

YANG models for ACTN & TE Performance Monitoring Telemetry and Network Autonomics

draft-ietf-teas-actn-pm-telemetry-autonomics-00

Young Lee, Dhruv Dhody, Satish K, Ricard Vilalta, Daniel King, Daniele Ceccarelli

Overview

YANG data models that support:

- Performance Monitoring (PM) Telemetry and scaling intent mechanism for TE-Tunnels and VNs to allow customers to subscribe to certain KPI PM.
 - *ietf-te-kpi-telemetry*
 - *ietf-vn-te-kpi-telemetry*
- Customer to subscribe and monitor KPI of interest on a particular TE tunnel or a VN.
- Customer could also program autonomic scaling intent

Recent Changes

- Based on the feedback during and before WG adoption
 - Added various clarifications in Introduction and Terminology to set the stage
 - Added clarity on terminology used
- Describe network autonomics in detail with examples.
- Reference to Notification mechanism in NETCONF.
- Updated security & IANA sections
- Yang model changes
 - Add references
 - Follow style guide for yang models
- Thanks to Dieter Beller, Lou Berger, Tom Petch and others who provided comments

Terminology

Key Performance Data

- This refers to a set of data the customer is interested in monitoring for their instantiated VNs or TE-tunnels.

Scaling

- This refers to the network ability to re-shape its own resources. Scale out refers to improve network performance by increasing the allocated resources, while scale in refers to decrease the allocated resources, typically because the existing resources are unnecessary.

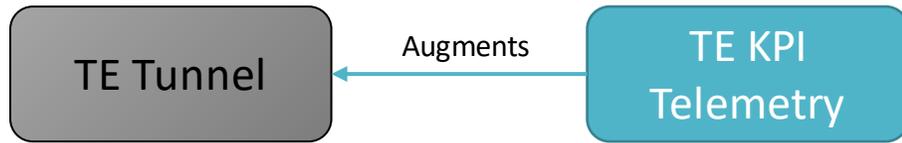
Scaling Intent

- To declare scaling conditions, scaling intent is used. Specifically, scaling intent refers to the intent expressed by the client that allows the client to program/configure conditions of their key performance data either for scaling out or scaling in. Various conditions can be set for scaling intent on either VN or TE-tunnel level.

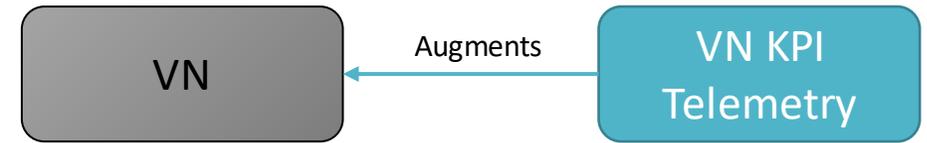
Network Autonomics

- This refers to the network automation capability that allows client to initiate scaling intent mechanisms and provides the client with the status of the adjusted network resources based on the client's scaling intent in an automated fashion.

Yang Models Design



- TE KPI Telemetry model provides the TE tunnel level performance monitoring.
- Augment the TE tunnel State with performance attributes
 - Use the notification subscription
- Scaling Intent configurations for scaling in/out programming by the customer.



- VN KPI Telemetry model provides the VN level aggregated performance monitoring.
- Augment the VN state as well as individual VN-member state with performance attributes.
 - Use notification subscription
- Scaling Intent configurations at the VN level for scaling in/out programming by the customer.

Autonomic Scaling Intent Mechanism

- Scaling intent configuration mechanism allows the client to configure automatic scale-in and scale-out mechanisms on both the TE-tunnel and the VN level.
- Various conditions can be set for auto-scaling based on the PM telemetry data.
 - scale-out-intent or scale-in-intent: whether to scale-out or scale-in.
 - performance-type: performance metric type (e.g., one-way-delay, one-way-delay-min, one-way-delay-max, two-way-delay, two-way-delay-min, two-way-delay-max, utilized bandwidth, etc.)
 - threshold-value: the threshold value for a certain performance-type that triggers scale-in or scale-out.
 - scaling-operation-type: in case where scaling condition can be set with one or more performance types, then scaling-operation-type (AND, OR, MIN, MAX, etc.) is applied to these selected performance types and its threshold values.
 - Threshold-time: the duration for which the criteria must hold true.
 - Cooldown-time: the duration after a scaling action has been triggered, for which there will be no further operation.

Illustration of Scaling Mechanism

```
module: ietf-te-kpi-telemetry
augment /te:te/te:tunnels/te:tunnel:
+rw te-scaling-intent
| +rw scale-in-intent
| | +rw threshold-time?      uint32
| | +rw cooldown-time?      uint32
| | +rw scale-in-operation-type? scaling-criteria-operation
| | +rw scaling-condition* [performance-type]
| |   +rw performance-type    identityref
| |   +rw threshold-value?    string
| |   +rw te-telemetry-tunnel-ref? -> /te:te/tunnels/tunnel/name
+rw scale-out-intent
| +rw threshold-time?      uint32
| +rw cooldown-time?      uint32
| +rw scale-out-operation-type? scaling-criteria-operation
| +rw scaling-condition* [performance-type]
|   +rw performance-type    identityref
|   +rw threshold-value?    string
|   +rw te-telemetry-tunnel-ref? -> /te:te/tunnels/tunnel/name
```

Let say the client wants to set the scaling out operation based on two performance-types (e.g., two-way-delay and utilized-bandwidth for a te-tunnel),

it can be done as follows:

1. Set **Threshold-time**: 3600 (sec) (duration for which the criteria must hold true)
2. Set **Cooldown-time**: 60 (sec) (the duration after a scaling action has been triggered, for which there will be no further operation)
3. Set **AND** for the **scale-out-operation-type**

List 1: Scaling Condition for Two-way-delay

- performance type: Two-way-delay
- threshold-value: 300 mile-seconds

List 2: Scaling Condition for Utilized bandwidth

- performance type: Utilized bandwidth
- threshold-value: 300 megabytes

The above two criteria have to meet at the same time to trigger scale-out operation.



Thanks!