

AI-ML for Network Problem Identification

Mentors:

Shawn McKee (University of Michigan)

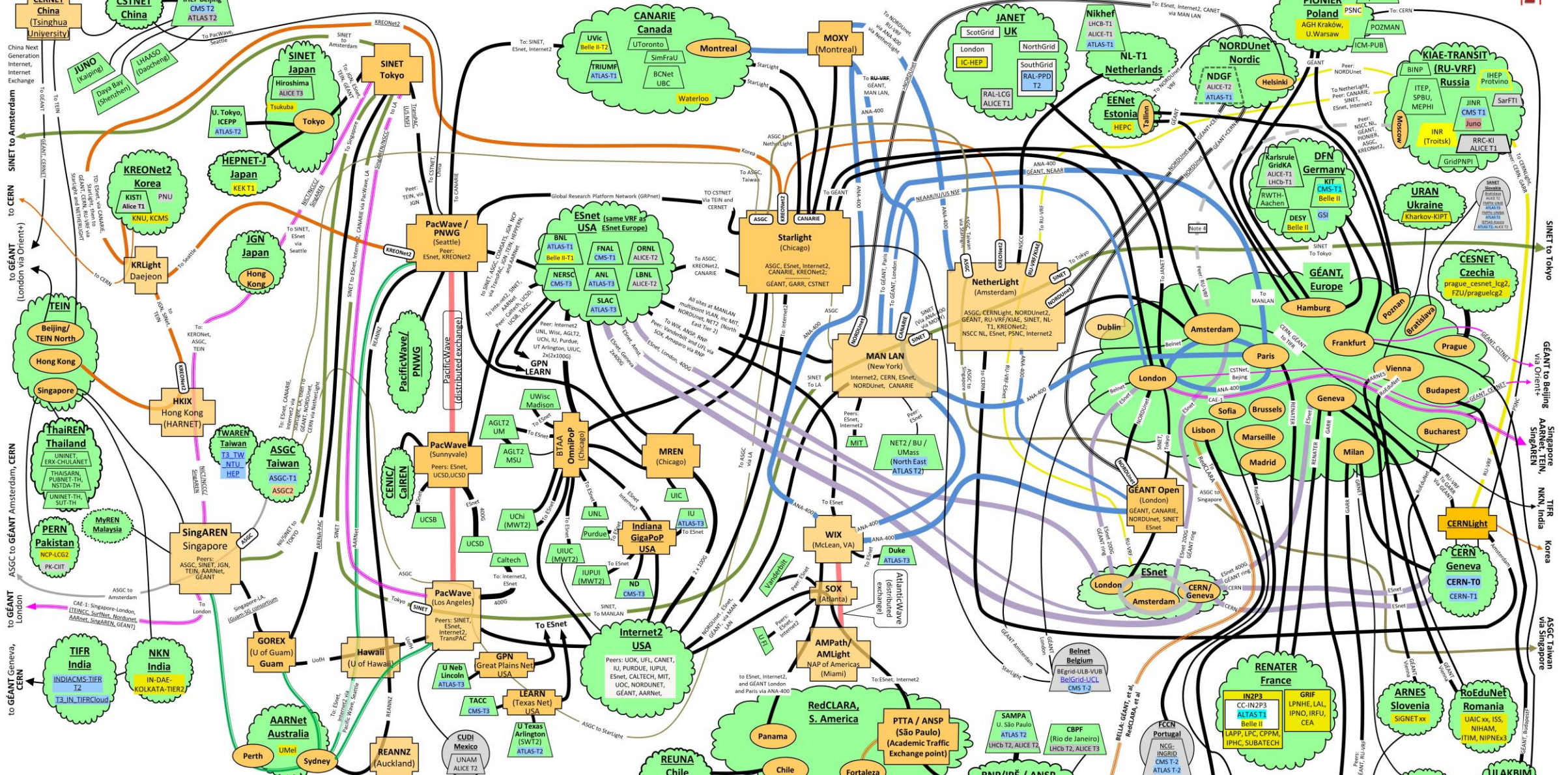
Petya Vasileva (University of Michigan)

Introduction

LHC data generation



LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON, JUNO)



LHCONE Map Ver. 6.0, 2022-11-15 – WEJohnston, ESnet, wej@es.net

Legend:

- Green circle:** LHCONE VRF domain/aggregator - A provider network.
- Orange circle:** Connector network - provides, e.g., an L2 path between VRFs.
- Blue circle:** Provider network PoP router.
- Light blue circle:** WLCG sites that are not connected to LHCONE.
- Black circle:** Exchange point.
- Green line:** NREN/site router at exchange point.
- Black line:** Communication links: $100G=1.5pt, 100G=4pt, 200G=5pt, 400G=6pt, 800G=7.5pt$
- Grey line:** Underlined link information indicates link provider, not use.
- Double dashed line:** Double dash outline indicates distributed site.
- Red dashed line:** Future site.

International infrastructure by provider/collaboration

- Black line:** various
- Green line:** AARNet
- Blue line:** GEANT
- Red line:** SINET, Japan, global ring
- Orange line:** ASGC, Taiwan
- Purple line:** ESnet transatlantic, USA
- Light blue line:** NICT/NCCC/SingAREN
- Dark blue line:** SINET
- Light green line:** NORDUnet
- Yellow line:** KIAE, Russia
- Light orange line:** KREONet2, Korea
- Light purple line:** BELLA: GEANT, et al
- Light red line:** RedCLARA, et al
- Light blue line:** ANA-300/400 - Various links provided by CANARIE, ESnet, GEANT, Internet2, NORDUnet, SURFnet, SINET, IU/NSF

Site abbreviations:

- LHCb-T1:** LHC ALICE or LHCb site
- CNAF-T1:** LHC Tier 1 ATLAS and CMS
- Uchi:** LHC Tier 2/3 ATLAS and CMS
- KEK:** Belle II Tier 1/2
- JUNO:** JUNO
- UNL:** Sites that are standalone VRFs

NOTES

- 1) ONLY links involved in LHCONE are shown
- 2) LHCOPN links are not shown on this diagram
- 3) For map explanation see "Interpreting the LHCONE Map" at <https://www.dropbox.com/sh/padfo88j01r3z/AA0B5K8RlSH9fCjA4eCtea7dl?dl=0>
- 4) GEANT and CANARIE have shutdown the peering between their VRF and KIAE, as a result of the Ukraine war.

Additional site information:

- RedIRIS Spain:** PIC-T1, CIEMAT-LCG2, UAM-LCG2
- GARR Italy:** INFN Bari, Catania, Frascati, Legnaro, Milano, Roma1, Torino
- RENATER France:** IN2P3, ALTAIS-T1, Belle II, LAPP, LPC, CPPM, IPHC, SUBATECH
- ARNES Slovenia:** SIGNET xx
- RoEduNet Romania:** UAIC xx, ISS, NIKHEK
- ULAKBIM Turkey:** Tubitak ATLAS-T2
- GRNET Greece:** Ioannina CMS-T2
- BREN Bulgaria:** BGGS-SUGrid CMS-T3
- KAUST Saudi Arabia:** DTN1

Performance monitoring with perfSONAR



Image: Hosts location on the world map

Aggregating and analyzing the data

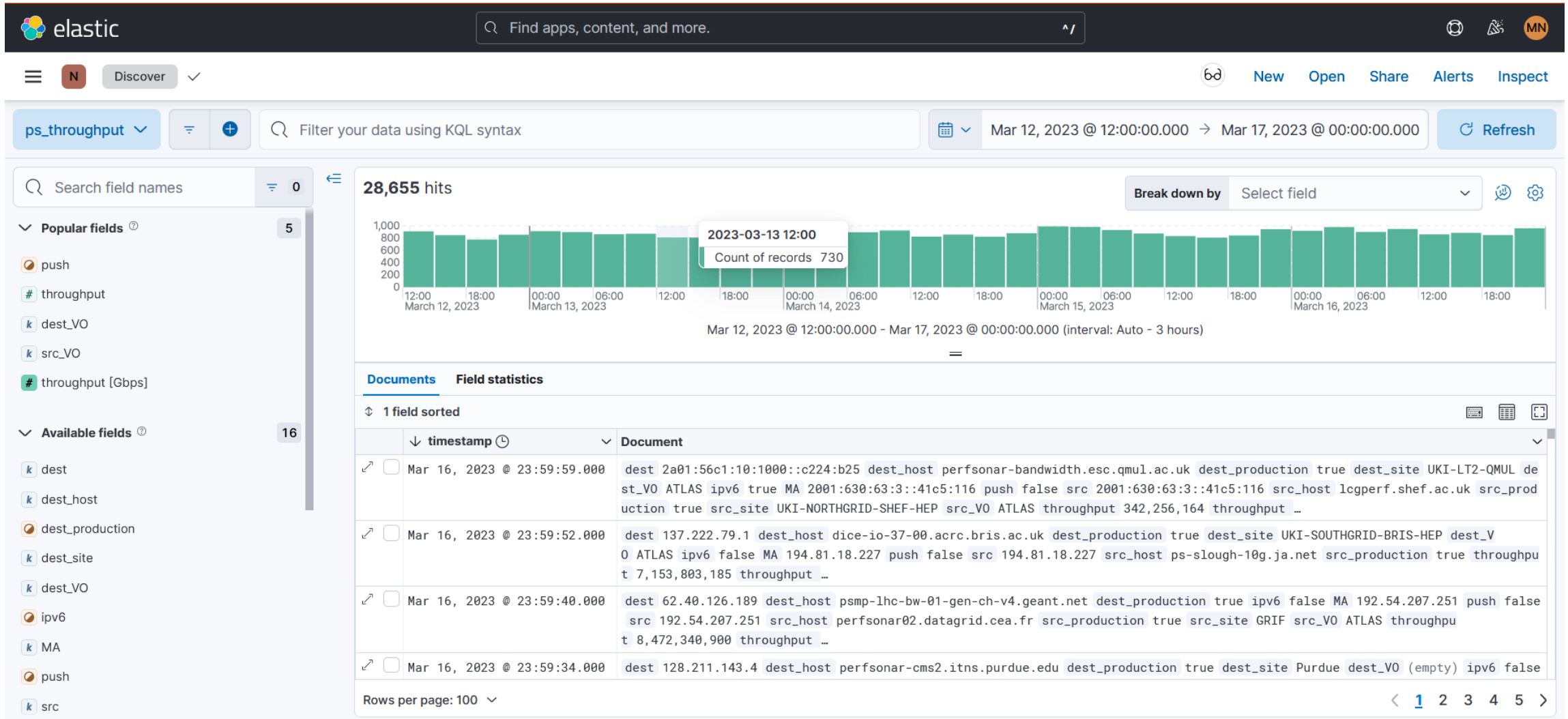


Image: Using Elasticsearch to monitor the status of the hosts

Network problems identification for
throughput measurements

Throughput alarms – dataset and one-hot encoding

	0	1	2
from	1672547280000	1672547280000	1672547280000
to	1672550880000	1672550880000	1672550880000
ipv6	False	False	False
value	6938.0	3191.0	9777.0
doc_count	1	1	1
dt	2023-01-01 04:28:00	2023-01-01 04:28:00	2023-01-01 04:28:00
ipv	ipv4	ipv4	ipv4
src_site	SWT2_CPB	SWT2_CPB	UNKNOWN
dest_site	ATLAS GREAT LAKES TIER-2	GRIF-LPNHE	T1_US_FINAL
src_host	psuta02.atlas- swt2.org	psuta02.atlas-swt2.org	perfonar.unl.edu
dest_host	psmsu02.aglt2.org	perfonar02.datagrid.cea.fr	psonar3.fnal.gov
alarm_created	0	0	0

12 rows × 595393 columns

one-hot
encoding
→

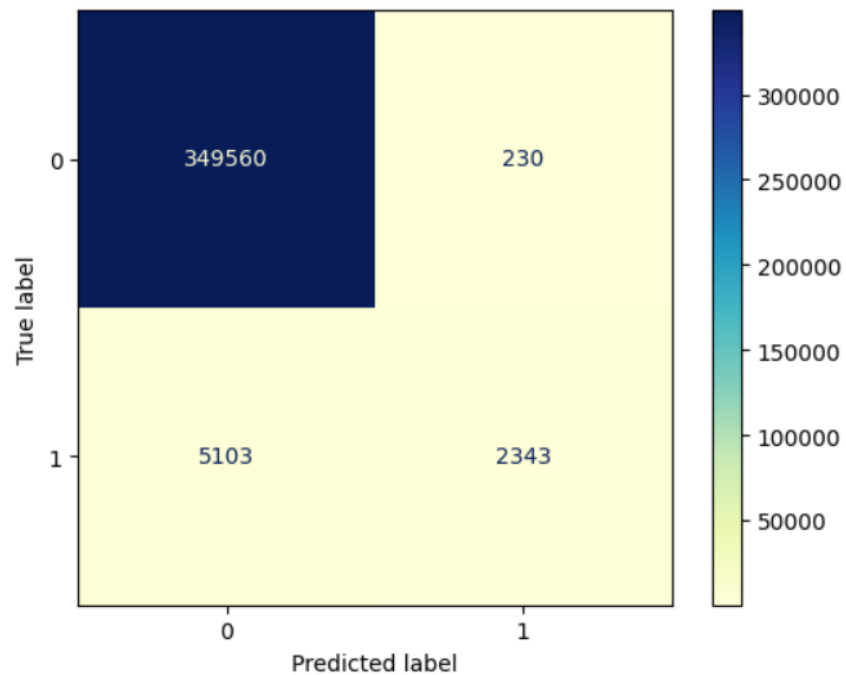
	0	1	2
from	1,672,547,280,000	1,672,547,280,000	1,672,547,280,000
to	1,672,550,880,000	1,672,550,880,000	1,672,550,880,000
ipv6	0	0	0
value	6,938	3,191	9,777
doc_count	1	1	1
...
dest_host_t2-pfsn1.jinr.ru	0	0	0
dest_host_t2-pfsn2.jinr.ru	0	0	0
dest_host_t2ps- bandwidth2.physics.ox.ac.uk	0	0	0
dest_host_tau.ijs.si	0	0	0
dest_host_uct2- net2.mwt2.org	0	0	0

367 rows × 595393 columns

Images: data on the throughput measurements fetched from the Elasticsearch for the period [from 2023-01-01 until 2023-05-31]

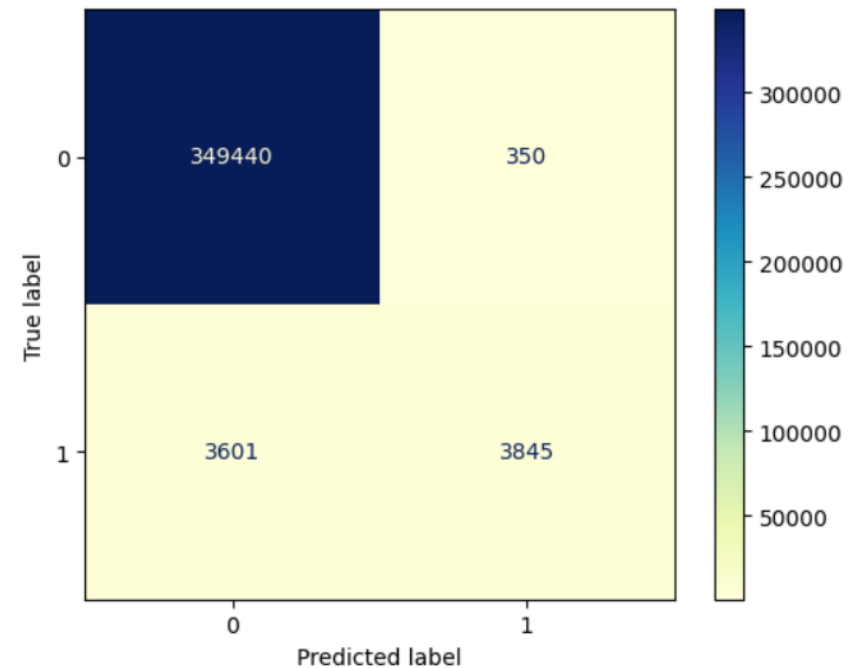
Throughput alarms – Random hyperparameters search (using f1 score)

Accuracy of the XGB Classifier: 98.51 %
F1 score of the XGB Classifier: 0.46771134843796786



RHS
→

Accuracy of the XGB Classifier: 98.89 %
F1 score of the XGB Classifier: 0.660596168714028



Images: Classification reports and confusion matrices for the test data

Throughput alarms – ML analysis site-specifically

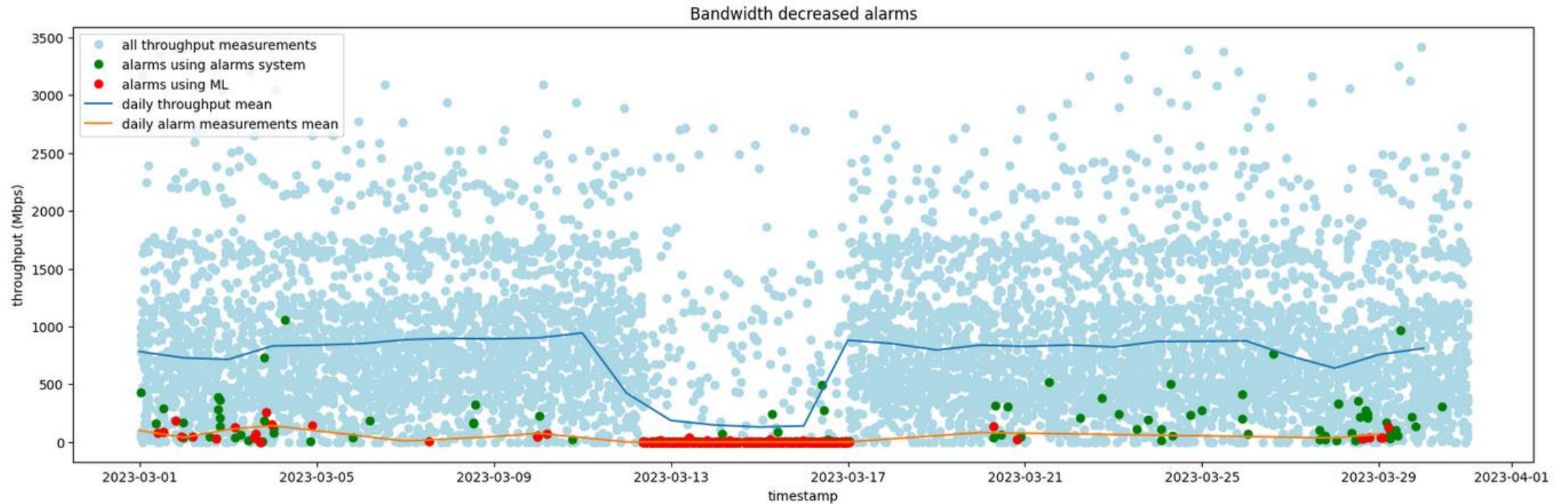


Image: TAIWAN-LCG2 site performance on the period of time where it had a known network issue

Throughput alarms – ML analysis site-specifically

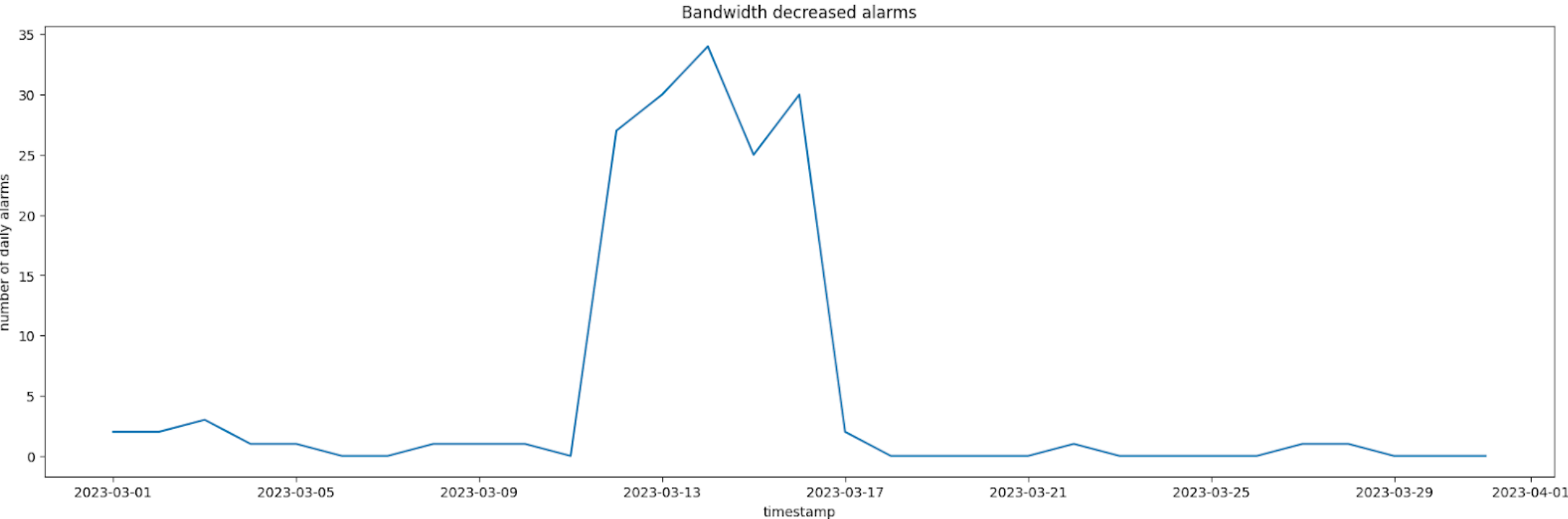


Image: TAIWAN-LCG2 site number of alarms aggregation by days

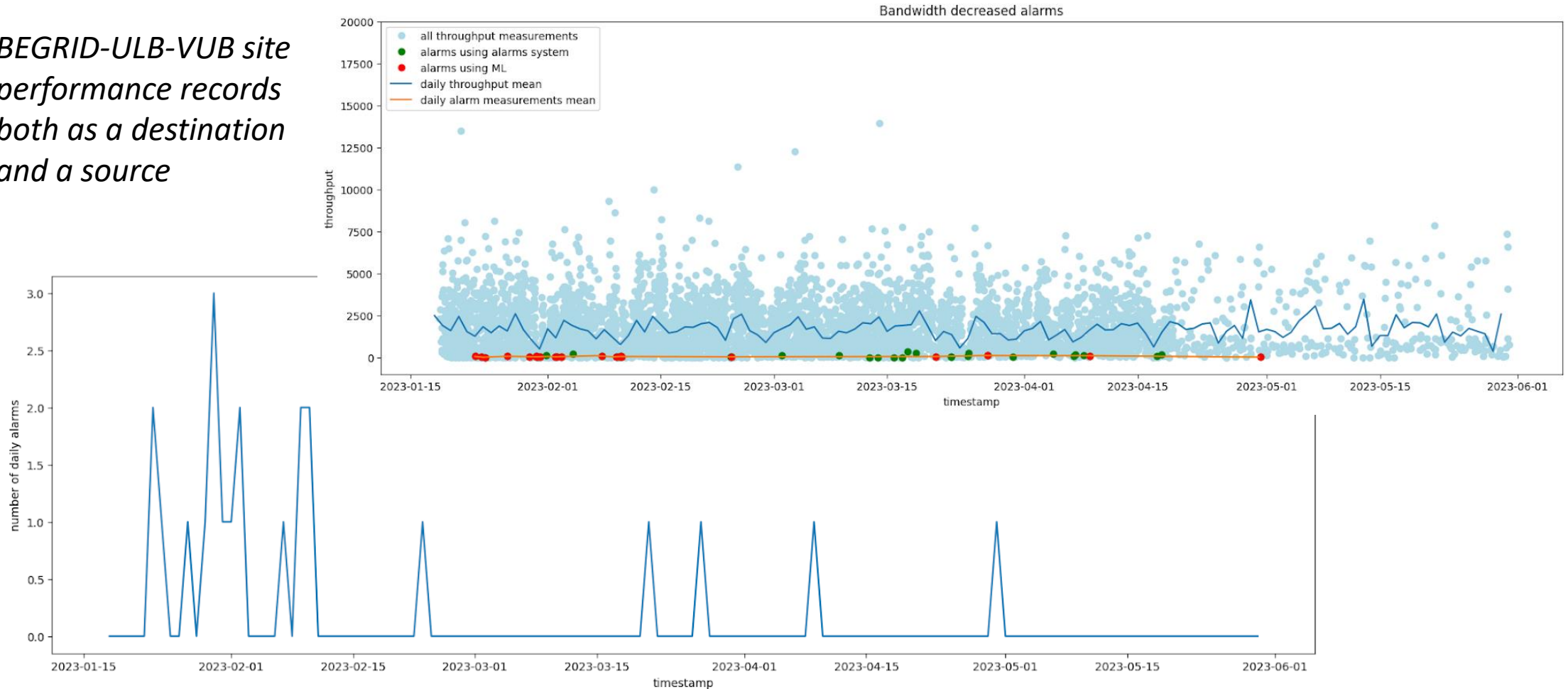
Throughput alarms – ML analysis through all sites

<i>USCMS-FNAL-WC1</i>	<i>45 alarms at 2023-02-05</i>	<i>28 alarms at 2023-02-06</i>	<i>27 alarms at 2023-02-07</i>	...
<i>UKI-LT2-QMUL</i>	<i>2 alarms at 2023-02-19</i>	<i>2 alarms at 2023-02-20</i>		
<i>TAIWAN-LCG2</i>	<i>64 alarms at 2023-03-14</i>	<i>62 alarms at 2023-03-15</i>	<i>62 alarms at 2023-03-16</i>	...
<i>T2_US_CALTECH</i>	<i>14 alarms at 2023-02-05</i>	<i>17 alarms at 2023-02-06</i>	<i>16 alarms at 2023-02-07</i>	...
<i>BEGRID-ULB-VUB</i>	<i>1 alarms at 2023-01-31</i>	<i>1 alarms at 2023-02-01</i>	<i>2 alarms at 2023-02-02</i>	...
<i>JINR-LCG2</i>	<i>37 alarms at 2023-02-08</i>	<i>19 alarms at 2023-02-09</i>		
<i>VANDERBILT</i>	<i>15 alarms at 2023-01-25</i>	<i>16 alarms at 2023-01-26</i>	<i>15 alarms at 2023-01-27</i>	...
...	<i>22 unique sites</i>

Diagram: Some of the 22 unique sites with 3-14 days of consecutive high number of daily alarms received from the ML analysis

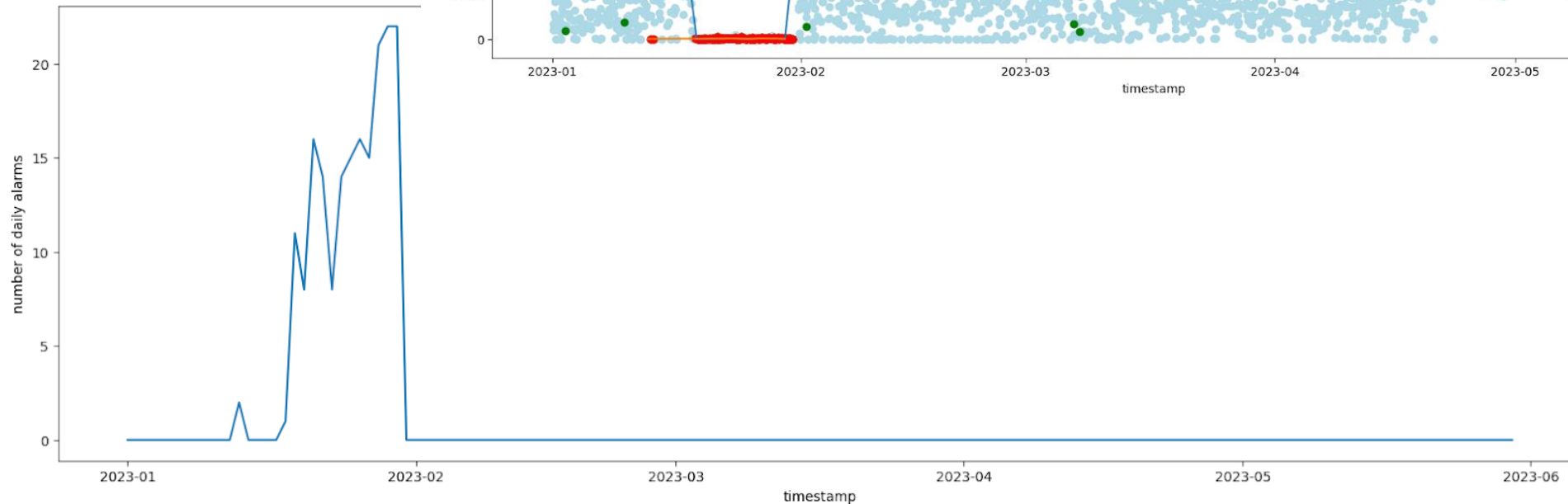
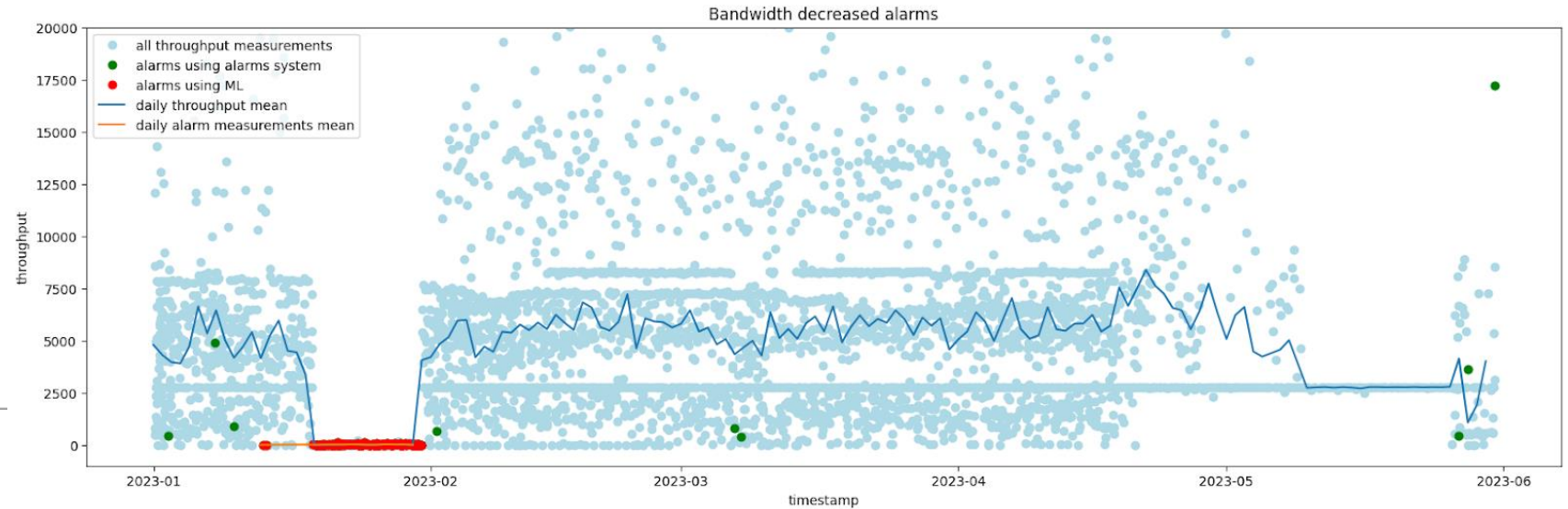
Throughput alarms – ML analysis through all sites (*BEGRID-ULB-VUB*)

BEGRID-ULB-VUB site performance records both as a destination and a source



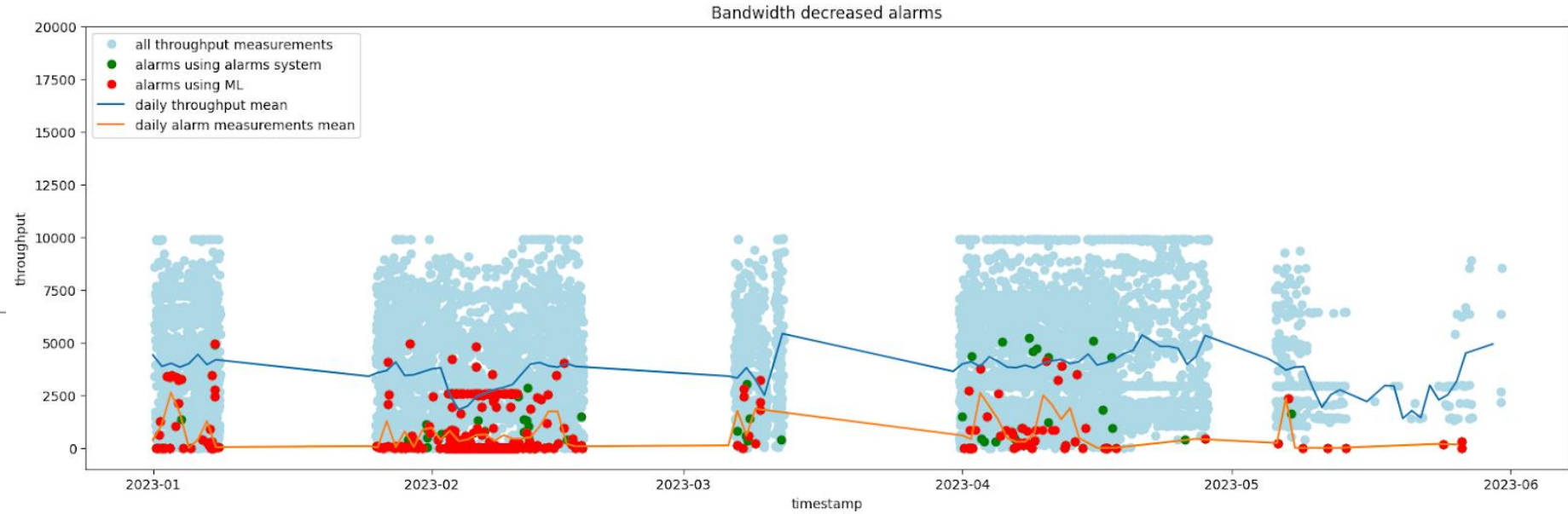
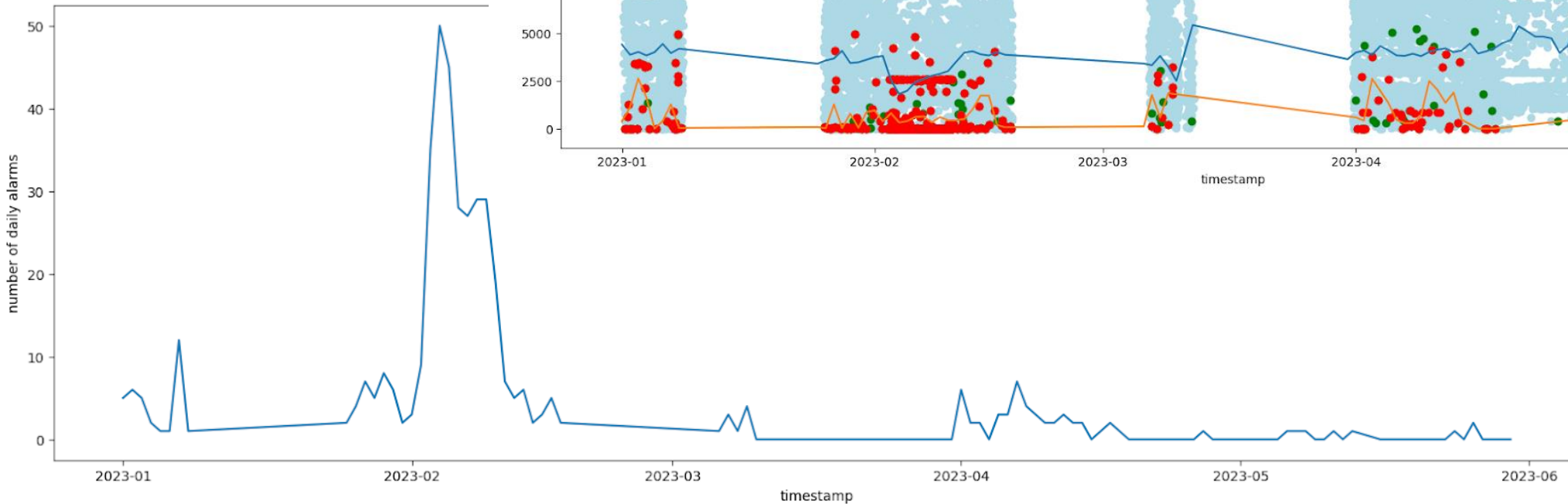
Throughput alarms – ML analysis through all sites (*VANDERBILT*)

VANDERBILT site performance records both as a destination and a source



Throughput alarms – ML analysis through all sites (*USCMS-FNAL-WC1*)

USCMS-FNAL-WC1 site performance records both as a destination and a source



Throughput alarms – ML analysis through all sites (*USCMS-FNAL-WC1*)

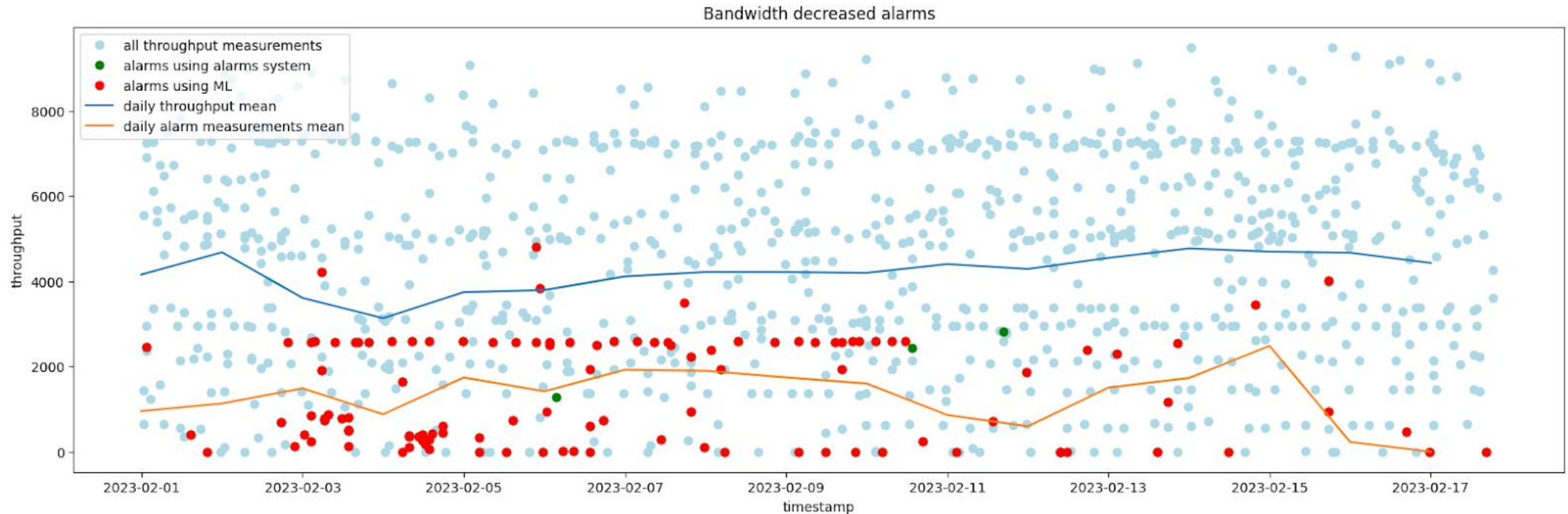
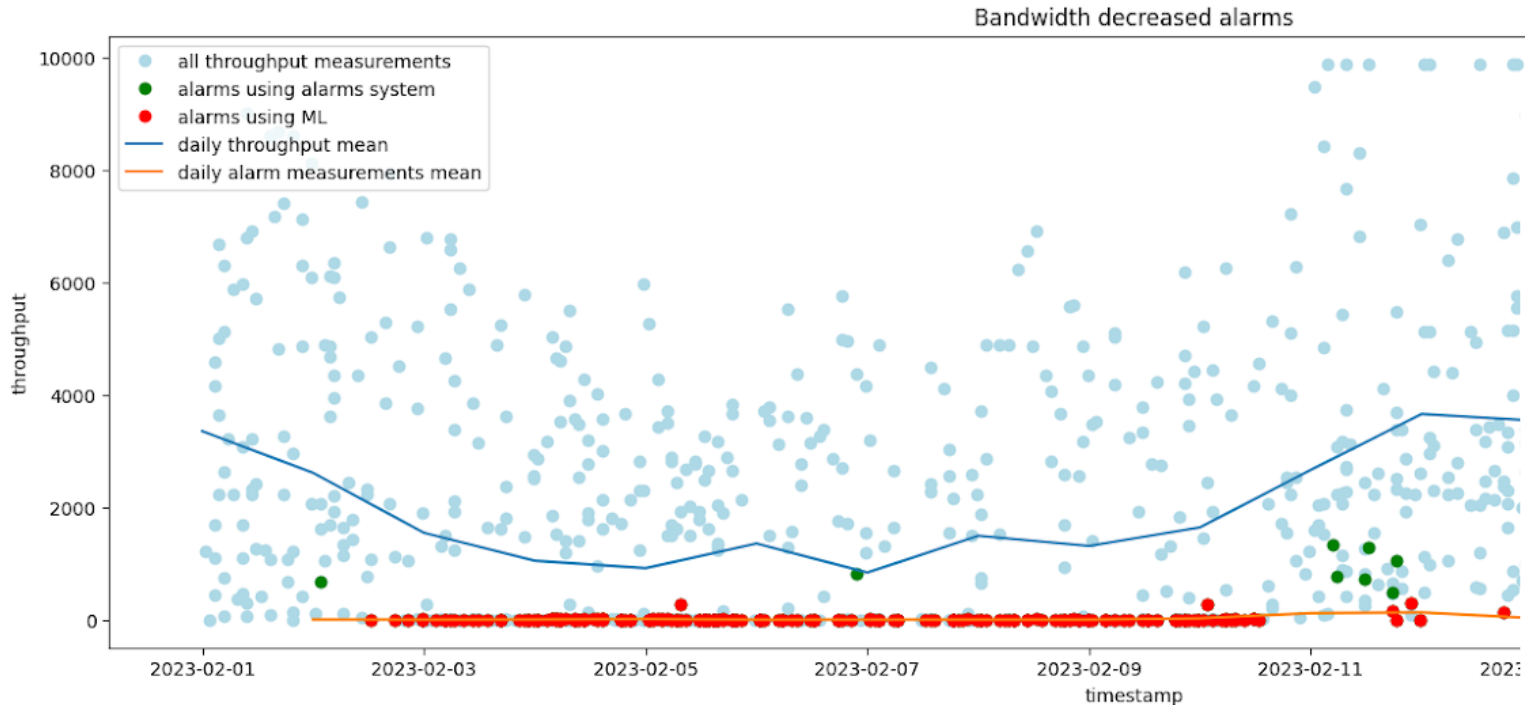


Diagram: *USCMS-FNAL-WC1* site performance records as a destination ONLY

Throughput alarms – ML analysis through all sites (*USCMS-FNAL-WC1*)



Amount of source - destination pairs for USCMS-FNAL-WC1 as a source:

Name of the destination site	Number of unique src-dest connections
CIT_CMS_T2	47
GLOW	40
IN2P3-CC-T3	33
NDGF-T1	29
RRC-KI-T1	3
RWTH-AACHEN	1
T2_US_FLORIDA	26
TAIWAN-LCG2	13

Diagram: *USCMS-FNAL-WC1* site performance records as a source ONLY

Creating alarms for packet loss measurements

Packet loss alarms – ML analysis through all sites

USCMS-FNAL-WC1	375 alarms at 2023-02-05	399 alarms at 2023-02-06	388 alarms at 2023-02-07	...
CBPF	58 alarms at 2023-03-25			
TAIWAN-LCG2	374 alarms at 2023-03-15	264 alarms at 2023-03-16		
CIT_CMS_T2	330 alarms at 2023-01-03			
GSI-LCG2	19 alarms at 2023-01-19			
SARA-MATRIX	218 alarms at 2023-04-18			
...

Diagram: Sites with a consecutive high number of daily alarms received from the ML analysis

Packet loss alarms – ML analysis through all sites (*TAIWAN-LCG2*)

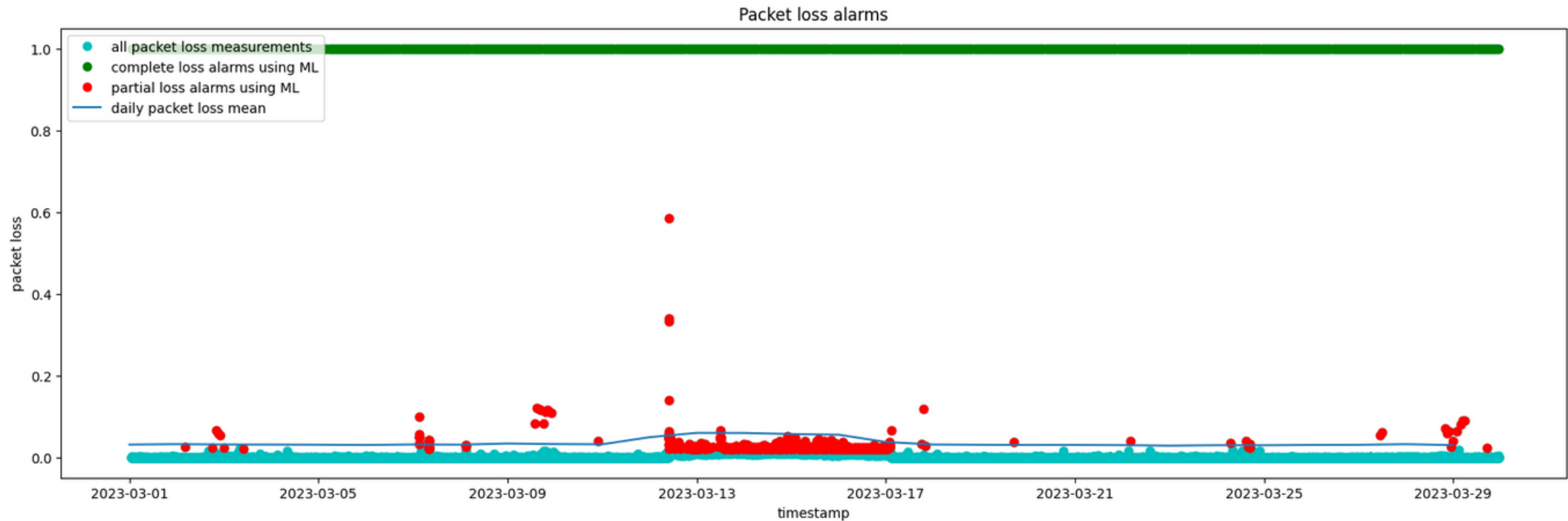


Image: TAIWAN-LCG2 site performance on the period of time where it had a known network issue

Packet loss alarms – comparison with throughput decrease alarms

VANDERBILT site had a very prominent alarm in the throughput analysis [from 2023-01-21 till 2023-01-30] but nothing here.

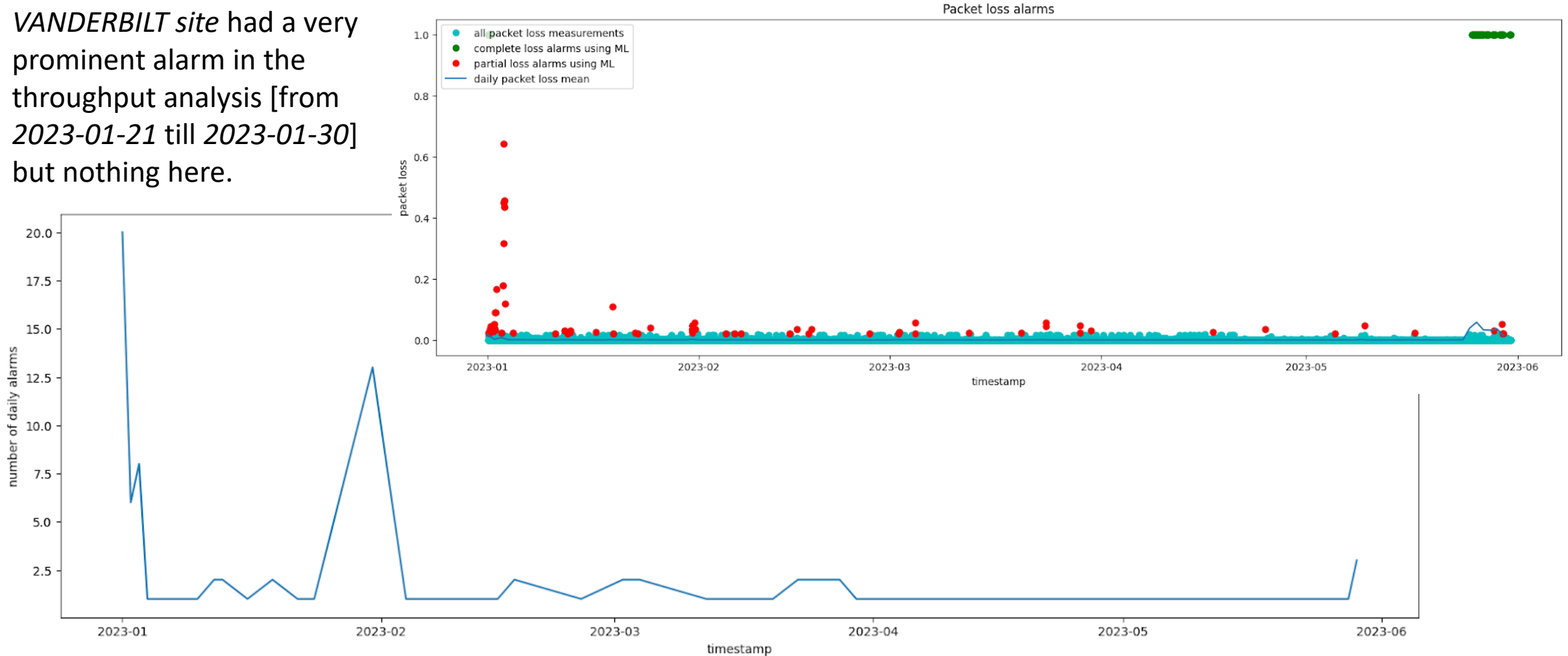
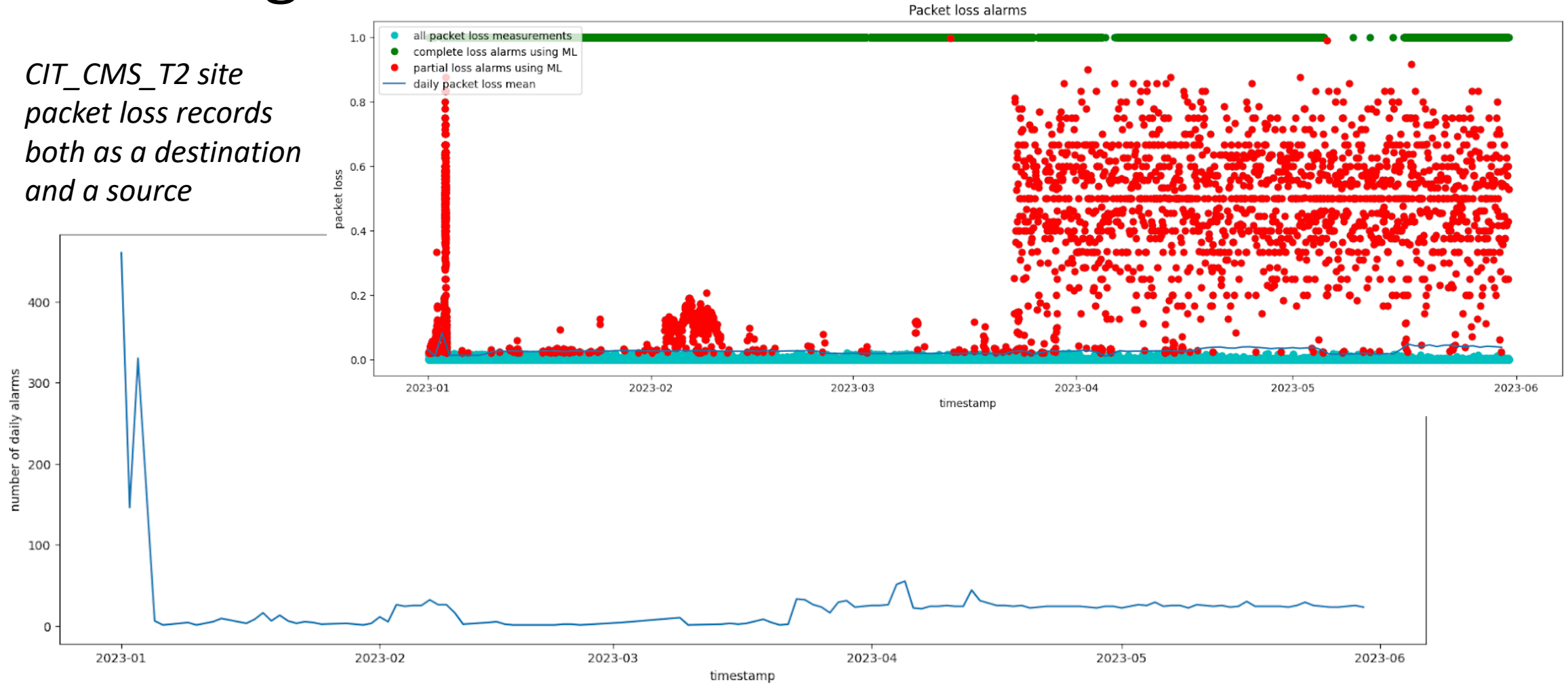


Image: *VANDERBILT* site packet loss records/alarms both as a destination

Packet loss alarms – comparison with throughput decrease alarms

*CIT_CMS_T2 site
packet loss records
both as a destination
and a source*



Making the project interactive using
Plotly Dash

Conclusions

- Using different techniques such as one-hot encoding suitable datasets were created from the data fetched from the Elasticsearch platform
- Using these datasets our own alarms were created using Machine Learning and existing automatic alarms system to label the faulty measurements
- Two Elasticsearch scopes were explored this way – throughput and packet loss measurements
- While comparing the results from both, instead of expected correlation between high packet loss and low throughput we had some cases of low throughput present and no high packet loss alarms and vice versa

Links

- LHCONE topology - <https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneMaps>
- PerfSONAR stats - <https://stats.perfsonar.net/d/spFwAQi4z/perfsonar-public?orgId=2>
- Elasticsearch data – [link](#)
- Dash page for exploring alarms – <https://ps-dash.uc.ssl-hep.org/search-alarms>

Thank you for attention!