacdd5ab3b6b	E1F6F857079336ED	CA	idizemert.ch	Quebeck	protectthegays	gay team	Monreal	TRUE
0b2d469d039371c3bcc2be00c9a9f41784092ced	BD72C1EF6AE6AD4C	CA	tusndtburi.id	Quebeck	protectthegays	gay team	Monreal	TRUE
83b8e5507ebc0adb6f172879f455328af6cc58f9	AE2063F6E5976236	CA	thhilayati.kg	Quebeck	protectthegays	gay team	Monreal	FALSE
bcf0cdb42301d73851c02dab06ecd2cb1627e089	9F26DE294DFD7CB9	CA	eatsictit.gs	Quebeck	protectthegays	gay team	Monreal	FALSE
ab40c12fbec9f22938b5a0ef30e09f9e1dcae5bb	A5ED9FF9F9C5937E	CA	moliveverisp.cg	Quebeck	protectthegays	gay team	Monreal	FALSE
4b34559bc3f5f1893ee6bb667ba59b7dd3b5198b	986C08A96DAD889B	CA	uscecyttpli.pf	Quebeck	protectthegays	gay team	Monreal	FALSE
99aec4b45eb95d712922b211019afb6587864c5e	C093A6AA3270A765	CA	istutmava.lv	Quebeck	protectthegays	gay team	Monreal	FALSE
d0873e96527027f3b2419c7f83ca9e531c175264	B10BCAA8103D4519	CA	peanghela.bd	Quebeck	protectthegays	gay team	Monreal	FALSE
c348f5c953f79671ee2bfac313e6df330255c763	C3664B5C61B6FFEC	CA	aultirranwad.pn	Quebeck	protectthegays	gay team	Monreal	FALSE
82a2cd12aca09eea29f5c1f6ef4dd1b5b2ad1ddb	DBE3E05CA19B9A25	CA	ondinthancer.fo	Quebeck	protectthegays	gay team	Monreal	FALSE
88262484ff061851f2d15996acc77bfdc3f916ba	CE3F221FB57FEA9C	CA	trstulebafrur.ac	Quebeck	protectthegays	gay team	Monreal	FALSE
ab6706bb401ebf7d7b3be5bc4ad9774f472385d6	915674E80AC82A34	CA	tatrorofonyb.cr	Quebeck	protectthegays	gay team	Monreal	FALSE
f6e1f71411ead03713d76d0cd65d82f6cd8b90fc	E464D2DE30D073C6	CA	bereasteobrt.cy	Quebeck	protectthegays	gay team	Monreal	FALSE

Result 1 | Export... | 13 rows | 28.4 s

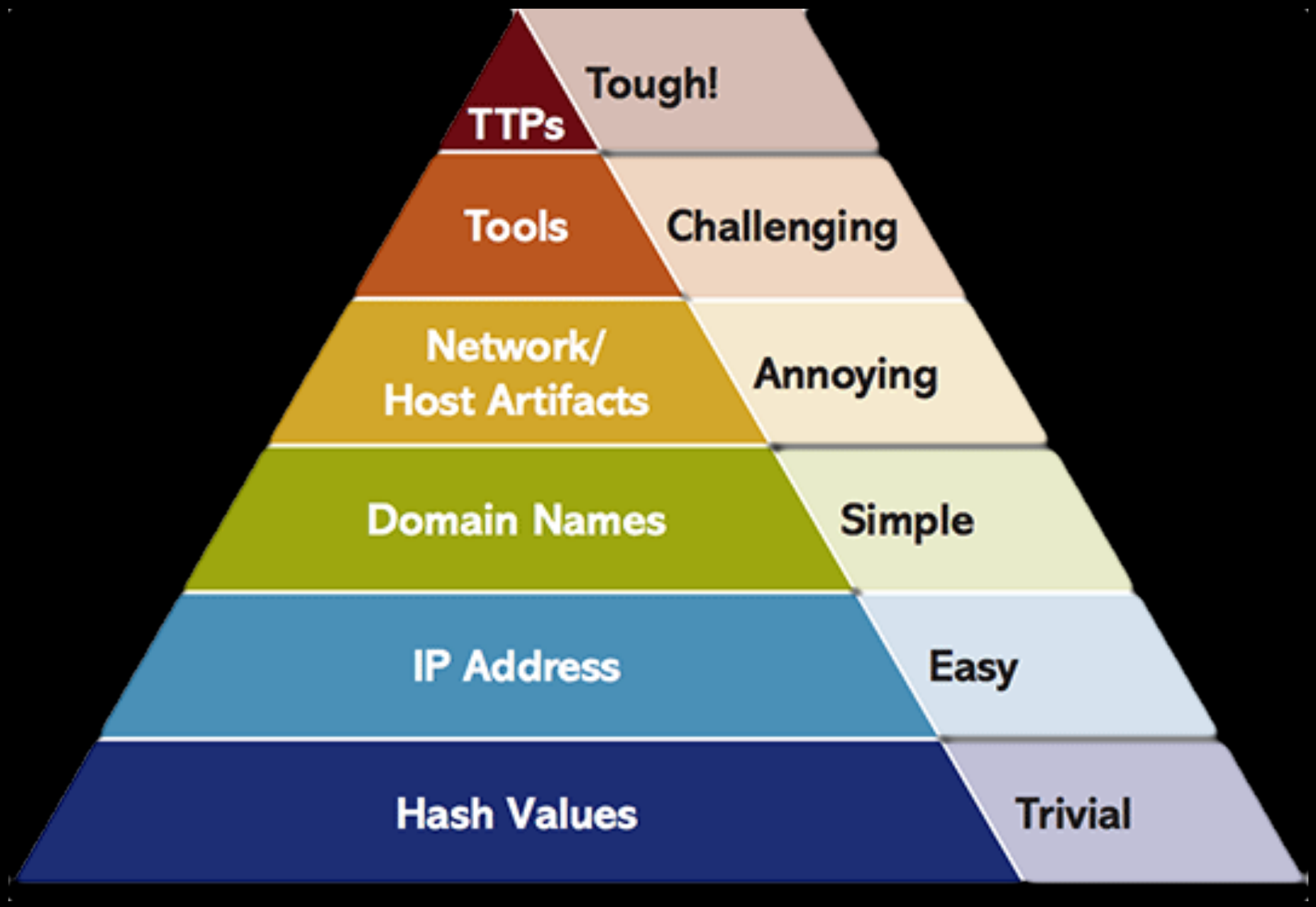
What can I do with this?

- OK: Search for patterns in the set of certificates you know about; produce a bigger blacklist.

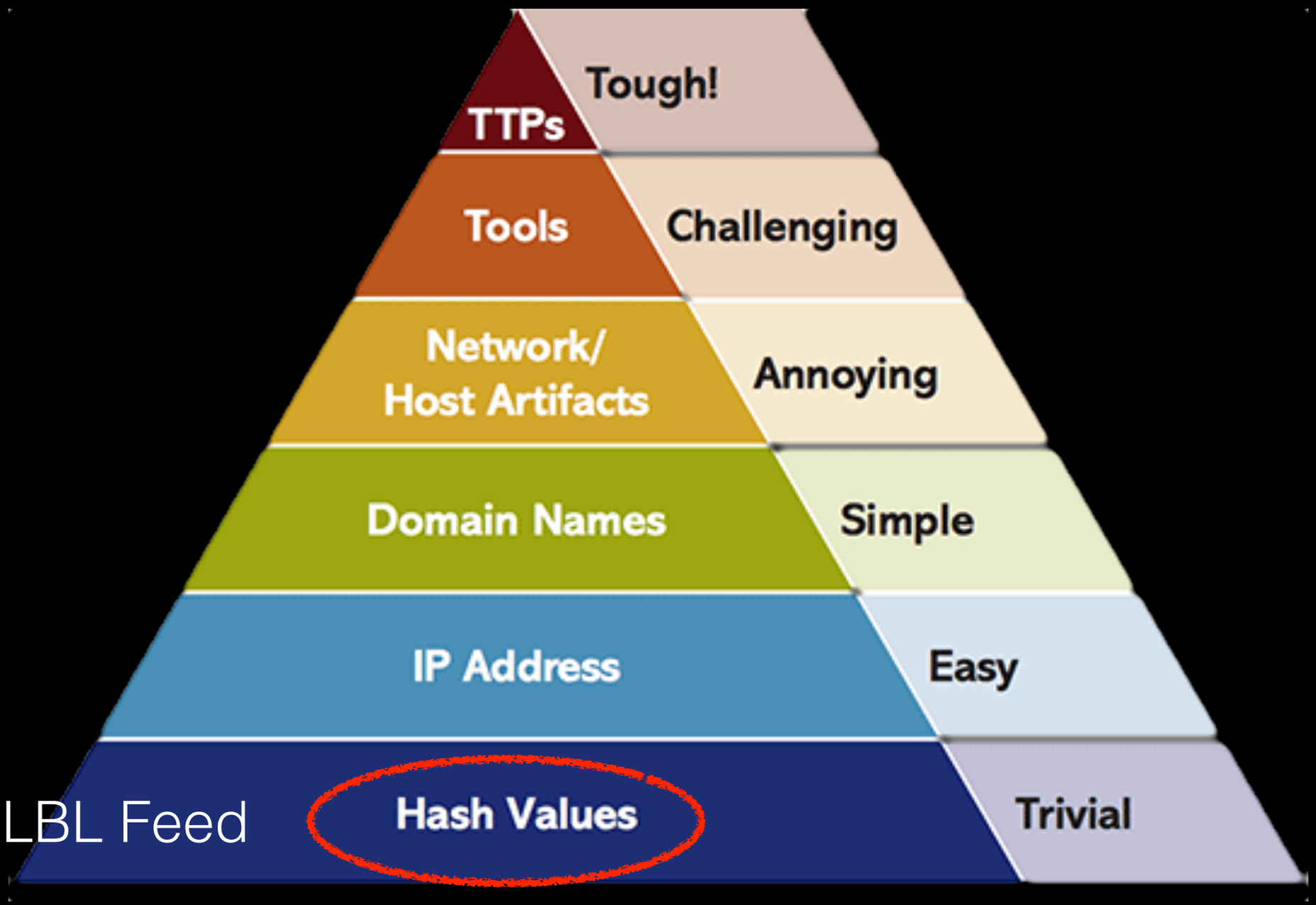
- Better: Look for relevant strings in SSL traffic using a packet-based IPS like Snort or Suricata

- Best: Look for patterns in your network traffic using a tool that decodes SSL metadata (like Bro)

David Bianco's ~~Pyramid~~ Triangle of Pain

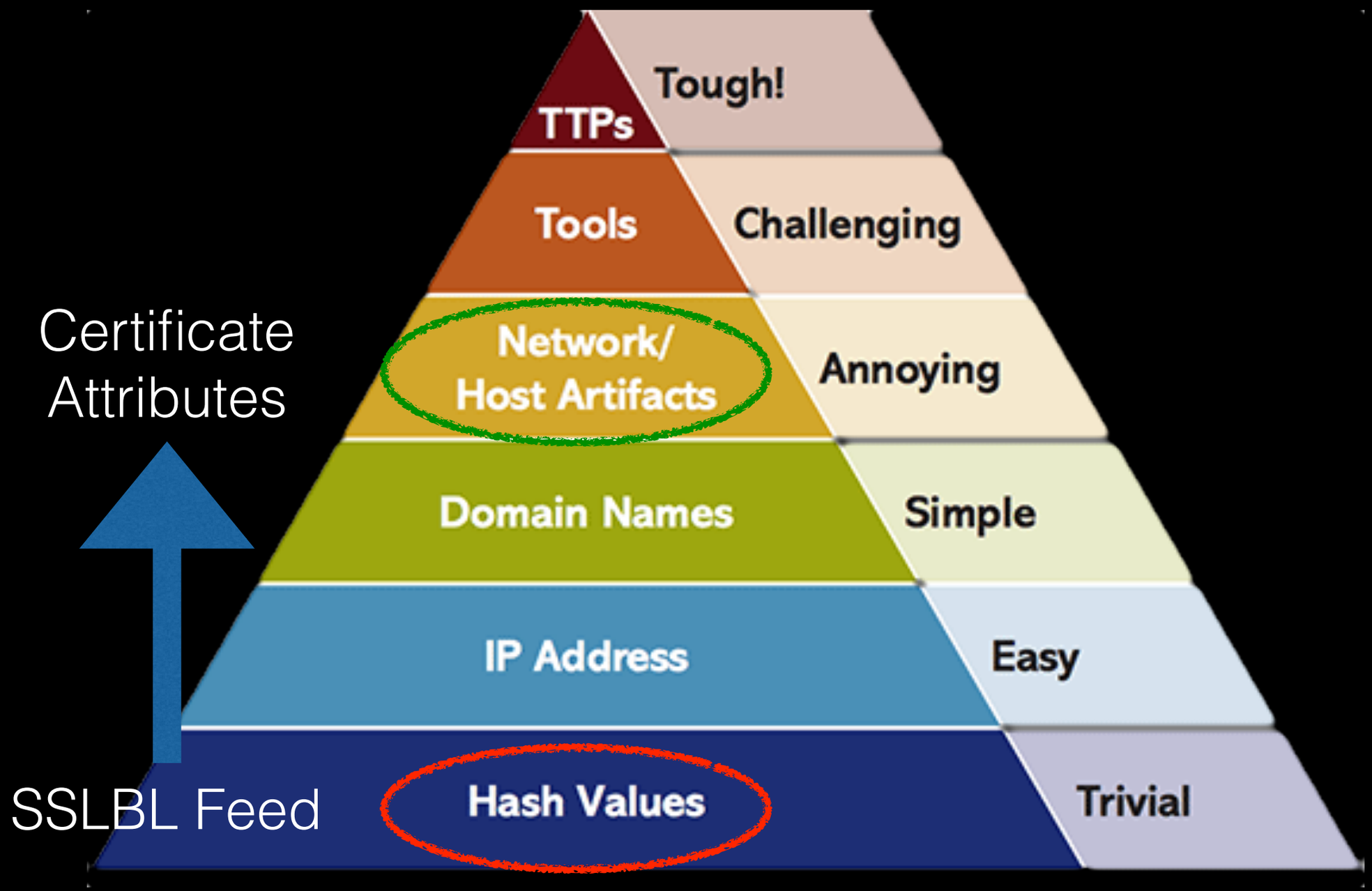


David Bianco's ~~Pyramid~~ Triangle of Pain



SSLBL Feed

David Bianco's ~~Pyramid~~ Triangle of Pain



Takeaways

- It’s possible to leverage large data sets with specific indicators to find more general patterns
- Choose your data sets carefully
- Machine Learning algorithms give you insight into your data. Sometimes they tell you to use something simpler.

Thanks to:

- BSides DC
- abuse.ch for creating the SSLBL feed
- Rapid7 Labs for providing Project Sonar data

Andrew Beard
andrew@atomicmole.com
@bearda24

Ajit Thyagarajan
ajit@atomicmole.com

Atomic Mole, LLC
https://github.com/atomicmole
@atomicmolellc
We’re hiring! careers@atomicmole.com

CA Breakdown

- Majority still self-signed
- Let's Encrypt barely registers (Green slice)
- A lot of our data is old, and COMODO has offered free trial certificates for years