TWAMP Extension to monitor Services KPIs
draft-spv-ippm-monitor-implementation-services-kpi

Authors:
Srivathsa Sarangapani
Peyush Gupta
Vinayak Hegde
Discuss implementation and use cases for extending TWAMP to calculate Services KPIs like:

• Service Latency
• Liveliness of an Application/Service
• Service load
• Service Throughput
• Packet statistics for a service
4. TWAMP Extensions

The TWAMP connection establishment follows the procedure defined in Section 3.1 of OWAMP [RFC4656] and Section 3.1 of TWAMP [RFC5357] where the Modes field is used to identify and select specific communication capabilities. At the same time the Modes field been recognized and used as an extension mechanism of TWAMP Reflect Octets and Symmetrical Size Features [RFC6038]. The new feature requires a new bit position to identify the ability of a Session-Reflector to monitor Services KPIs. There are changes in both the Control-Client and TWAMP-Test packet formats to support this functionality.
Use Cases

Three use cases are explained here for L4 to L7 services in the network to provide multiple aspects of the said extension.

1. To Calculate service latency of a service chain in the network

2. To detect the application/server liveliness for http server in the network

3. To calculate the service load for any specific service in terms of number of sessions, number of packets processed, number of flows and like.
**Service Latency**

![Diagram of Service Latency](image)

- **TWAMP Client**
  - TWAMP HDR
  - Service PDU

- **TWAMP Server**
  - TWAMP HDR
  - Service PDU

- **Service 1**
  - IPSEC

- **Service 2**
  - DPI

- **Service 3**
  - JFLOW

- **Router**

RTT = (T4-T1) - (T3-T2)

Service Latency = T6-T5

*FIG 2. Service latency measurement in the network with service chaining*
Service Latency Continued

- FIG 1. in slide 5 capture the use case with 3 services IPSEC, DPI and JFLOW running in the network as part of a service chain.

- The Session-Sender sends the Service PDU as part of the TWAMP-Test Packet Padding.

- When Session-Reflector receives the TWAMP-Test message, it extract the Service PDU and inject that service PDU in the Service Block (here Service 1).

- TWAMP server injects the packet to the service1, and then sequence service1->service2->service3 is followed on the same packet. TWAMP server records the time stamp T5 and T6 [refer Fig 1.]

- Service Latency using T5 and T6 is calculated and session reflector sends the response with these TS values T5 and T6 to the session sender.
Liveliness of an App/Service

Fig 3. Liveliness detection of HTTP/DNS server
Liveliness of HTTP server

- The Session-Sender sends the Service PDU with http request embedded in it, as part of the TWAMP-Test Packet Padding.

- When Session-Reflector receives the TWAMP-Test packet, it extract the Service PDU. Then Session-Reflector inject the http request embedded in service PDU to the http server for checking liveliness.

- In case the session reflector receives a successful response, that means the http server is ALIVE.

- If the http response timeout the http server can be marked as unresponsive or DEAD.

- The session reflector sends the response back to the session sender indicating the keep alive state of the http server in the TWAMP test packet response.
Service load

Service here can mean: Session/Subscriber/Tunnel/Bandwidth/Flows etc.

Multiple RS (Real Server) VNFs in Cloud

TLB (Traffic Load Balancer) /
SLB (Server Load Balancer)

Fig 4. Service load calculation using TWAMP Test Session
Next Steps

• Seek Reviewers from ippm community
• Comments are welcome
• Call for WG adoption
Q&A

THANK YOU
(IETF ippm WG chair and members)