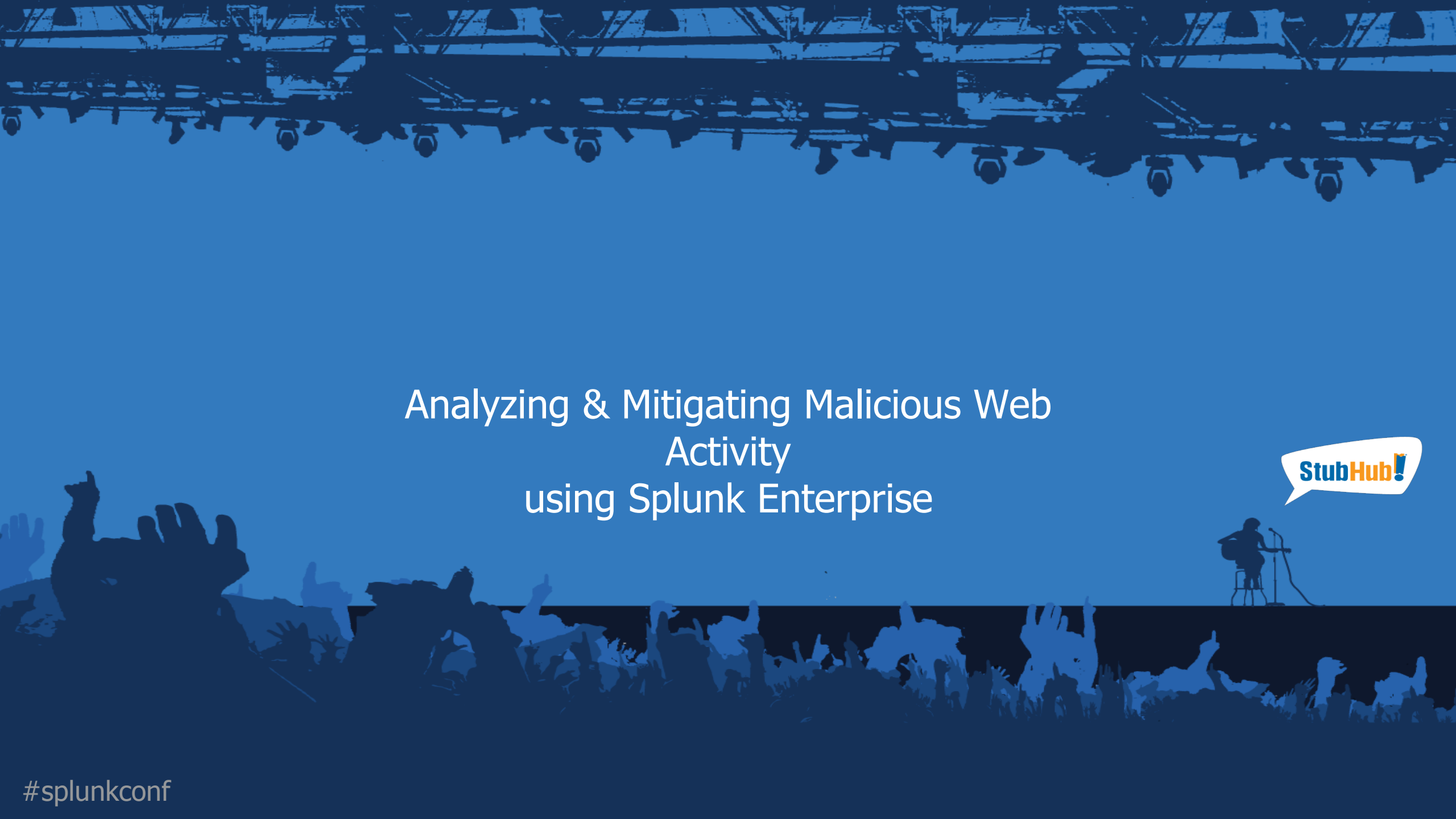
Analyzing & Mitigating Malicious Web Activity using Splunk Enterprise



StubHub – The World's Largest Fan-to-Fan Marketplace

- At StubHub, our mission is simple: provide fans a safe, convenient place to get tickets to the games, concerts, and theater shows they want to see, and an easy way to sell their tickets when they can't go.

Brief Intro

- Who am I?
  - Joined StubHub 2007 as part of application support team
  - In 2011 moved to lead for Tools & Automation team
  - Bit of a Splunk nerd
- www.linkedin.com/in/nathanpratt/
- npratt@ebay.com

Agenda

- There is a constant stream of malicious web hits, poorly written scripts, and overly aggressive web crawlers. By collecting all web access logs into Splunk, you have the power to catalog and trend this activity in real time

Why Attack StubHub?

- Why not?

- Tickets are very liquid – cash!

What Are Web Access Logs?

- Web access logs are the data points generated by a web server when you visit the content that it serves
- These logs establish a historical record of visitor activity
- Traffic patterns can be established and analyzed from this data

Common Attacks

- Usual suspects
  - Sql injection
  - Trolling for admin pages
  - Malformed parameters & parameter walking
- Scripts
  - Shady: scripts that interact with web forms
  - Abusive: scraping data
- Fraud

What Can We Learn From Web Access Logs?

- 161.69.14.159 - [11/Jul/2013:12:19:27 +0000] "GET /admin/default.\"Xx<XaXaXXaXaX>xX/ HTTP/1.1" 301 - "-" "Mozilla/5.0 (compatible; MSIE 7.0; MSIE 6.0; ScanAlert; +http://www.scanalert.com/bot.jsp) Firefox/2.0.0.3" + 14370
- IP address lets us know who made a request: 161.69.14.149
- The other half of who is the user agent: Mozilla/5.0 ... ScanAlert
- What was requested?: GET /admin/default.\"Xx<XaXaSSaXaX>xX/
- We also know the 'where', as Splunk is aware of the endpoints that generated the logs

What IP is Hogging Resources?

1. Write a search to find your access logs

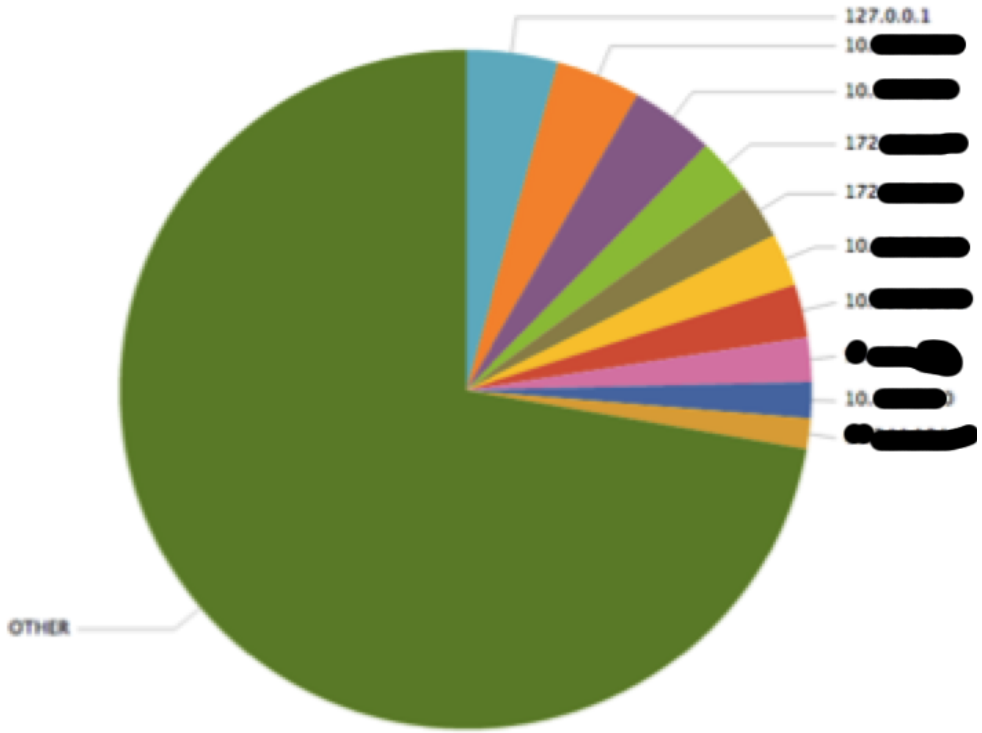
   "index=web"

2. Identify the ten most frequent values of the field 'ip address'

   "| top ipAddress"

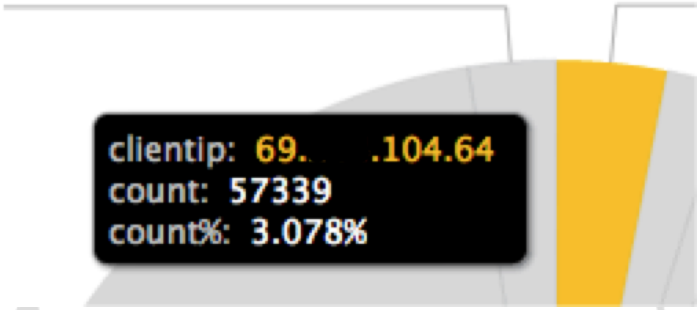
3. View the chart created!

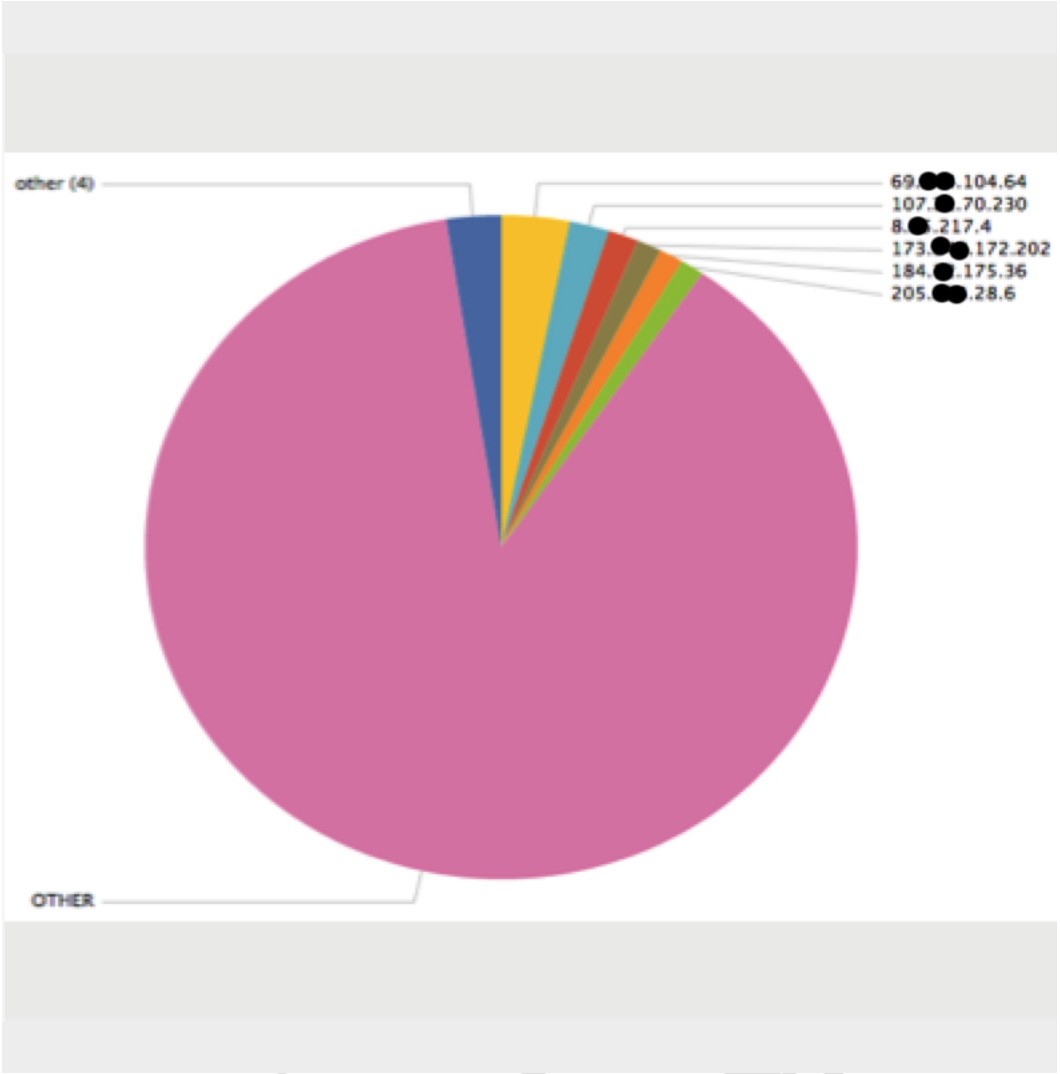
4. Wait... Those are internal addresses…



What IP is Hogging Resources?

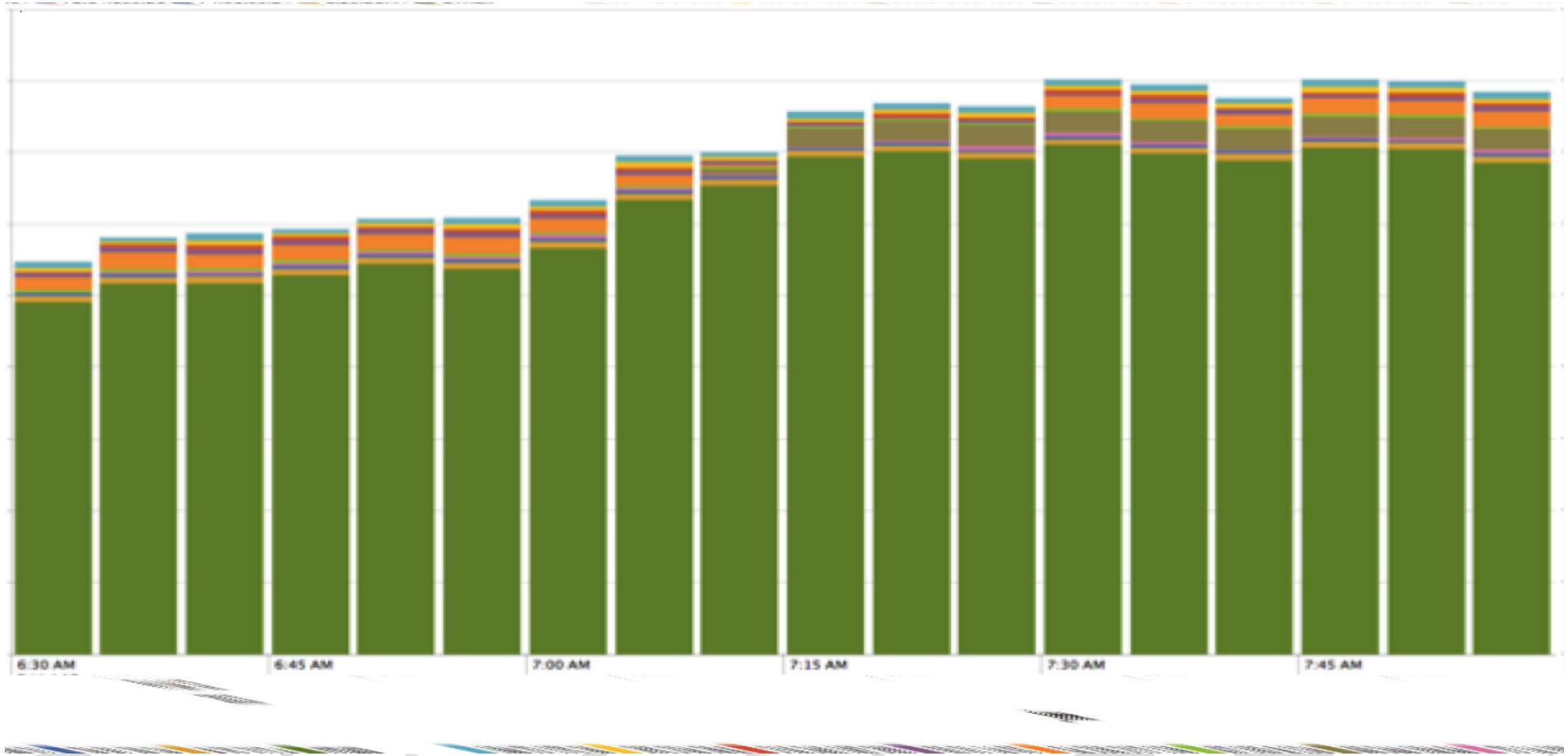
- Add filters to the base search:
  ipAddress!=10.*
- Index=web
  ipAddress!=10.* OR ipAddress!=xxx |
  top ipAddress





When Did This Start?

- Use "| timechart count by ipAddress" instead of "top"

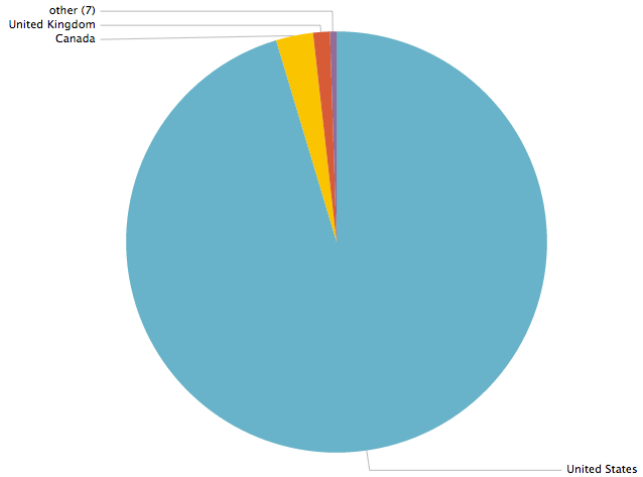


We Need Metadata!

Location	| geoip ipAddress
Who owns the IP?	| lookup whois ipAddress
What does the IP resolve to?	| nslookup ipAddress
Is it a threat?	Project Honeypot: | lookup threatscore ipAddress On the internal whitelist/blacklist?: | lookup ip_whitelist ipAddress

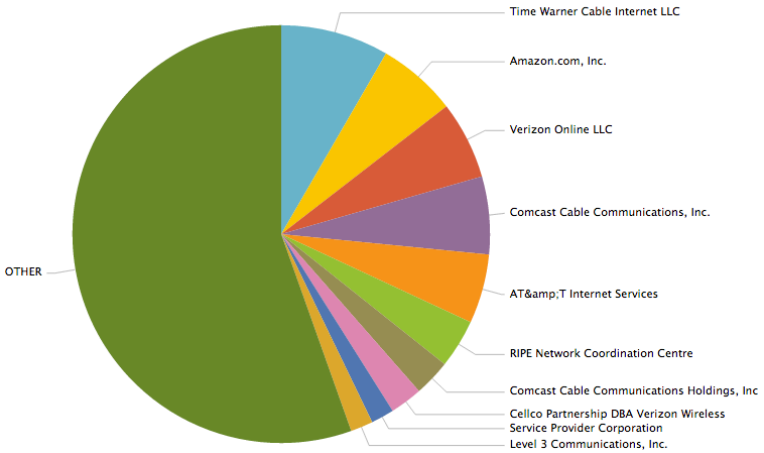
Search Result With Metadata…

Geographically – by City & Country



City/Country	Percent
United States	18.4758
Philadelphia, United States	3.05866
New York, United States	2.772872
Seattle, United States	2.447025
Chicago, United States	1.430892
Houston, United States	1.350529
Santa Clara, United States	1.343202
San Francisco, United States	1.141196
Woodbridge, United States	1.110908
Los Angeles, United States	1.096496
OTHER	65.77242

IP Address owner



IP Owner	percent
Time Warner Cable Internet LLC	.008926
Amazon.com, Inc.	.694768
Comcast Cable Communications, Inc.	.92611
Verizon Online LLC	.876519
AT&T Internet Services	207042
RIPE Network Coordination Centre	.37218
Cellco Partnership DBA Verizon Wireless	.235805
Comcast Cable Communications Holdings, Inc	.603521
Service Provider Corporation	.591123
Level 3 Communications, Inc.	.735681
OTHER	54.748326

Project Honeypot

threatscore	percent
45	0.031158
36	0.039227
35	0.04909
30	0.027571
27	0.069264
24	0.044159
21	0.02645
19	0.041469
17	0.035417
16	0.050659
8	0.380169
6	0.032727
5	1.054208
4	0.027123
0	97.778161
OTHER	0.313147

Project Honeypot is a log based score!!!

What Can We Do with User Agent?

Scenario:

- Load is spiked on all servers
- Hundreds of IP addresses are hitting hard and fast at multiple endpoints
- No pattern to who owns the IP address
- | top ipAddress is wildly askew…
- Hmmm, that looks funny in Splunk

Not a DDoS… 'web crawler' Gone Wild

Custom Web Crawling

Setup your own web crawl in minutes and run it on over 50,000+ computers

Free to sign up. Upgraded accounts available starting at $99 / month.

If you have a good handle of what web crawling you want to do, take advantage of our vast crawling network that enables fast and powerful web crawling. Our custom

How did the company come about?

Shion Deysarkar: Actually, we have a sister company called Plura Processing. Plura is a distributed grid computer--we're able to gather a whole lots of nodes across the Internet for

Shion Deysarkar: Basically, the computer power we use comes from nodes in Plura's network--which is made up of people's home computers all over the world. Plura integrates with a couple of desktop applications, like a pretty popular chat client. People can opt in, and help support the chat client by allowing Plura to run. When they have excess or idle bandwidth and computing power, we can use that for our own purposes.

http://lifehacker.com/5336382/digsby-joins-the-dark-side-uses-your-pc-to-make-money

http://www.texastechpulse.com/interview_with_brad_wilson_and_shion_deysarkar_8_legs/s-0020904.html

We Can Identify Who… What Else Can We Do? Let's Review Data Available

From Logs themselves

- Who: IP Address & User Agent
- What: Request string & GET/POST
- Where: URL

Metadata

- IP owner
- IP DNS name (if available)
- IP Geographical location
- Reputation scores
  - SANS
  - Project Honeypot
  - Internal Whitelist/blacklist

Time to Analyze the Request Itself…

- Malicious requests
  - `<website url>/yankees-tickets/../../../../etc/passwd`
  - `/administrators/index.php (StubHub is not a PHP shop…)`
  - `/;DROP`
- High frequency requests
  - Hits are <1 second apart
- Malformed requests
  - Might be made to avoid caching by CDN

Malicious Request

```
2013-07-12T13:48:45  111.142.171.254  /join_form.php/
2013-07-12T13:48:44  111.142.171.254  /join_form.php
2013-07-12T13:48:43  111.142.171.254  /join.php
2013-07-12T13:48:44  111.142.171.254  /join.php/
```

- StubHub does not have a valid 'join_form.php' URL
- Requests came in seconds apart
- IP originates from China

Eventtypes for Known Bad Requests!

Eventtypes in Splunk are a way to categorize data in Splunk

- Naming convention + wildcard = WIN!
- Search: "index=web web_threat*"
- To add another type of bad request, we simply add another eventtype
- Alerts & dashboards that use the search above will automatically begin using it

- Web_threat_php
  - `url=*.php*`
- Web_threat_aspx
  - `url=*.aspx*`
- Web_threat_admin
  - `url=*admin*`
- Web_threat_passwd
  - `url=*/etc/passw*`
- Web_threat_jmx
  - `url=*/jmx/* OR url=*/jmx-console/*`

Malicious Requests – Creating a Smart Search

- `"index=web web_threat* | stats count dc(eventtype) as attackCount by ipAddress,useragent"`
  - Returns count of bad web hits by unique IP address & useragent combinations, as well as a count of distinct types of bad requests made
- `"| eval threatscore=attackCount"`
  - Creates numerical score driven by the unique type of attacks
- `"| lookup (geo|honeypot|whois|etc)"` *(syntax is not correct here)*
  - Add metadata
- `"| eval threatscore=if(match(country,"USA|Canada|UK"),threatscore,threatscore*2)"`
  - Skew the score against countries that StubHub does NOT have a presence in
- `"| eval threatscore=if(match(useragent,"linux|wget|curl|-") OR isnull(useragent),threatscore*2,threatscore"`
  - Skew the score against known 'interesting' user agents, or missing user agents as these are signs that this is a bad actor

Malicious Requests – Creating a Smart Search

- `"| eval threatscore=if(match(method,"POST") AND match(status,"200"),threatscore*5,threatscore)`
- `"| where threatscore>5"`
  - Eliminate low scoring IP addresses
- `"| sort –threatscore"`
  - Sorts from highest score to the lowest
- `"index=apache web_threat* | stats count dc(eventtype) as attackCount by ipAddress,useragent | eval threatscore=attackCount | lookup (geo|honeypot|whois|etc) | eval threatscore=if(match(country,"USA|Canada|UK"),threatscore,threatscore*2) | eval threatscore=if(match(useragent,"linux|wget|curl") OR isnull(useragent),threatscore*2,threatscore | eval threatscore=if(match(method,"POST") AND match(status,"200"),threatscore*5,threatscore) | where threatscore>5 | sort –threatscore"`

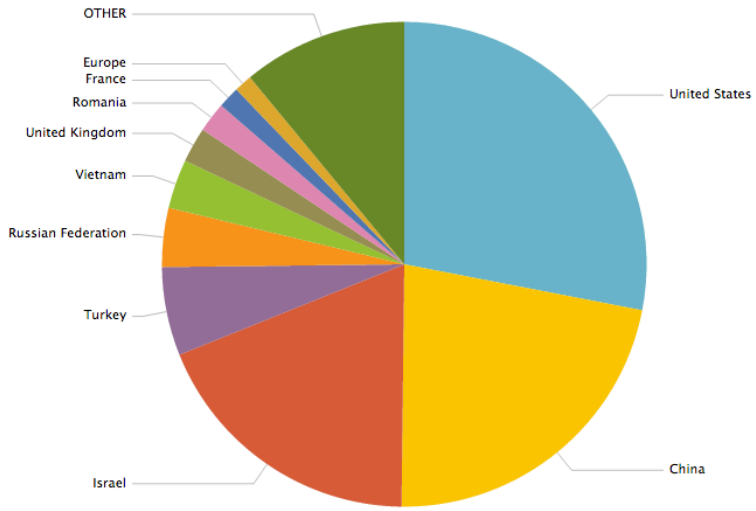
Malicious Requests – Creating a Smart Search

- This can be run from a dashboard in Splunk, or as an alert
- The alert can be set to record the results into a 'summary index'
- Email result:

src_ip	clientip_city	clientip_country_name	whois lookup	Hit Count	Threat Rating	unique url count	unique useragent count	Threat Policy	HTTP Request Method	PH ThreatScore
78.163.60.125	Mersin	Turkey	RIPE Network Coordination Centre	10	6	2	1	apache-uri-php	GET	0

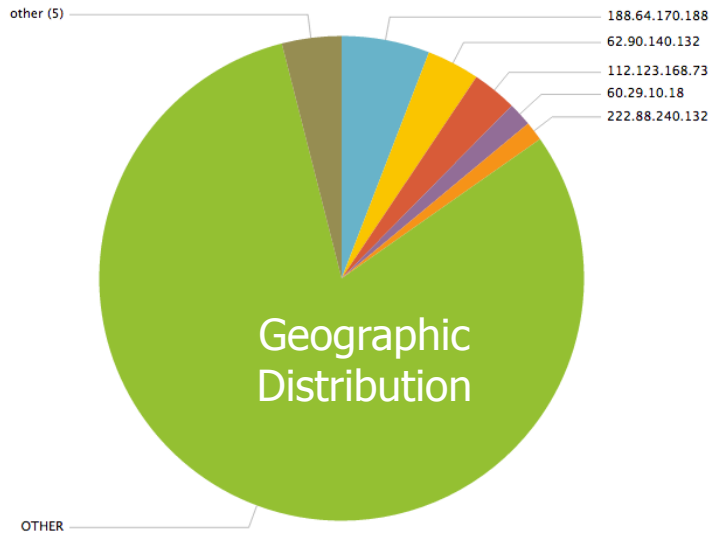
Interesting Stats from the Database of Maliciousness

Geographically – by City & Country



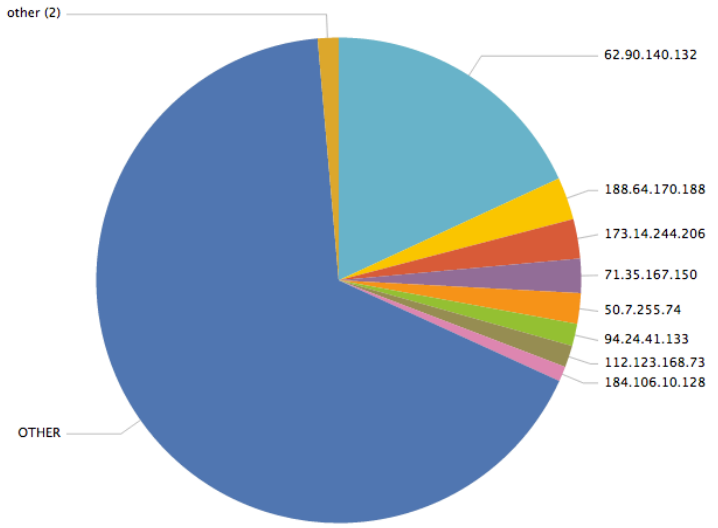
Country	count	percent
United States	43202	28.03195
China	34130	22.145513
Israel	28876	18.736415
Turkey	9097	5.902658
Russian Federation	6058	3.93078
Vietnam	5043	3.272189
United Kingdom	3621	2.349514
Romania	3085	2.001726
France	2225	1.443708
Europe	1851	1.201036
OTHER	16929	10.984512

IP Distribution -1 week



- 188.64.170.188
  - >4500 bad requests
  - Project Honeypot score is **37**
  - From Russian Federation

IP Distribution



- 62.90.140.132
  - >28,000 bad requests
  - No Honeypot score
  - From Israel

Analyzing & Mitigating Malicious Web Activity

Identify Your Visitors

User ID in the Web Access Logs

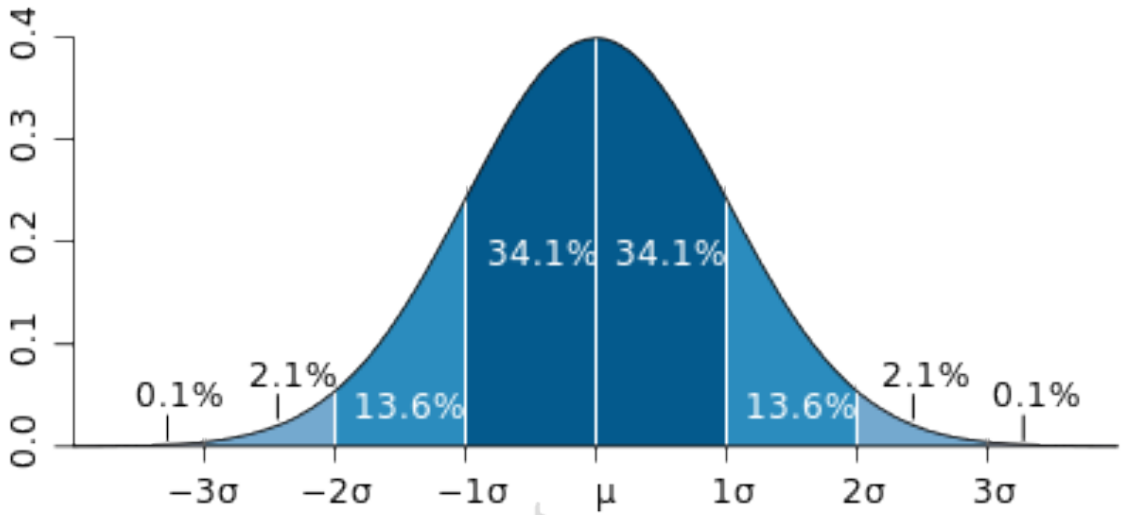
- With identifying information present, you can apply all the existing alerts/ dashboards/queries, but enrich with far more intelligence.
- Form parameters, cookie values, unique URL's, etc are some methods that could be used to accomplish this.

Analyze User Behavior
| stats count dc(unique identifier) as userCount

clientip	count	userCount	timeRange	threatscore	orgname	useragent
[redacted]203	7	1	394	0	[redacted] University	Mozilla/5.0 (Macintosh; Intel
[redacted].34	6	3	429	0	Optimum Online (Cablevision Systems)	Mozilla/4.0 (compatible; MSI
[redacted].254	5	2	387	0	Verizon Online LLC	Mozilla/5.0 (iPhone; CPU iPh
[redacted]17	4	4	10655	0	Comcast Cable Communications Holdings, Inc	Mozilla/5.0 (Linux; Android 4
[redacted].75	4	1	497	0	PSINet, Inc.	Mozilla/5.0 (Windows NT 5.1
[redacted].6	4	1	84	0	Suddenlink Communications	Mozilla/5.0 (Windows NT 6.2
[redacted]213	4	1	633	0	Comcast Cable Communications, Inc.	Mozilla/4.0 (compatible; MSI
[redacted].234	4	1	707	0	Cellco Partnership DBA Verizon Wireless	Mozilla/5.0 (iPhone; CPU iPh
[redacted].226	4	2	945	0	RIPE Network Coordination Centre	Mozilla/5.0 (Macintosh; Intel
[redacted].55	4	1	257	0	RIPE Network Coordination Centre	Mozilla/5.0 (Windows NT 6.1
[redacted]238	4	1	445	0	RIPE Network Coordination Centre	Mozilla/5.0 (Windows NT 5.1
[redacted].50	3	1	195	0	[redacted]	Mozilla/5.0 (Macintosh; U; In
[redacted].162	3	2	944	0	[redacted]	Mozilla/5.0 (Windows NT 6.1
[redacted]98	3	1	116	0	[redacted]	Mozilla/5.0 (compatible; MSI
[redacted]148	3	2	159	0	[redacted]	Mozilla/5.0 (compatible; MSI
[redacted].108	3	1	29	0	Comcast Cable Communications, Inc.	Mozilla/5.0 (Windows NT 6.0
[redacted].185	3	1	119	0	Comcast Business Communications, LLC	Mozilla/5.0 (compatible; MSI
[redacted].210	3	3	16256	0	Comcast Business Communications, LLC	Mozilla/5.0 (Macintosh; Intel
[redacted].84	3	1	422	0	Comcast Cable Communications, Inc.	Mozilla/5.0 (Macintosh; Intel

Detect Brute Force Attacks

- If I want to detect this activity, what do I do? Splunk search!
  - `'index=web | stats dc(unique identifier) as uniqueCount by ipAddress,useragent | sort –uniqueCount'`
- Hmm, that's a lot of data…how do I filter the search results?
  - Hard limit? Attacker will find limit and attack until 'limit-1'
  - Let's use statistics
  - See Splunk Blog for inspiration



http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

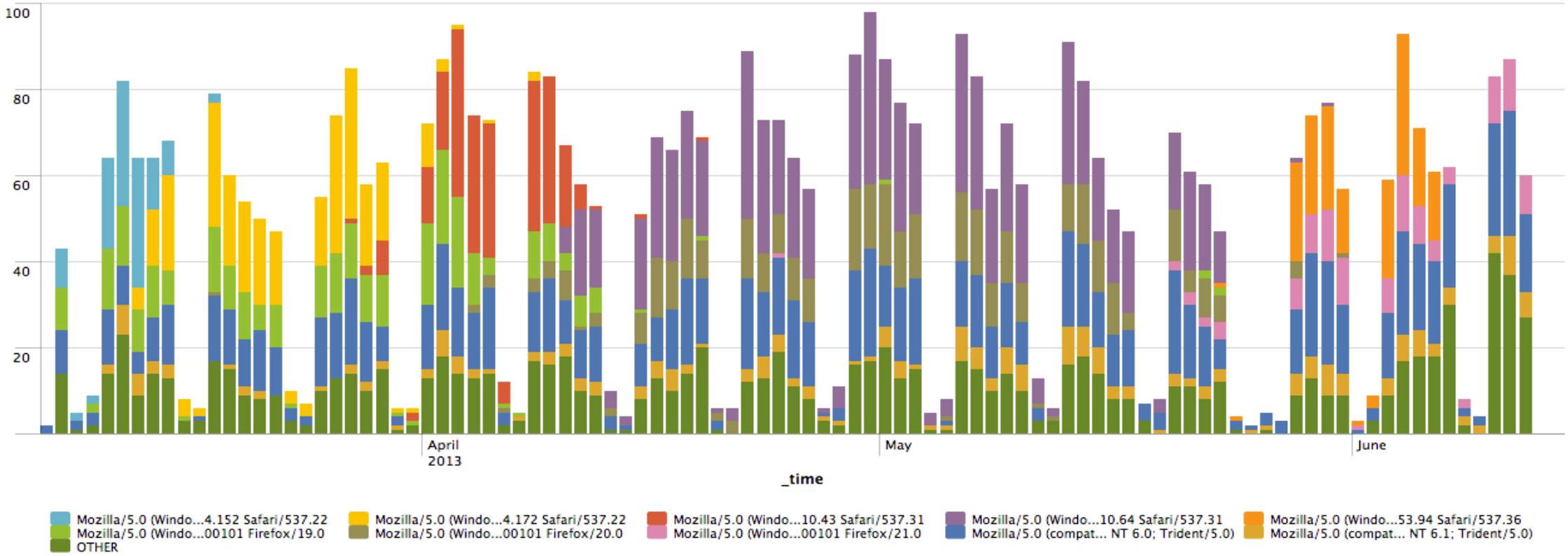
Detect Brute Force

- Base search: 'index=web'
- Statistics: `'| stats dc(unique identifier) as userCount by ipAddress,useragent'`
- Calculate average number of users touched by all IP/user agent combinations:
  - `'| eventstats avg(userCount) as avg stdevp(userCount) as stdev'`
- Add Metadata

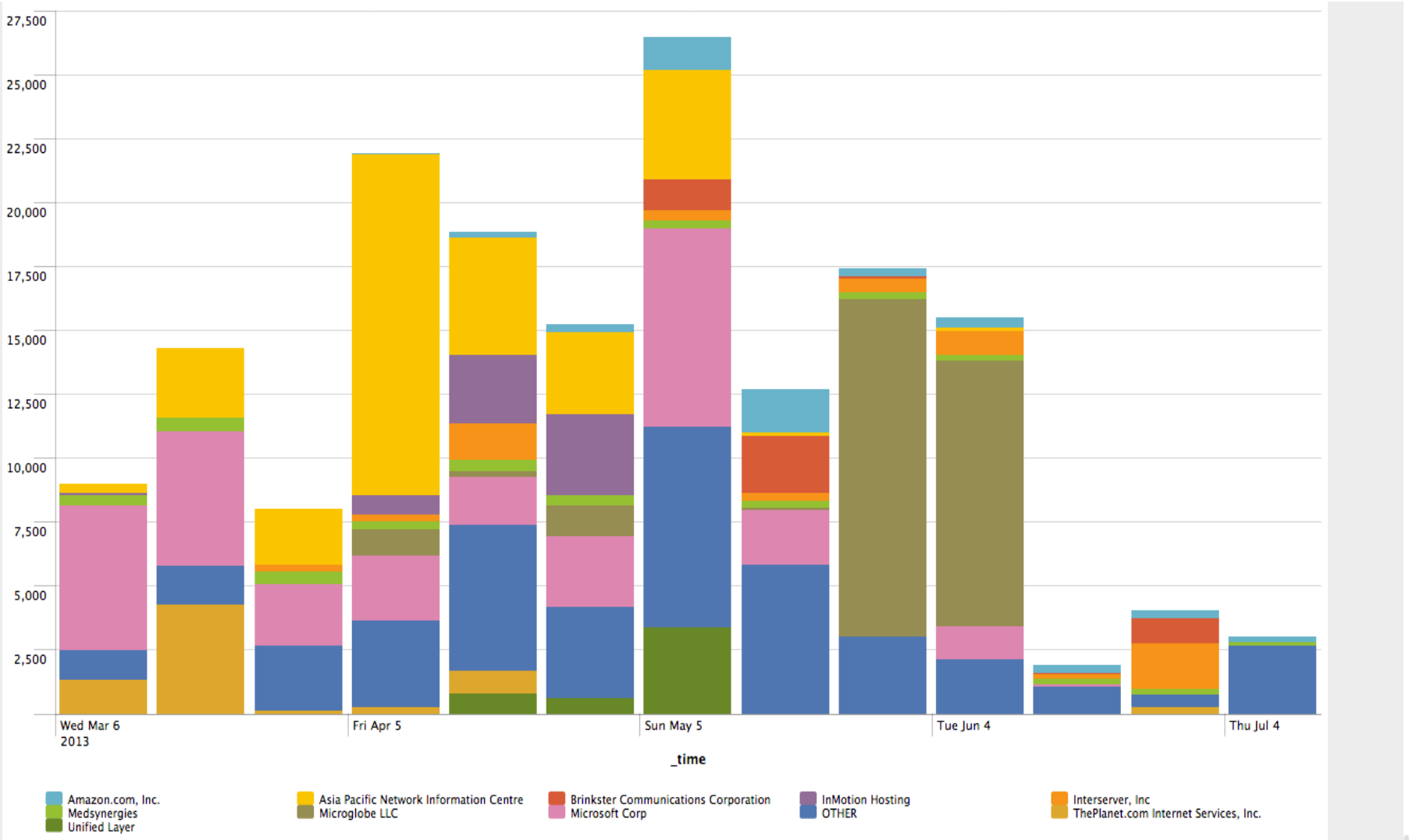
Detect Brute Force

- A ‘z’ score (number of standard deviations from mean) of +4 is above the 99th percentile; let’s pick an incredibly high ‘z’ score as a limit
  - `‘| where userCount>avg+(stdev*20)’`
- USAF is account peeking? (Hi NSA!)
  - Organizations with tightly controlled computers & single exit points to the internet show up frequently
- Used Splunk to compress 5,000:1, with a granularity of 1 day so we can do long term reporting in <5 minutes.

Corporate IT Upgrade Art

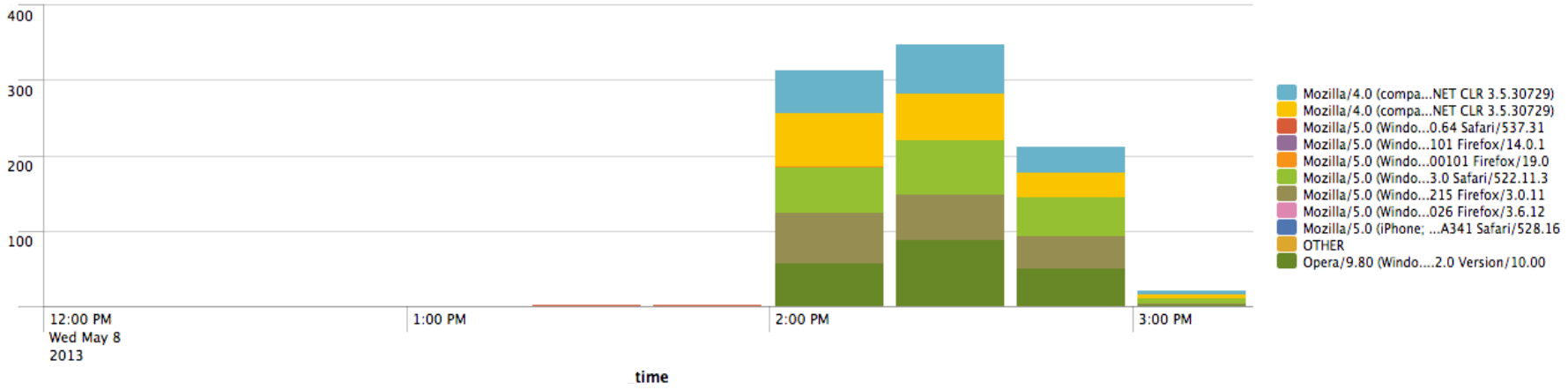


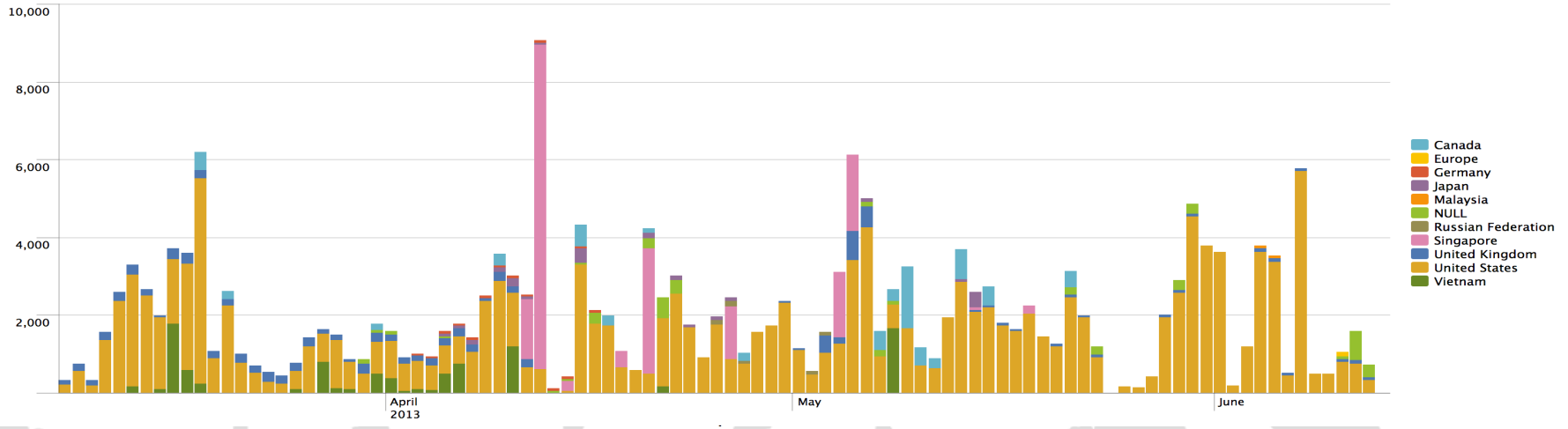
Brute Force by IP Owner



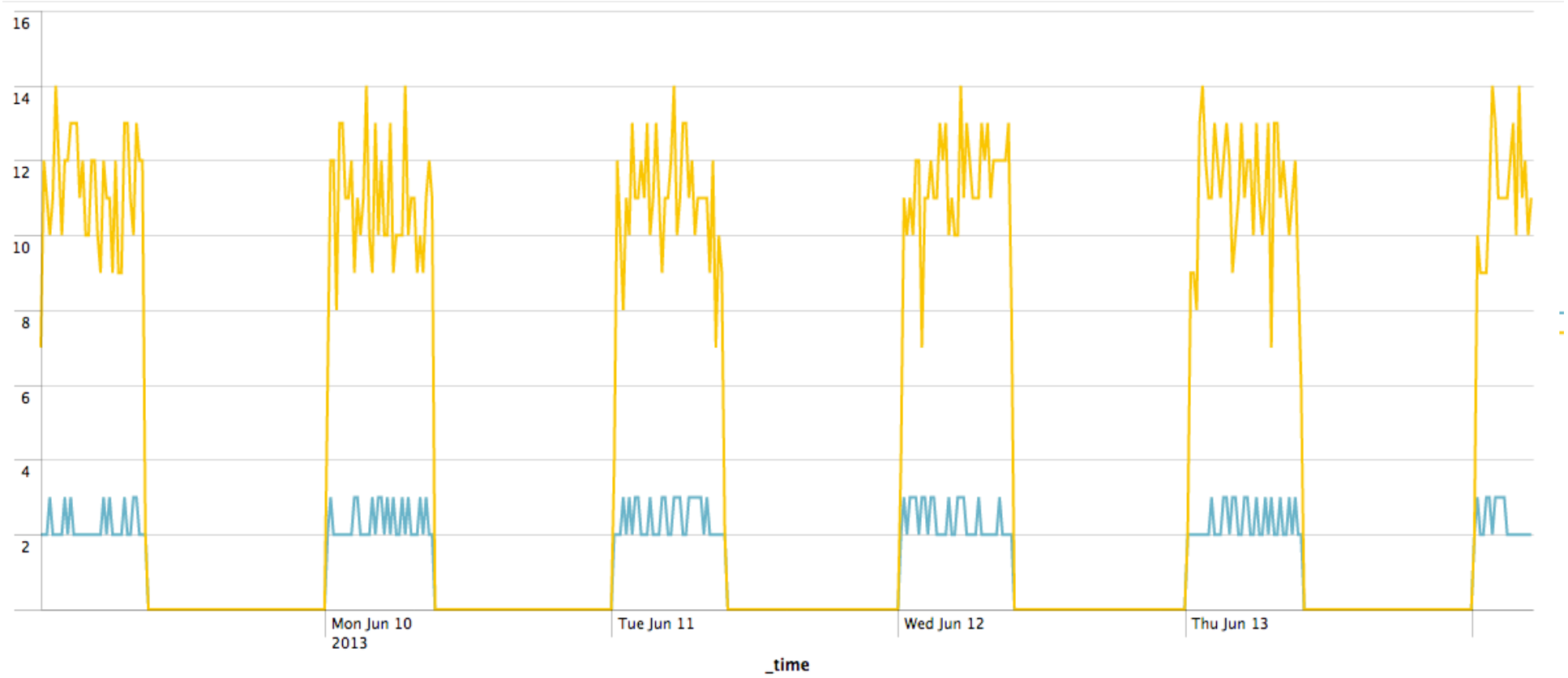
65% of Traffic to an Application Pool!

Swing the Ban Hammer!

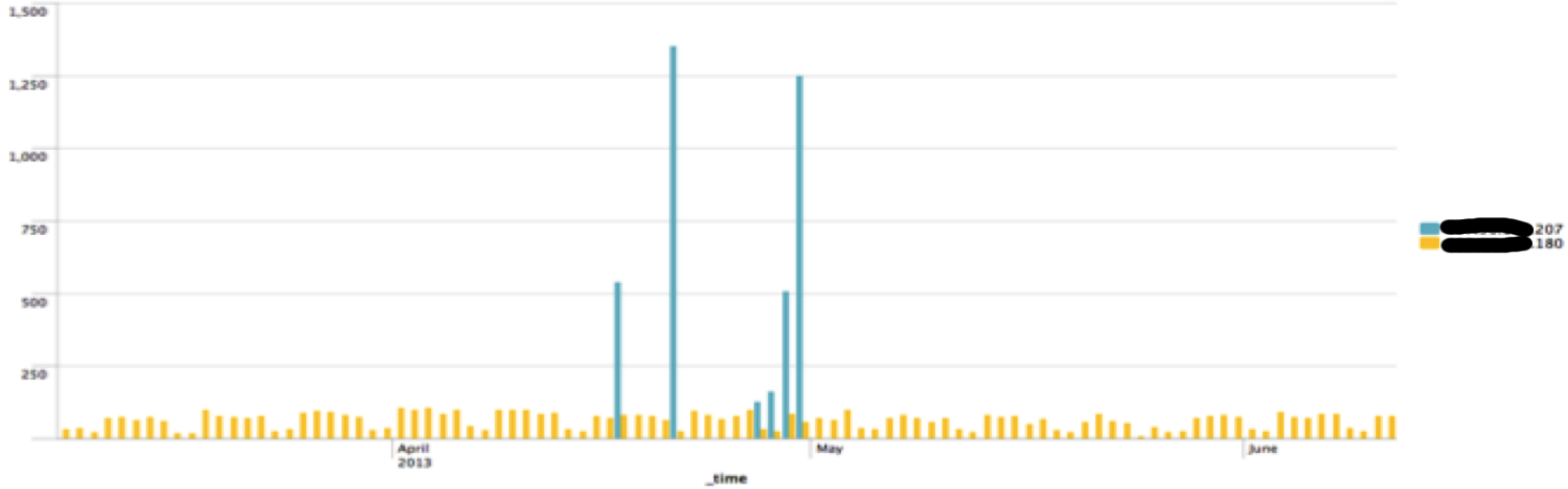




Splunk Doesn't Sleep

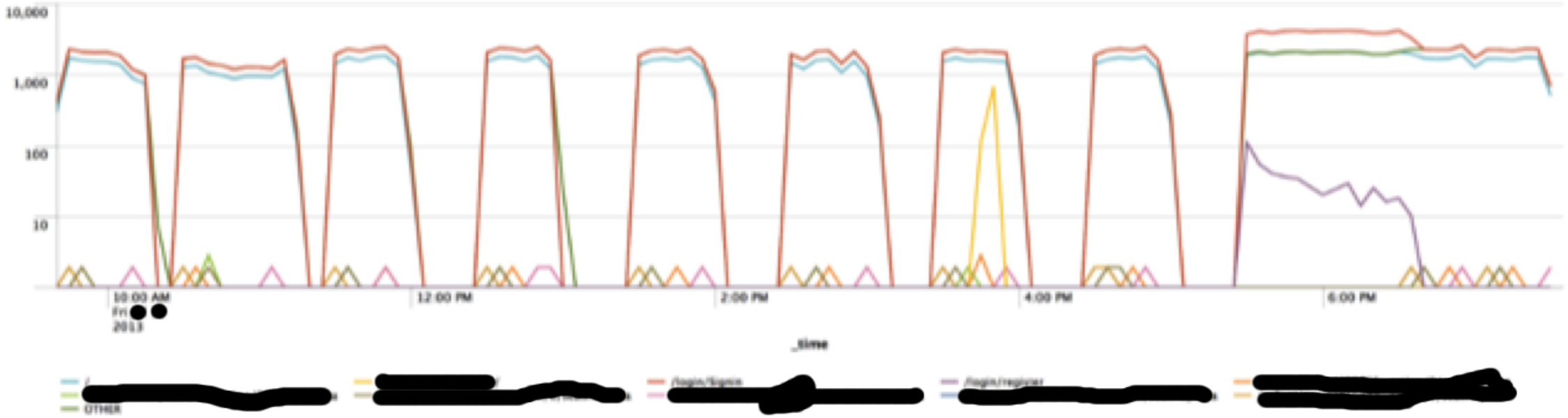


Low & Slow, or Hard & Fast?



	src_ip	[redacted]Count
1	[redacted].207	3578
2	[redacted].180	2905

Anatomy of an Attack



Mitigation

Manual Mitigation

- Block the IP or range
- Block the user agent
- Captcha
- Etc.

Automatic

- Report to your risk/fraud/etc teams to review users
- Automate the steps to the left with a script…

Next Steps

1. **Download the .conf2013 Mobile App**
   If not iPhone, iPad or Android, use the Web App

2. Take the survey & **WIN A PASS FOR .CONF2014**…
   Or one of these bags!

3. **Go to "Securing Splunk for the Enterprise – How to keep Accreditors Away from Splunk"**
   Mont-Royal 1, Level 4
   Today, 10:15-11:15pm