



# ***IPSERVER ONE***<sup>®</sup>

**NetFlow Data Analytics with ELK Stack**

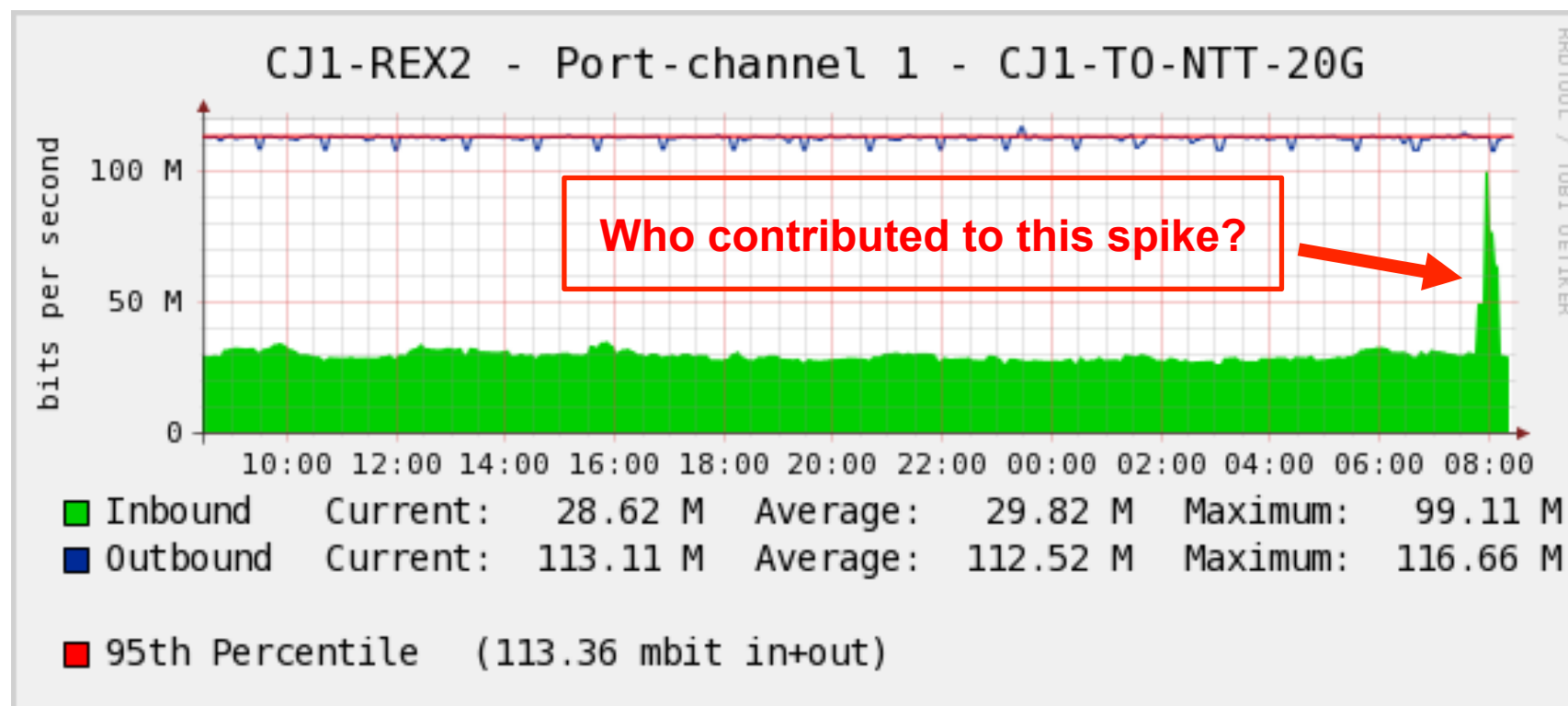
# About IP ServerOne

- Founded in **2003**
- Over **60** employees
- Managing over **5000 physical servers**
- Total **250** racks at **5** data centers across  
**MY, SG and HK**
- Contributing **10%** of Malaysia's  
**domestic traffic**
- Approximately **6.8 Gbit/s** total traffic  
sending to the Internet at peak
- **Up to 1.2TB** DDoS mitigation capacity



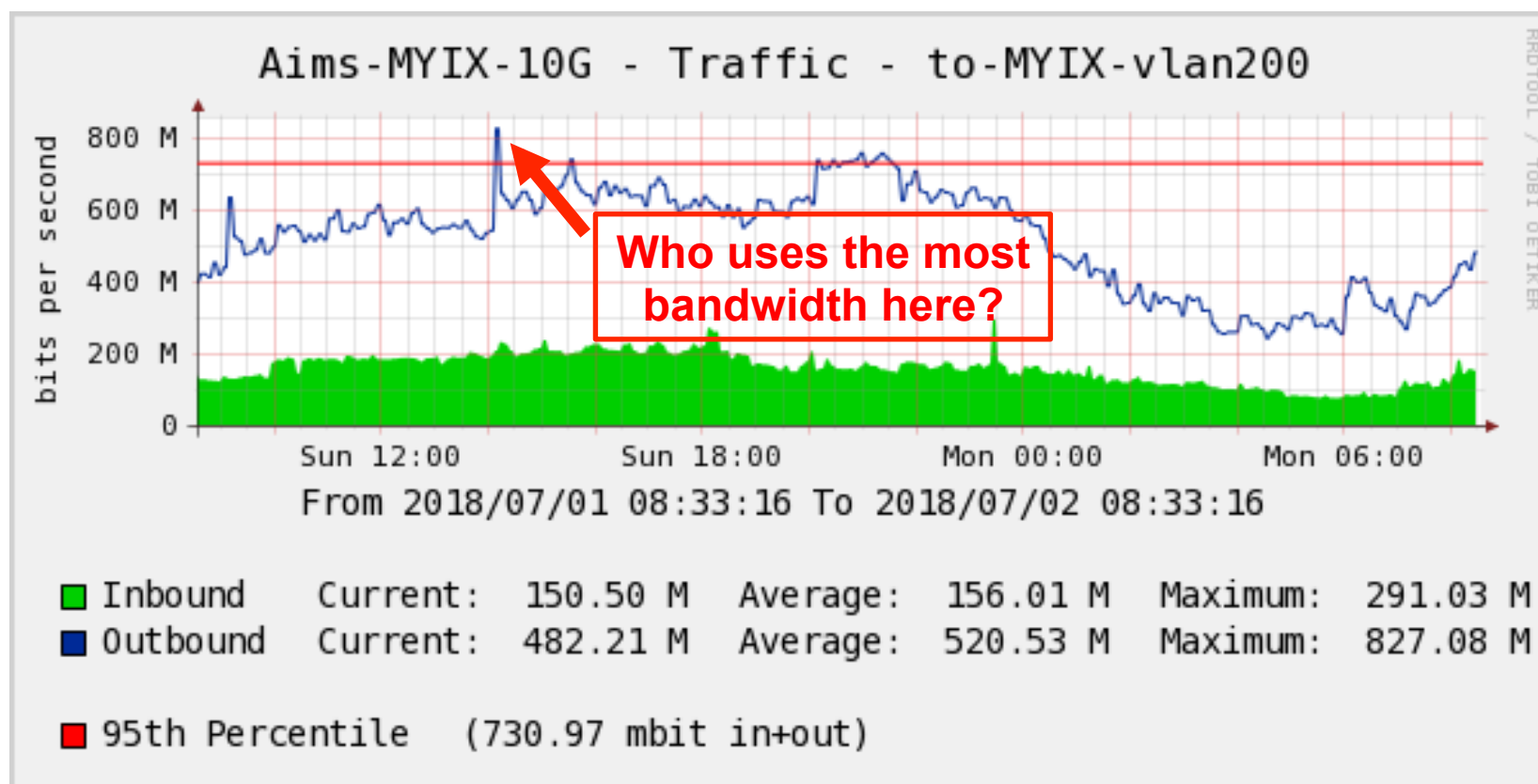
Why do we need to use  
**NetFlow?**

# Most companies have their MRTG configured



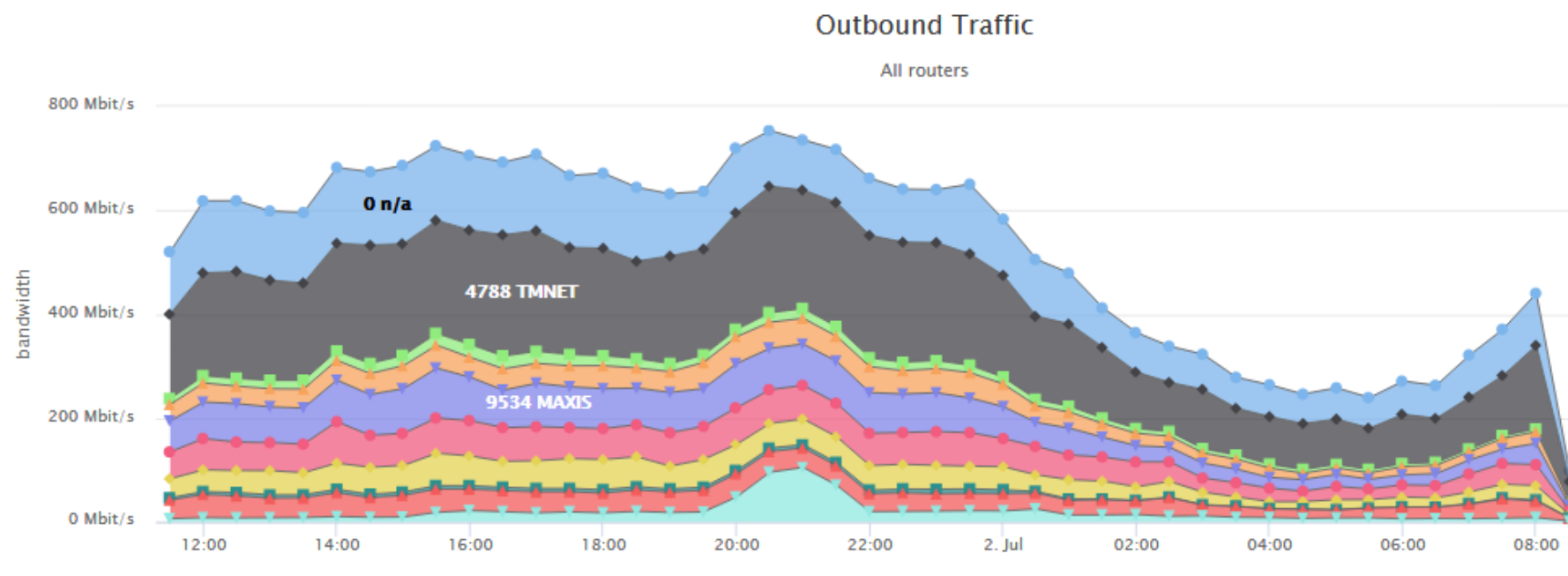
But MRTG cannot tell you which IP is receiving a spike traffic (such as the above graph)

You probably may need to know where the majority of your traffic comes from, right?

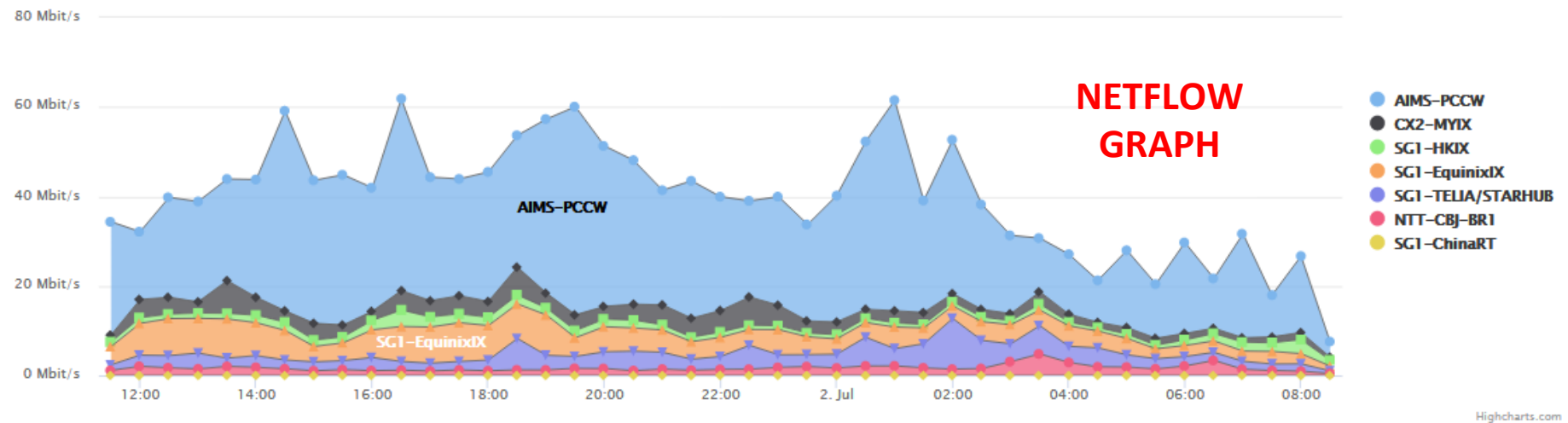
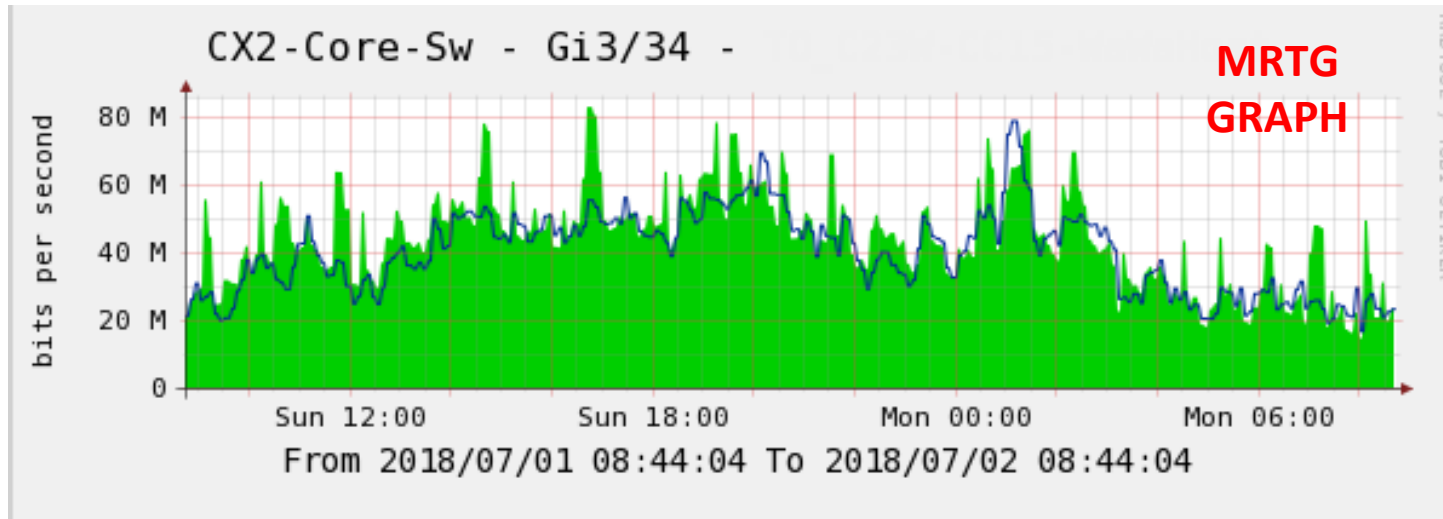


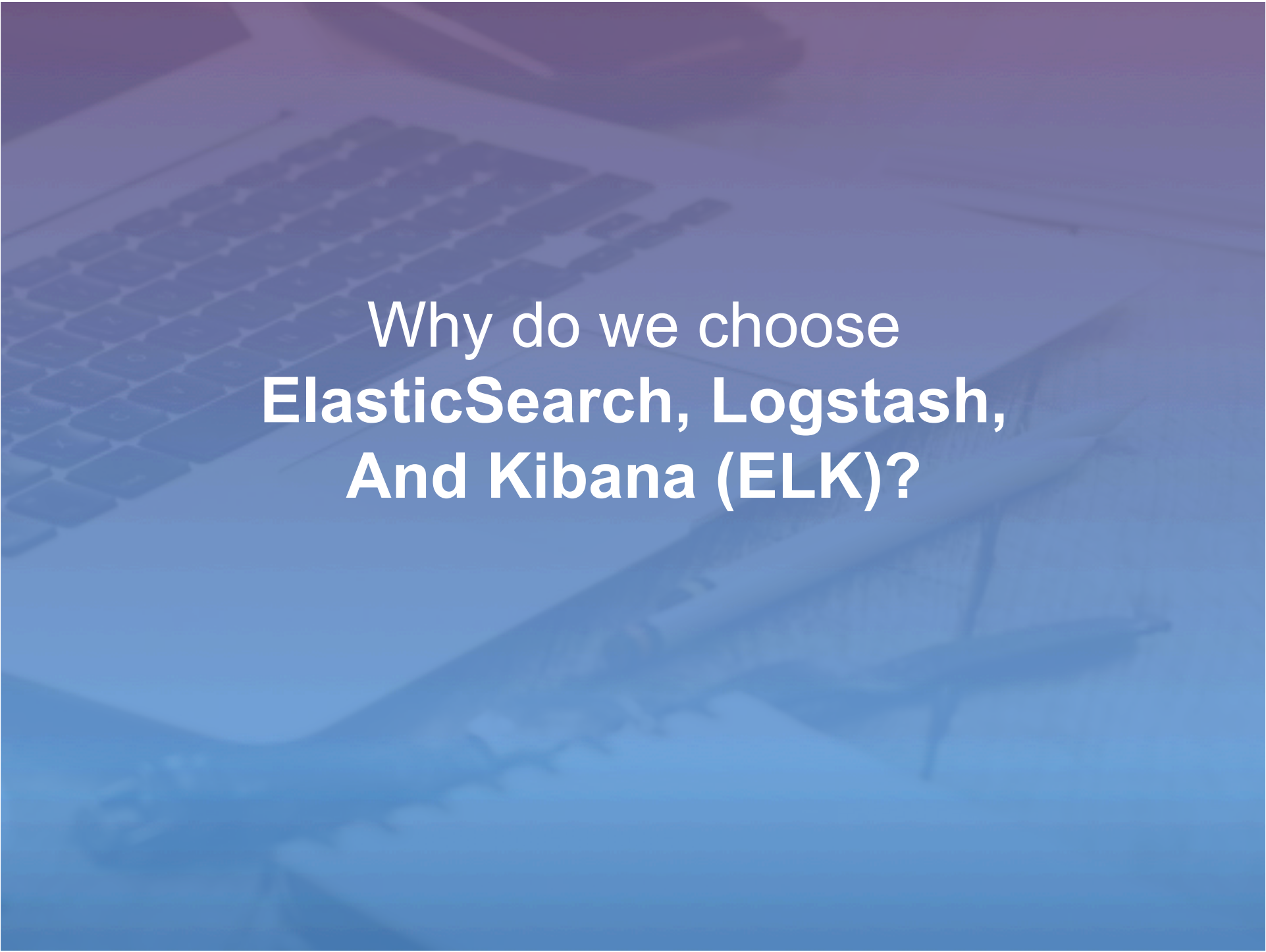


# A NetFlow graph would be able to breakdown the usage for your outbound / inbound traffic



# Replacing your MRTG with a NetFlow graph



The background of the slide is a blurred image of a laptop keyboard and a document with a diagram. The keyboard is on the left, and the document with a diagram is on the right. The text is centered over the image.

Why do we choose  
**ElasticSearch, Logstash,  
And Kibana (ELK)?**



## Why ELK?

- Before I get to know ELK stack, I was using MySQL to store all the NetFlow information.
- I wrote a PHP application that converts NetFlow information into a MySQL statement.
- That was too slow on the conversion performance and the data retrieval was a complete nightmare.
- There is no function / feature to get traffic statistic in the histogram form.

**It's just too difficult to run this in MySQL**

## Why ELK?

- **Speed is the primary reason that I have chosen ELK**
- **It has a lot of codec, which I can just plug and play**
- **COST; it runs on commodity hardware and it works just fine with Nearline SAS Hard drives**
- **Open Source**
- **Support Clustering**
- **It has SQL like syntax, so data searching is much more easier**
- **It has a very high performance; we had a working environment of 100Kflows per second**

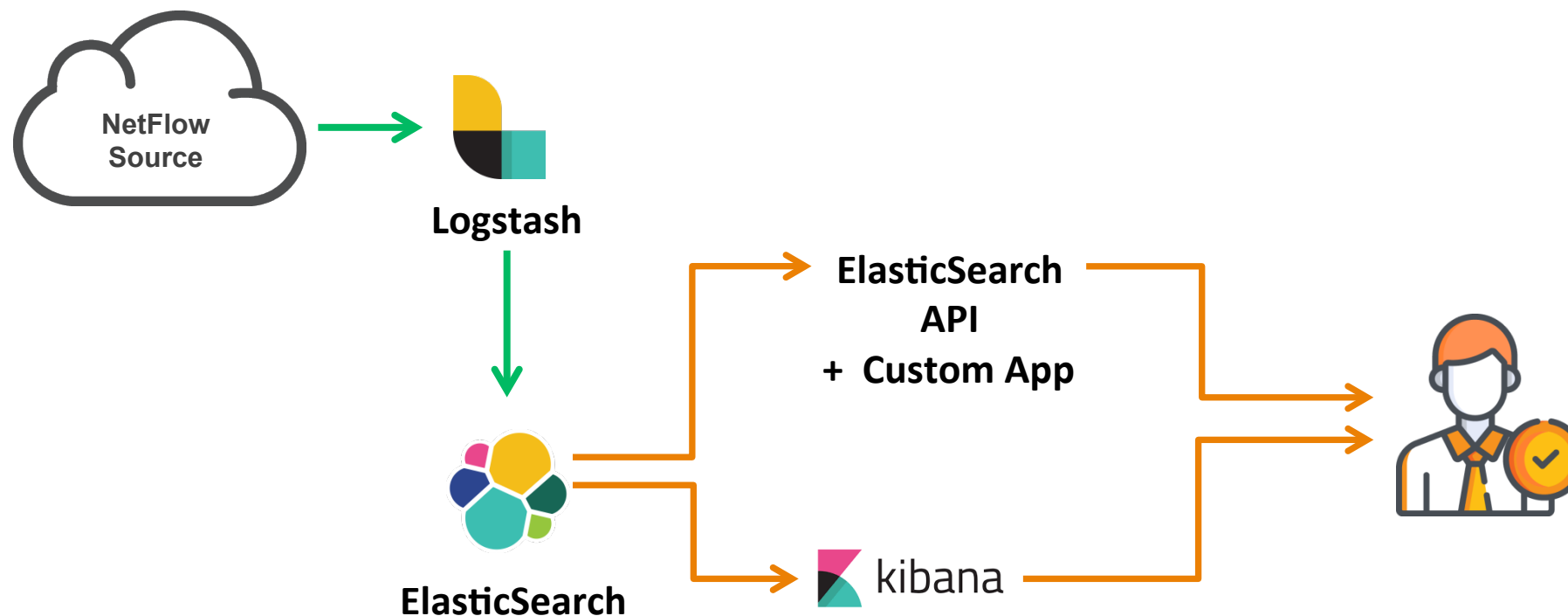
## Alternative to ELK

- **We did consider to use InfluxDB**  
The OpenSource edition doesn't support clustering.
- **OpenTSDB**  
The setup is very time-consuming.
- **MongoDB.**  
This is a great DB; however, we still prefer to use ElasticSearch.



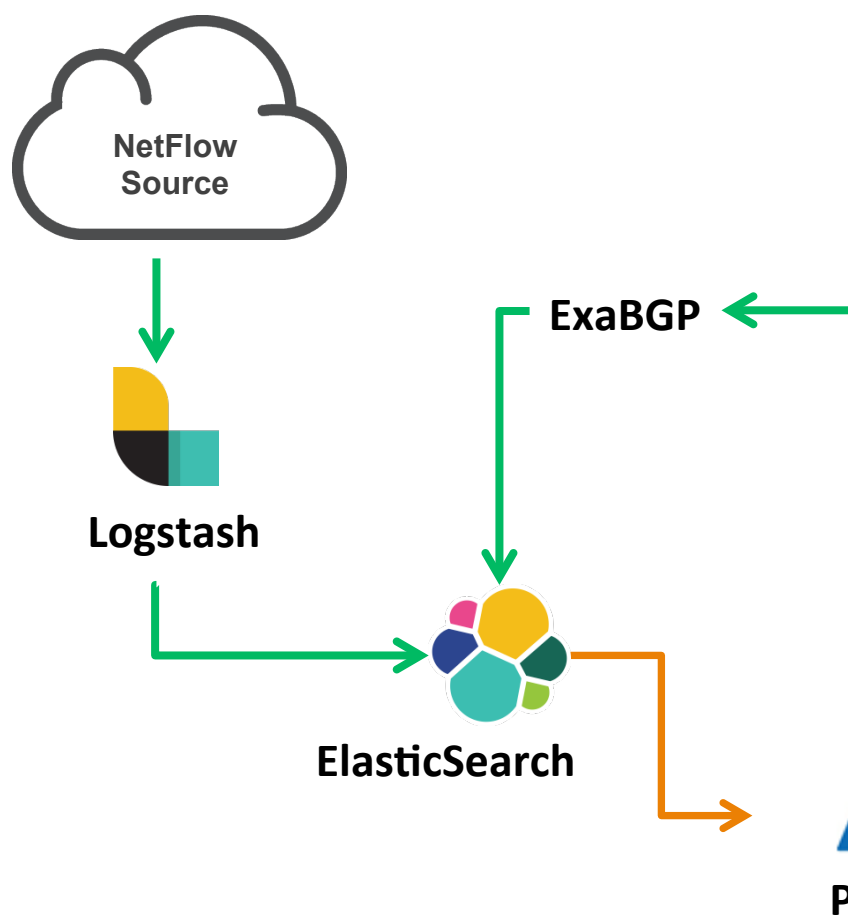
# How to record the **NetFlow Data?**

# The NetFlow is being collected with the following setup





# Adding BGP table information into the ElasticSearch



## BGP Routing Table

```

BGP routing table entry for 103.3.174.0/24, version
737937
Paths: (34 available, best #21, table default)
  Not advertised to any peer
  Refresh Epoch 1
3356 3491 45352
    4.69.184.193 from 4.69.184.193 (4.69.184.193)
      Origin IGP, metric 0, localpref 100, valid,
external
      Community: 3356:666 3356:2012 3491:400 3491:413
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
3549 3356 2914 45352
    208.51.134.254 from 208.51.134.254 (67.16.168.191)
      Origin IGP, metric 0, localpref 100, valid,
external
      Community: 3356:3 3356:86 3356:575
3356:666 3356:2011 3356:11940 3549:2581 3549:30840
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
20912 1267 45352
    212.66.96.126 from 212.66.96.126 (212.66.96.126)
      Origin incomplete, localpref 100, valid, external
      Community: 1267:167 1267:200 20912:65001
      rx pathid: 0, tx pathid: 0
  Refresh Epoch 1
route-views>
  
```

**INi**  
PHP + Golang



# We use NetFlow v9 in our projects

## *Here is the field that we keep*

Field Type	Value	Length (bytes)	Description
IN_BYTES	1	N (default is 4)	Incoming counter with length N x 8 bits for number of bytes associated with an IP Flow.
IN_PKTS	2	N (default is 4)	Incoming counter with length N x 8 bits for the number of packets associated with an IP Flow
FLows	3	N	Number of flows that were aggregated; default for N is 4
PROTOCOL	4	1	IP protocol byte
SRC_TOS	5	1	Type of Service byte setting when entering incoming interface
TCP_FLAGS	6	1	Cumulative of all the TCP flags seen for this flow
L4_SRC_PORT	7	2	TCP/UDP source port number i.e.: FTP, Telnet, or equivalent
IPv4_SRC_ADDR	8	4	IPv4 source address
SRC_MASK	9	1	The number of contiguous bits in the source address subnet mask i.e.: the submask in slash notation

INPUT_SNMP	10	N	Input interface index; default for N is 2 but higher values could be used
L4_DST_PORT	11	2	TCP/UDP destination port number i.e.: FTP, Telnet, or equivalent
IPv4_DST_ADDR	12	4	IPv4 destination address
DST_MASK	13	1	The number of contiguous bits in the destination address subnet mask i.e.: the submask in slash notation
OUTPUT_SNMP	14	N	Output interface index; default for N is 2 but higher values could be used
IPv4_NEXT_HOP	15	4	IPv4 address of next-hop router
SRC_AS	16	N (default is 2)	Source BGP autonomous system number where N could be 2 or 4
DST_AS	17	N (default is 2)	Destination BGP autonomous system number where N could be 2 or 4
BGP_IPv4_NEXT_HOP	18	4	Next-hop router's IP in the BGP domain

# Hardware vs Software

The **hardware specification**  
used for keeping our NetFlow



**1 x Intel Xeon 8 cores**  
**2.1Ghz Processor**



**4 x 2TB HDD**



**32GB RAM**



**1 x Gigabit**  
**Network Card**

The **software** used to  
run our NetFlow



**CentOS 7**  
64bit Operating System



**Java**



**MySQL**



**PHP**



**ElasticSearch, Logstash**

## How to put up the software?

### CentOS Installation

You can follow the way you do normally; but please remember to keep most of the free space into /var.



# CentOS

# ElasticSearch Installation

**ElasticSearch** is a search engine based on Lucene. It provides a distributed architecture, support multi-tenancy and full-text search engine with an HTTP web interface.

## Installing from the RPM repository



Create a file called `elasticsearch.repo` in the `/etc/yum.repos.d/` directory for RedHat based distributions, or in the `/etc/zypp/repos.d/` directory for OpenSuSE based distributions, containing:

```
[elasticsearch-6.x]
name=Elasticsearch repository for 6.x packages
baseurl=https://artifacts.elastic.co/packages/6.x/yum
gpgcheck=1
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch
enabled=1
autorefresh=1
type=rpm-md
```

And your repository is ready for use. You can now install Elasticsearch with one of the following commands:

```
sudo yum install elasticsearch ⓘ
```



# Start Elasticsearch

```
[root@elk-stack ~]# systemctl daemon-reload
[root@elk-stack ~]# systemctl start elasticsearch
[root@elk-stack ~]# systemctl enable elasticsearch
```

To check what are the indexes available in the Elasticsearch:

```
[root@elk-stack ~]# curl -XGET 'http://localhost:9200/_cat/indices?v'
```

health	status	index	uuid	pri	rep	docs.count	docs.deleted	store.size	pri.store.size
yellow	open	stat-20180603	byH89tWFQSS_R9kS_QPGPw	5	1	54822544	0	6.9gb	6.9gb
yellow	open	stat-20180616	qZYSua4CQDa18GGMc8uiHQ	5	1	51830338	0	6.6gb	6.6gb
yellow	open	stat-20180604	PYdGUxX7SZ2aaFRV-ng4NQ	5	1	57828976	0	7.3gb	7.3gb
yellow	open	stat-20180630	FwrBuf6FQ-6SlyZhknATLQ	5	1	50014372	0	6.4gb	6.4gb
yellow	open	stat-20180618	_Nloca3jROCQ2vChWmDoGw	5	1	54976264	0	7gb	7gb
yellow	open	stat-20180526	ObGvcFbfTDuuk_MtZN1CQA	5	1	51836183	0	6.6gb	6.6gb
yellow	open	stat-20180615	t_CxQoauRUiVRTaJRPz2eQ	5	1	55490519	0	7gb	7gb

# Logstash Installation

**Logstash** is one of the softwares inside the ELK stack. The main objective for this software is to convert NetFlow data into ElasticSearch acceptable format.

## YUM

Download and install the public signing key:



```
rpm --import https://artifacts.elastic.co/GPG-KEY-elasticsearch
```

Add the following in your `/etc/yum.repos.d/` directory in a file with a `.repo` suffix, for example

`logstash.repo`

```
[logstash-6.x]
name=Elastic repository for 6.x packages
baseurl=https://artifacts.elastic.co/packages/6.x/yum
gpgcheck=1
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch
enabled=1
autorefresh=1
type=rpm-md
```

And your repository is ready for use. You can install it with:

```
sudo yum install logstash
```

# Configure Logstash to decode NetFlow

```
LS_HOME/bin/logstash-plugin install logstash-codec-sflow
LS_HOME/bin/logstash-plugin update logstash-codec-netflow
LS_HOME/bin/logstash-plugin update logstash-input-udp
LS_HOME/bin/logstash-plugin update logstash-filter-dns
```

## Create a netflow.conf /etc/logstash/

```
input {
  udp {
    port => 2055
    codec => netflow
  }
}

output {
  elasticsearch {
    protocol => "http"
    host => "127.0.0.1"
  }
  stdout { codec => rubydebug }
}
```

# Kibana Installation

**Kibana** is one of the GUI tools that helps retrieve data from Elasticsearch. It can also come with the graphing capability to manipulate the Doc in Elasticsearch to be something more meaningful to system engineers.

## Installing from the RPM repository



Create a file called `kibana.repo` in the `/etc/yum.repos.d/` directory for RedHat based distributions, or in the `/etc/zypp/repos.d/` directory for OpenSUSE based distributions, containing:

```
[kibana-6.x]
name=Kibana repository for 6.x packages
baseurl=https://artifacts.elastic.co/packages/6.x/yum
gpgcheck=1
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch
enabled=1
autorefresh=1
type=rpm-md
```

And your repository is ready for use. You can now install Kibana with one of the following commands:

```
sudo yum install kibana ❶
```

# Kibana Configuration

**Kibana** does not listen to any IP besides 127.0.0.1; you will need to update the configuration file to make the Kibana accessible from outside the host.

**vi /etc/kibana/kibana.yml**

```

root@l3-ini-ips1:/etc/kibana
# Kibana is served by a back end server. This setting specifies the port to use.
#server.port: 5601

# Specifies the address to which the Kibana server will bind. IP addresses and host names are both valid values.
# The default is 'localhost', which usually means remote machines will not be able to connect.
# To allow connections from remote users, set this parameter to a non-loopback address.
server.host: 0.0.0.0
# Enables you to specify a path to mount Kibana at if you are running behind a proxy. This only affects
# the URLs generated by Kibana, your proxy is expected to remove the basePath value before forwarding requests
# to Kibana. This setting cannot end in a slash.
#server.basePath: ""

# The maximum payload size in bytes For incoming server requests.
#server.maxPayloadBytes: 1048576

# The Kibana server's name. This is used for display purposes.
#server.name: "your-hostname"

# The URL of the Elasticsearch instance to use for all your queries.
#elasticsearch.url: "http://localhost:9200"

# When this setting's value is true Kibana uses the hostname specified in the server.host
# setting. When the value of this setting is false, Kibana uses the hostname of the host
# that connects to this Kibana instance.
#elasticsearch.preserveHost: true

# Kibana uses an index in Elasticsearch to store saved searches, visualizations and
# dashboards. Kibana creates a new index if the index doesn't already exist.
#kibana.index: ".kibana"

# The default application to load.
#kibana.defaultAppId: "home"

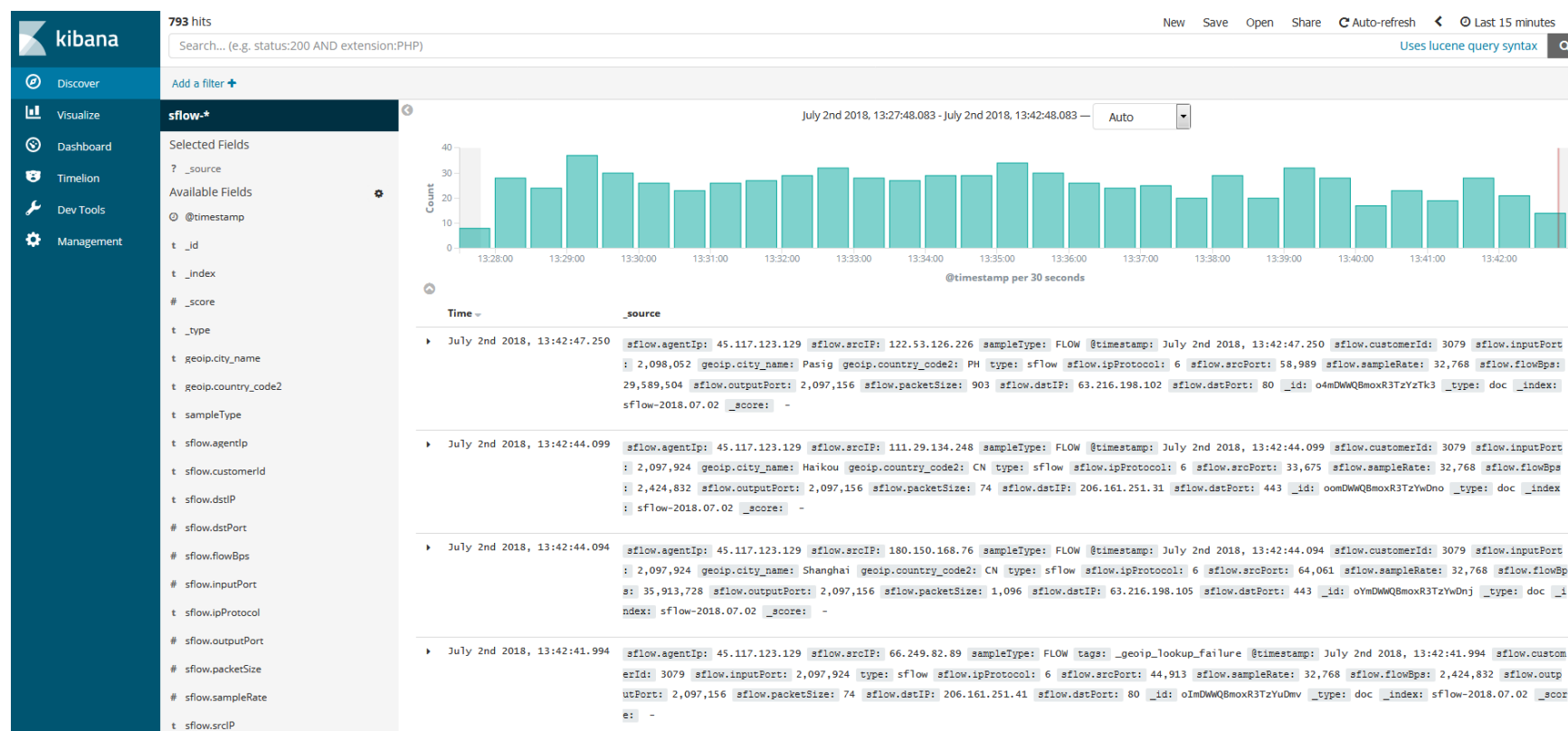
# If your Elasticsearch is protected with basic authentication, these settings provide
# the username and password that the Kibana server uses to perform maintenance on the Kibana

```



# A quick look on the data stored in ElasticSearch

If the data is successfully collected by Logstash, this is what will be shown in Kibana:



The background of the slide is a blurred image of a laptop keyboard and a document. The document appears to contain a diagram or flowchart, but the details are not clear due to the blur. The overall color scheme is a gradient of blue and purple.

# How to query Elasticsearch for top 10 IP talkers?

# ElasticSearch has it's own Query Language called Query DSL

Here is a sample query command for the IP range 103.64.13.0/24 at the specific time period. **(formatted in epoch milliseconds)**

The screenshot shows the Kibana Dev Tools console. On the left, the 'Console' tab is active, displaying a Query DSL query. The query is a GET request to the \_search endpoint, with a size of 1. It includes a range query for the 'ts' field, with a minimum value of 1532382430693 and a maximum value of 1532388430693, formatted as epoch milliseconds. The query also includes a term query for the 'sa' field with the value '103.64.13.0/24'. The response is displayed on the right, showing the search results in JSON format. The response includes the search status (took: 52, timed\_out: false, shards: {total: 231, successful: 231, skipped: 196, failed: 0}), the total number of hits (30187), and the maximum score (3). The first hit is shown, with fields like \_index, \_type, \_id, \_score, and \_source.

```
1 GET _search
2 {
3   "size": 1,
4   "query": {
5     "bool": {
6       "must": [
7         {
8           "range": {
9             "ts": {
10              "gte": 1532382430693,
11              "lte": 1532388430693,
12              "format": "epoch_millis"
13            }
14          },
15          {
16            "term": {
17              "sa": {
18                "value": "103.64.13.0/24"
19              }
20            }
21          }
22        ],
23        "filter": [
24          {
25            "range": {
26              "dp": {
27                "gte": 0,
28                "lte": 65535
29              }
30            }
31          }
32        ]
33      }
34    }
35  },
36  "aggs": {
37    "router_ip": {
38      "terms": {
39        "field": "ra"
40      },
41      "aggs": {
42        "router_interface": {
43          "terms": {
44            "field": "in"
45          }
46        }
47      }
48    }
49  }
50 }
```

```
1 {
2   "took": 52,
3   "timed_out": false,
4   "_shards": {
5     "total": 231,
6     "successful": 231,
7     "skipped": 196,
8     "failed": 0
9   },
10  "hits": {
11    "total": 30187,
12    "max_score": 3,
13    "hits": [
14      {
15        "_index": "stat-20180724",
16        "_type": "f",
17        "_id": "LVAqyWQBkSWHIMPJyur5",
18        "_score": 3,
19        "_source": {
20          "ts": "1532383261727",
21          "ra": "103.3.172.226",
22          "din": "I",
23          "sa": "103.64.13.104",
24          "sp": "11000",
25          "da": "120.29.117.50",
26          "dp": "55021",
27          "sas": "0",
28          "das": "17639",
29          "pr": "6",
30          "flg": ".A....",
31          "tos": "0",
32          "pps": "1000",
33          "bps": "40000",
34          "in": "49",
35          "out": "47",
36          "smk": "24",
37          "dmk": "23",
38          "nh": "27.111.229.79",
39          "ssn": "103.64.13.0/24",
40          "dsn": "120.29.116.0/23",
41          "asp": "-"
42        }
43      }
44    ],
45    "aggregations": {
46      "router_ip": {
47        "terms": {
48          "field": "ra"
49        },
50        "aggs": {
51          "router_interface": {
52            "terms": {
53              "field": "in"
54            }
55          }
56        }
57      }
58    }
59  }
60 }
```

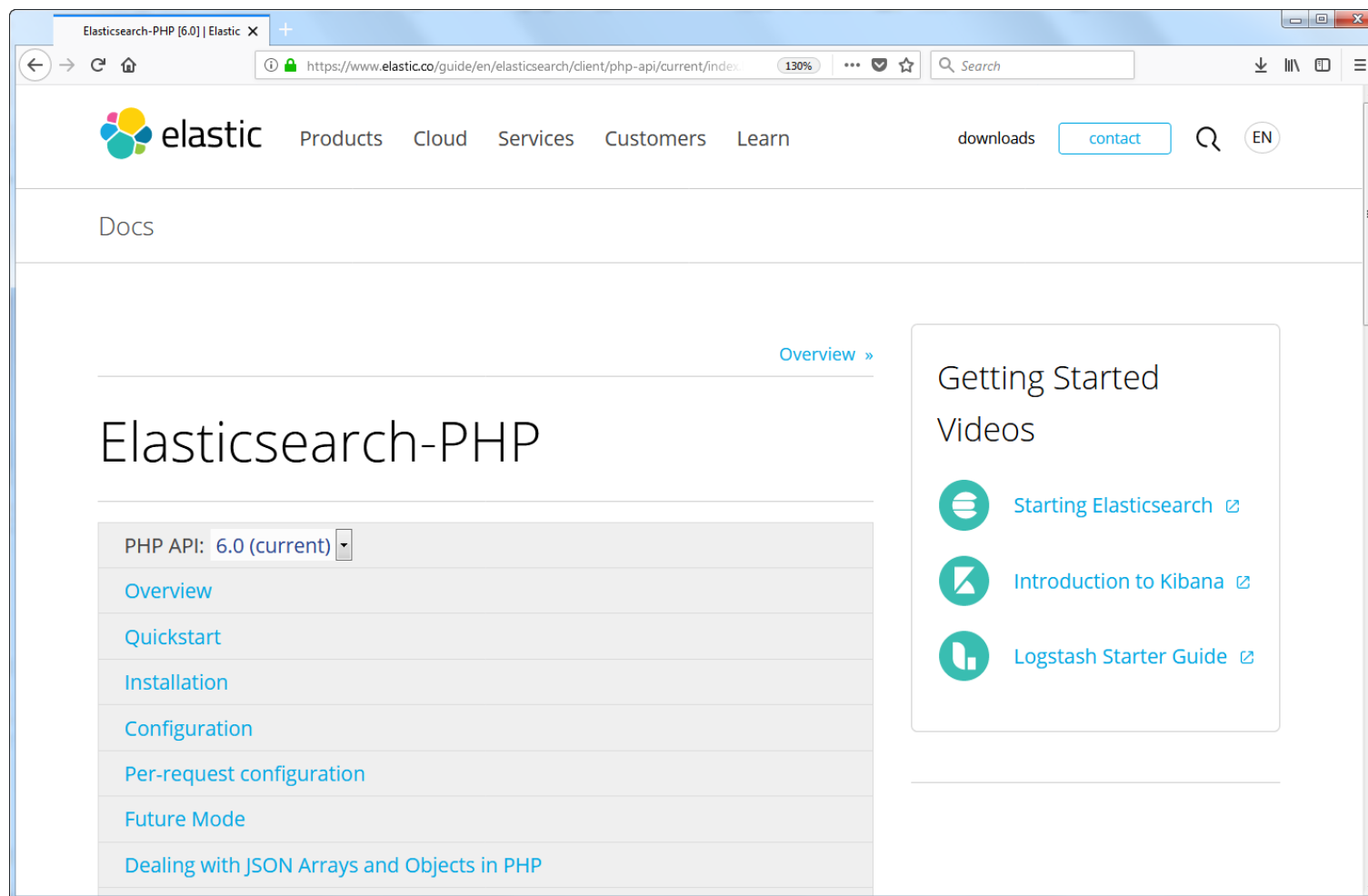
# Kibana is easy to use...

## However, it's still complicated for my NOC team

We make use of Elasticsearch Client API for PHP, to make a query interface so that they can do the job quicker and simplify the learning curve.

# To integrate with PHP, we use Elasticsearch-PHP

It works quite well with our PHP environment





# A Query screen for the NOC engineer

Here is the result of what we have developed, which makes our engineers' life easier



DASHBOARD
TRAFFIC REPORT
DDOS DETECTION
CLOUD DDOS MITIGATION
CUSTOMER REGION
SETTING

Usage By Interface
Usage by ASN
Usage by Prefixes / Subnet
Usage by IP Address
Usage by Conversation

Download

Traffic by Interface

Start:

2018-07-24 05:06:00

End:

2018-07-24 08:06:59

Customer Region:

Router:

Interface:

Source IP:

Source IP

Destination IP:

Destination IP

Source ASN:

Source ASN

Destination ASN:

Destination ASN

Result size:

10

Show Report

Source Port:

80,443

Destination Port:

1024-65535

Resolution:

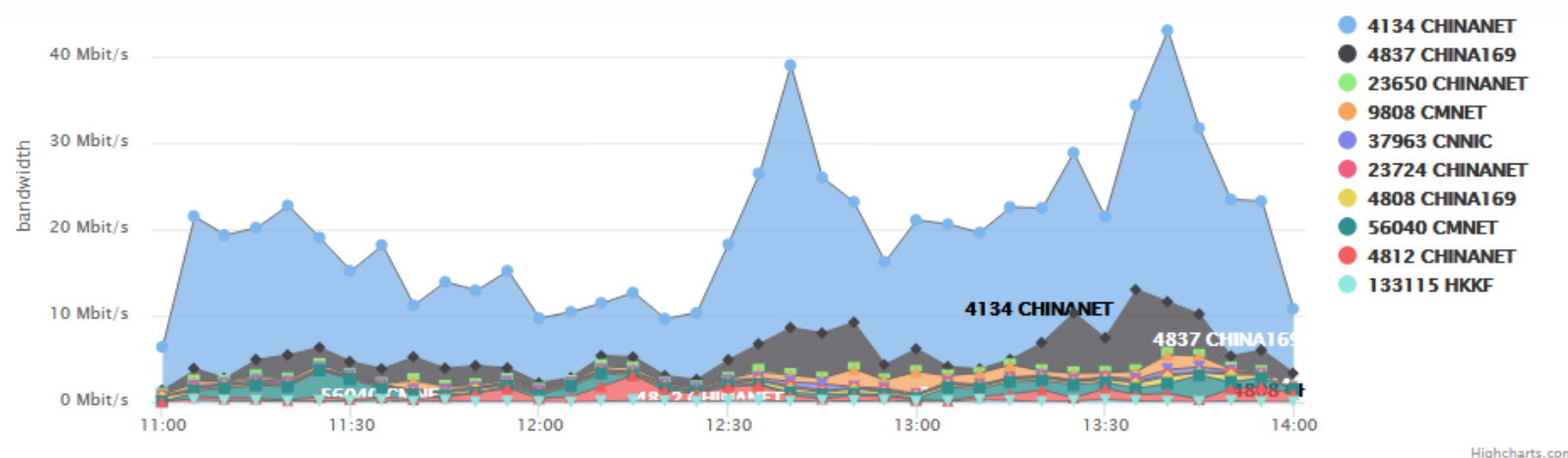
5 Minutes

The background of the slide is a blurred image of a laptop keyboard and a document. The document appears to contain a diagram or flowchart, with lines and shapes visible but out of focus. The overall color scheme is a gradient of blue and purple.

# **Samples on how we use the NetFlow Data**

## Outgoing traffic by ASN and it's AS-PATH

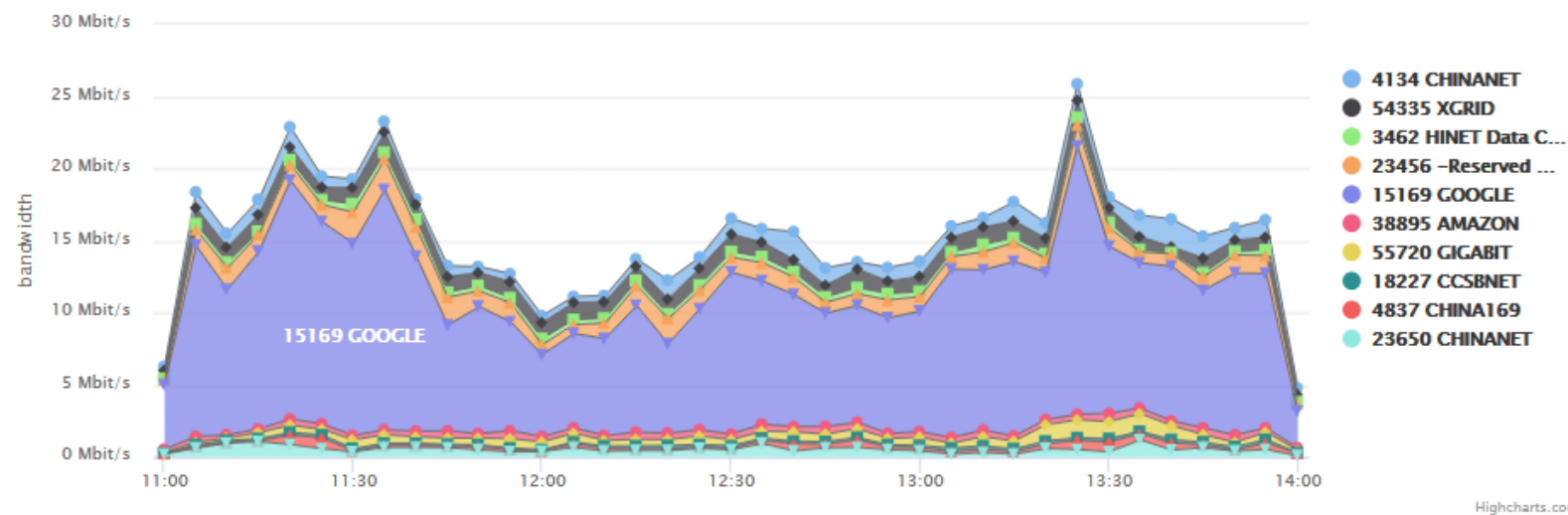
This allows us to know which ASN the traffic flows; and helps us optimize the planning and traffic engineering according to AS Number.



#	ASN	Name	Max	Avg	Min	95%	AS-PATH
1	4134	CHINANET	31.40	14.20	5.01	23.30	3491 4134
2	4837	CHINA169	9.36	2.47	0.28	6.12	2914 1239 4837
3	56040	CMNET	3.09	0.87	0.01	2.38	58453 9808 56040
4	4812	CHINANET	2.93	0.81	0.10	1.80	3491 4809 4812
5	9808	CMNET	2.58	0.62	0.10	1.80	58453 9808
6	4808	CHINA169	0.83	0.30	0.09	0.64	3491 9929 4808

# Incoming traffic by Source ASN

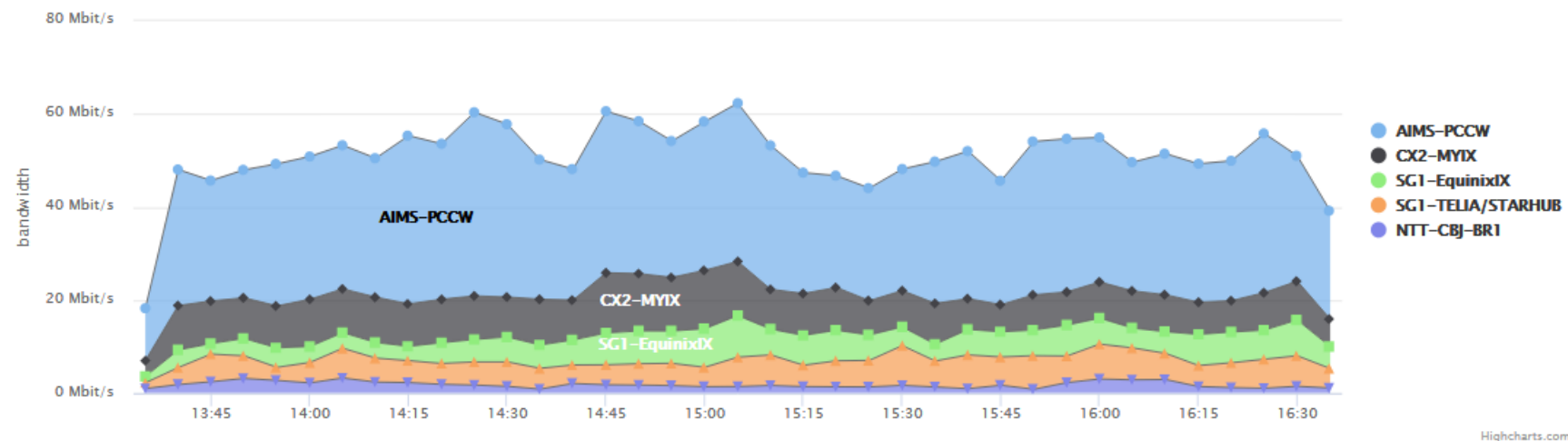
This is also helpful when it comes to traffic engineering



#	ASN	Name	Max	Avg	Min	95%
1	15169	GOOGLE	18.62	10.04	2.54	16.56
2	23456	-Reserved AS-, ZZ	2.17	1.19	0.34	2.04
3	54335	XGRID	1.40	1.00	0.40	1.23
4	4134	CHINANET	1.95	0.92	0.30	1.60
5	55720	GIGABIT	1.24	0.55	0.05	1.19

# Identify customer traffic profile

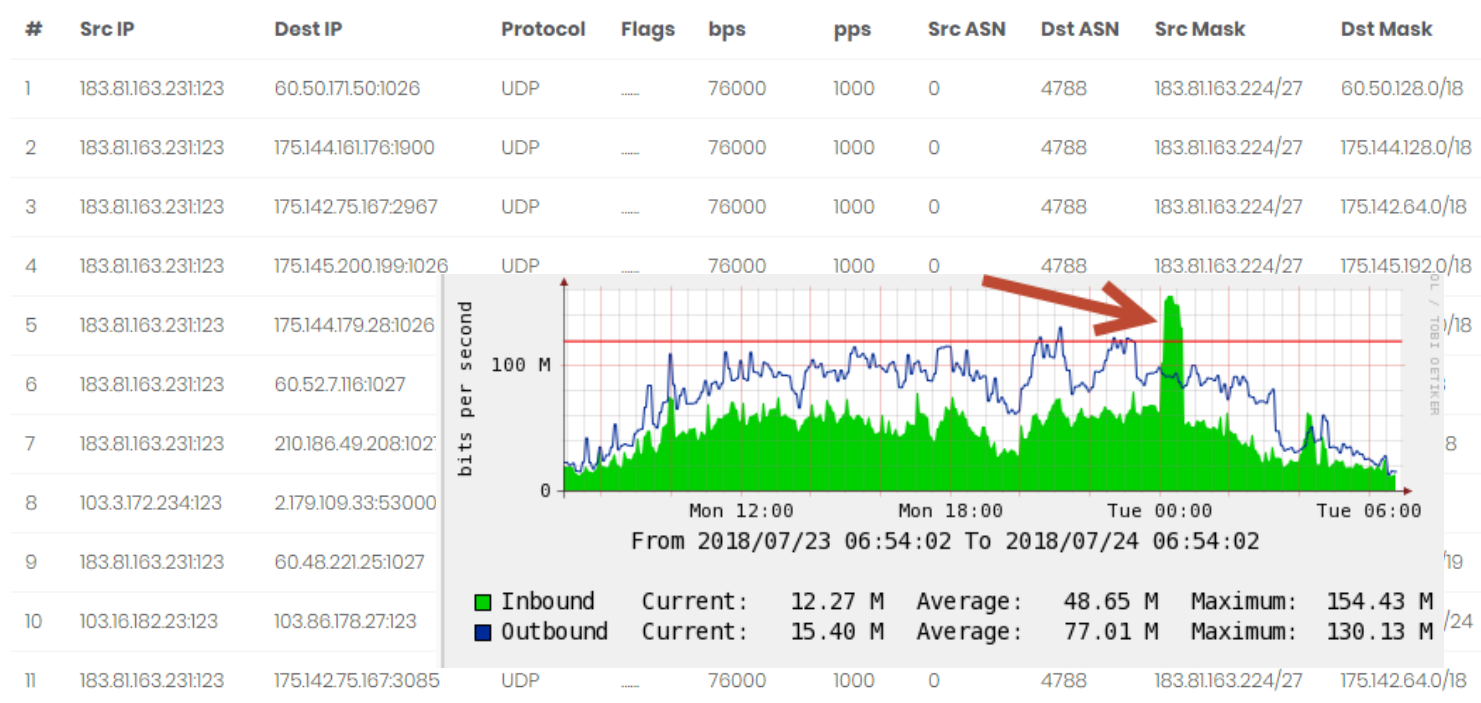
Identify the estimated bandwidth cost for each customer.  
See if the customer traffic utilization is more towards international or local bandwidth.



#	Router IP	Name	Max	Avg	Min	95%
1	210.5.40.21	AIMS-PCCW	39.46	29.79	11.26	36.34
2	103.10.156.73	CX2-MYIX	12.93	8.70	3.38	12.35
3	103.3.172.227	SG1-TELIA/STARHUB	8.54	5.30	1.25	7.32
4	103.3.172.226	SG1-EquinixIX	8.94	5.22	1.38	7.84
5	103.21.181.2	NTT-CBJ-BR1	3.25	1.81	0.83	3.07

# IP Conversation History

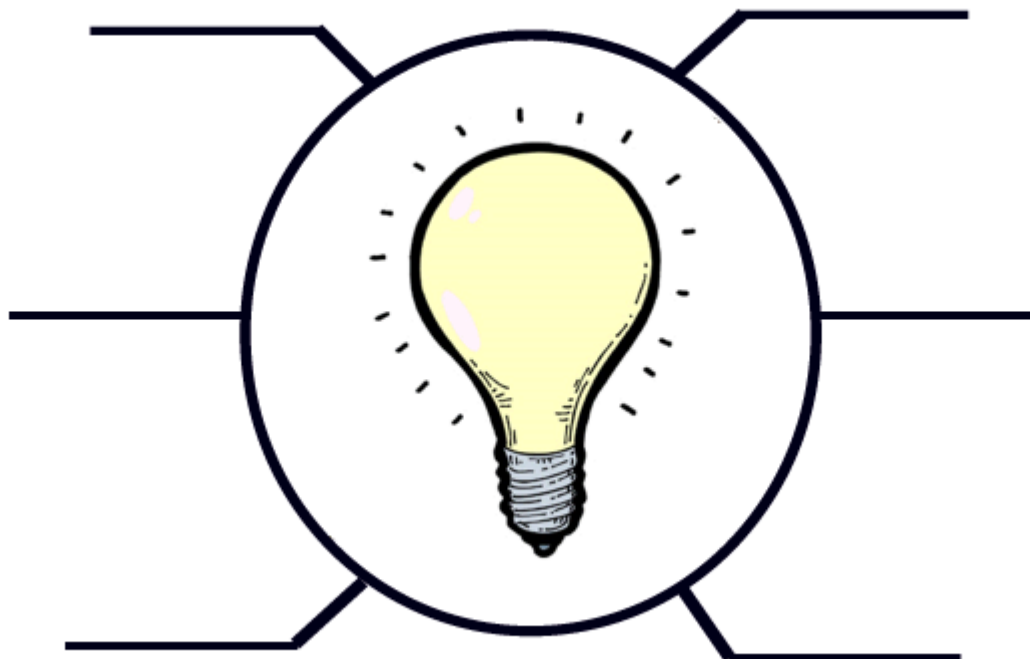
It's something really useful for troubleshooting a network related issue, such as **spamming activity**, **NTP attack** within the network, and ability to **identify the compromised host** quickly.



## Conclusion

ElasticSearch, Logstash and Kibana is a powerful tool to keep and analyze the NetFlow traffic.

In addition, it's not too difficult to deploy and run.





A dark, low-key photograph of a business meeting. In the foreground, a person's hand holds a pen over a document featuring a pie chart and bar graphs. Another person's hand is visible, pointing at the same document. In the background, a laptop is open, and another person's hand is near its keyboard. The overall atmosphere is professional and collaborative.

**ANY  
QUESTIONS?**

# Thanks

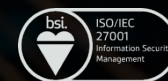
OUR INFRASTRUCTURE; YOUR GROWTH

E-mail: [cllee@ip.my](mailto:cllee@ip.my)

Mobile: +6 012-331 9286

03 2026 1688

[www.ipserverone.com](http://www.ipserverone.com)



ISO Certificate No: IS 651738