

Semantic Metadata **Annotation** for Network **Anomaly** Detection

draft-netana-opsawg-nmrg-network-anomaly-semantics-01

Helps to test, validate and compare outlier detection, supports supervised and semi-supervised machine learning development, enables data exchange among network operators, vendors and academia, and make anomalies for humans apprehensible

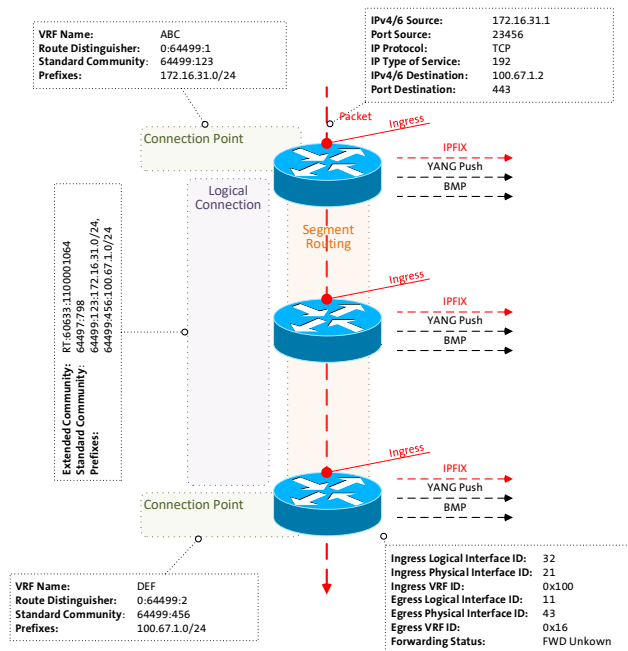
thomas.graf@swisscom.com
wanting.du@swisscom.com
alex.huang-feng@insa-lyon.fr
vincenzo.riccobene@huawei-partners.com
antonio.roberto@huawei.com

04. November 2023

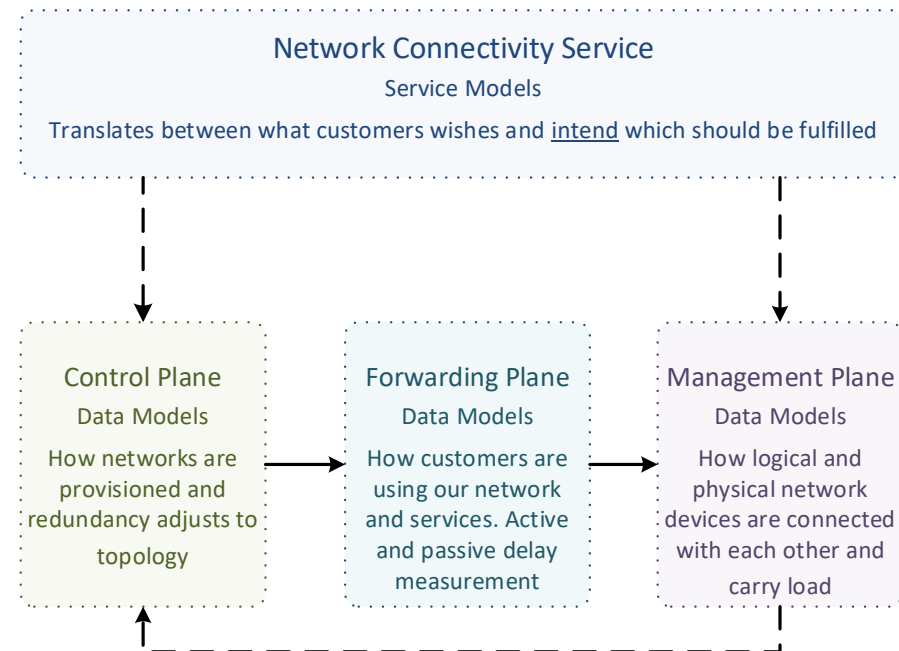
What to monitor

Which operational metrics are collected

« Network operators **connect customers in** routing tables called **VPN's** »

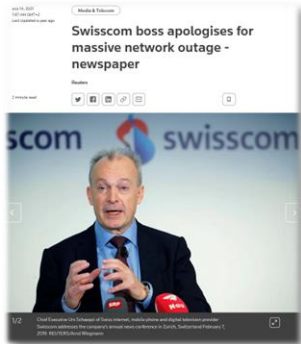
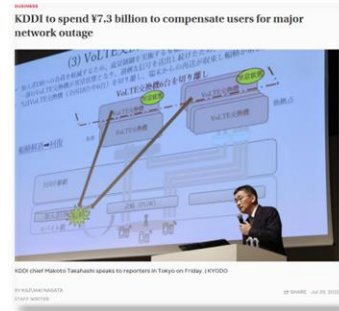


« Network Telemetry (RFC 9232) describes how to collect data from all 3 network planes efficiently »



Why to automate monitoring

Recognize network incidents faster than humans can



05 FEB 2023 | 08:23 AM UTC
Italy: TIM internet services interruption reported nationwide Feb. 5

TIM internet services interruption reported in Italy Feb. 5. Likely communication disruptions.

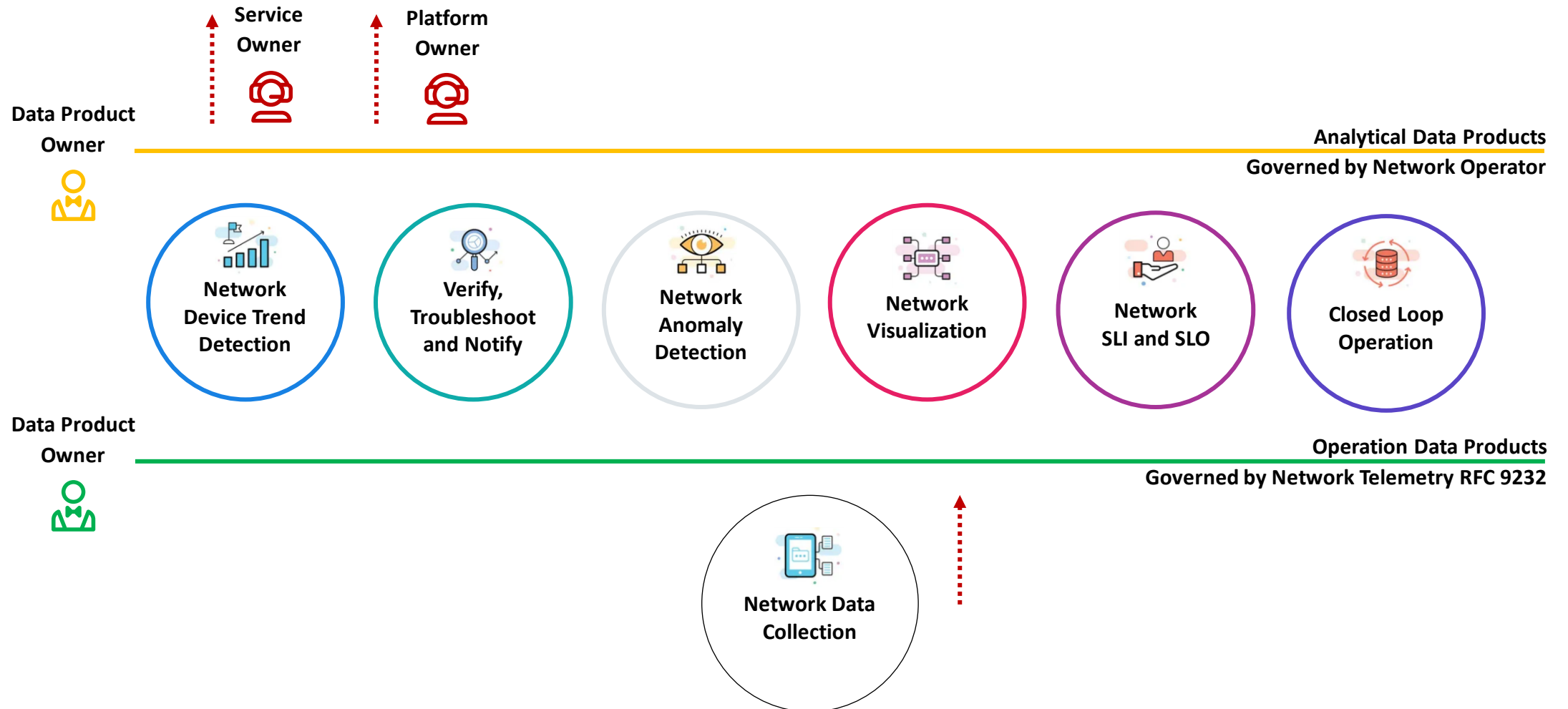
Informational Communications/technology Transportation ITA



« Customers are **always connected**, when VPN's **changing**, regardless due to operational or configurational reasons, network operators are **late to react** due to **missing visibility and automation** »

How to organize and collaborate with data

The Data Mesh Architecture enables Network Analytics use



What does Network Anomaly Detection mean

Monitor changes



Network Anomaly Detection

For VPNs, Network Anomaly Detection **constantly monitors and detects any network or device topology changes**, along with their associated forwarding consequences for customers as outliers. Notifications are sent to the Network Operation Center before the customer is aware of service disruptions. **It offers operational metrics for in-depth analysis**, allowing to understand on which platform the problem originates and facilitates problem resolution.



Answers

What changed and when, on which connectivity service, and how does it impact the customers?



Focuses

Provides meaningful connectivity service impact information before customer is aware of and support in root-cause analysis.



Data Mesh

Consumes operational real-time Forwarding Plane, Control Plane and Management Plane metrics and produces analytical alerts.



Direction

From connectivity service to network platform.

Presented in ANRW 2023
At IETF 117 San Francisco

« A more detailing paper
will be submitted soon to
IEEE Transactions on
Network and Service
Management»

The screenshot shows a web browser window with the URL <https://anrw23.hotcrp.com/doc/anrw23-paper8.pdf>. The page content includes the title, authors, abstract, and introduction. The authors listed are Alex Huang Feng, Pierre Francois, Stéphane Frenot, Thomas Graf, Wanting Du, and Paolo Lucente. The abstract describes an architecture for anomaly detection in large VPN networks. The introduction discusses the challenges of service level agreements and the need for anomaly detection.

Daisy: Practical Anomaly Detection in large BGP/MPLS and BGP/SRv6 VPN Networks

Alex Huang Feng
alex.huang-feng@insa-lyon.fr
CITI Laboratory, INSA Lyon
Lyon, France

Pierre Francois
pierre.francois@insa-lyon.fr
CITI Laboratory, INSA Lyon
Lyon, France

Stéphane Frenot
stephane.frenot@insa-lyon.fr
CITI Laboratory, INSA Lyon
Lyon, France

Thomas Graf
thomas.graf@swisscom.com
Swisscom
Zurich, Switzerland

Wanting Du
wanting.du@swisscom.com
Swisscom
Zurich, Switzerland

Paolo Lucente
paolo@pmacct.net
pmacct.net
Spain

ABSTRACT
We present an architecture aimed at performing Anomaly Detection for BGP/MPLS VPN services, at scale. We describe the challenges associated with real time anomaly detection in modern, large BGP/MPLS VPN and BGP/IPv6 Segment Routing VPN deployments. We describe an architecture required to collect the necessary routing information at scale. We discuss the various dimensions which can be used to detect anomalies, and the caveats of the real world impacting the level of difficulty of such anomaly detection and network modeling. We argue for rule-based anomaly detection assisted with machine learning based customer classification is best suited given the current state of the art. Finally, we review the current IETF contributions which are required to benefit from a fully open, standard, architecture.

ACM Reference Format:
Alex Huang Feng, Pierre Francois, Stéphane Frenot, Thomas Graf, Wanting Du, and Paolo Lucente. 2023. Daisy: Practical Anomaly Detection in large BGP/MPLS and BGP/SRv6 VPN Networks. In *Proceedings of ACM Conference (Conference '17)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/nmmmmn.nmmmmn>

1 INTRODUCTION

Customers subscribing to BGP/MPLS VPN services usually come along with stringent Service Level Agreements. Consequently, Service Providers must be capable of detecting

anomalies in their services in a timely fashion, while accommodating for scale. Around 10 thousand L3 VPNs in our Swisscom use case. Long-lasting outages, detected by the customer before the service provider, are detrimental to the perception of service quality, and may dramatically impact the customer business.

The goal of the presented architecture is to provide an anomaly detection solution that scales while, being flexible on the following aspects: (i) the dimensions that must be used to detect anomalies are multiple; (ii) VPN customers wear different profiles in terms of normal and abnormal values for such dimensions; (iii) the amount of information collected to produce values for such dimensions is extremely large in such deployments: around 175 thousand messages/second in our use case; (iv) the operating costs for managing an anomaly detection solution must be kept low; and (v) the networking platforms providing the service may come from different vendors and have different monitoring capabilities.

The remainder paper is structured as follows. In section 2, we define what is considered a network anomaly and present the associated challenges behind its detection. In Section 3, we describe the Daisy architecture. In Section 4, we review the ongoing IETF efforts aimed at filling the gaps for a fully open, standard, Anomaly Detection (AD) implementation. And finally, in section 5, we present the first results of Daisy deployment at Swisscom.

2 PROBLEM STATEMENT

We describe some of the challenges associated with customer diversity, and a non-exhaustive list of anomalies targeted by the base recipes from our limited proof of concept deployment setup.

2.1 What is an Anomaly?

An anomaly is defined in this project as follows: *Whatever would let an operator frown and investigate when looking*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference '17, July 2017, Washington, DC, USA
© 2023 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nmmmmn.nmmmmn>

What our motivation is

Automate learn and improve

From network incidents postmortems we network operators **learn and improve** so does network anomaly detection and supervised and semi-supervised machine learning.

The more network incidents are observed, the more we can improve. With more incidents the **postmortem process needs be automated, let's get organized** first by defining human and machine-readable metadata semantics and annotate operational and analytical data.

Let's get further organized by exchanging standardized labeled network incident data among network operators, vendors and academia to **collaborate on academic research**.

« **The community working on Network Anomaly Detection** is probably the only group **wishing for more network incidents** »

What is a symptom and how to categorize them

From action to reason to cause

Action: Which action the network node performed for a packet in the forwarding plane, a path or adjacency in the control plane or state or statistical changes in the management plane.

Reason: For each reason one or more actions describing why this action was used. From drop unreachable, administered, and corrupt in forwarding plane, to reachability withdraw and adjacency teared down in control plane, to Interface down, errors or discard in management plane.

Cause: For each reason one or more causes describes why the action was chosen. From missing next-hop and link-layer information in forwarding plane, to reachability withdrawn due to peer down or path no longer redistributed.

« Symptoms are categorized in **which plane** they have been **observed**, their **action, reason and cause** »

Questions to the audience

Do you care?

Network Operators: Do you agree that today's actions; traffic is dropped, path is withdrawn and interface down, are always exposed through Network Telemetry. But reasons and causes, dropped due to unreachable next-hop, withdrawn due to peer down, interface down due to missing signal, are rarely exposed to telemetry would be most interesting?

Network Vendors: Is the assumption correct that a when network service process, routing process and withdrawing a path occur, most of the time the vendor knows why it acts that way, and could potential make this reason and cause information available?

Academia: Would it help if network operators would provide well defined labeled operational and analytical data to enable and validate their research?

Everybody: Should these symptoms be clearly described and standardized for a common terminology so that operators, researchers and anomaly detection systems alike understand their meaning and learn and act accordingly?

Outliers in Anomaly Detection

From global to contextual to collective

Global outliers: An outlier is considered "global" if its behavior is outside the entirety of the considered data set.

Contextual outliers: An outlier is considered "contextual" if its behavior is within a normal (expected) range, but it would not be expected based on some context. Context can be defined as a function of multiple parameters, such as time, location, etc.

Collective outliers: An outlier is considered "collective" if the behavior of each single data point that are part of the anomaly are within expected ranges (so they are not anomalous, it's either a contextual or a global sense), but the group taking all the data points together, is.

« **Collective outliers** are important because networks are connected. Through **different planes interconnected** symptoms from various angles can be observed »

Annotate Operation Data

YANG Module

```
module: ietf-symptom-semantic-metadata
+--rw symptom
  +--rw id          yang:uuid
  +--rw event-id   yang:uuid
  +--rw description          string
  +--rw start-time yang:date-and-time
  +--rw end-time   yang:date-and-time
  +--rw confidence-score    float
  +--rw concern-score?     float
+--rw tags* [key]
  | +--rw key    string
  | +--rw value  string
+--rw (pattern)?
  | +--:(drop)
  | | +--rw dropempty
  | +--:(spike)
  | | +--rw spike          empty
  | +--:(mean-shift)
  | | +--rw mean-shift     empty
  | +--:(seasonality-shift)
  | | +--rw seasonality-shift empty
  | +--:(trend)
  | | +--rw trend          empty
  | +--:(other)
  |   +--rw other          string
+--rw source
  +--rw (source-type)
  | +--:(human)
  | | +--rw human          empty
  | +--:(algorithm)
  |   +--rw algorithm      empty
  +--rw name?              string
```

- **Symptoms** describe what changed in the network for what reason and cause with which concern score from when to when.
- **Tags** describes in which network plane, which action, reason and cause was observed.
- **Pattern** describes the measurement pattern over time of the time series data.
- **Source** describes which system **observed** the outlier. A human or a network anomaly detection system.

Annotate Analytical Data

YANG Module

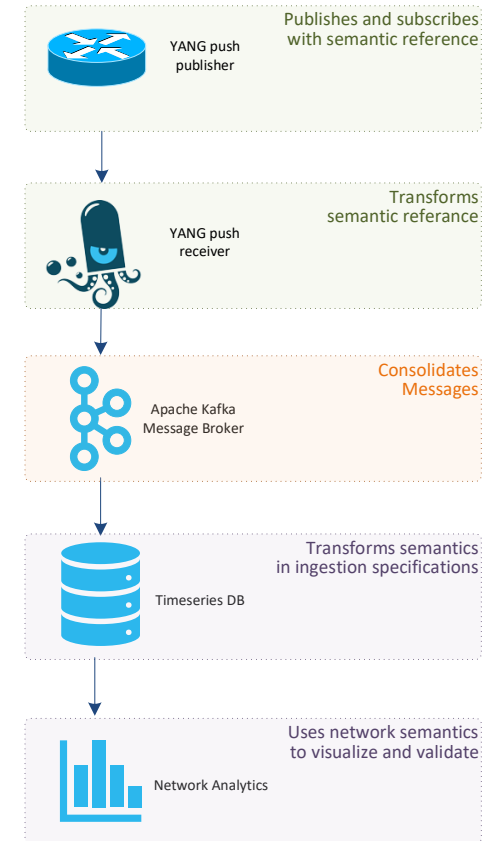
```
module: ietf-incident-semantic-metadata
+--rw incident
  +--rw id          yang:uuid
  +--rw description string
  +--rw start-time  yang:date-and-time
  +--rw end-time    yang:date-and-time
+--rw symptoms* []
  | +--rw symptom
  |   +--rw id          yang:uuid
  |   +--rw event-id   yang:uuid
  |
  <snip>
+--rw source
  +--rw (type)
  | +--:(human)
  | | +--rw human      empty
  | +--:(algorithm)
  |   +--rw algorithm  empty
  +--rw name?         string
```

- **Incidents** has a unique ID and description with a start and end time and a concern score.
- **Symptoms** describe what changed in the network for what reason and cause with which concern score from when to when.
- **Source** describes which system **reported** the outlier. A human or a network anomaly detection system.

Semantic Metadata Annotation for Network Anomaly Detection

Next steps

- Do you realize the benefit of having standardized semantic metadata annotation for Network Anomaly Detection and how it helps network operators, vendor and academia to collaborate?
- -> **What are your thoughts and comments?**
- This document looks for a community and working group who have interest in Network Anomaly Detection, bridging network and data engineering, operator, vendors and academia, by writing the **semantics and ontology of network symptoms for operational and analytical data**.
- This work will unveil what is missing in Network Telemetry data and provide input for other documents to enable a more detailed and holistic view from networks.



thomas.graf@swisscom.com
wanting.du@swisscom.com
alex.huang-feng@insa-lyon.fr
vincenzo.riccobene@huawei-partners.com
antonio.roberto@huawei.com

04. November 2023