

Find A Hay In Haystack! How Splunk Help Recruit To Detect 0.000001% Threat And More...

Mitsuhiro Nakamura

Senior Security Engineer, Recruit Technologies Co., Ltd

.conf2016

splunk>

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.



Agenda

Bio & About Company

1. Heuristic Analysis Using Splunk
2. Data Visualization Using Splunk D3.js Apps
3. Deep Learning (Extra Challenge)



Bio

Mitsuhiro Nakamura (hiro)

- 10+ years experience in cyber security specializing Pen-Test, Forensics, and Incident-Response.
- Built Web Application Vulnerability Assessment Methodology.
- Data analysis for Threat Detection using Splunk.

like: Windsurfing (Wave)



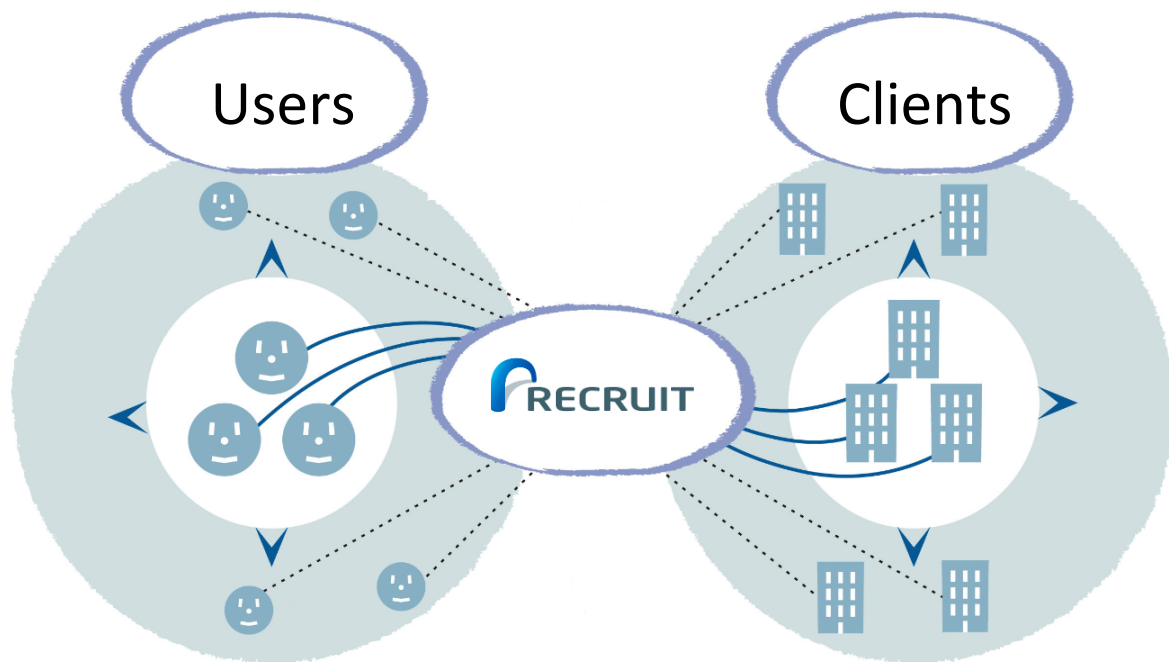
Recruit Holdings Co., Ltd.



Founded	1960
Employees	38,000
IPO	Oct/2014
Annual Sales	\$ 14.0 bn



Business Model



Housing and
Real Estate

JPY **83.9Bn**

Bridal

JPY **53.6Bn**

Travel

JPY **53.4Bn**

Dining

JPY **34.3Bn**

Beauty

JPY **39.9Bn**

Domestic
Recruiting

JPY **239.8Bn**

Domestic
Staffing

JPY **389.5Bn**

We are committed to creating chances to discover “Opportunities for Life,” connecting users and clients through new channels.

Domestic

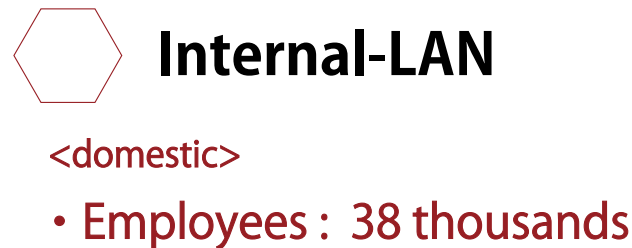
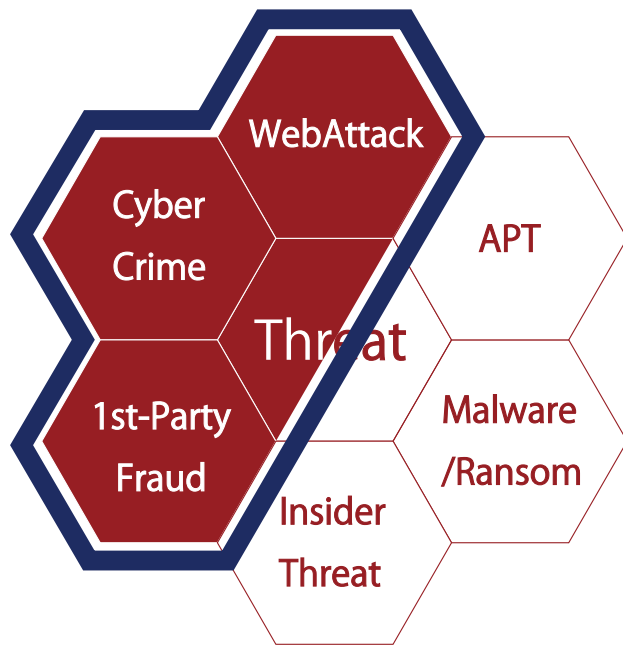


Abroad (M&A)



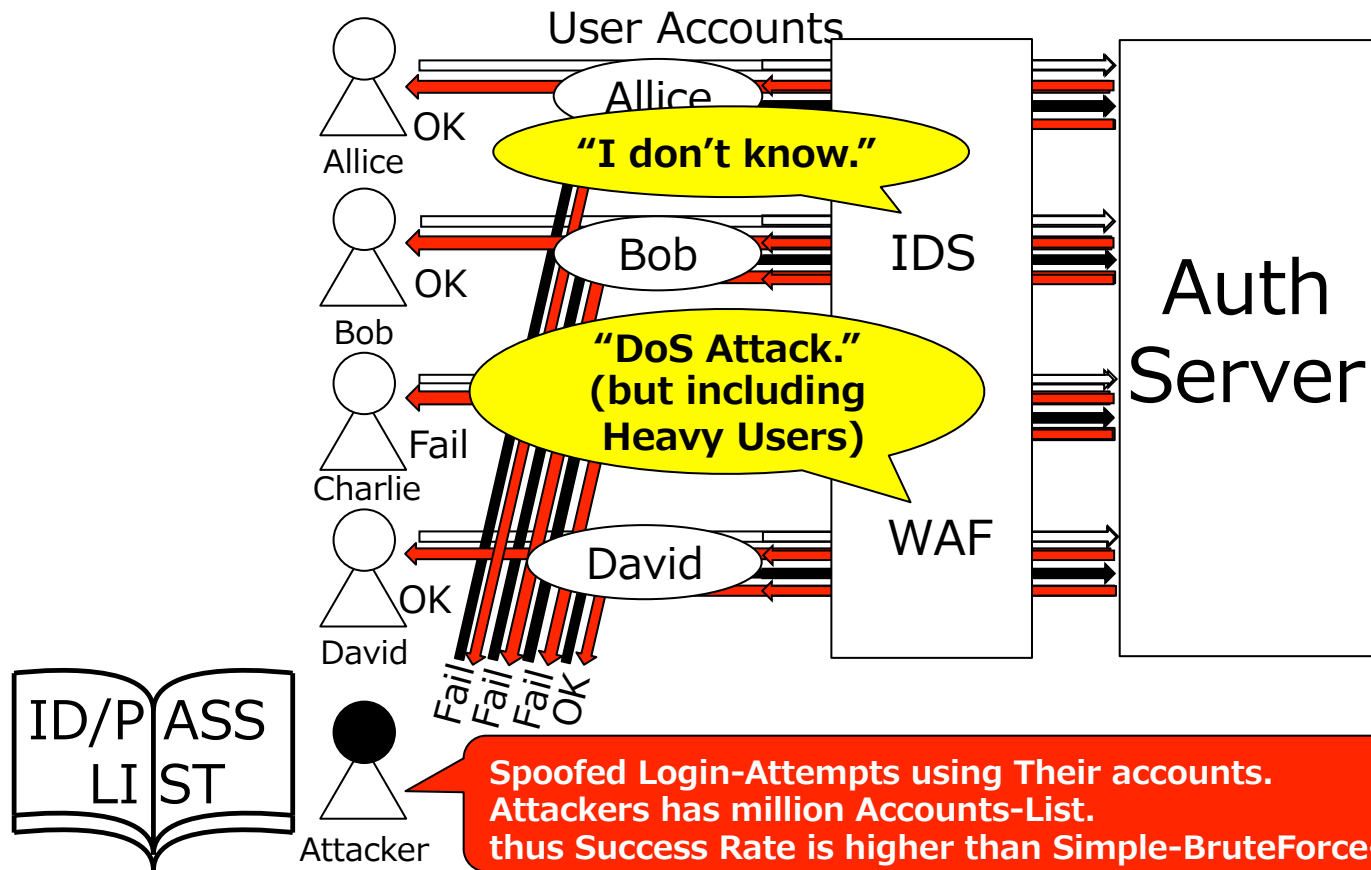
200+ Web Commercial Sites

What Is “Threat” On The Recruit?



Today's my topic is..
“List-Type Account Hacking”

What Is The “List-type Account Hacking”?



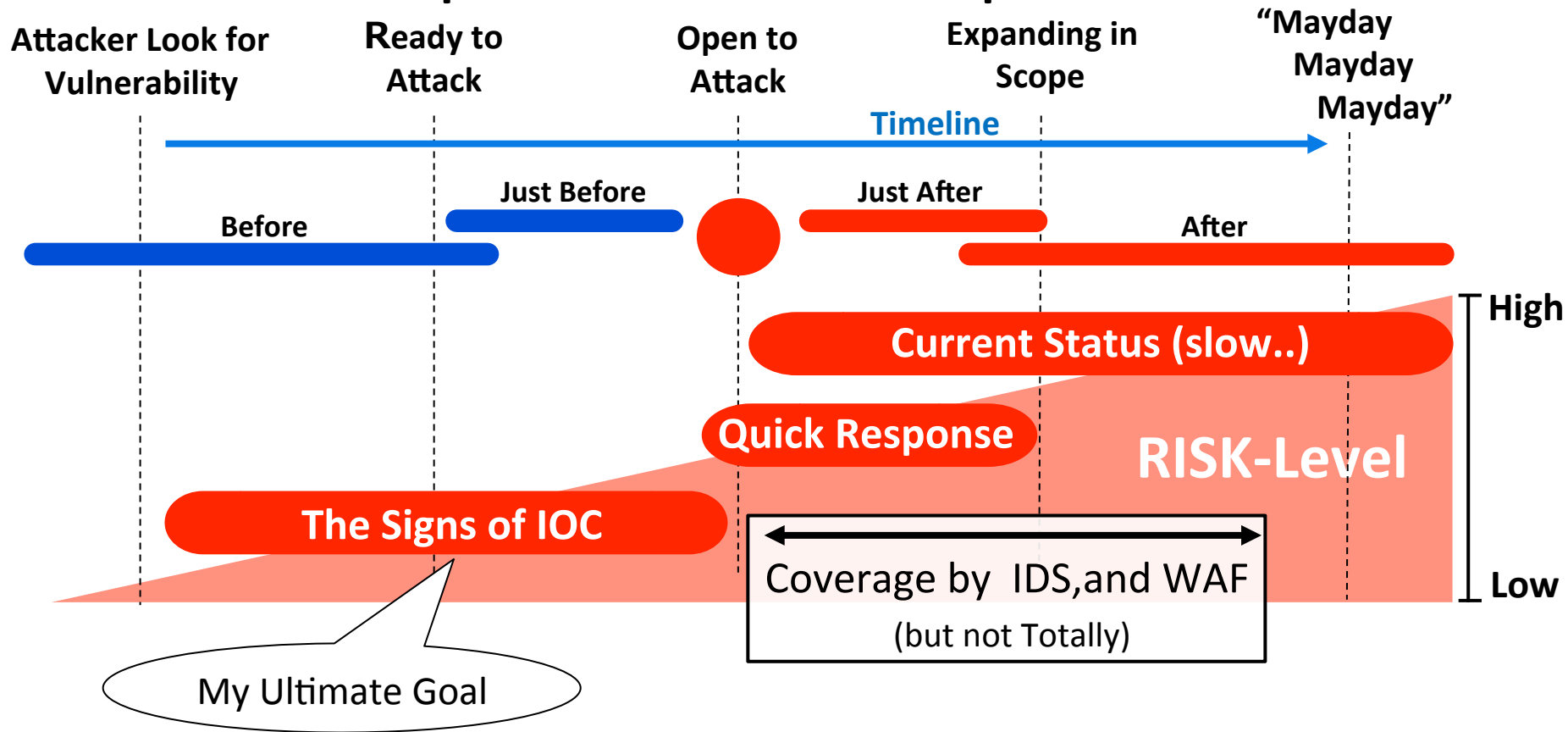
attacked at
Sep/2014

damaged
9749 ID

We were wild-samurai
at that time!



Rapid Incident Response



The Reason I Chose Splunk

1. Log says Everything.

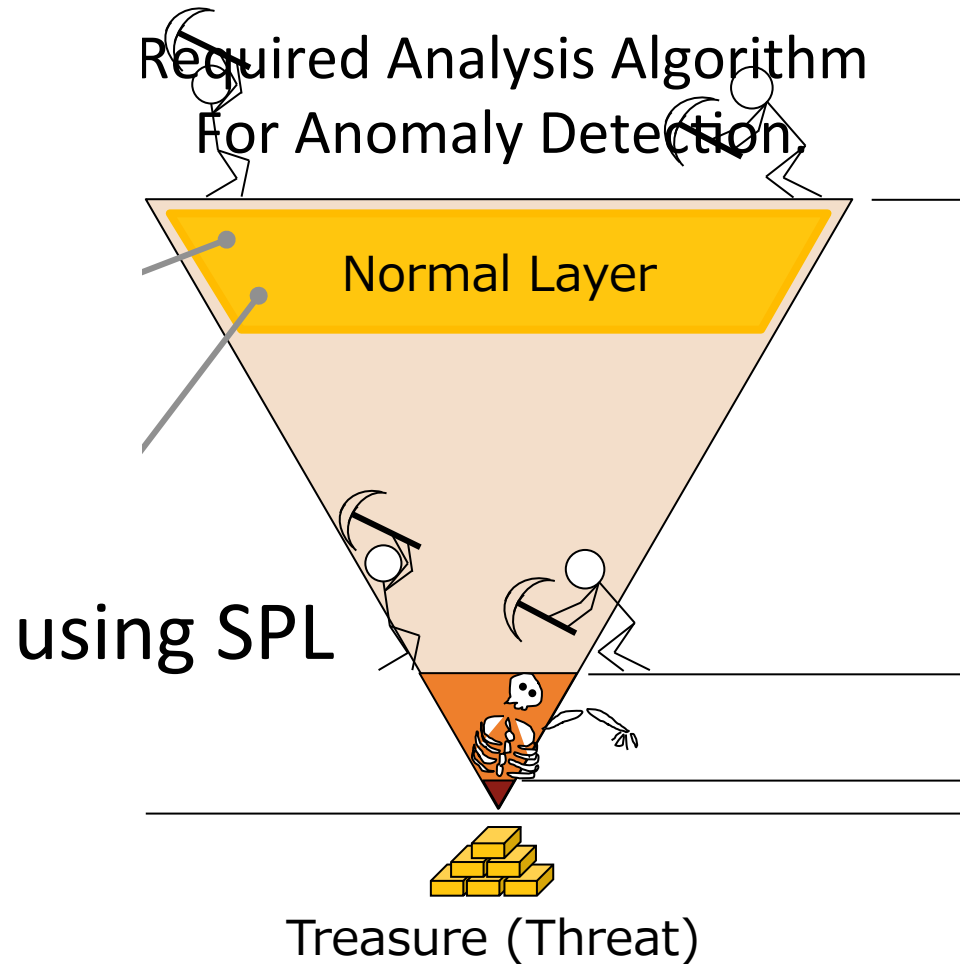
2. Real-time
analysis Capability.

3. SPL.





Required Analysis Algorithm For Anomaly Detection



Challenge 1: Recovering Behavior From Fragment-data. “Heuristic Analysis”

.conf2016

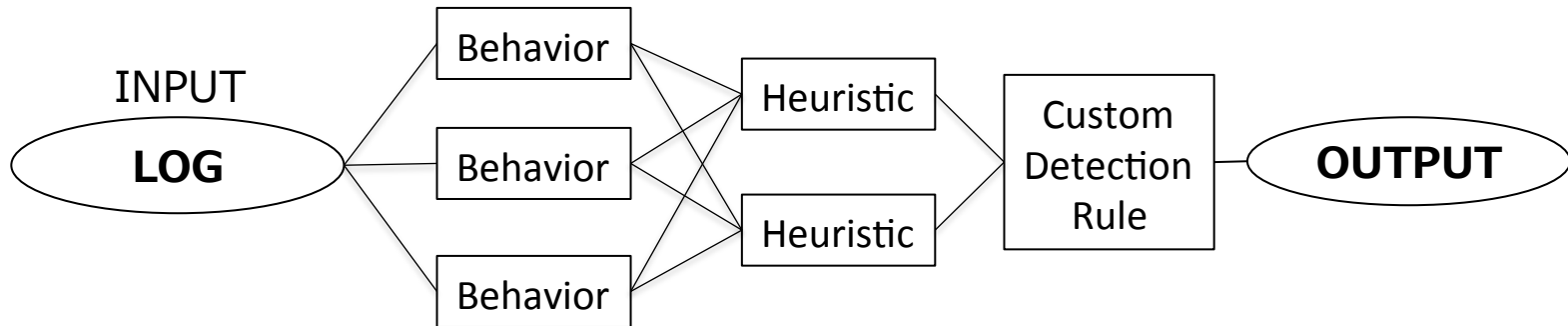
splunk>

Index	source	Source type	Key-Value Pair	By
Recruit	Apache	Apache	URL,HttpStatus, etc..	SrcIP
	Tomcat	Tom_Act	All of Manipulations,etc..	SrcIP,BrowserHash
	Auth-DB	DB_Login_Act	Results, RecruitID, UniqID	
Summary	⌘ All of the Statistics from above			

& Lookup

iplocation / dnslookup / Tor-exitnode (update per day with crontab)/InternalAPI

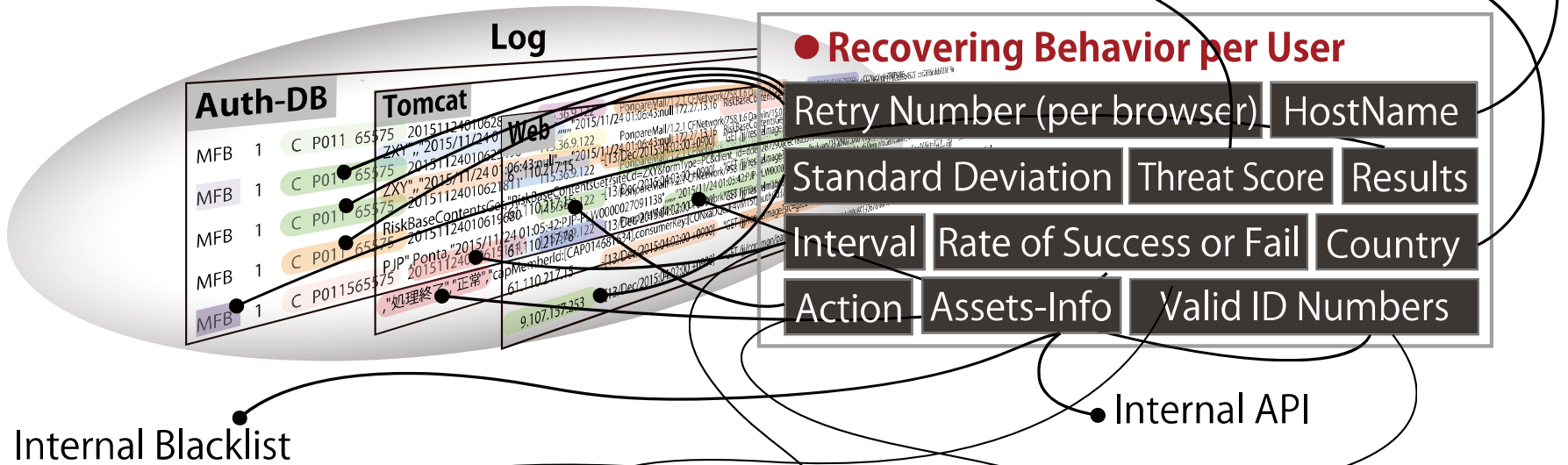
▼ SPL (Heuristic)



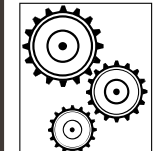
①Building BaseQuery
(Summary Index)

Whois,TorExitNode,WebAPI,etc

Reverse DNS
Country



②Heuristic Analysis



AnomaryData + Statistics + Feature + StandardDeviation + Score

Heuristic Method using Pre and After the Point of Incident

Sample Base Query

Sample BaseQuery (Building 5min ~ Hourly Summary Index into index=summary label=base)

index=Recruit sourcetype=DB_Login_ACT | fillnull UniqID

- ① |streamstats window=0 current=f last(UniqID) as NextUniqID
last(_time) as epochNextTime by SrcIP,BrowserHash **Change-Point Detection**
- ② | fillnull NextUniqID |search NextUniqID!=0 | where NextUniqID!=UniqID **Threshold**
- ③ |eval Interval=epochNextTime-_time **Intervals**
- ④ |stats count as Total stdev(Interval) as StDev count(eval(AuthCode=1)) as Success
min(_time) as min _time max(_time) as max _time by SrcIP,BrowserHash **Statistics**

Another BaseQuery (Building 5min ~ Hourly Summary Index into index=summary label=base_si)

index=Recruit sourcetype=DB_Login_ACT_APF | fillnull UniqID

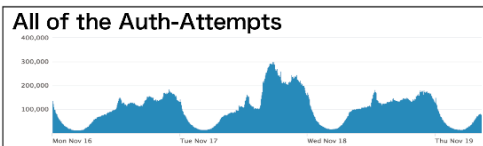
- ⑥ |streamstats window=0 current=f last(UniqID) as NextUniqID by SrcIP,BrowserHash
|fillnull NextUniqID |search NextUniqID!=0 | where NextUniqID!=UniqID **make Macro as `Anomaly`**
- ⑤ |sistats values(UniqID) values(RecruitID) by SrcIP,BrowserHash **Statistics & for Merge**

Sample Heuristic Analysis

Sample Hourly or Daily Heuristic Analysis from BaseQuery (Using Summary Index)

index=summary label=base

- ① |stats count as slot sum(dc_Success) as sum_dc_Success sum(Total) as sum_Total
avg(StDev) as StDev min(min_time) as Start max(max_time) as End by SrcIP,BrowserHash
 - ② |JOIN type=left SrcIP,UA [search index=summary label=base_si
stats dc(UniqID) as ss_dc_UniqID dc(RecruitID) as ss_dc_RecruitID by SrcIP,BrowserHash
fields SrcIP,UA, ss_dc_UniqID,ss_dc_RecruitID]
 - ③ |eval SuccessRate=round(sum_dc_Success/sum_Total*100,2)
|eval ValidIDRate=round(ss_dc_RecruitID/ss_dc_UniqID*100,2)
 - ④ |search ss_dc_RecruitID>=10 | iplocation SrcIP | lookup dnslookup SrcIP
 - ⑤ |table Start End SrcIP clienthost Country slot *Rate sum_* StDev ss_* |sort-SuccessRate
- Feature Extraction
- More Statistics
- As you like..
- Output



100%

Anomaly Extraction by Heuristic Analysis



0.0001%

100 Browsers

Single Values of Top Attacker

いつから	いつまで	どこから
00:00		Japan
ID Numbers		Success Rate
1,505		99.3%
How much IDs	Success Num	Fail Num
		Success Rate

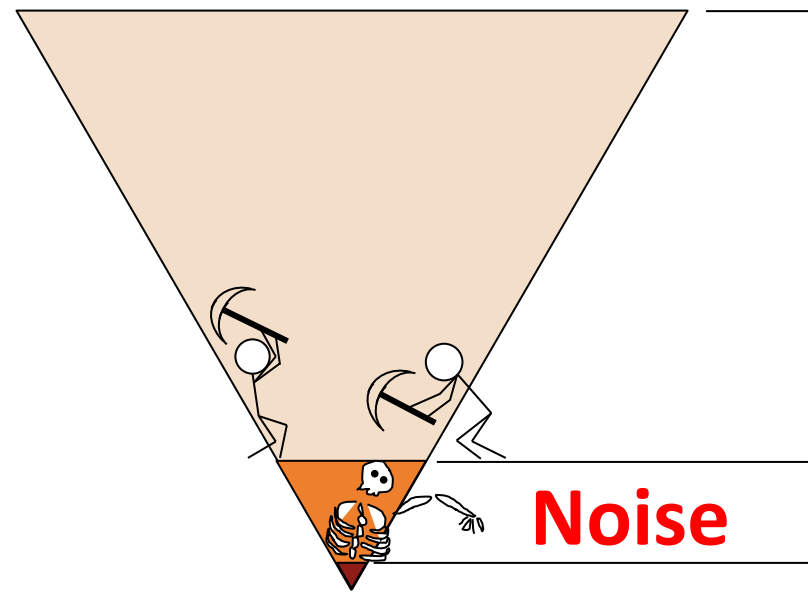
No Way!
Kidding me?



Splunk Helps Identify Cyber Crimes & Cheats.

Mystery Data follow-up survey

Using SPL.



Treasure
(Account hacking)

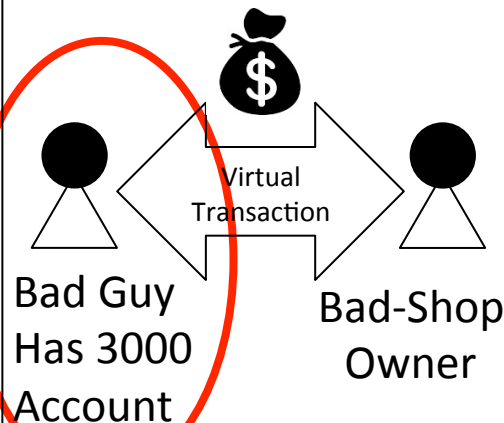
The One is..



1st Party Frauds

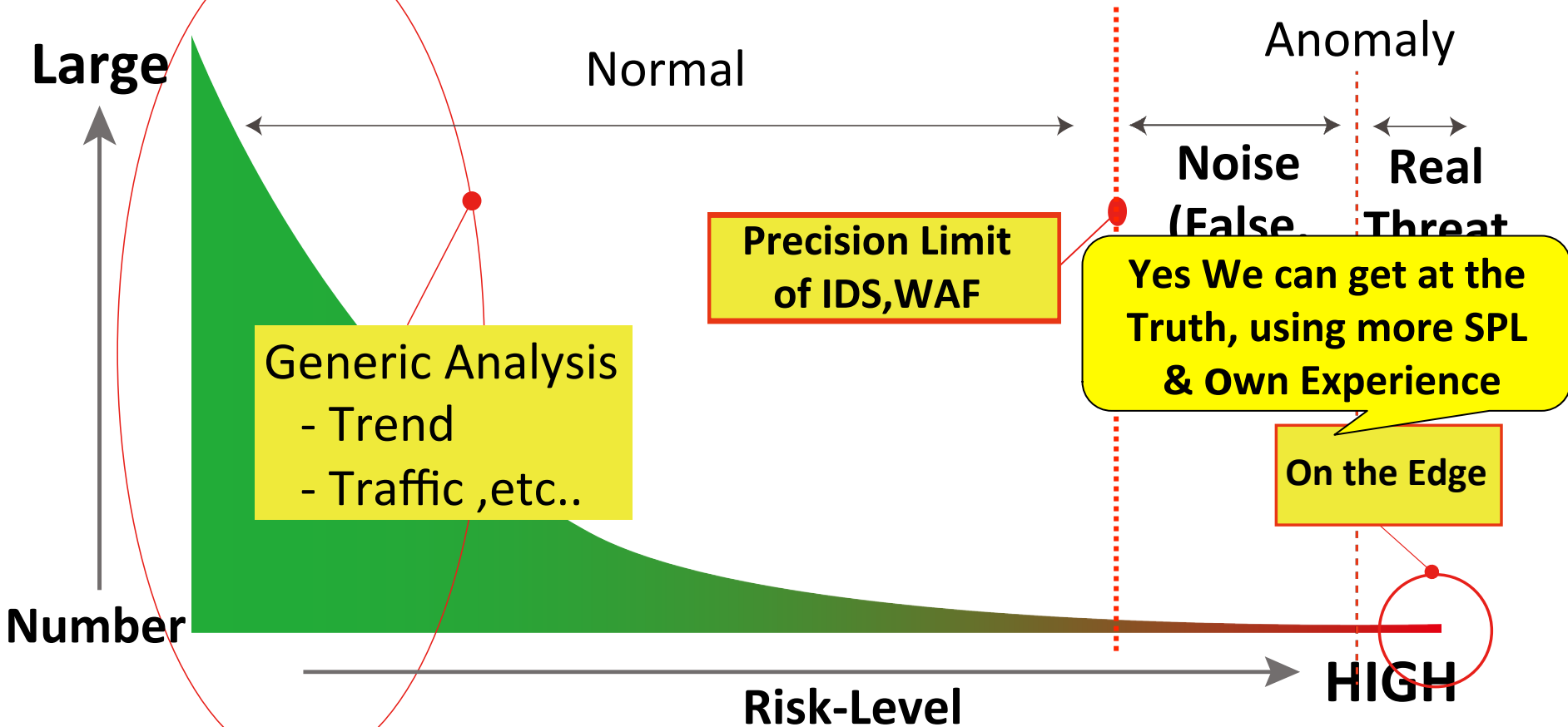
The Other is..

Illegal Transaction in
Our Commercial Site



Cyber Crime

Threats Remain On The Edge Of The Long Tail



Real-threat Extraction With Customdetectionrule

Sample Threat Extraction from AnomalyBaseQuery (Using Summary Index)

① `index=summary label=AnomalyBase NOT [search index=Recruit sourcetype=Apache URL="/pointGive/doComp/" [search `anomaly` ] | streamstats last(_time) as NextTime by SrcIP | eval Interval=NextTime-_time|search Interval<=10 |fields SrcIP]`

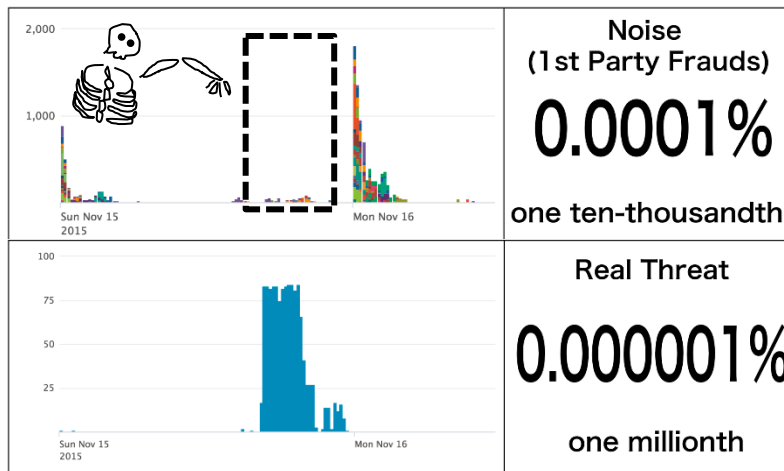
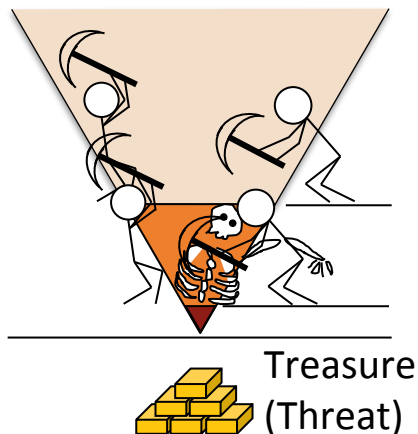
Sift out NOISE
using Custom-Rule,
and throw away

② `|JOIN SrcIP,BrowserHash [search index=Recruit sourcetype=Tom_Act [search `anomaly` ] |stats count(eval(action="RegID" AND method="done")) as ct_ID_reg by SrcIP,BrowserHash |fields SrcIP,BrowserHash,ct_ID_reg]`

Feature Extraction
for specify Threat

③ `|search StDev<=0.9 ct_ID_reg>=10 | timechart span=10m count by SrcIP`

Final Output



Registrations of 5000+ IDs
by one person.

Another Observed Threat.

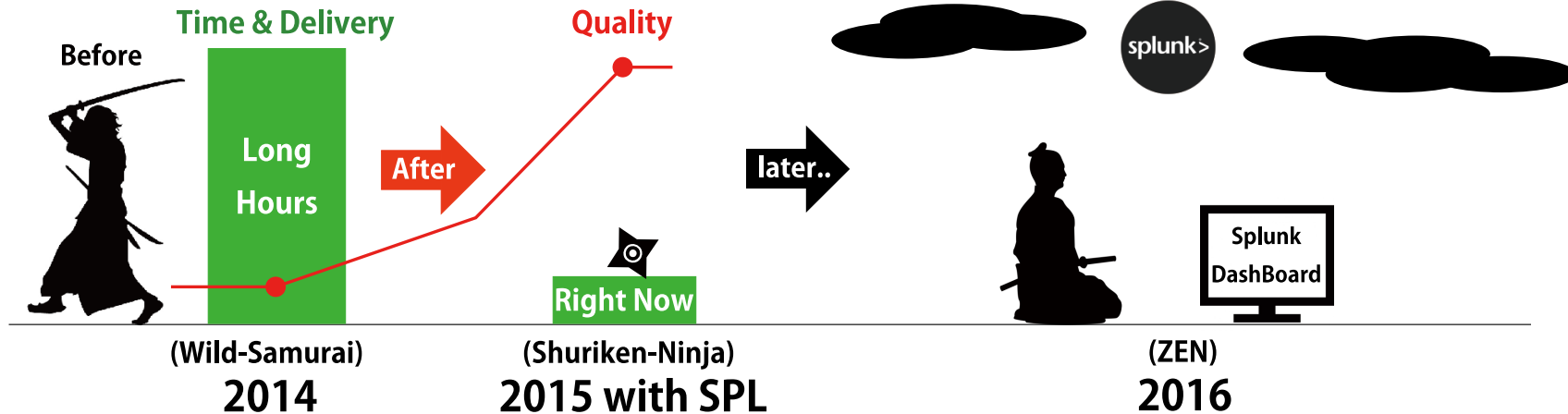
- List-Type Account Hacking.
 - The Act of removing Dummy ID's.
 - Online Virtual-Coin Laundering.
- etc...

Weekly Results (Nov/2015)

SPL			527	NOISE
	12	REAL THREAT	1st Party Fraud	
WAF	5	ALERT (but False..)	WAF didn't catch the Fraud!	

Their DeepAnalysis has changed dramatically.

They can judge things calmly..



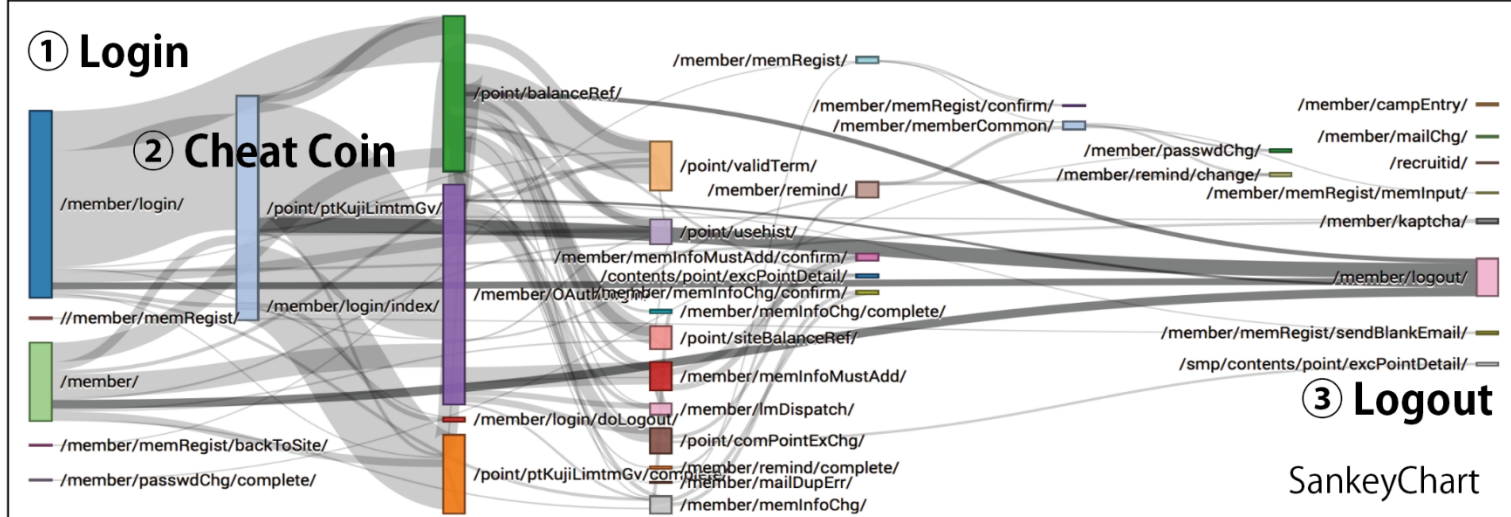
Challenge 2: Data Visualization Using d3.js

.conf2016

splunk>

Data-Driven Documents

The LINKs of 1st Party Frauds. They make sure choose shortest Path.



D3.js is a JavaScript library for manipulating documents based on data. **D3** helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

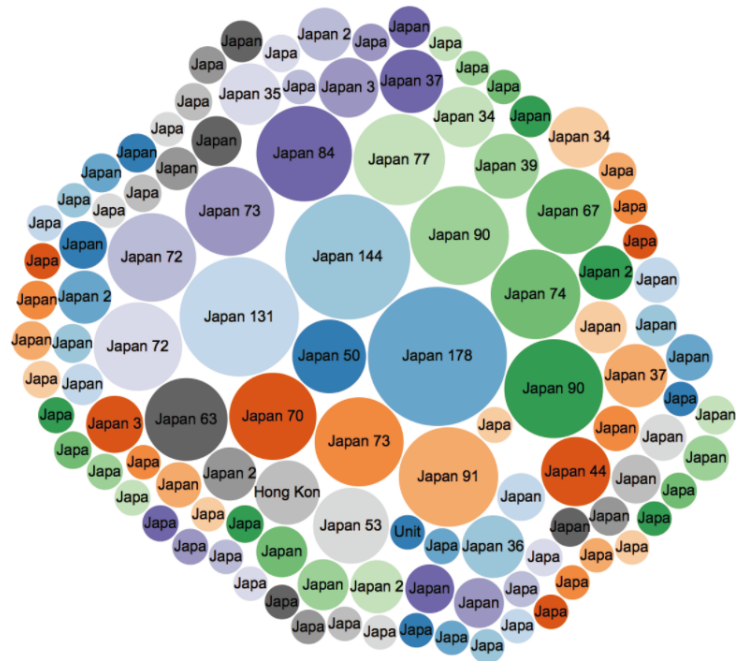
See [more examples](#).

<https://d3js.org/>

Data-visualization Of 1st Party Frauds (Noise)

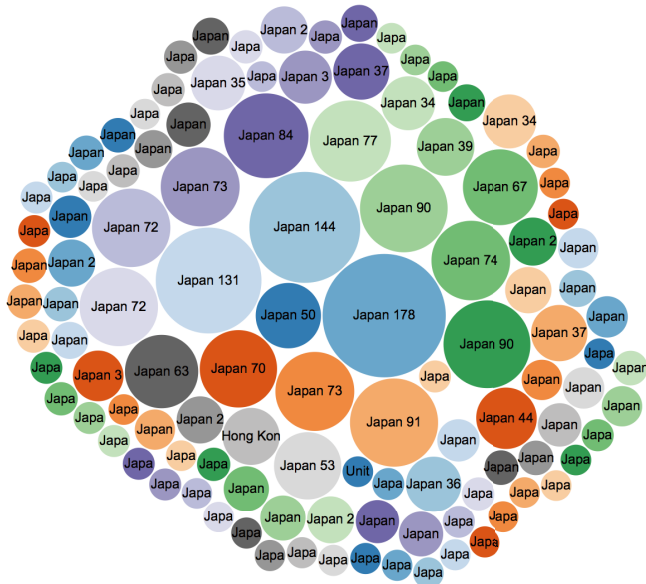
同一ブラウザから異常な数のIDでログイン成功率高い人(ス)											
成功率	SrcIP	clienthost	Country	Sum	Success	失敗	ロック	Black	開始日時	終了日時	
99.3%	211.243		Japan	5362	5323	39	0	0	04/15 00:00	04/22 02:59	
98.9%	231.160		Japan	4849	4794	55	0	0	04/15 00:00	04/22 04:46	
80.0%	154.238		Japan	2420	1935	485	0	0	04/15 00:00	04/22 00:51	
96.7%	138.214		Japan	1827	1766	59	2	0	04/15 10:08	04/22 02:08	
99.7%	93		Japan	1746	1741	5	0	0	04/15 00:00	04/22 00:38	
98.3%	11.41		Japan	1728	1698	30	0	0	04/15 00:00	04/22 02:31	
99.6%	167		Japan	1439	1433	6	0	0	04/15 00:00	04/22 00:52	
99.0%	107.12		Japan	1263	1250	13	0	0	04/15 20:20	04/22 00:41	
99.7%	58.33		Japan	1127	1124	3	0	0	04/15 00:00	04/22 00:20	
67.4%	85.134		Japan	1437	968	469	0	0	04/17 00:00	04/17 01:41	

Table

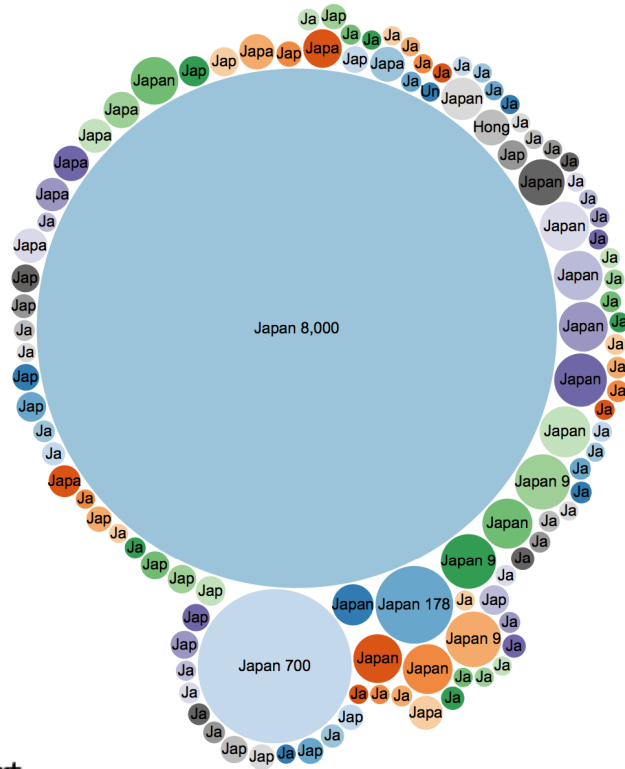


BubbleChart

Normal (Including Noise)



List-Type Account Hacking



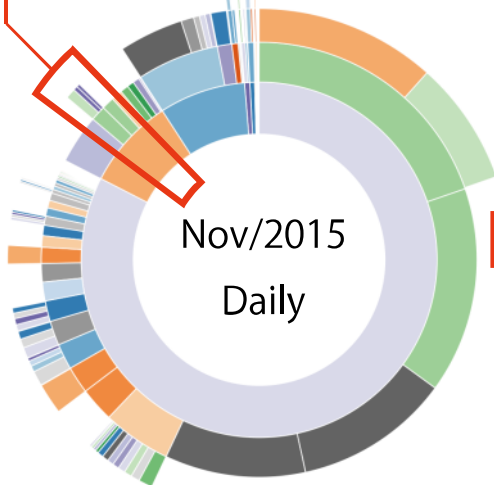
BubbleChart

Visible All from the Above

100%



Noise Extraction

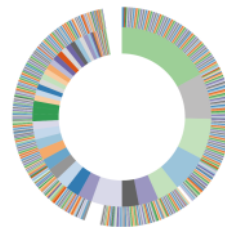


Sunburst

Noise Extraction

0.0001%

Identify
1st-Party'

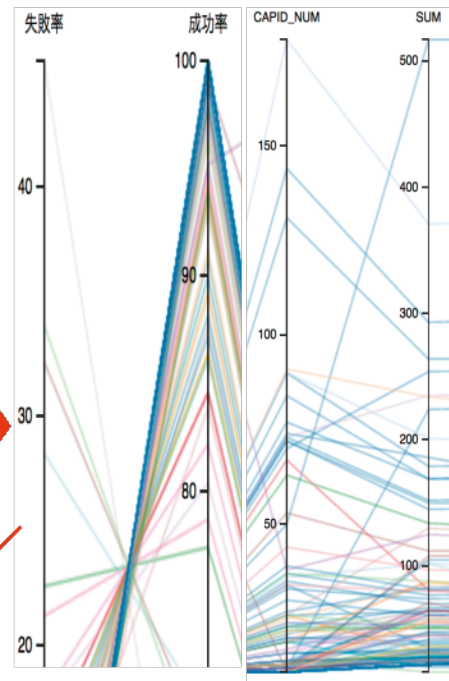


threat
Extraction

BubbleChart

Real Threat

0.000001%



Parallel Coordinates

Case Studies

.conf2016

splunk>

Identify and Visualize Virtual Currency Threat

Visualize Black-N



■ CAPOTR	965214
■ CAPOTR	327720
■ CAPOTR	173462
■ CAPOTR	507065
■ CAPOTR	482095
■ CAPOTR	466992
■ CAPOTR	466958
■ CAPOTR	994277
■ CAPOTR	788685
■ CAPOTR	788521
■ CAPOTR	788497
■ CAPOTR	788460
■ CAPOTR	788426
■ CAPOTR	788396
■ CAPOTR	788365
■ CAPOTR	788320
■ CAPOTR	788282
■ CAPOTR	788193
■ CAPOTR	788000
■ CAPOTR	966650

56000614

n Laundering.

and Many Other-Site ID's.

HUB ID

84866700214

SankeyChart

Two Dashboards

For extract Threat-Seeds

For extract Real Threat



For Future IR

For Rapid IR

Wrap Up

.conf2016

splunk>

Wrap Up

1. Recovering Behavior from Fragment Data.

Using SummaryIndex, using Stats, StreamStats, and Sstats

2. Classify Anomaly Data as to NOISE and THREAT by Heuristic Analysis.

Extraction Accuracy had reached a millionth

3. Stuff who responding Incident in Recruit make rapid progress.

Splunk defeat WAF in the Recruit!!

4. Splunk is not only for Rapid-IR, but also for Future-IR

5. Data-Visualization has infinite possibilities



Extra Challenge Deep Learning

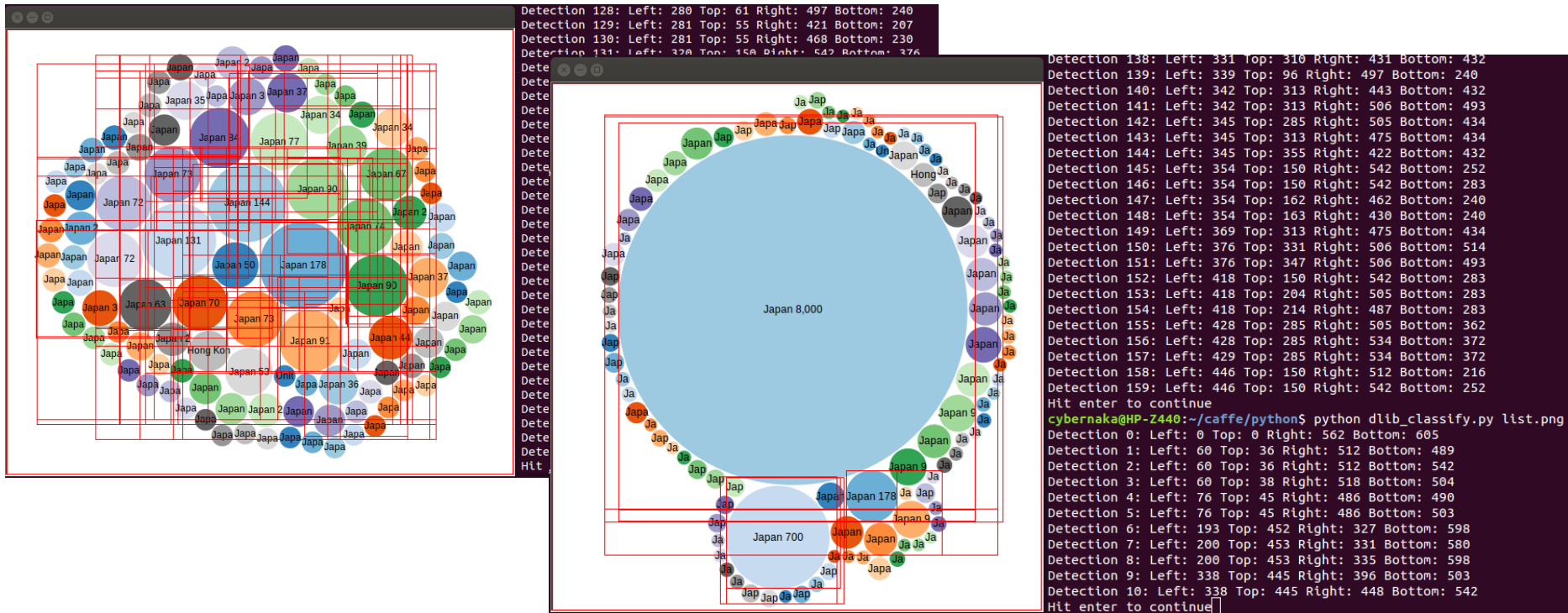
d3.js-Image Recognition

.conf2016

splunk>

Threat Detection With Deeplearning (Without Training-data)

Object Recognition using “Dlib” & ” Scikit-Image”



Threat Detection With Deeplearning (With Training-data)

Input : d3.js Images

CNN



▽ Technologies

- CUDA
- Caffe
- SelectiveSearch
- PyData, etc..

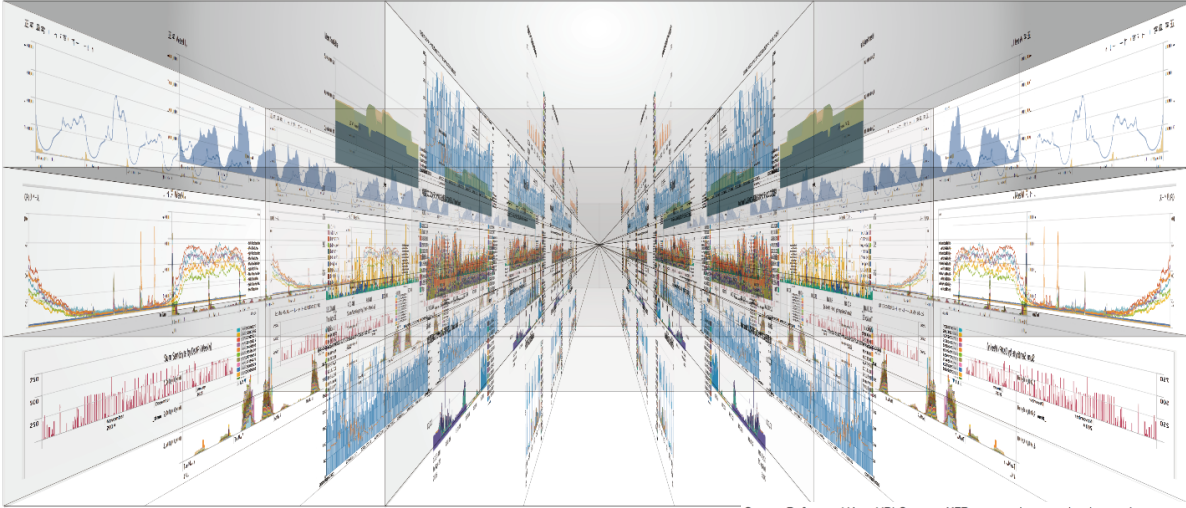
▽ Who gains benefit ?

24/7 SOC



Output : Whether Man or Machine, Anyone can see Attack-Image is Threat !

```
/home/cybernaka/images/normal_011.png noise [ 1.35602495e-05 9.99962211e-01 2.41894377e-05]
/home/cybernaka/images/normal_010.png normal [ 2.91483525e-06 9.99987364e-01 9.69005850e-06]
/home/cybernaka/images/attack_003.png attack [ 9.99996305e-01 2.66407096e-06 1.04353694e-06]
/home/cybernaka/images/noise_013.png noise [ 2.74476929e-06 9.99974489e-01 2.27220935e-05]
/home/cybernaka/images/attack_004.png attack [ 9.99952674e-01 2.53877643e-05 2.18860405e-05]
/home/cybernaka/images/normal_009.png normal [ 1.18293997e-06 9.99989390e-01 9.38913399e-06]
/home/cybernaka/images/noise_015.png noise [ 9.11432835e-06 6.66544220e-05 9.99924302e-01]
/home/cybernaka/images/noise_016.png noise [ 4.90193088e-07 9.99991179e-01 8.27781787e-06]
/home/cybernaka/images/noise_006.png noise [ 8.99720419e-07 9.99989390e-01 9.72105954e-06]
/home/cybernaka/images/noise_005.png noise [ 1.20140055e-05 9.99982933e-01 4.99830730e-06]
/home/cybernaka/images/normal_007.png normal [ 2.85860751e-05 9.99961257e-01 1.01217247e-05]
/home/cybernaka/images/normal_010.png normal [ 2.49328377e-06 4.35543464e-05 9.99954104e-01]
/home/cybernaka/images/normal_010.png noise [ 3.07662481e-06 9.99965429e-01 3.15485049e-05]
/home/cybernaka/images/normal_008.png normal [ 1.45854501e-05 5.93026634e-05 9.99926090e-01]
cybernaka@HP-Z440:~/caffe/python$
```

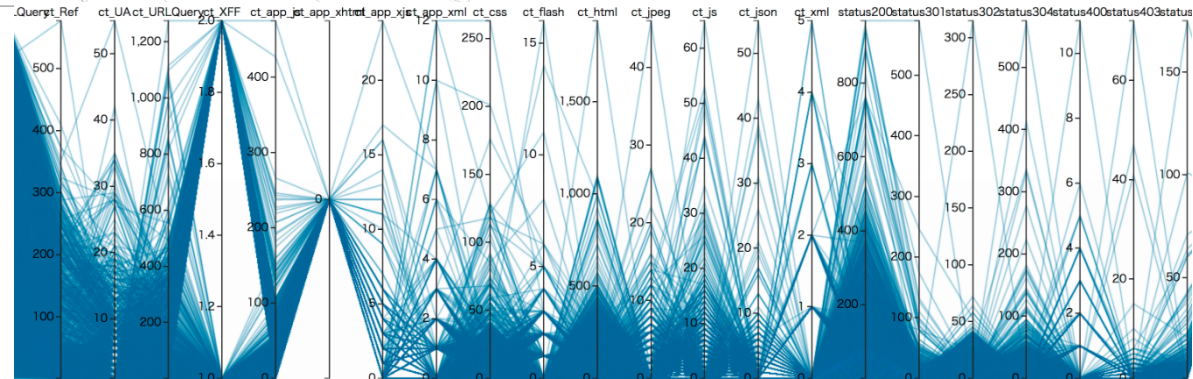


Experts & Long Time



Deep Learning
and Training-Data

Data-Visualization



Parallel Coordinates

A visual toolkit for multidimensional detectives.

<https://github.com/syntagmatic/parallel-coordinates>

TotalDur...	HttpMet...	TOP_URL...	ct_Ref	ct_UA	ct_URLQ...	ct_XFF	ct_app_j...	ct_app_...	ct_app_...	ct_app_...	ct_css	ct_flash	ct_html	ct_jpeg	ct_js	ct_json	ct_xml	status20...
159	GET	/	1	1	68	1	0	0	0	0	0	0	68	0	0	0	0	1
506	GET HEAD	/secure/...	1	1	23	1	0	0	0	0	0	0	24	0	0	0	0	1
-7326	GET	/wp-con...	1	1	39	1	0	0	0	0	0	0	44	0	0	0	0	2

THANK YOU

Special Thanks: Mifune-san, Splunk-Japan, and Macnica

hiro:<nakamura@r.recruit.co.jp>

Facebook : <https://www.facebook.com/hiro.goahead>

.conf2016