Log Analysis as a Service using open source scalable systems

Gurvinder Singh Dahiya, Uninett AS Belgrade Security Workshop, 20.03.2015



Motivation

- Distributed Systems
- Centeralized interface to logs
- Easier access
- Detection of hidden pattern
- Access to logs across organization
- Cenralized alerts and anomaly detection across services





http://www.themeparkreview.com/tatsumediaday/tatsumediaday57.jpg

Challenges

- Different Formats and logging methods
- Different requirements for processing
- Differnt Dashboards
- Various Alerts requirments





http://img72.imageshack.us/img72/3885/nephew2logs.jpg

What is Log ?

Jun 2 07:40:34 scintilla kernel: [77262.488918] hid-generic 0003:046D:0A15.001C: input,hidraw4: USB HID v1.00 Device [Logitech Logitech G35 Headset] on usb-0000:00:1d.0-1.5.4/input3

[2014-06-02 14:08:39,870][INFO][cluster.metadata update_mapping [mail] (dynamic)] [pltrd003] [mail-2014.06.02]

2014-06-02T12:11:25.271Z 158.36.2.74 https://idp.feide.no feide:sso ntnu.no [u'urn:mace:feide.no:services:no.ntnu.ssowrapper'] 1401711085.27

158.38.213.3 - - [02/Jun/2014:14:12:48 +0200] "POST /__es/logstash-2014.05.26/_search HTTP/1.1" 200 329628 "https://logs.uninett.no/"; "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/34.0.1847.137 Safari/537.36"

DEBUG 2014-06-02 14:15:13,371 [Thread-110743] no.uninett.agora.user.UserRoleSiteUtil Querying for liferay sites for roles: [fs]

TIMESTAMP + DATA = LOG



Time Formats

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE. CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS THE CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13 20130227 2013.02.27 27.02.13 27-02-13 27.2.13 2013. II. 27. $\frac{27}{2}$ -13 2013.158904109 MMXIII-II-XXVII MMXIII $\frac{LVII}{CCCLXV}$ 1330300800 $((3+3)\times(111+1)-1)\times 3/3-1/3^3$ 2013 $\frac{1}{2},\frac{3}{3},\frac{1}{2},\frac{4}{3},\frac{7}{8}$



A log is human readable...

Jun 2 07:40:34 scintilla kernel: [77262.488918] hid-generic 0003:046D:0A15.001C: input,hidraw4: USB HID v1.00 Device [Logitech Logitech G35 Headset] on usb-0000:00:1d.0-1.5.4/input3



65.55.215.109 [18/Jan/2013:19:57:13 -0500] "GET /robots.txt HTTP/1.1" 301 303 "-" "msnbot-medi	ia/1.1 (+http://search.msn.com/msnbot.htm)"
65.55.215.109 [18/Jan/2013:19:57:13 -0500] "GET /gallery/gallery/Subaru/ormeau25_April_200!	5/thumbs/IMG_2854.JPG HTTP/1.1" 301 303 "-" "msnb
ot-media/1.1 (+http://search.msn.com/msnbot.htm)"	
178.255.215.69 [18/Jan/2013:20:02:41 -0500] "GET /robots.txt HTTP/1.1" 301 303 "-" "Mozilla/5	.0 (compatible; Exabot/3.0; +http://www.exabot.co
m/go/robot)"	
178.255.215.69 [18/Jan/2013:20:02:42 -0500] "GET /wordpress/tag/smoke/ HTTP/1.1" 301 303 "-" '	"Mozilla/5.0 (compatible; Exabot/3.0; +http://www
.exabot.com/go/robot)"	있는 것이 이미 가 많았다. 것이 많이 한 것이 같은 것이 가지 않았다. 이미가 가지 않는 것이 가 있었다. 가지 같은 것 같은 것
66.249.73.60 [18/Jan/2013:20:08:19 -0500] "GET /wordpress/cooking/mozambique-style-piri-piri-	chicken/feed/ HTTP/1.1" 301 351 "-" "Mozilla/5.0
<pre>(compatible; Googlebot/2.1; +http://www.google.com/bot.h</pre>	
24.225.83.49 [18/Jan/2013:20:17:53 -0500] "GET /favio	nginabox.com/2009/10/foodblogger-event-zoes-kitc
hen.html" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT	
66.249.73.60 [18/Jan/2013:20:21:19 -0500] "GET /word	nta/trackback/ HTTP/1.1" 301 348 "-" "Mozilla/5.
0 (compatible; Googlebot/2.1; +http://www.google.com/bot	
205.169.30.2 [18/Jan/2013:20:23:44 -0500] "GET /favi	nginabox.com/2011/10/paleo-diet-bento.html" "Moz
illa/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.17 (KHTI	/537.17"
72.14.199.133 [18/Jan/2013:20:23:45 -0500] "GET /work	tcher-Google; (+http://www.google.com/feedfetche
r.html; 18 subscribers; feed-id=12636598490283692241)"	
66.249.73.60 [18/Jan/2013:20:34:00 -0500] "GET /word	les/comment-page-1/ HTTP/1.1" 301 348 "-" "Mozil
la/5.0 (compatible; Googlebot/2.1; +http://www.google.com	
157.55.33.80 [18/Jan/2013:20:48:51 -0500] "GET /robo	(compatible; bingbot/2.0; +http://www.bing.com/b
ingbot.htm)"	
157.55.33.80 [18/Jan/2013:20:50:50 -0500] "GET /word]	ozilla/5.0 (compatible; bingbot/2.0; +http://www
.bing.com/bingbot.htm)"	
66.249.73.60 [18/Jan/2013:21:00:01 -0500] "GET /word]	-pate-stuffed-chicken-legs/trackback/ HTTP/1.1"
301 350 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +ht	
58.22.10.92 [18/Jan/2013:21:00:04 -0500] "GET /wordp	on-and-pastrami-pizza-with-garlic-pizza-fritta/
HTTP/1.0" 301 352 "http://xesla.ro/wordpress/cooking/pati	pizza-with-garlic-pizza-fritta/" "Opera/9.80 (Wi
ndows NT 6.1; Win64; X64; U; ru) Presto/2.10.289 Version	
24.220.125.231 [18/Jan/2013:21:06:00 -0500] "GET /WO	301 332 "http://honest-food.net/wild-game/venis
on-recipes/venison-stews/wild-game-gumbo/" "Mozilla/5.0	DKit/536.26.17 (KHIML, like Gecko) Version/6.0.2
72.14.199.133 [18/Jan/2013:21:23:45 -0500] "GET /WORK	tcher-Google; (+http://www.google.com/teedfetche
F. ntml; 18 subscribers; Teed-td=12030598490283092241)	D/4 4" 201 240 " " "Warilla/5 0 (compatible: Com
66.249.73.60 [18/Jan/2013:21:26:07 -0500] "GET /Word]	P/1.1 301 340 - Mozilla/5.0 (compatible; Goo
glebot/2.1; +http://www.google.com/bot.ntml)	01 222 " " "Decome /2 @ NOOFS (=100. TD. 124116) (==
00.249.73.00 [18/Jan/2013:21:28:57 -0500] GET /WOFd[01 332 - DOCOMO/2.0 N905((C100;TB;W24H10) (CO
<pre>Mpattole; Googlebot-Mobile/2.1; +nttp://www.google.com/bot.ntml) of 100 150 235 [10/350/2012:21:20:07 0500] "CFT /sobots tyt UTTD/1 1" 201 202 " " "Mostilla/5</pre>	A (compatible, Vander Pat/2 A. abtter //wander com
/bets)"	.0 (Compartible; YandexBor/3.0; +http://yandex.com
/DULS) DE 100 1E0 22E [10/Jon/2012:21:20:07 DE00] "CET /cohote tyt HTTD/1 1" 201 207 " " "Mozillo/E	@ (compatible: VandexPot/2 @: ubttp://wandex.com
/bots)"	.o (compartible; YandexBoc/3.0; +http://yandex.com
/UULS) 05 108 150 225 [18/]=n/2012:21:20:08 _0500] "CET / HTTP/1 1" 201 202 "_" "Mozilla/5 0 (compati	ble: VandexPot/2 0: +http://wandex.com/bots)"
95 108 150 235 [18/Jan/2013.21.29.08 -0500] CET / HTTP/1.1 301 305 - MOZILLA/S.0 (COMPAL	ible: VandexBot/3 0: +http://vandex.com/bots)
94 228 34 233 [18/Jan/2013:21:30:08 -0500] "GET /wordpress/cooking/chocolate-broad/food/ HTTP	/1 1" 301 332 "-" "magnie-crawler/1 1 (1) linux a
md64: ep-GB: +http://www.brandwatch.pet)"	



But machine parsable ... maybe ?

Jun 2 07:40:34 scintilla kernel: [77262.488918] hid-generic 0003:046D:0A15.001C: input,hidraw4: USB HID v1.00 Device [Logitech Logitech G35 Headset] on usb-0000:00:1d.0-1.5.4/input3







Apache Regex ...

`(?:\\.|[^\\`])*`))))))

(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-]{0,62})(?:\.(?:[0-9A-Za-z][0-9A-Za-z-]{0,62}))*(\.?|\b))|((?<![0-9])(?:(?:25[0-5]|2[0-4][0-9]|[0-1]? $[0-9]{1,2}[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2]$ $[0-4][0-9][0-1]?[0-9]{1,2}[.](?:25[0-5]]2[0-4][0-9][0-1]?[0-9]$ $\{1,2\})(?![0-9])))(([a-zA-Z0-9_-]+))(([a-zA-Z0-9_-]+)))(((2-zA-Z0-9_-]+))))$ [01]][1-2]?[0-9]|0?[1-9]))/(\b(?:Jan(?:uary)?|Feb(?:ruary)?|Mar(?:ch)?|Apr (?:il)?|May|Jun(?:e)?|Jul(?:y)?|Aug(?:ust)?|Sep(?:tember)?|Oct(?:ober)?|Nov(?: ember)?|Dec(?:ember)?)\b)/([0-9]+):((?!<[0-9])((?:2[0123]|[01][0-9])):((?:[0-5][0-9]))(?::((?:(?:[0-5][0-9]|60)(?:[.,][0-9]+)?)))(?![0-9])) ((?:[+-]?(?:[0-9]+)))))] "(\b\w+\b) (((?:/[A-Za-z0-9\$.+!*'(),~:#%_-]*)+)(?:(\?[A-Za-z0-9\$.+!*'(),~#%&/=:;_-]*))?) $HTTP/((?:((?<![0-9.+-])(?>[+-]?(?:(?:[0-9]+(?:\backslash.[0-9]+)?)|(?:\backslash.[0-9]+))))))" ((?:((?<![0-9.+-])(?>[+-]?)))))))))$) "(?:(([A-Za-z]+(\+[A-Za-z+]+)?)://(?:(([a-zA-Z0-9_-]+))(?::[^@]*)?@)?(?:(((?:(\b(?:[0-9A-Za-z][0-9A-Za-z-] $[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2})[.](?:25[0-5]|2[0-4][0-9]|[0-1]?[0-9]{1,2}))(?![0-9]))))(?::(b(?:[0-9]+)b))?))?$

• Stage 1 – Single Host

\$ grep -I "Invalid user " /var/log/auth.log* | awk '{ print \$10; }'

• Stage 2 - A handful of hosts

```
#!/bin/sh
USER=root
KEY=/root/public_key.pub
for HOST in server1 server2 server3 server4
do
    ssh -I $USER -i $KEY $HOST grep -I "Invalid user " \
    /var/log/auth.log* | awk '{ print $10; }'
done
```



• Stage 3 – Lot of servers



\$ grep -I "Invalid user " /rsyslog/*/auth.log* | awk '{ print \$10; }'



• Stage 4 – Start using splunk

Splunk	Splunks - Pre	ferences + 05/01/2006 00:00:00	Duration	Minutes 🛨 End 0	05/04/2006 23:59:59
illir Events by Time 80- 40-	Click any bar to set	a time range. Shift-click an	other to extend the range		1 bar = 1 hour ⊙, ⊕, -80 -40
Monda Events (686)	y May 01 Event Types (18)	Tuesday May 43 ev 05/03 05/03 Source Types (1)	ents between 3 2006:11:00:00 and 2006:11:59:59 Hoats:12	ý May 03 Sources (1)	Thursday May 04
Events 1-10 of 686, so	rted by time. Page: 1	2 3 4 5 6 68 69 N	ext >		Results per page 10 -
May 4 16:44:51 (EVENTTYPE: 74993 SOURCETYPE: syst Look up event Sh	gvzk1 sshd(pam_unix)[2 0 05/04/2006, 16:44:51 og 0 HOST:gwrk1 0 are event	2277]: session closed Similar Related SOURCE: top:127.0.0.1:9	for user jgibbons 998 Show source		
May 4 14:47:28 rhost=splurk-2.er EVENTTYPE:71059 SOURCETYPE:syst Look up event Sh	gvrk1 sshd(pam_unix)[4 net.interop.net 9 📰 05/04/2006, 14:47:2 og 🖾 HOST:gwrk1 🛅 are event	1572]: 2 more authenti 16 Similar Related 160URCE top:127.0.0.1:9	ootion failures; log 998 Show source	pname≕ uid≕0 euid≕0	tty=ssh ruser=
May 4 14:47:28 (EVENTTYPE 79036 SOURCETYPE syst Look up event Sh	gwrk1 sshd[4572]: Fail 05/04/2006, 14:47:25 og HOST:gwrk1 are event	ed password for inval 3 Similar Related BOURCE top:127.0.0.1:9	lid user dien from :: 998 Show source	ffff:45.106.2.28 po	art 53398 ssh2



• The first one is free – 500 MB per day





• Incoming invoice from Splunk





- Stage 5 Open source scalable solutions
 - Logstash
 - Elasticsearch
 - Kibana
 - ZeroMQ
 - Logstash-forwarder
 - Rsyslog
 - Statsd



Logstash

• Turns this:

{

- «192.168.0.74 - [13/May/2014:04:28:55 -0500] "GET /robots.txt HTTP/1.1" 301 303 "-" "Mozilla/5.0 (compatible; DSASE/1.0; bot@gmail.com)"»
- Into:

```
"client address": "192.168.0.74",
"user": null,
"timestamp": "2014-05-13T14:04:28-0500",
"verb": "GET",
"path": "/robots.txt",
"query": null,
"http version": 1.1,
"response code": 301,
"bytes": 303,
"referrer": null
"user agent": "Mozilla/5.0 (compatible; DSASE/1.0; bot@gmail.com)"
```



Elasticsearch

- Document Oriented Free Text Search & Analytics Engine
- JSON
- Apache Lucene
- No Schema
- Mapping Types
- Horizontally Scaleable, Distributed
- REST API
- Vibrant Ecosystem
- Completely Open Source



Elasticsearch

- Index
 - Logical collection of data; might be time based
 - Analogous to a database
- Sharding
 - Split logical data over several machines
 - Write scalability
 - Control data flows
- Replication
 - Read scalability
 - Removing SPOF



Elasticsearch Shards and Replication

• Index allocation with 10 shards and 3 replication factors





Kibana

- Javascript based web application
- No dependency except a web server (Changing a bit in Kibana 4)
- Visualize data stored in Elasticsearch
- Dynamic panel and dashboard creation support
- Multiple panel types support
 - Tables
 - Pie charts
 - Maps
 - Text
 - Trends
 - Histograms



Kibana Visualization

Gun Death Map 🛛 🌣

This dashboard is updated daily from data provided by slate.com. Every data point is a single human life taken by a gun in the United States since December 13th, 2012. Click on clusters to drill down. Hover over blue markers to see the names of the deceased.

Since		
12/13/2012	00:00:00	~

Relative | Absolute | Since | Auto-refresh

Dashboard Control



Kibana 3 milestone 2

People killed by Guns in the United States





Kibana Visualization





UNINETT



TOP DEST PORTS	•	• + ×	TOP SRC POR	TS 🔁	• +
-	01	0 - 41	Term	Count	Actio
Term	Count	Action	3544	2608	Qe
445	24789	Q Ø	6000	2197	90
23	6317	Q Ø	80	633	90
22	2754	Q Ø	5000	479	90
3389	1965	90	40000	357	90
80	1752	Q Ø	5168	205	90
53	1116	Q Ø	55951	194	00
1900	1040	Q Ø	57094	185	0.0
8080	771	Q Ø	442	164	0.0
			443	164	40

Lego Brick-Built





Authentication & Authorization

- Sensitive information
- Privacy concerns
- Support LDAP groups
- Support Feide, Norwegian single-sign on solution
- A nodejs app, forked from another kibana-proxy
- Access is allowed on per service basis by service owner



Data Model

- Normalize the diversity of information
- Logsource field for host information
- Username
- src_ip,dest_ip, src_port, dest_port
- _service

UNINETT

• type

on	@timestamp	೦⊘≣	2014-06-04T19:04:28.000Z
	@version	Q Ø Ⅲ	1
	_id	Q Ø Ⅲ	87f2ed1bf7a0806f5ee6732a860bb1f4
	_index	Q Ø Ⅲ	dns-2014.06.04
	_service	Q Ø Ⅲ	dns
	_type	Q Ø Ⅲ	dns
	direction	Q Ø Ⅲ	IN
	flags	Q Ø Ⅲ	
	geoip.asn	Q Ø Ⅲ	UNINETT, The Norwegian University & Research Network
	geoip.continent_code	Q Ø Ⅲ	EU
	geoip.country_code2	Q Ø Ⅲ	NO
	geoip.country_code3	Q Ø Ⅲ	NOR
	geoip.country_name	Q Ø Ⅲ	Norway
	geoip.ip	Q Ø Ⅲ	158.38.100.171
	geoip.latitude	Q Ø Ⅲ	62
	geoip.location	Q Ø Ⅲ	10,62
	geoip.longitude	Q Ø Ⅲ	10
	geoip.number	Q Ø Ⅲ	AS224
	geoip.timezone	Q Ø Ⅲ	Europe/Oslo
	host	Q Ø Ⅲ	2001:700:1:8:0:0:180:57:58234
	logsource	Q Ø Ⅲ	moholt
	message	Q Ø Ⅲ	Jun 4 21:04:28 moholt named[2124]: client 158.38.100.171#45933: query: puppet.uninett.no.uninett.no IN AAAA + (158.38.180.57)
	pid	Q Ø Ⅲ	2124
	program	Q Ø Ⅲ	named
	qtype	Q Ø Ⅲ	AAAA
	query	Q Ø Ⅲ	puppet.uninett.no.uninett.no
	src_ip	Q Ø Ⅲ	158.38.100.171
	src_port	Q Ø Ⅲ	45933
	timestamp	Q Ø Ⅲ	Jun 4 21:04:28
	tld	Q Ø Ⅲ	no
	type	Q Ø Ⅲ	dns

Log Analysis as Service (LAAS)

- It's up and running
- Pilot is offered till April 14, 2015 and production from April 15, 2015
- In pilot with Geant SA7-T1 acitivity
 - Colloboration with GarrNet, Heanet
- Following institutions are sending logs
 - Høgskolen i Oslo-Akerhus
 - Høgskolen i Hedemark
 - Høgskolen i Østfold
 - Høgskolen i Ålesund
 - Handelshøyskolen
 - Uninett AS
- Currently receiving 60 GB data per day
- Currently total sliding data size is 2TB



Log Analysis as Service (LAAS)

- It's becoming an eco system
 - Suricata IDS service coming up and can send logs from IDS service to LaaS
 - Edudbg service used LaaS as backend to provide debugging information to all eduroam users in Norway.
- Intstitutions can share data with each other without giving access to servers itself
- Can share data with external consultants as well
- Collaborate on making parsers which can help saving resources and share benefit
- Collaborate on dashboards as well for similar log services





