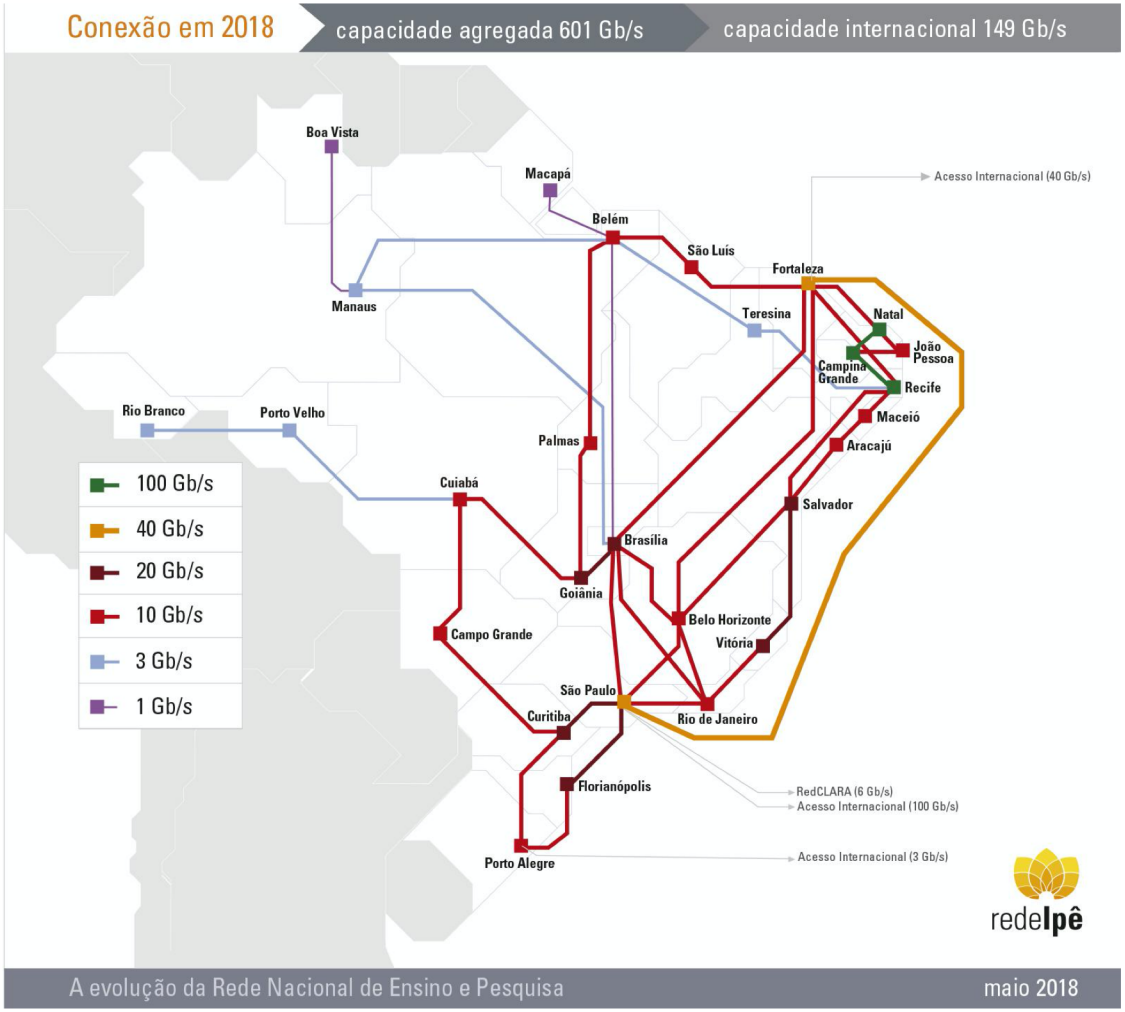
Análise de fluxos de rede e machine learning para detecção de anomalias com ELK

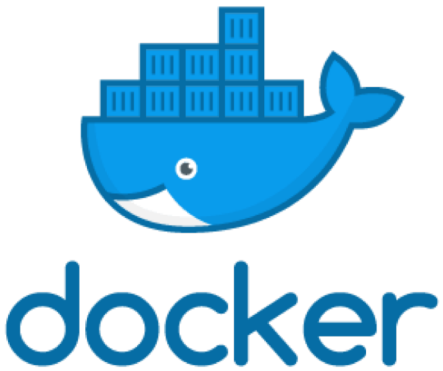
Rodrigo Bongers
Rodrigo Pescador

DEO – Diretoria de Engenharia e Operações

O backbone nacional da RNP



A solução: ELK + elastiflow + docker



Recursos computacionais utilizados

No desenvolvimento da solução foram alocados recursos computacionais na nuvem da RNP, sendo:

Logstash

- 24 CPUs
- 16 GB Memória RAM
- 10 GB Disco

7

Elasticsearch (Elastiflow)

- 24 CPUs
- 32 GB Memória RAM
- 2TB Disco (LUN dedicada)

6

Kibana + Elasticsearch Client

- 12 CPUs
- 12 GB Memória RAM

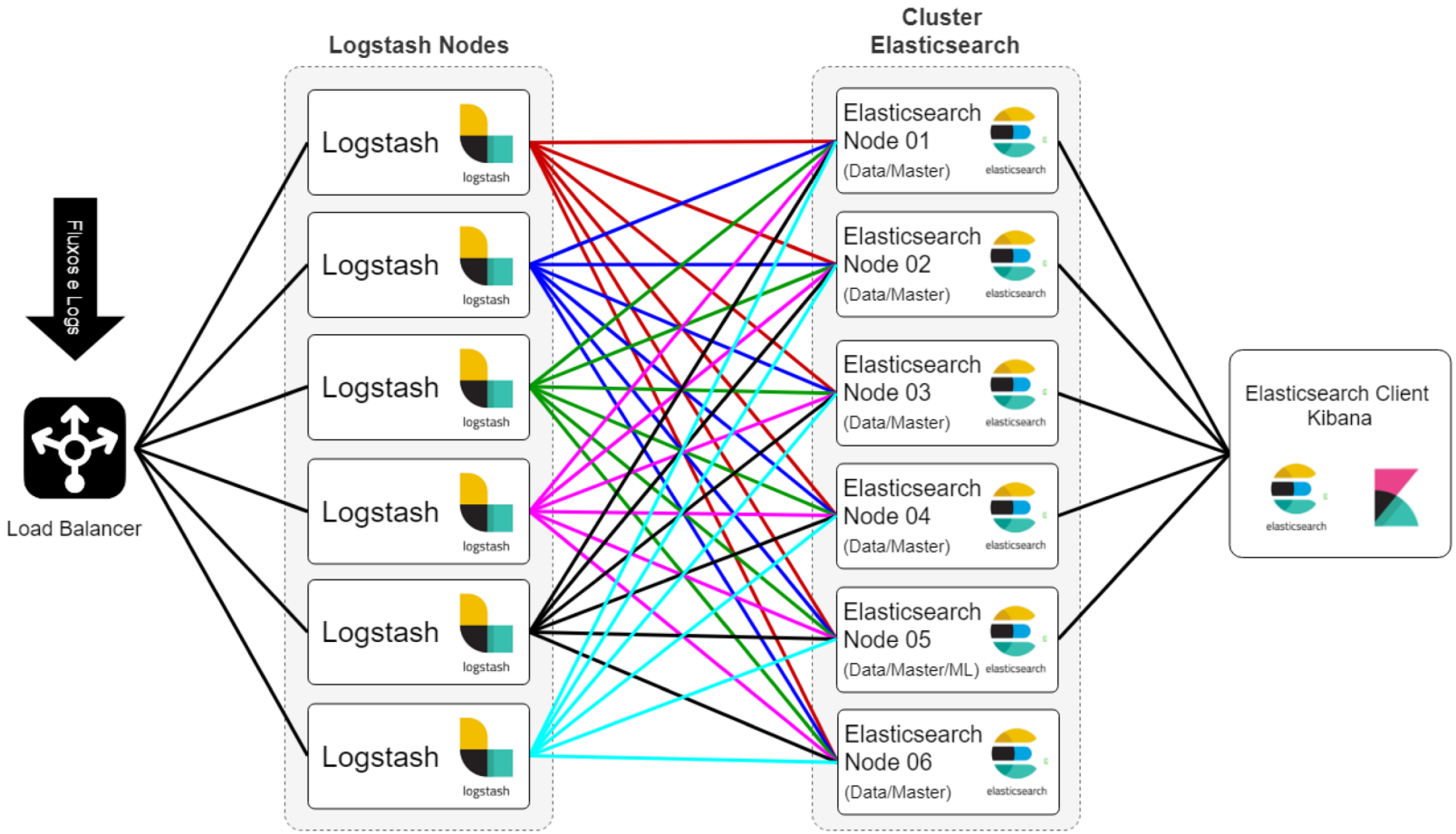
1

Capacidade Computacional Total

- 324 CPUs @2.3GHz
- 316 GB de memória
- 12 TB de disco para armazenamento

Infraestrutura da solução

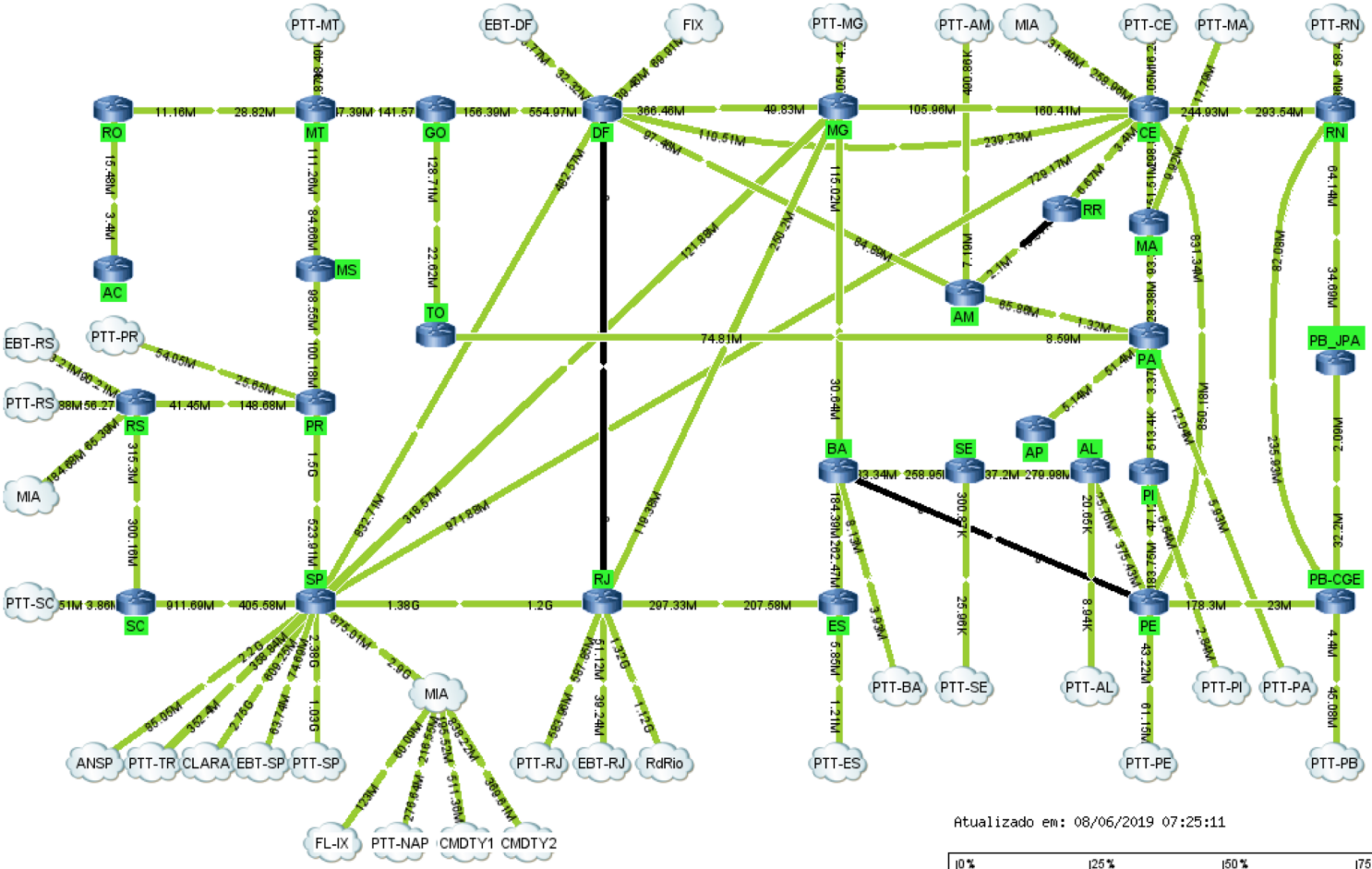
Todos os serviços rodam sob Docker



Tipos de dados recebidos e processados

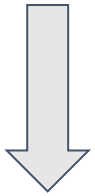
- Recebimento e processamento de fluxos do backbone IP
  - IPFIX
- Recebimento e processamento de logs
  - SYSLOG

MACHINE LEARNING TRAZ INTELIGÊNCIA PARA OS DADOS PROCESSADOS!


PTT-MT
EBT-DF
FIX
PTT-MG
PTT-AM
MIA
PTT-CE
PTT-MA
PTT-RN
RO
MT
GO
DF
MG
CE
RN
AC
MS
TO
RR
MA
AM
EBT-RS
PTT-PR
PA
PB_JPA
PTT-RS
RS
PR
AP
BA
SE
AL
PI
MIA
SP
RJ
PB-CGE
PTT-SC
SC
ES
PE
ANSP
PTT-TR
CLARA
EBT-SP
PTT-SP
MIA
PTT-RJ
EBT-RJ
RdRio
PTT-BA
PTT-SE
PTT-AL
PTT-PI
PTT-PA
PTT-ES
PTT-PE
PTT-PB
FL-IX
PTT-NAP
CMDTY1
CMDTY2
Atualizado em: 08/06/2019 07:25:11
0%
25%
50%
75%
100%
Carga de Tráfego

Números

- Nós exportadores de dados: 37
- Dados processados pelo sistema:
  - Logs: ~ 28 / segundo
  - Fluxos: ~ 20.000 / segundo



O tráfego total do backbone da RNP em horários de pico, incluindo tráfego nacional e internacional, é de 300 Gbps

Novos enlaces de 100Gbps vão incrementar estes valores

Documentos por índice (rotação diária):

- logs: ~ 2.5 milhão
- fluxos: ~ 900 milhões

~ 750GB de dados por dia

GROK

Parser de uma mensagem syslog enviada ao logstash para pré-processamento com GROK e posterior envio ao elasticsearch

[... restante da configuração omitida ...]

```
 "message", "%{MONTH:syslog_month} %{MONTHDAY:syslog_day}
%{HOUR:syslog_hour}:%{MINUTE:syslog_minute}:%{SECOND:syslog_second}
(?:%{HOSTNAME:syslog_hostname}|%{IP?syslog_ip})
%{PROG:syslog_program}(?:\[%{POSINT:pid}\]): %{WORD:syslog_tag_juniper}:
%{GREEDYDATA:syslog_message}",
```

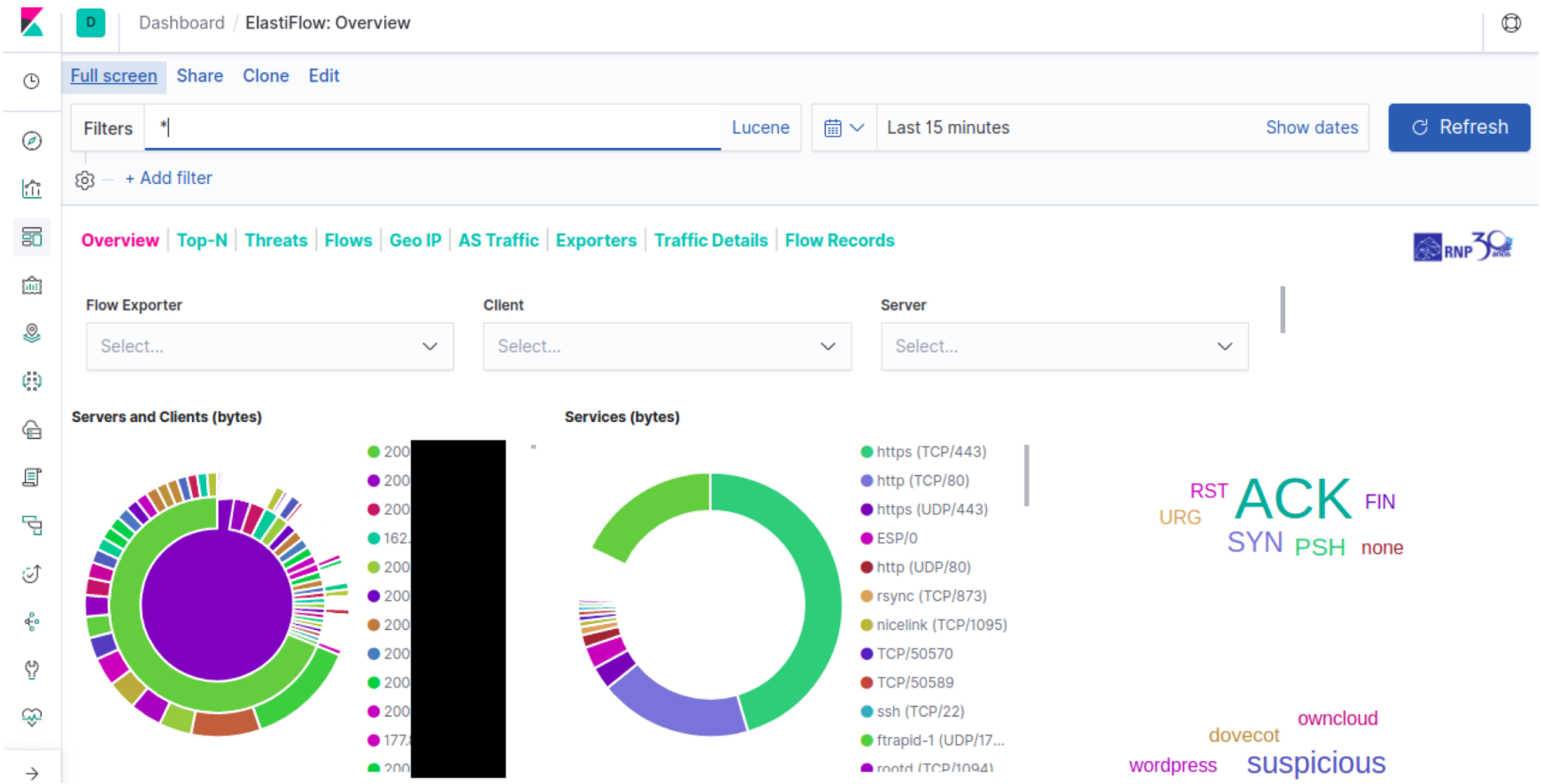
```
 "message", "%{MONTH:syslog_month} %{MONTHDAY:syslog_day}
%{HOUR:syslog_hour}:%{MINUTE:syslog_minute}:%{SECOND:syslog_second}
(?:%{HOSTNAME:syslog_hostname}|%{IP?syslog_ip})
%{PROG:syslog_program}(?:\[%{POSINT:pid}\]): %{GREEDYDATA:syslog_message}",
```

[... restante da configuração omitida ...]

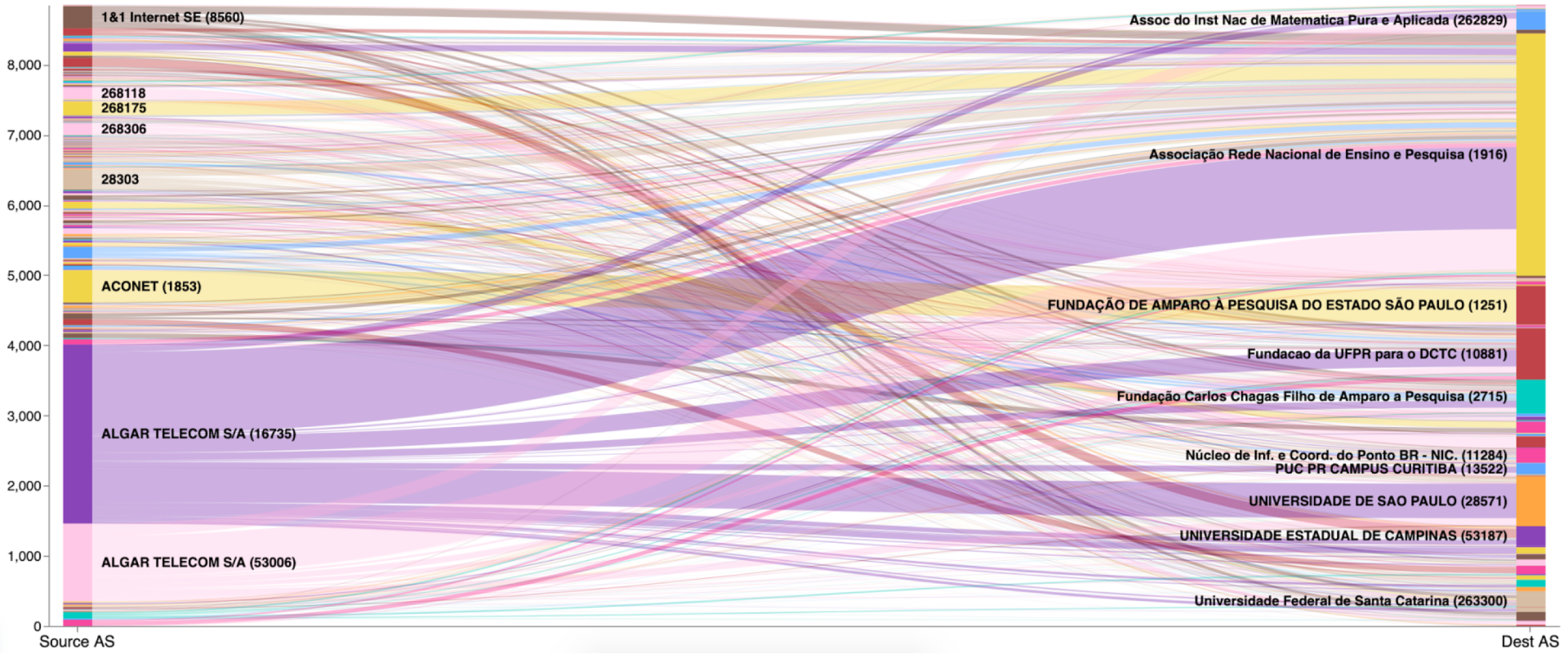
MUTATE

- Altera e remove tags
- Cria informações adicionais para facilitar a busca e agregação do elastic
  - Exemplo: GeoIP, ASN Lookup

O ambiente


Dashboard / ElastiFlow: Overview
Full screen
Share
Clone
Edit
Filters
*
Lucene
Last 15 minutes
Show dates
Refresh
+ Add filter
Overview
Top-N
Threats
Flows
Geo IP
AS Traffic
Exporters
Traffic Details
Flow Records
Flow Exporter
Client
Server
Select...
Servers and Clients (bytes)
Services (bytes)
https (TCP/443)
http (TCP/80)
https (UDP/443)
ESP/0
http (UDP/80)
rsync (TCP/873)
nicelink (TCP/1095)
TCP/50570
TCP/50589
ssh (TCP/22)
RST
ACK
FIN
URG
SYN
PSH
none
owncloud
dovecot
wordpress
suspicious

O ambiente


1&1 Internet SE (8560)
268118
268175
268306
28303
ACONET (1853)
ALGAR TELECOM S/A (16735)
ALGAR TELECOM S/A (53006)
Assoc do Inst Nac de Matematica Pura e Aplicada (262829)
Associação Rede Nacional de Ensino e Pesquisa (1916)
FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO SÃO PAULO (1251)
Fundacao da UFPR para o DCTC (10881)
Fundação Carlos Chagas Filho de Amparo a Pesquisa (2715)
Núcleo de Inf. e Coord. do Ponto BR - NIC. (11284)
PUC PR CAMPUS CURITIBA (13522)
UNIVERSIDADE DE SAO PAULO (28571)
UNIVERSIDADE ESTADUAL DE CAMPINAS (53187)
Universidade Federal de Santa Catarina (263300)
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
Source AS
Dest AS

Monitoramento do ambiente

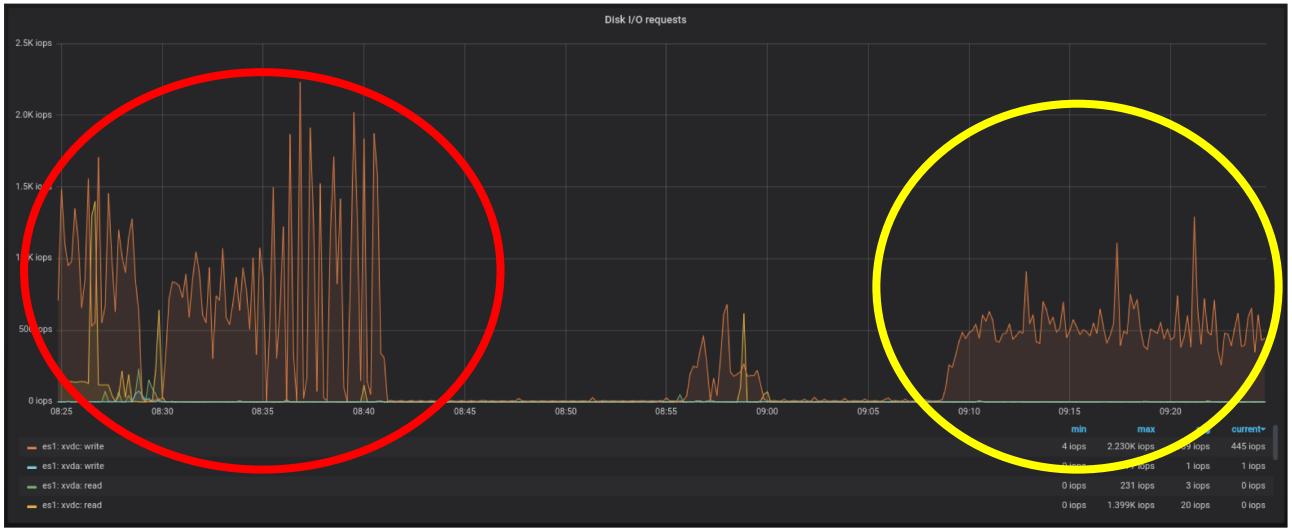
Todos os hosts que fazem parte da infraestrutura são monitorados em detalhes em três diferentes níveis:

- **Hosts Linux**
  - Métricas relacionadas ao host (CPU, memória, load average, I/O disco, informações de rede, etc)
- **Containers Docker**
  - Métricas relacionadas ao container (CPU, memória, informações rede, etc)
- **Cluster ELK**
  - Métricas relacionadas a todo ambiente ELK (CPU, memória, informações da JVM, taxa de indexação dos índices, shards, saúde do cluster, JVM heap, threads, I/O de leitura/escrita, etc)

Monitoramento do ambiente: Hosts Linux

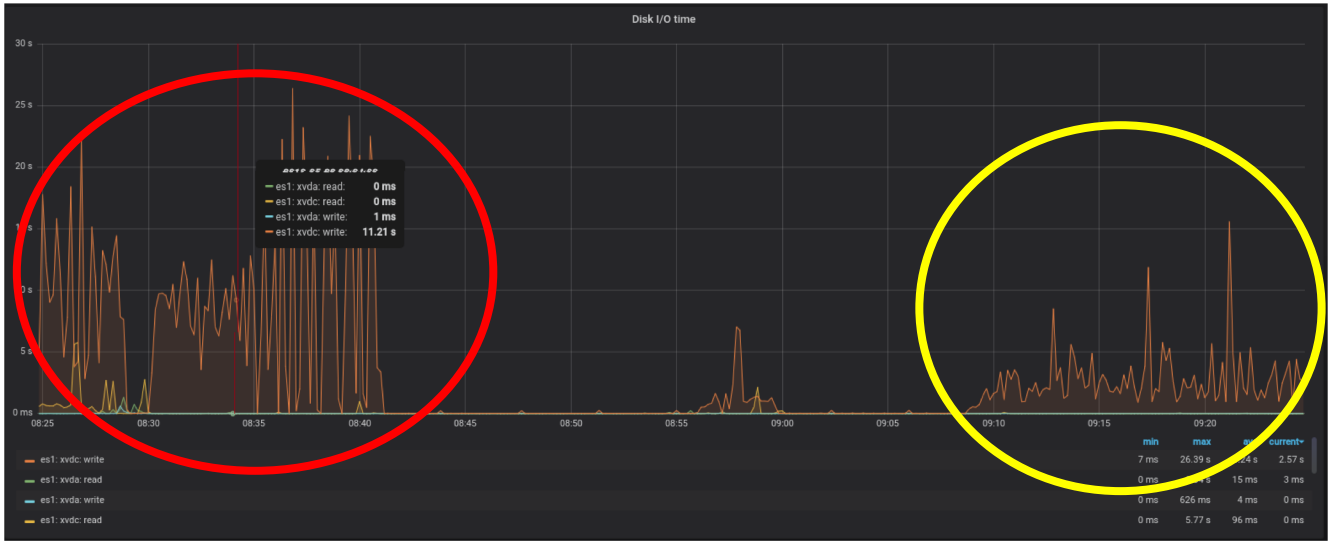
Monitoramento do ambiente: Cluster ELK

- **IOPS nos discos do Elasticsearch**

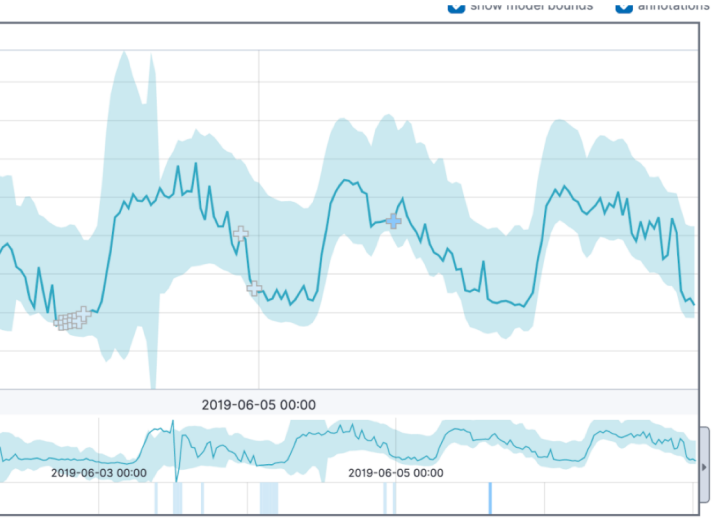


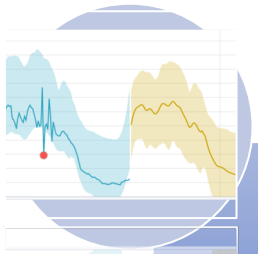
O armazenamento dos dados em dois discos físicos (data paths) distintos aumenta a performance do sistema

Com somente um disco físico (data path) disponível o número de IOPS é alto devido a concorrência no acesso

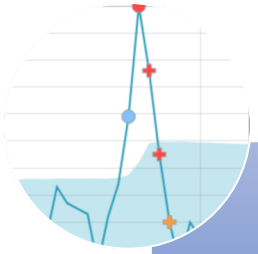


Machine Learning





Cria Forecast



Detecta Anomalias

Busca Padrões

Analisa Historico

Detectores

Define campos e funções utilizadas para a análise

Influenciadores

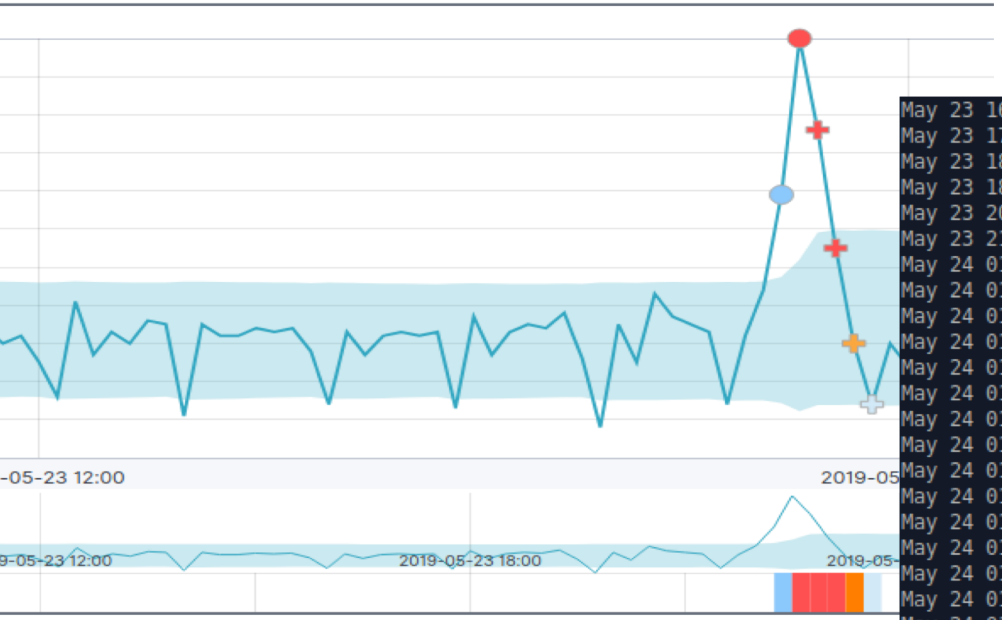
Define campos que influenciam os resultados

- O tempo de processamento do ML é diretamente proporcional ao tamanho de dados a serem analisados
    - Criar filtros para buscar as informações de maneira precisa
        - Evita falso positivo
    - Otimizar buscas e processamento criando filtros
    - Evitar excesso de influenciadores (Ideal 1 a 3)

- Quanto maior o forecast, maior a chance de erro

- Índice do sistema exclusivo para ML

GMT-3
22h27m

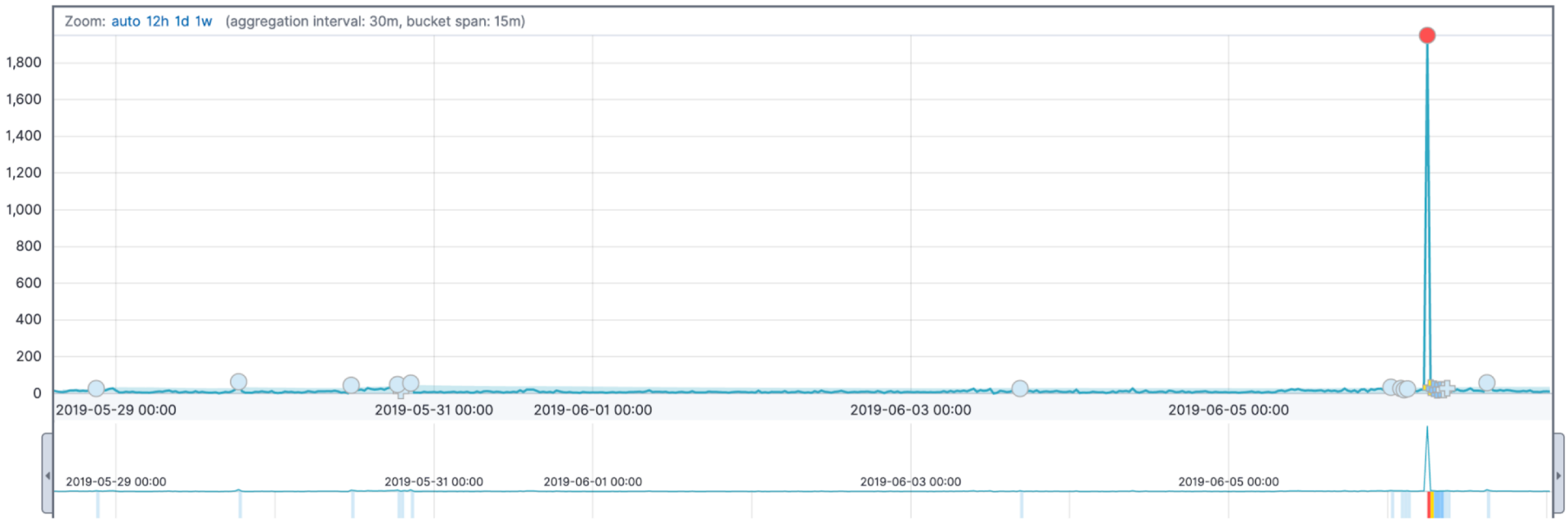


influenced by	actual	typical
syslog_program.keyword: xntpd	86	30.2
syslog_hostname.keyword: mdf2 syslog_hostname.keyword: mpr syslog_hostname.keyword: mxmia2 syslog_program.keyword: xntpd	55	30.4

UTC 1h26m

```
May 23 16:41:28.139  jmx_pr-0 xntpd[19023]: kernel time sync enabled 6001
May 23 17:32:44.124  jmx_pr-0 xntpd[19023]: kernel time sync enabled 2001
May 23 18:41:02.216  jmx_pr-0 xntpd[19023]: kernel time sync enabled 6001
May 23 18:58:06.184  jmx_pr-0 xntpd[19023]: kernel time sync enabled 2001
May 23 20:06:23.128  jmx_pr-0 xntpd[19023]: kernel time sync enabled 6001
May 23 21:14:40.117  jmx_pr-0 xntpd[19023]: kernel time sync enabled 2001
May 24 01:17:38.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:07.098  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:09.108  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:11.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:13.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:15.102  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:17.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:19.102  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:21.108  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:26:23.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:24.100  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:26.098  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:28.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:30.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:32.100  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:34.100  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:36.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:38.100  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 01:43:40.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:44.098  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:46.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:48.098  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:50.098  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:52.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:54.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:56.105  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:00:58.099  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:01:00.105  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
May 24 02:18:04.101  jmx_pr-0 xntpd[19023]: NTP Server 200.1          is Unreachable
```

UDP Amplification Attack



time	max severity	detector	found for	influenced by	actual	typical	description	actions
June 6th 2019	97	sum("flow.packets") partitionfield="flow.dst_port"	2	flow.src_addr: 178.	3,885	9.08	More than 100x higher	

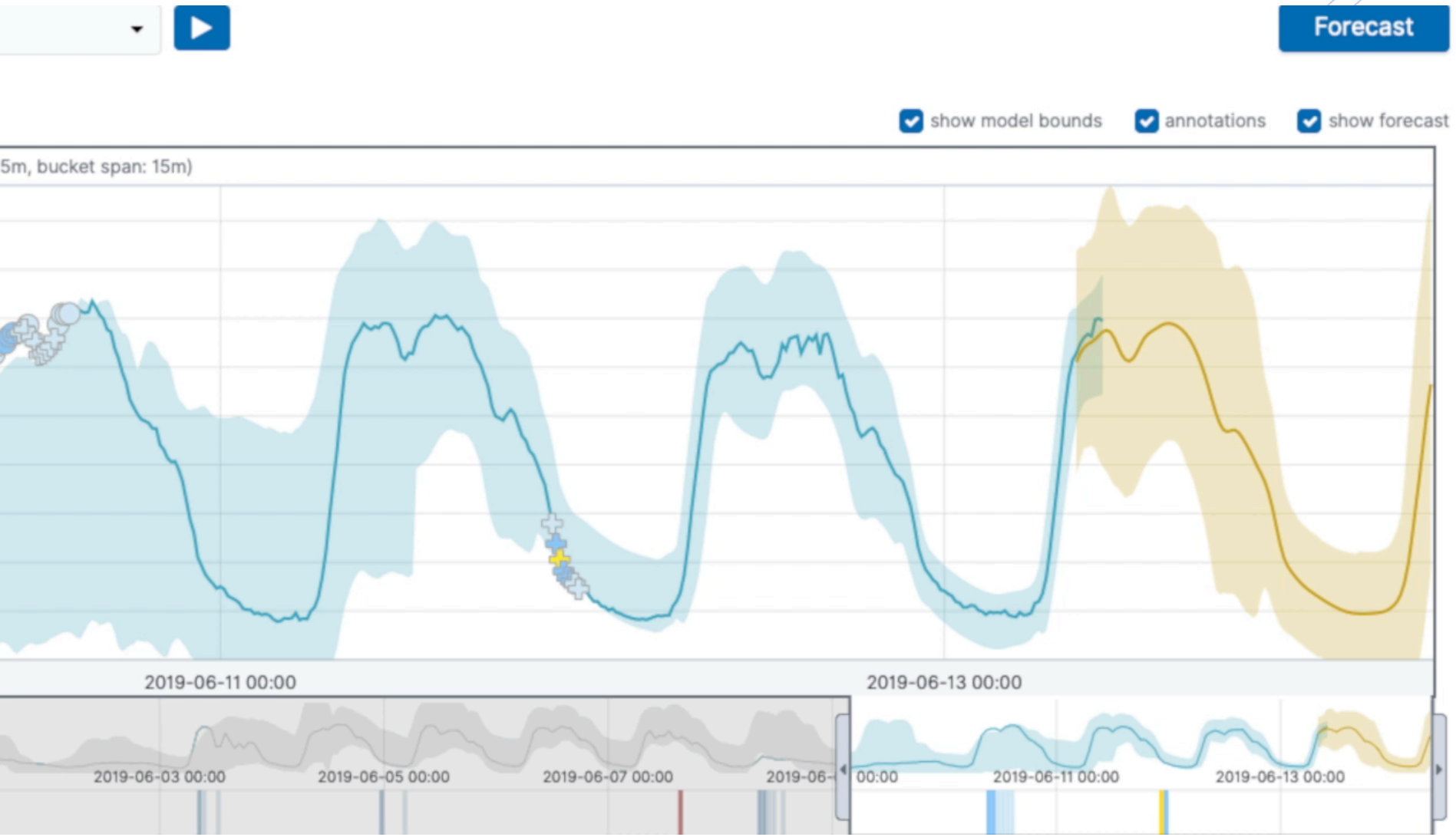
Description
critical anomaly in sum("flow.packets") partitionfield="flow.dst_port" found for flow.dst_port 2

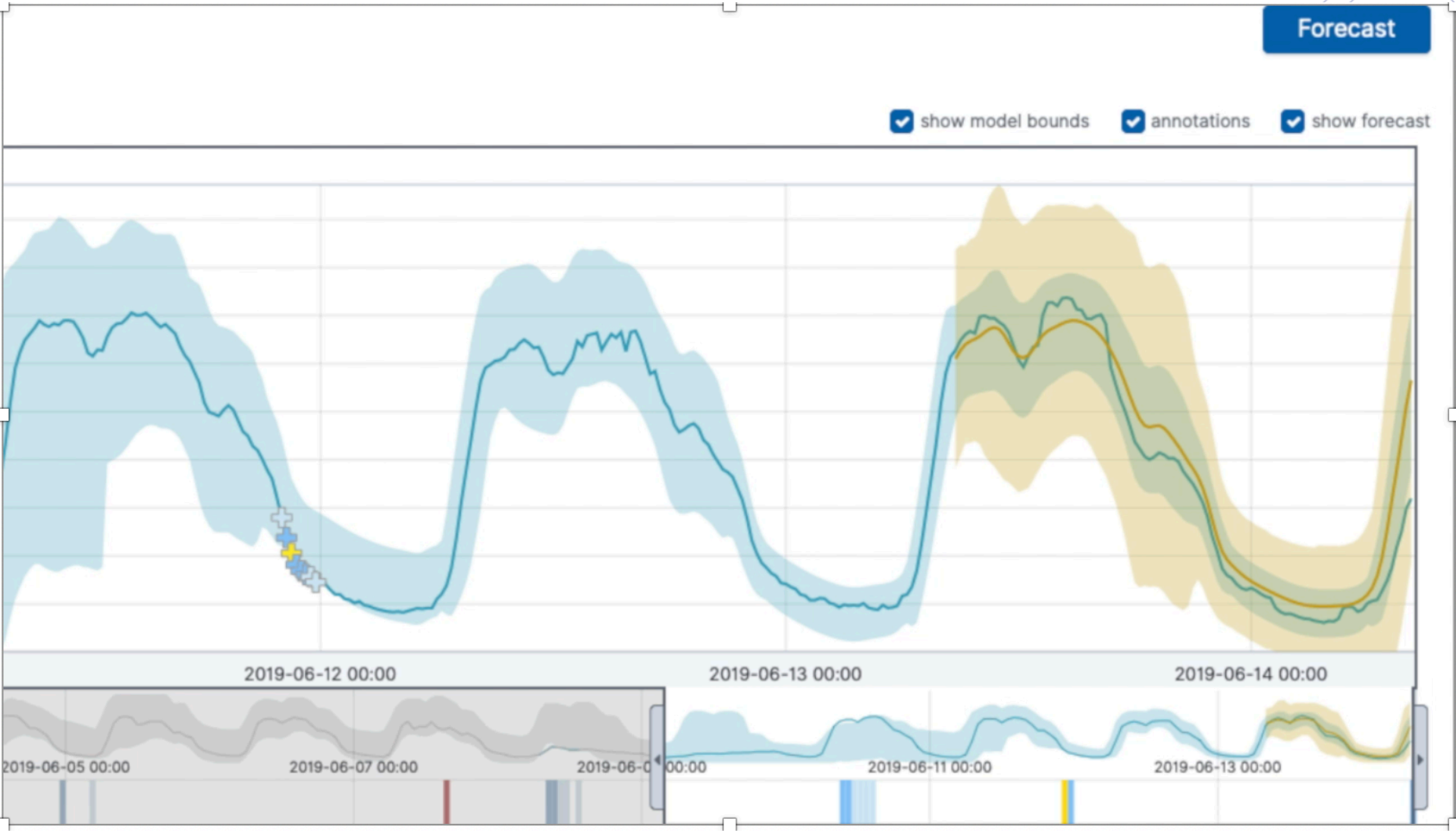
Details on highest severity anomaly

flow.dst_port	2
time	June 6th 2019, 06:15:00 to June 6th 2019, 06:30:00
function	sum
fieldName	flow.packets
actual	3885
typical	9.08
job ID	adv-udp-amp-attack-dst-addr-pps-v3
probability	8.73732000688166e-16

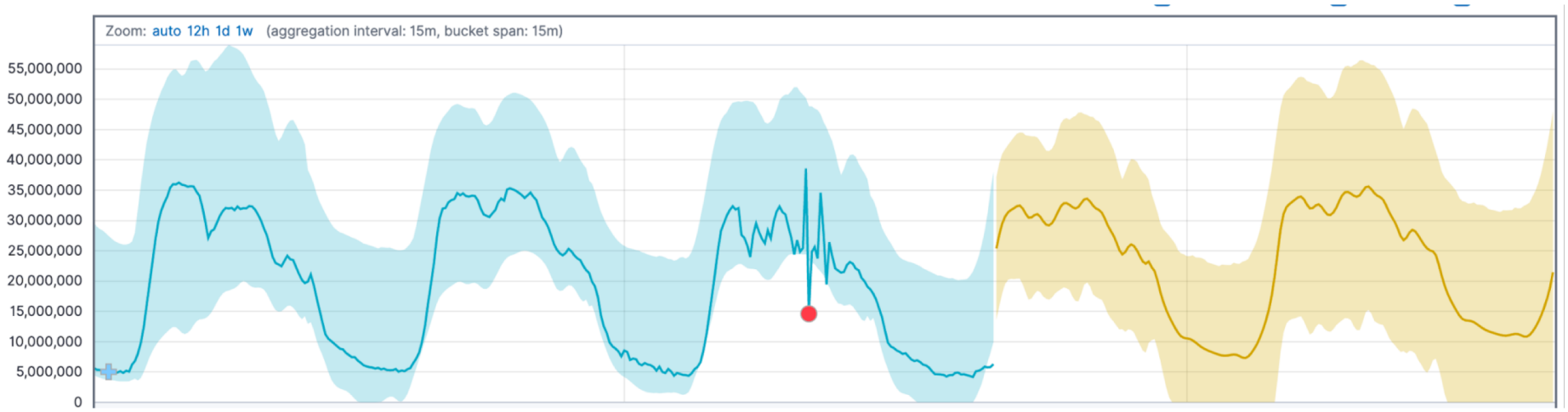
Influencers

flow.src_addr	178.


Forecast
show model bounds
annotations
show forecast
5m, bucket span: 15m)
2019-06-11 00:00
2019-06-13 00:00
2019-06-03 00:00
2019-06-05 00:00
2019-06-07 00:00
00:00
2019-06-11 00:00
2019-06-13 00:00


Forecast
show model bounds
annotations
show forecast
2019-06-12 00:00
2019-06-13 00:00
2019-06-14 00:00
2019-06-05 00:00
2019-06-07 00:00
2019-06-11 00:00
2019-06-13 00:00

Forecast


Zoom: auto 12h 1d 1w (aggregation interval: 15m, bucket span: 15m)
55,000,000
50,000,000
45,000,000
40,000,000
35,000,000
30,000,000
25,000,000
20,000,000
15,000,000
10,000,000
5,000,000
0

Próximos passos

- Utilização de servidores bare metal com discos SSD 1 TB NVMe
- Adição de nós para atuarem exclusivamente como masters
- Kubernets para orquestração e administração do cluster

Obrigado!

Rodrigo Bongers
rodrigo.bongers at rnp.br

Rodrigo Pescador
rodrigo.pescador at rnp.br