OPSAWG Working Group
Internet Draft
Category: Experimental

Ram(Ramki) Krishnan
Support Vectors

Expires: April 2017

December 23, 2016

In-band Telemetry for a Proactive SLA Monitoring Framework

draft-krishnan-opsawg-in-band-pro-sla-00

Abstract

The goal of in-band telemetry is to drive per packet, per hop real-time monitoring for the infrastructure towards achieving a programmable proactive SLA monitoring framework. Some of the key aspects from a switch/NIC perspective are - ingress/egress timestamp (latency), queue depth, bandwidth etc. Some of the key aspects from a server perspective are - cache/memory statistics etc. This document summarizes the current work in the industry in this area and identifies key requirements for a comprehensive solution. Towards addressing the requirements, this document describes uses cases and defines reusable monitoring packet formats across all layers in the OAM hierarchy.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April, 2014.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC 2119].

Table of Contents

1. Introduction

Proactive SLA monitoring is key for enabling DevOps in a converged infrastructure. As new services are continuously enabled using DevOps methodologies, it is critical to make sure that the users are delivered the promised SLAs through proactive SLA monitoring; in the case where SLAs are violated, the system should be able to automatically fix the issue or revert back to the old configuration as needed.

Standards-based monitoring schemes [ietf-twamp] are coarse grained - first, based on injected packets and not on customer data packets and next, lack of per hop visibility while monitoring end-to-end and last, lack of coverage for network functions.

New proposed monitoring schemes focus on switches/routers end-to-end in the DC - in-band network telemetry [p4-in-band] is to enable per packet, per hop monitoring for timestamp (latency), queue depth, bandwidth etc., Data-plane probe for in-band telemetry collection [ietf-in-band-dpp] is to enable the above per injected packet, [ietf-sfc-monitor] describes one-way latency monitoring for service chaining nodes using timestamps.

Given the above landscape, the key requirements for a comprehensive proactive SLA monitoring framework are as follows

. Ability to monitor selective flows, e.g. monitor only low latency traffic

. Ability to mirror selective flows which are monitored, e.g. mirror only low latency traffic (mirroring all flows may not scale)

. Ability to strip monitoring information in the network edge since the application network stacks may not be able to process the additional monitoring information

. Ability to handle encrypted packets, e.g. enterprise cloud VPN across WAN, secure IaaS tunnels within a DC

. Ability to monitor individual network function paths, e.g. VNF service chaining where several VNFs/VMs are sharing the same physical server

. Ability to address each layer in the OAM hierarchy in a generic way using a common monitoring format. Within a DC,

the various OAM layers could be Service Function, Overlay and Underlay.

. Ability to pre-construct the space for monitoring headers [telemetry-header-options] to guarantee deterministic performance especially for virtual network functions which are subject to a cache hierarchy in an industry standard server

. Ability to programmably select the hops being monitored to make sure the monitoring header size is bounded

Towards addressing the key requirements, this document describes uses cases and packet formats for handling encrypted data packets (e.g. IPSEC for IaaS deployment) and service chaining and also describes options for maintaining deterministic application performance while performing elaborate monitoring.

1.1. Acronyms

DPI: Deep Packet Inspection

MPLS: Multiprotocol Label Switching

NVGRE: Network Virtualization using Generic Routing Encapsulation

OAM: Operations, Administration, and Maintenance

SF: Service Function

SFC: Service Function Chain

SFP: Service Function Path

VXLAN: Virtual Extensible LAN

2. In-band Telemetry for IPSEC tunnel packets

The following describes in-band telemetry for IPSEC tunnels which is the most popular WAN tunneling protocol for secure communication.

Use Cases:

. Cloud VPN: IPSEC tunnel between Enterprise branch and Enterprise/Cloud DC

oPrimary use case for IPSEC is inter-domain, for example enterprise branch office to PoP could be one network domain

(operator A) and PoP to Enterprise/Cloud DC could be another network domain (Operator B), e.g. Google Cloud Interconnect

oValue proposition:

. Real-time visibility/Service assurance for high priority tunnels carrying applications such as real-time voice/video

. Minimal WAN switch/router buffer overprovisioning for all classes of traffic and maximizing WAN link utilization

. Intra-DC: IPSEC tunnel between overlay end points for a private multi-tenant environment in a converged infrastructure (vlan, VXLAN provide isolation but not privacy)

. Real-time visibility/Service assurance for high priority tunnels carrying applications such as transactional storage, real-time big data

. Minimal DC switch/router buffer overprovisioning for all classes of traffic

There are several possible packet formats for achieving the above use cases. They are described below.

2.1. Packet Format 1 - Geneve

. Outer MAC Header

. Outer IP Header

oIP protocol - UDP

oDestination IP, Source IP, other fields

. Outer UDP Header

oDestination UDP port - Geneve (6081)

. Outer Geneve Header

oProtocol type - 0x6558 (RFC 1701- trans ethernet bridging)

oOption Length - greater than zero

oOption "INT"

. Option Class (16 bits) - INT

. Option Class needs to sync up with [ietf-geneve]

oOption "Next Protocol" - new option (total length including data is 8 bytes)

. Option class (16 bits) - Next Protocol

. Overrides protocol type in base Geneve header

. Type (8 bits) - Critical bit is set, Lower 8 bit byte in 4 bytes of data is protocol

. Reserved (3 bits)

. Length (5 bits) - set to 0x1 (4 bytes of data)

. Data (32 bits) - for IPSEC - set to 0x0000032 (ESP) or 0x00000033 (AH)

. Encrypted or Authenticated payload

2.2. Packet Format 2 - VXLAN GPE

. Outer MAC Header

. Outer IP Header

oIP protocol - UDP

oDestination IP, Source IP, other fields

. Outer UDP Header

oDestination UDP port - VXLAN GPE (4790)

. Outer VXLAN GPE Header

oNext Protocol - 0x5 - INT

. Outer INT Header(s)

oNext Protocol - ESP (0x7) or AH (0x8)

- Need to create two new next protocols, aligning with [ietf-nsh] and [p4-in-band]

- Encrypted or Authenticated payload

2.3. Packet Format 3 - IP options

Just like Geneve option format, IP options could be leveraged for in-band telemetry data.

- Outer MAC Header

- Outer IP Header

  - IP protocol - ESP (0x7) or AH (0x8)

  - Destination IP, Source IP, other fields

  - IP Header length > 5 (indicate presence of IP options)

- Outer IP options Header

  - Option-type

    - Copied Flag - 1

    - Option Class - 2

    - Option Number

      - 10 - In-band Telemetry (new)

    - Option-Length - variable

    - Option-Data - in-band telemetry data

- Encrypted or Authenticated payload

3. In-band Telemetry for Service Chaining

Use cases:

- 1) Monitoring of the networking interconnect. This would typically involve monitoring the overlay/underlay across the individual service chain nodes and the service chaining header ([ietf-nsh] etc.) across the entire service chain at the entry and exit points.

. 2) Monitoring of the individual network functions comprising a service chain using the service chaining header ([ietf-nsh] etc.). The network functions could be virtual (VMs etc.) or physical. Typically, monitoring of the virtual network functions will bring additional value since they share resources such as caches, memory etc. in an industry standard server.

. Combination of above two use cases involving simultaneous monitoring of networking interconnect and individual network functions.

Typical elements involved in service chain monitoring are vSwitches/NIC/ToR. For each individual network functions comprising a service chain, vSwitch/NIC/ToR will monitor ingress traffic to the network function for one or more of the INT [p4-in-band] parameters such as timestamp, queue depth, bandwidth and egress traffic to the vSwitch/NIC/ToR for one or more of the aforementioned INT parameters. Monitoring of the entire service chain at the entry point involves monitoring traffic sent to the first network function from vSwitch/NIC/ToR and exit point involves monitoring traffic from the last network function to the vSwitch/NIC/ToR for one of the aforementioned INT parameters. For highly accurate monitoring, it is recommended to use HW NIC/ToR vs a software based vSwitch. For example, HW implementations can measure timestamps to a nanosecond accuracy and can synchronize accurately with the master clock using protocols like IEEE 1588 PTP. A useful reference is [odl-nsh] which describes NSH service chaining operations from a ToR perspective.

Typical elements involved in underlay monitoring are ToR, Aggregation and Core switches/routers.

There are several possible packet formats for achieving the above the above use cases. Some are described here. More packet formats are work in progress.

3.1. NSH for service chaining Packet Format

. NSH Header

oNext Protocol - 0x5 - INT

. Needs to sync with next protocol in [ietf-nsh]

. NSH INT Header(s) (processed in vSwitch/NIC/ToR at each configured service chaining hop besides entry and exit points)

oNext Protocol - 0x3 - Ethernet

- Inner Ethernet payload

3.2. VXLAN-GPE for overlay and NSH for service chaining Packet Format

- Outer MAC Header
- Outer IP Header
  - oIP protocol - UDP
  - oDestination IP, Source IP, other fields
- Outer UDP Header
  - oDestination UDP port - VXLAN GPE (4790)
- Outer VXLAN GPE Header
  - oNext Protocol - 0x5 - INT
- Outer INT Header(s)
  - oNext Protocol - 0x4 - NSH
- NSH Header
  - oNext Protocol - 0x5 - INT
    - Needs to sync with next protocol in [ietf-nsh]
- NSH INT Header(s) (processed in vSwitch/NIC/ToR at each configured service chaining hop besides entry and exit points)
  - oNext Protocol - 0x3 - Ethernet
- Inner Ethernet payload

3.3. VXLAN-GPE for overlay and NSH for service chaining Packet Format

- Outer MAC Header
- Outer IP Header
  - oIP protocol - UDP
  - oDestination IP, Source IP, other fields

- Outer UDP Header
  - Destination UDP port - VXLAN GPE (4790)
- Outer VXLAN GPE Header
  - Next Protocol - 0x5 - INT
- Outer INT Header(s)
  - Next Protocol - 0x4 - NSH
- NSH Header
  - Next Protocol - 0x5 - INT
  - needs to sync with next protocol in [ietf-nsh]
- NSH INT Header(s) (processed in vSwitch/NIC/ToR at each configured service chaining hop besides entry and exit points)
  - Next Protocol - 0x3 - Ethernet
- Inner Ethernet payload

4. IANA Considerations

This draft does not have any IANA considerations.

5. Security Considerations

Flexibility must be provided to preserve/strip the in-band telemetry information across multiple operator domains to address privacy concerns.

6. Acknowledgements

The authors would like to thank Anoop Ghanwani, Jack Harwood from Dell EMC and Mukesh Hira, Sumit Verdi from VMware for all the discussions.

7. References

7.1. Normative References

7.2. Informative References

[RFC 2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.

[RFC 6291] Andersson, L. et al., "Guidelines for the Use of the "OAM" Acronym in the IETF," June 2011

[p4-in-band] "In-band Network Telemetry (INT)," http://p4.org/wp-content/uploads/fixed/INT/INT-current-spec.pdf

[ietf-twamp] "A Two-Way Active Measurement Protocol (TWAMP)," RFC 5357

[ietf-in-band-dpp] "Data-plane probe for in-band telemetry collection," https://tools.ietf.org/html/draft-lapukhov-dataplane-probe-01

[ietf-sfc-monitor] "Network Service Header KPI Stamping," https://datatracker.ietf.org/doc/draft-browne-sfc-nsh-kpi-stamp/

[ietf-nsh] "Network Service Header," https://datatracker.ietf.org/doc/draft-ietf-sfc-nsh/?include_text=1

[odl-nsh] "Creating a Service Plane using NSH," https://www.opennetworking.org/images/stories/downloads/sdn-resources/IEEE-papers/service-function-chaining.pdf

[ietf-geneve] "Geneve: Generic Network Virtualization Encapsulation," https://datatracker.ietf.org/doc/draft-ietf-nvo3-geneve/

[telemetry-header-options] "In-band Telemetry - Header Options," https://drive.google.com/file/d/0B2rg72wXZMMVUGxiRV9NYXJ4WDg/view?usp=sharing

Authors' Addresses

Ram (Ramki) Krishnan
Support Vectors
Fremont, CA
Email: ramkri123@gmail.com