Predicting intermittent network device failures based on network metrics from multiple data sources

Authors:
H.P.M. van Doorn
C.H.J. Kuipers

University of Amsterdam

Supervisors:
P. Boers
M. Kaat

SURFnet

Tuesday 3 juli
RP91

Introduction

Collected Data over 2 years

~690 Million Device Events

~163 Billion Device Metrics

Introduction

Relevance

Failures impacting **connectivity**

Introduction

Research question

To what extent is it possible to **predict** intermittent network device **failures** based on network metrics from **multiple data sources**?

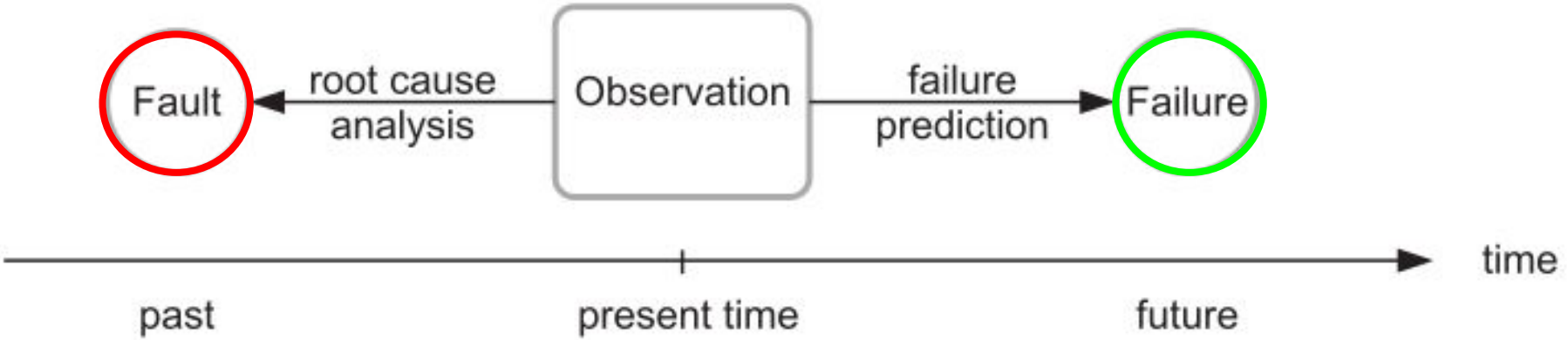
Introduction

Sub questions

- Which metrics are relevant?
- Patterns between failures?
- Correlation between data sources?

Introduction

Fault vs Failure



Source: Salfner et al. *"A Survey of Online Failure Prediction Methods"*.

Methodology

Identifying outages

Startingpoint: Big outages in the past 2 years:

Big: multiple customers losing connectivity

Based on:

- Ticketing System
- Network operators

Methodology

Categorizing outages

- Intermittent failure (Spontaneous reboots)
- Permanent failure (Line-card malfunctioning)

Methodology

Metrics at hand

Switch chassis metrics

- CPU and Memory utilization
- Temperature
- Uptime

Metrics per interface:

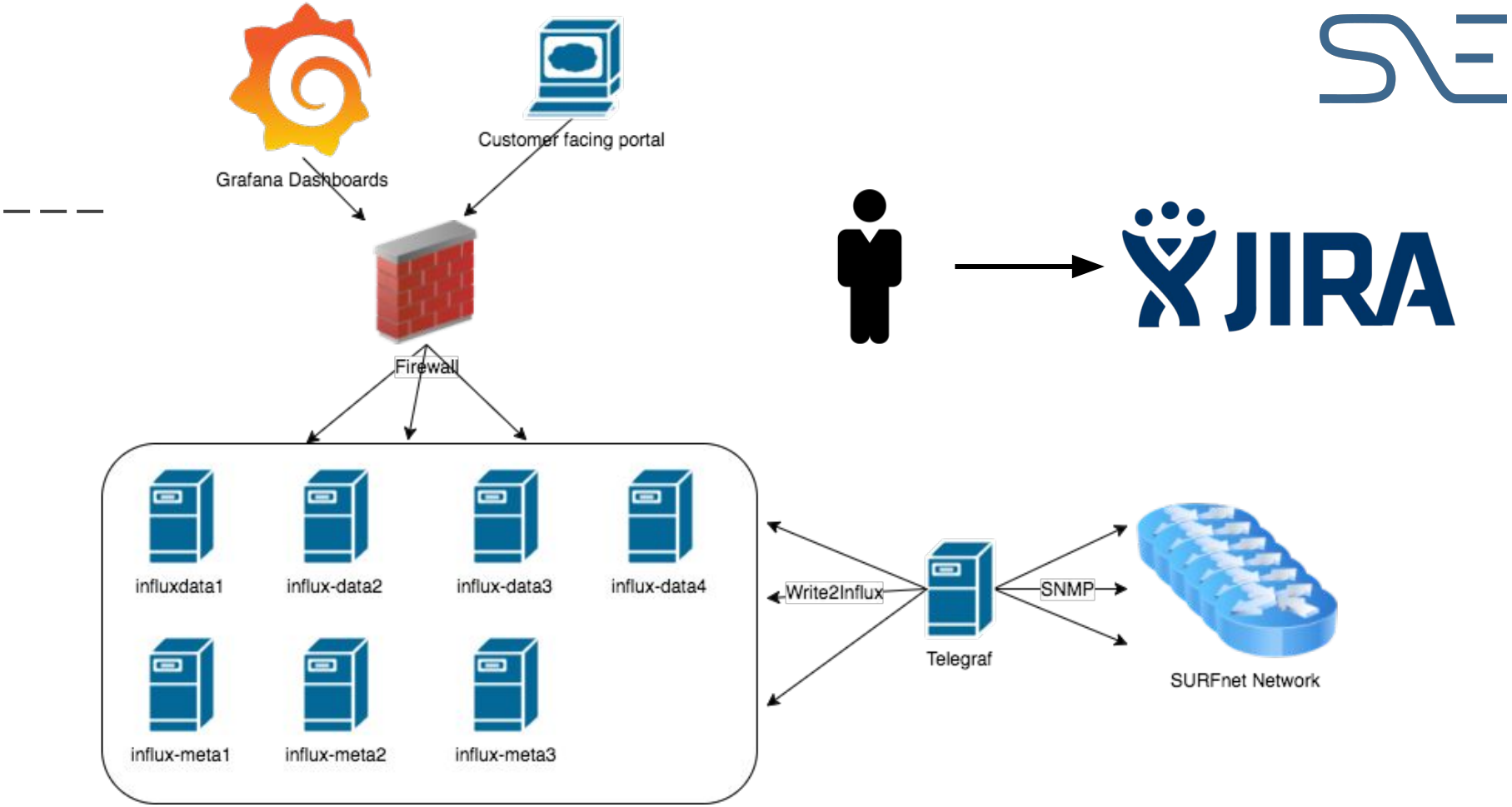
- Throughput
- Unicast packets
- Multicast packets
- Broadcast packets
- In/Out Errors

Data Sources

Overview Device Data

Device Metrics:

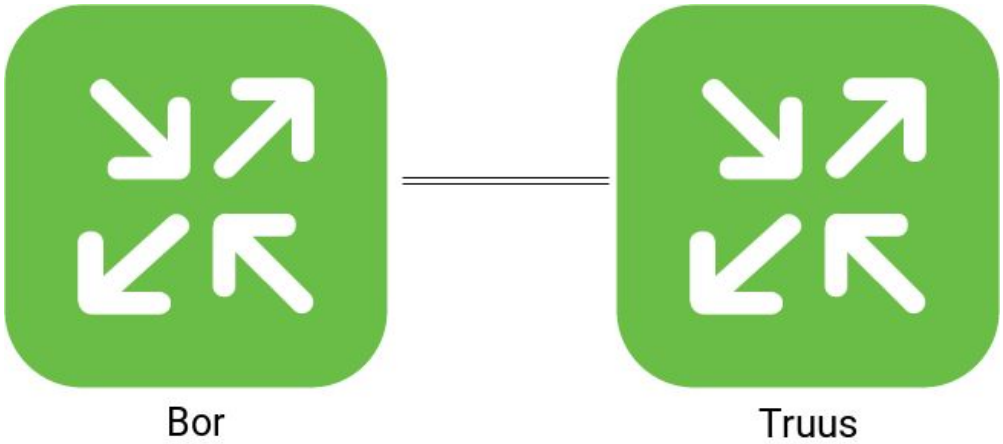
Device Events:


Grafana Dashboards
Customer facing portal
Firewall
JIRA
influxdata1
influx-data2
influx-data3
influx-data4
influx-meta1
influx-meta2
influx-meta3
Write2Influx
Telegraf
SNMP
SURFnet Network

Methodology

Line-card failure

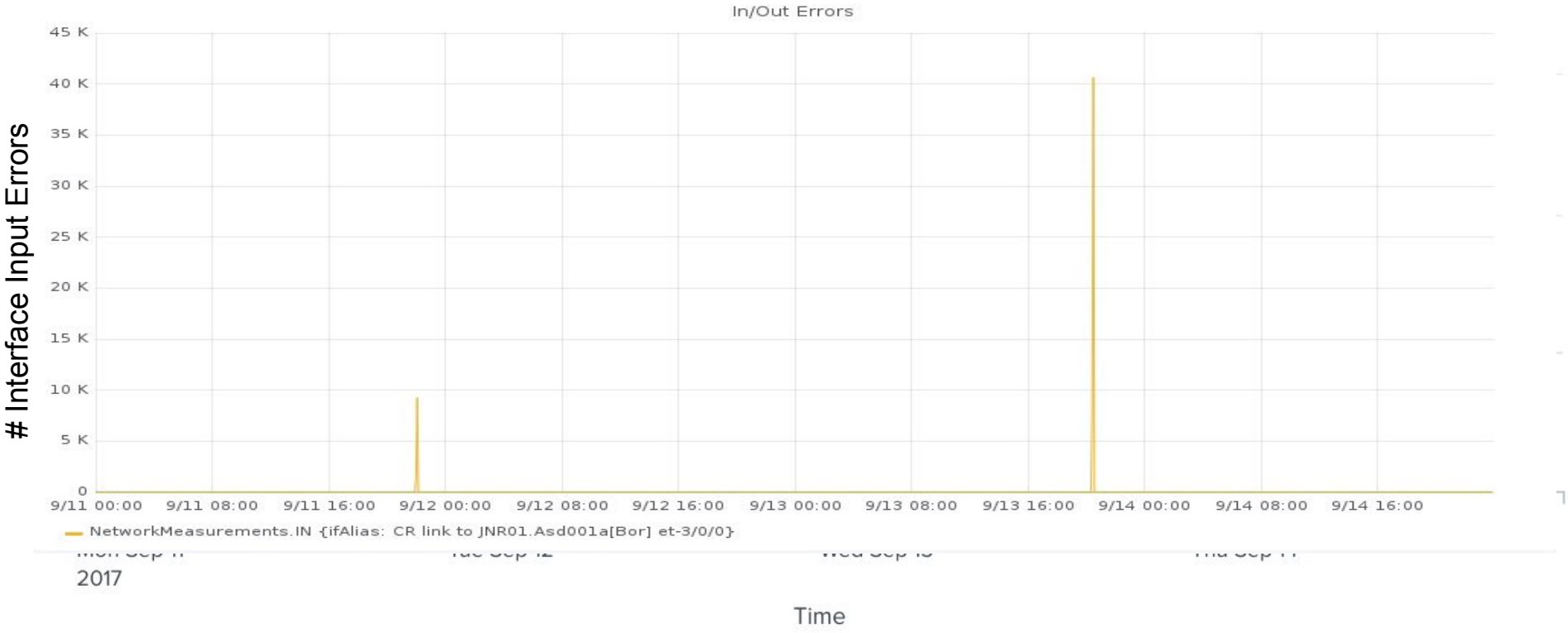
- Line-card *Bor* malfunctioning



Findings

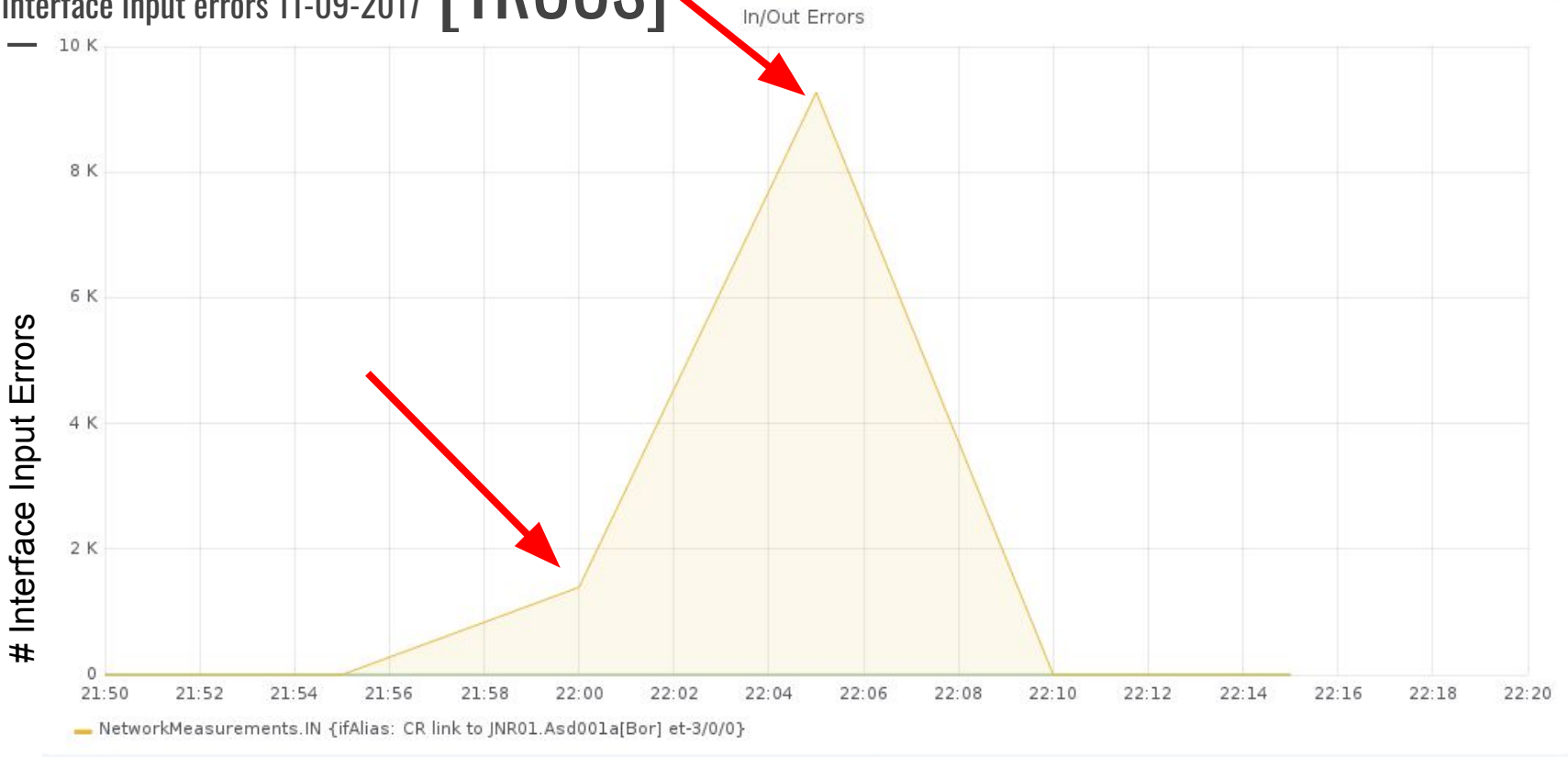
Line Card fault

Results


In/Out Errors
45 K
40 K
35 K
30 K
25 K
20 K
15 K
10 K
5 K
0
Interface Input Errors
9/11 00:00
9/11 08:00
9/11 16:00
9/12 00:00
9/12 08:00
9/12 16:00
9/13 00:00
9/13 08:00
9/13 16:00
9/14 00:00
9/14 08:00
9/14 16:00
NetworkMeasurements.IN {ifAlias: CR link to JNR01.Asd001a[Bor] et-3/0/0}
2017
Time

Findings

Interface Input errors 11-09-2017 [TRUUS]

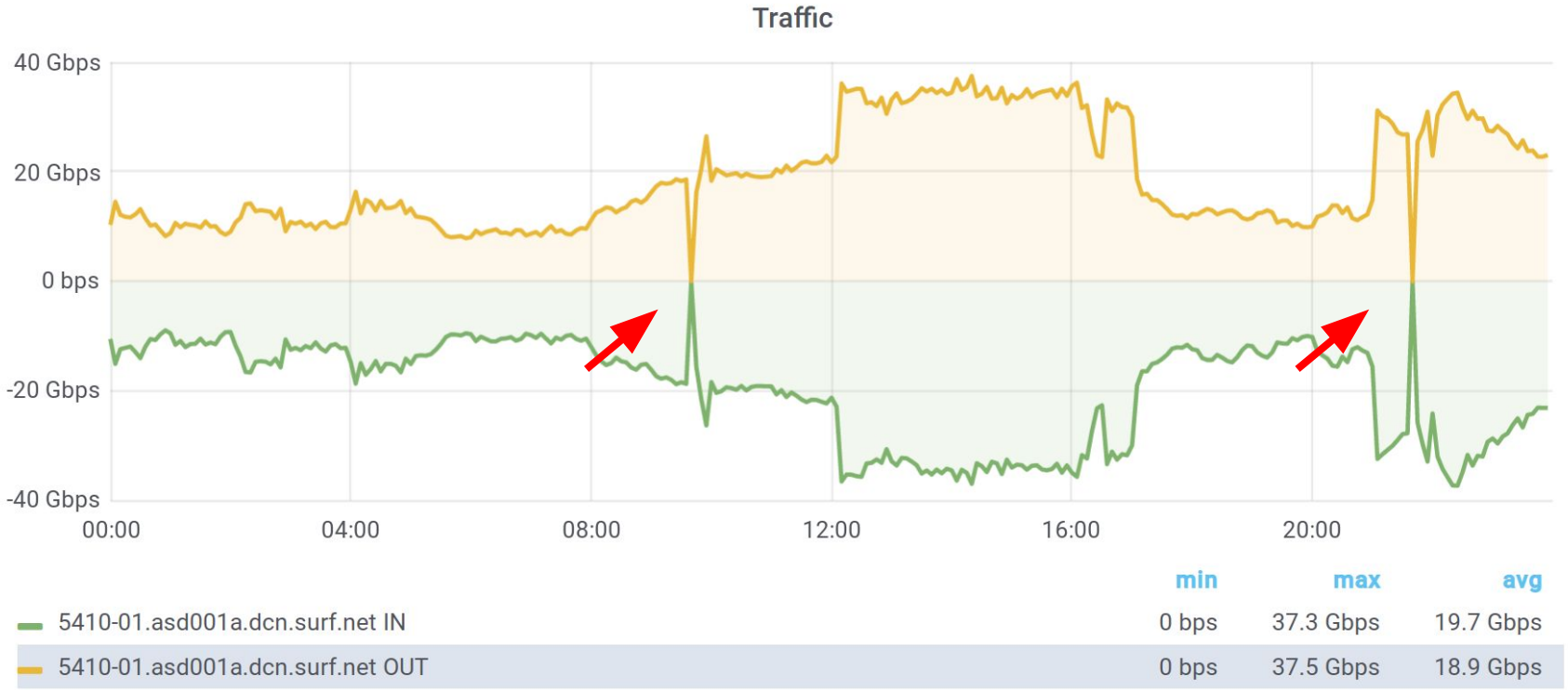


Findings

Loss of throughput

Findings

Spontaneous throughput loss (1)



Findings

Spontaneous throughput loss (2)

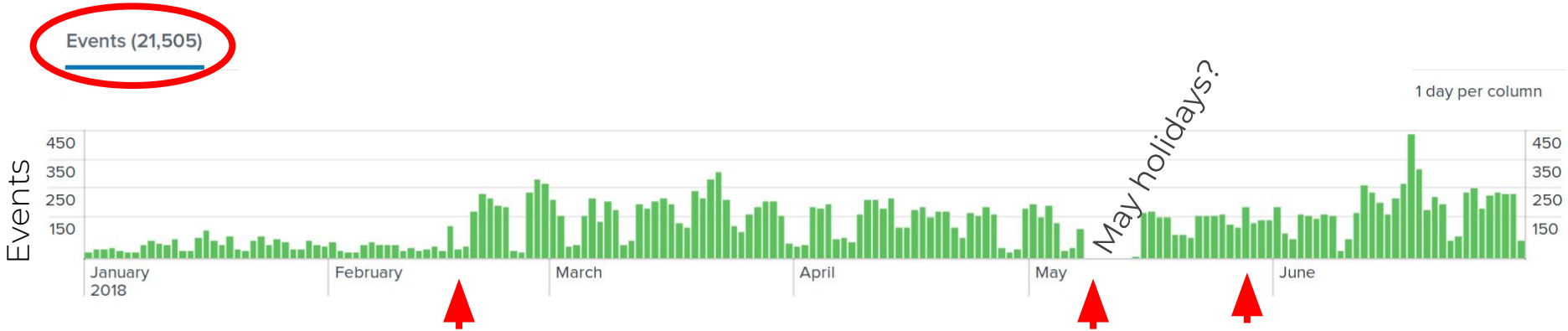
- Syslog event

```
2018 May 24 09:50:33
active.5410-01.Asd001A.dcn.surf.net
DATAPLANE-4-FLOOD_CONTAINMENT_THRESHOLD: chassis(1):
:Flood Containment Threshold Event Container LIMIT_2
on l2-ucast EXCEEDED
```

Findings

Spontaneous throughput loss (3)

- So is this a real problem?



Roughly 21.000 events for **this switch alone**

Findings

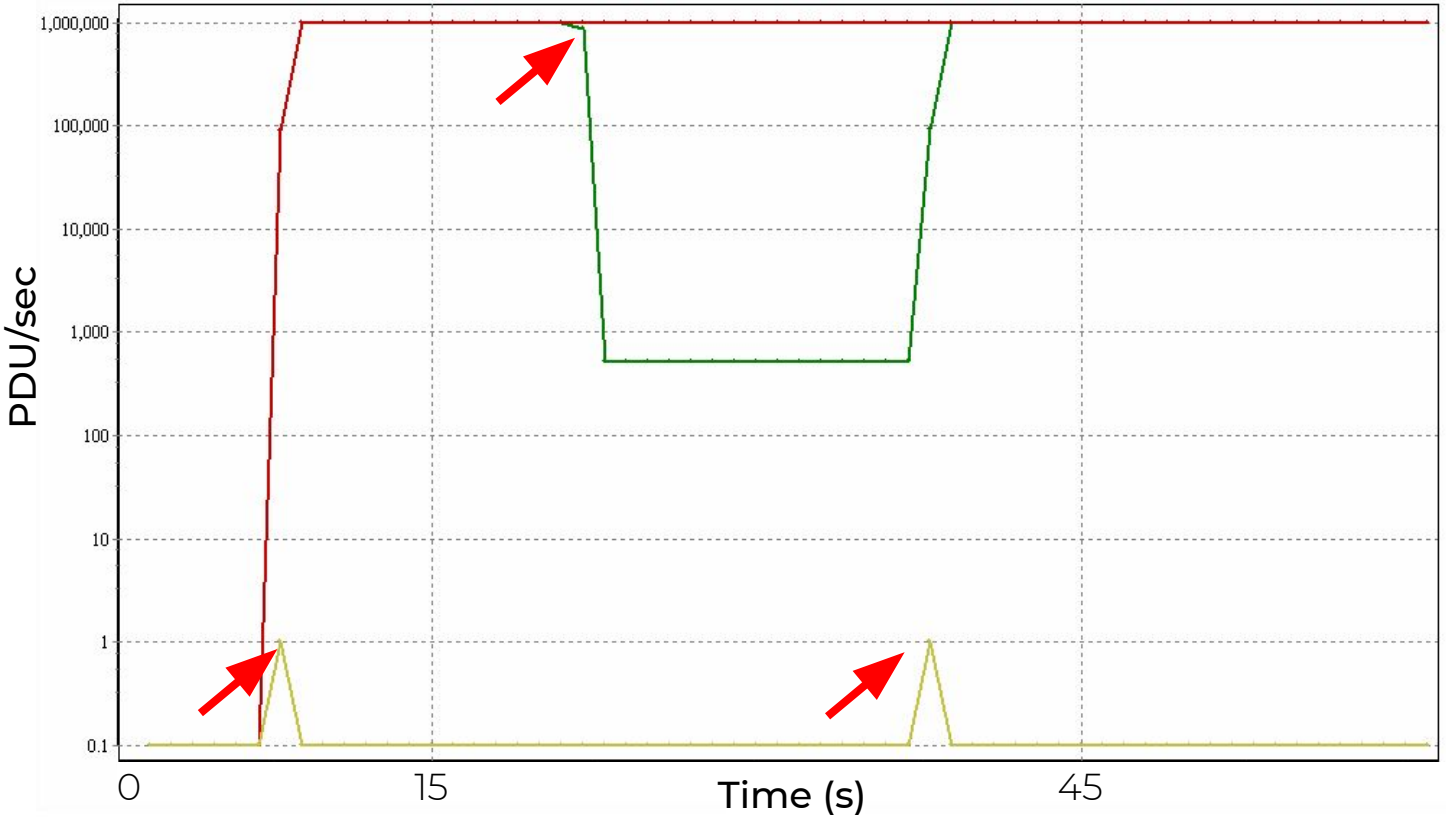
Spontaneous throughput loss (2)

- Syslog event

```
2018 May 24 09:50:33
active.5410-01.Asd001A.dcn.surf.net
DATAPLANE-4-FLOOD_CONTAINMENT_THRESHOLD: chassis(1):
:Flood Containment Threshold Event Container LIMIT_2
on l2-ucast EXCEEDED
```

Findings

Validating our hypothesis



Transmitted PDUs

Received PDUs

Opposite link

Discussion

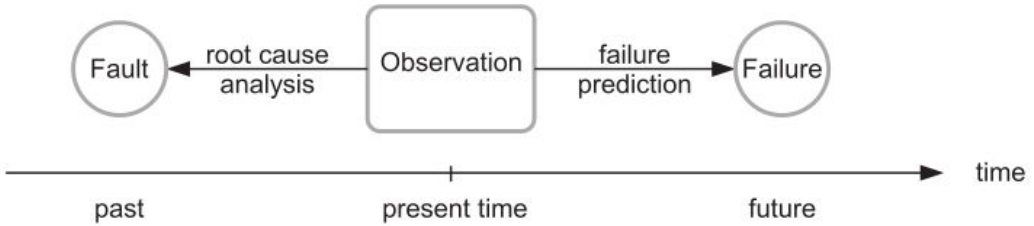
– – –

Identified:

- 2 cases of permanent line-card faults
- Thousands of flood containment events

Challenges:

- Data inconsistencies
- Measurement errors
- No labeled dataset



Conclusion

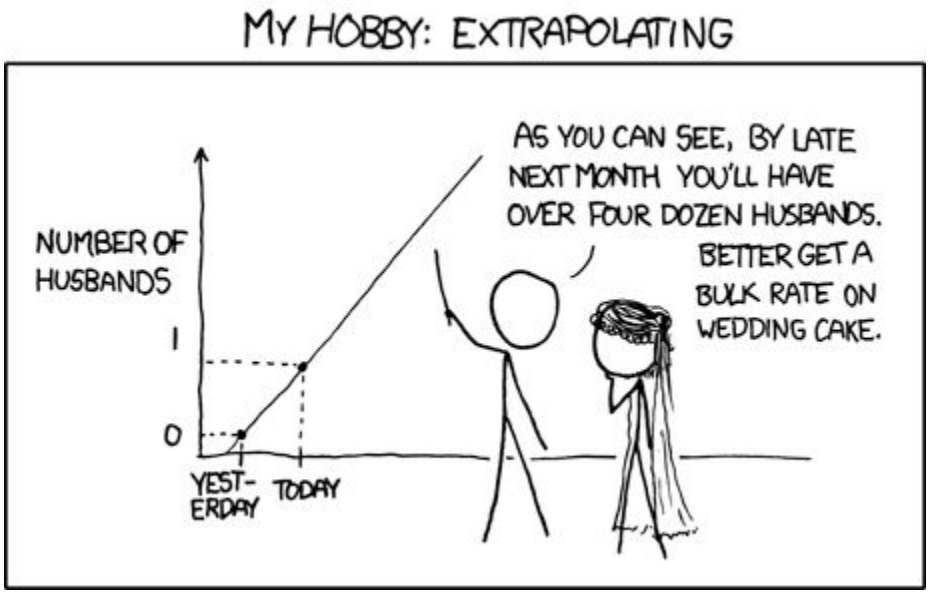
– – –

- Dataset not (yet) suitable for automated predictions
- No data that could indicate failure beforehand
- Proved link between two datasets
- Validated hypothesis

Future Work

- Normalizing datasets
- Create labeled dataset
- Other areas:
  - Capacity Management
  - Service Level Specification

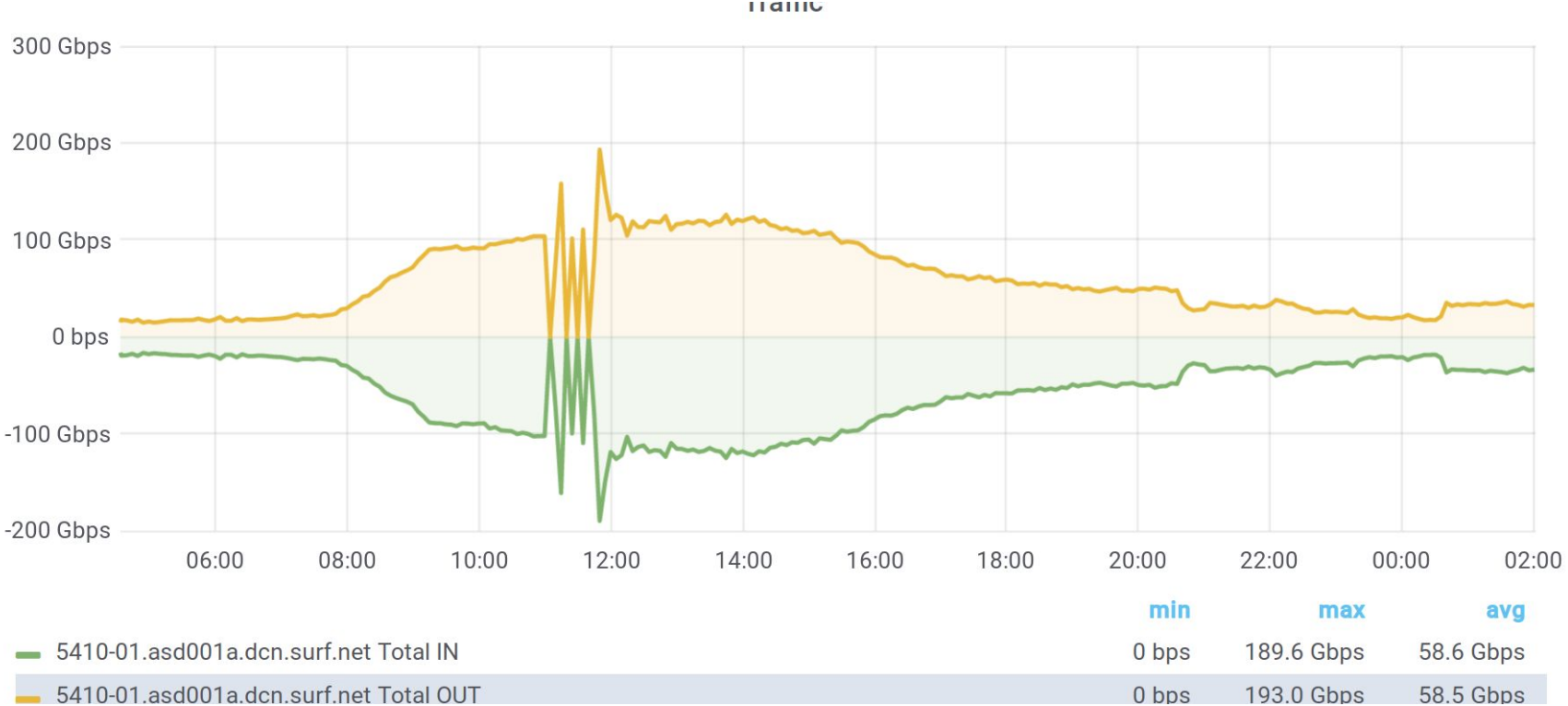
Questions?


MY HOBBY: EXTRAPOLATING
NUMBER OF HUSBANDS
1
0
YEST-ERDAY
TODAY
AS YOU CAN SEE, BY LATE NEXT MONTH YOU'LL HAVE OVER FOUR DOZEN HUSBANDS.
BETTER GET A BULK RATE ON WEDDING CAKE.

Backup slides

Bonus

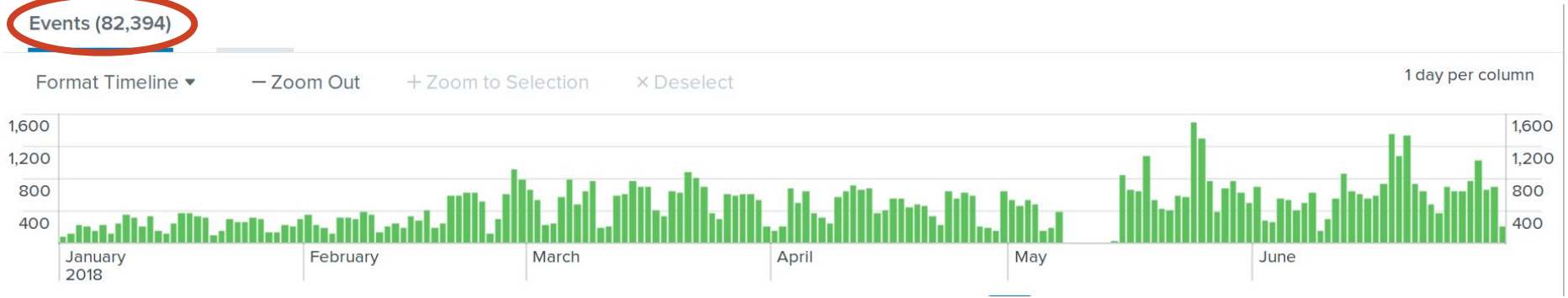
Spontaneous throughput loss



Bonus

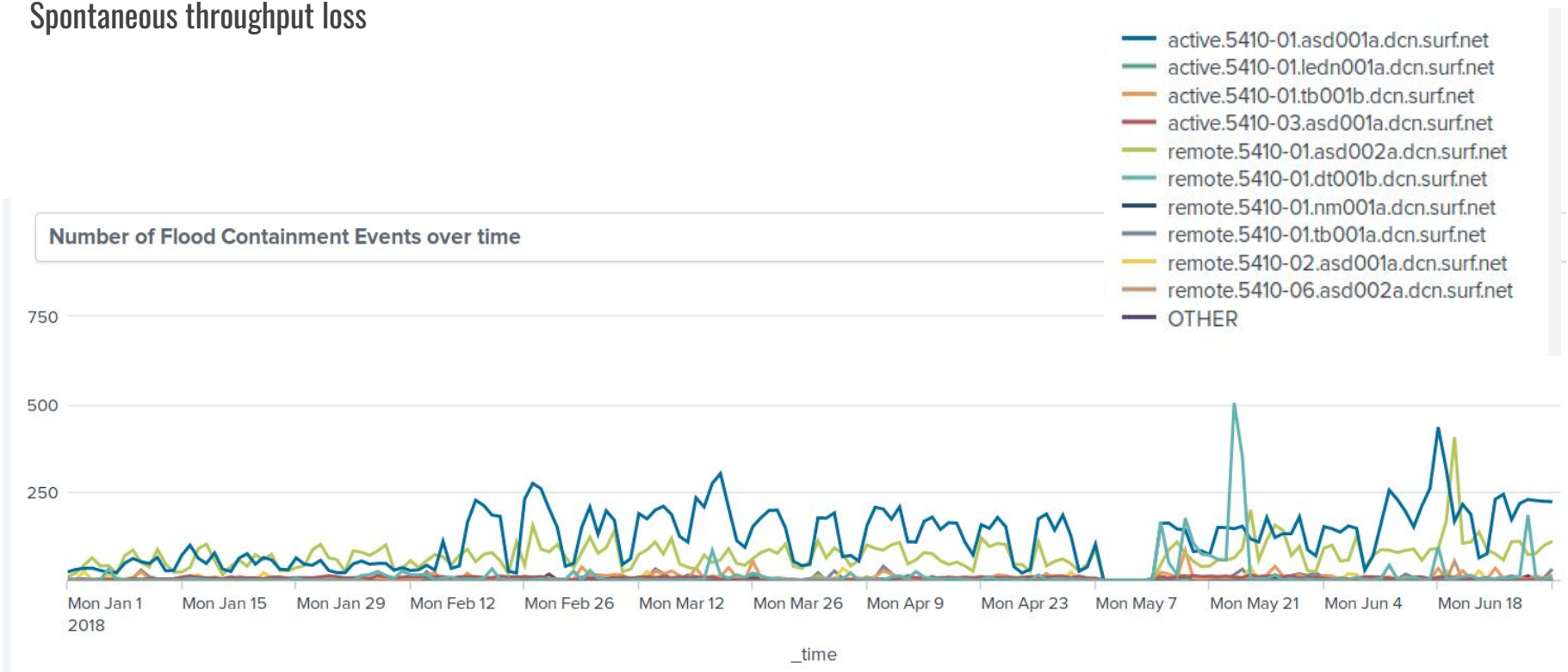
Spontaneous throughput loss

- So is this a real problem?



Bonus

Spontaneous throughput loss



Bonus

Spontaneous reboots