Semantic Metadata Annotation for Network Anomaly Detection
draft-netana-opsawg-nmrg-network-anomaly-semantics-01

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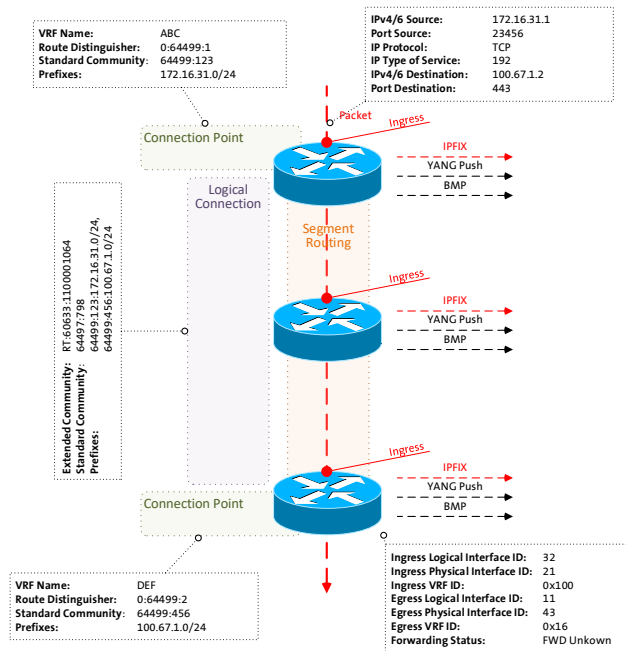and make anomalies for humans apprehensible

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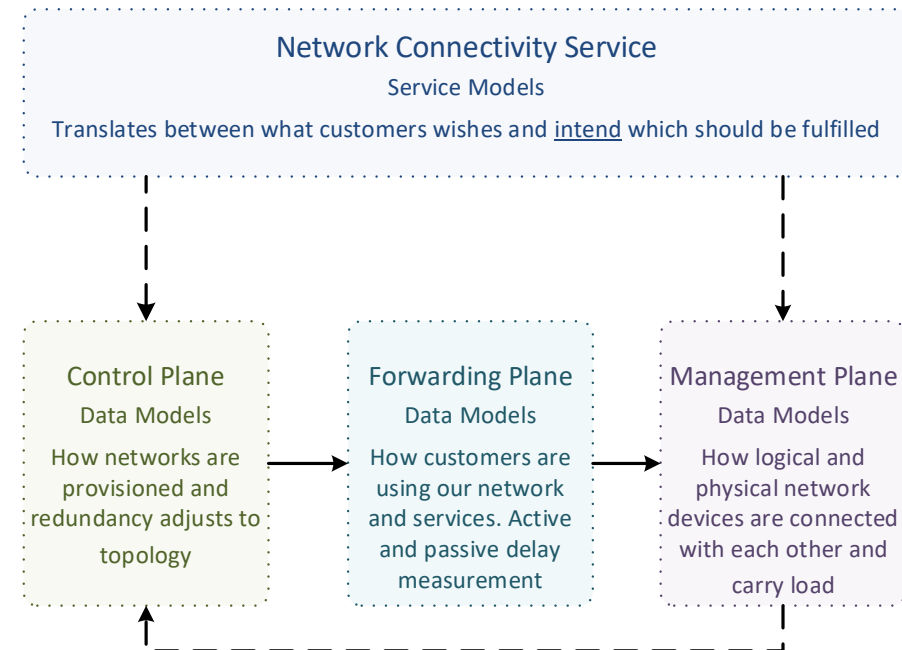antonio.roberto@huawei.com

01. November 2023

1

What to monitor
Which operational metrics are collected

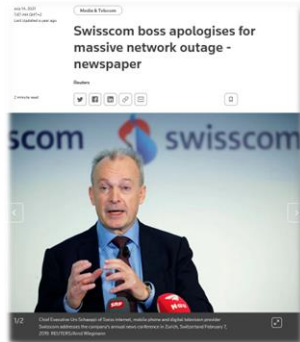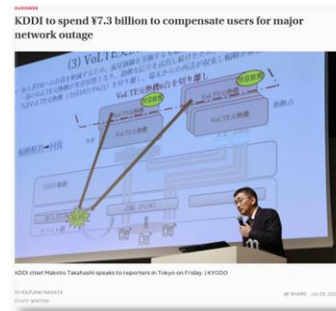« Network operators connect customers in routing tables called VPN's »

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

VRF Name:	ABC
Route Distinguisher:	0:64499:1
Standard Community:	64499:123
Prefixes:	172.16.31.0/24

IPv4/6 Source:	172.16.31.1
Port Source:	23456
IP Protocol:	TCP
IP Type of Service:	192
IPv4/6 Destination:	100.67.1.2
Port Destination:	443

Connection Point

Packet · Ingress

IPFIX
YANG Push
BMP

Logical Connection

Segment Routing

Ingress

IPFIX
YANG Push
BMP

Extended Community: RT:60633:1100001064
Standard Community: 64997:798
Prefixes: 64499:123:172.16.31.0/24, 64499:456:100.67.1.0/24

Ingress

IPFIX
YANG Push
BMP

Connection Point

VRF Name:	DEF
Route Distinguisher:	0:64499:2
Standard Community:	64499:456
Prefixes:	100.67.1.0/24

Ingress Logical Interface ID:	32
Ingress Physical Interface ID:	21
Ingress VRF ID:	0x100
Egress Logical Interface ID:	11
Egress Physical Interface ID:	43
Egress VRF ID:	0x16
Forwarding Status:	FWD Unkown

Network Connectivity Service

Service Models

Translates between what customers wishes and intend which should be fulfilled

Control Plane
Data Models

How networks are provisioned and redundancy adjusts to topology

Forwarding Plane
Data Models

How customers are using our network and services. Active and passive delay measurement

Management Plane
Data Models

How logical and physical network devices are connected with each other and carry load

Why to automate monitoring
Recognize network incidents faster than humans can

« 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

Service
Owner

Platform
Owner

**Data Product
Owner**

Analytical Data Products

Governed by Network Operator

**Network
Device Trend
Detection**

**Verify,
Troubleshoot
and Notify**

**Network
Anomaly
Detection**

**Network
Visualization**

**Network
SLI and SLO**

**Closed Loop
Operation**

**Data Product
Owner**

Operation Data Products

Governed by Network Telemetry RFC 9232

**Network Data
Collection**

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

« A more [detailing paper](link) was submitted to IEEE Transactions on Network and Service Management»

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 an outlier and how to categorize them
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 »

What is a symptom and how to categorize them
From action to reason to relation

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 action one or more reasons 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.

Relation: For each reason one or more relation describes the cause 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?

Annotate Operation Data
YANG Module

```
module: ietf-symptom-semantic-metadata
  +--rw symptom
    +--rw id             yang:uuid
    +--rw description              string
    +--rw (plane)?
    |  +--:(forwarding-plane)
    |  |  +--rw forwarding-plane    empty
    |  +--:(control-plane)
    |  |  +--rw control-plane       empty
    |  +--:(management-plane)
    |     +--rw management-plane    empty
    +--rw condition?string
    +--rw action?    string
    +--rw cause?     string
    +--rw entity_type?             yang:uuid
    +--rw (outlier-type)?
    |  +--:(global)
    |  |  +--rw global             empty
    |  +--:(contextual)
    |  |  +--rw contextual         empty
    |  +--:(collective)
    |  |  +--rw collective         empty
    |  +--:(none)
    |     +--rw other              empty
    +--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
```

- **Symptoms** describe what changed in the network for what reason and cause with which concern score from when to when.

- **Plane** describes in which network plane it was observed; Forwarding, Control or Management Plane.

- **Outlier-Type** describes which type of outlier it is; Global, Contextual or Collective.

- **Pattern** describes the measurement pattern over time of the time series data.

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 concern              uint8
    +--rw events* [id]
    |  +--rw id            yang:uuid
    |  +--rw symptoms* [id]
    |  |  +--rw id            yang:uuid
    |  |  +--rw symptom
    |  |  |  +--rw id        yang:uuid

<continues>

    |  |  +--rw entity-id?    yang:uuid
    |  |  +--rw start-time    yang:date-and-time
    |  |  +--rw end-time      yang:date-and-time
    |  |  +--rw concern?      uint8
    |  +--rw source
    |     +--rw (source-type)
    |     |  +--:(human)
    |     |  |  +--rw human        empty
    |     |  +--:(algorithm)
    |     |     +--rw algorithm    empty
    |     +--rw name?               string
    +--rw remediation-action?   string
```
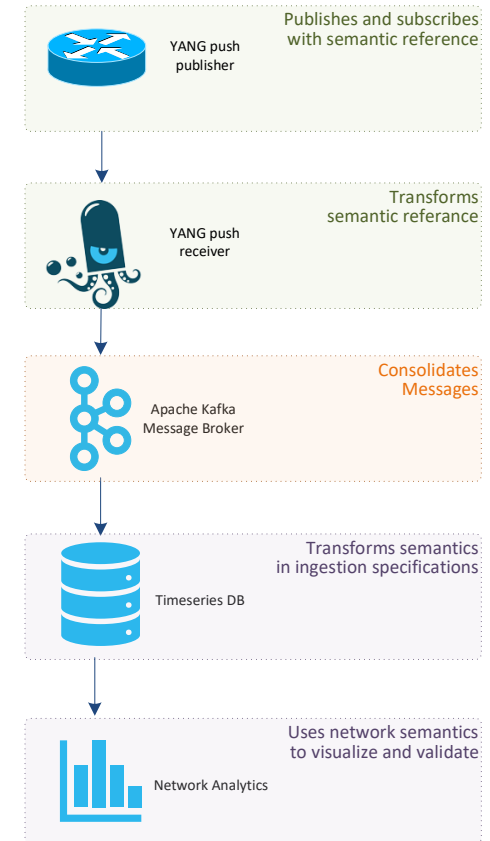
- **Incidents** has a unique ID and description with a start and end time and a concern score.

- **Symptoms** describe what changed in the network for what reason and cause with which concern score from when to when.

- **Source** describes who detected 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.

YANG push publisher — Publishes and subscribes with semantic reference

YANG push receiver — Transforms semantic referance

Apache Kafka Message Broker — Consolidates Messages

Timeseries DB — Transforms semantics in ingestion specifications

Network Analytics — Uses network semantics to visualize and validate

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

01. November 2023