Service Assurance for Cloud Service Providers: Qosmos Probe Monitors Virtualized Web Conferencing Services

Web conferencing requires a combination of voice, video and chat services (to name a few) which themselves may use a set of different network microservices. This situation creates a growing need for monitoring and service assurance solutions able to deliver granular visibility down to the network microservice; and to do it independently of the vendor-specific equipment that actually delivers those generic services.

### The Challenge

Web conferencing service providers need to monitor and troubleshoot to maintain a certain quality of service. Providing service assurance in a cloud environment requires not only visibility into the individual service components but also the ability to provide a logical view of the entire service by considering metrics from the individual microservices.

The majority of existing performance monitoring and troubleshooting tools have been developed for enterprise applications with generic Key Performance Indicators (KPIs) that provide a limited view and too many irrelevant metrics for most service providers.

The challenge is made even more formidable when you add virtualization. With SDN architectures, there is no longer a single point from which to monitor activity and performance. When resources are virtualized, usually across vast geographic areas, it becomes very difficult to get an end-to-end view of the service at the subscriber and at the session levels.

For example, imagine that the users of a web conferencing service call tech support to complain about slow response times and poor voice quality. The cloud datacenter application team, responsible for creating workloads and running applications, gets the trouble ticket and finds no bug with the application. However, they notice that the web conferencing application does not receive all the packets it should, and they infer that the problem must come from the virtualization software. The virtualization team, responsible for the OpenStack, VMware, SCVMM / Hyper-V software, investigates the problem, but it takes them a long time to find a solution. During this time, customer satisfaction worsens and creates a business risk.

### Benefits

- Non-intrusive solution which does not affect service performance
- Supports hybrid infrastructures (physical network, virtualized datacenter, cloud-based)
- Generates flow-based data records in real-time (enriched NetFlow-like records)
- Provides full visibility on network traffic, up to the application level
- Can enrich records with external information (VM names, container names, host names resolution, etc.)
- Standard management APIs (e.g. NETCONF)
- Can be deployed in OpenStack or VMware platforms

### Application Use Case

- An agent is running on each compute node, listening for nova messages
- When this agent detects the creation of a new VM, it dynamically creates a new tap to feed the vProbe with the traffic of this new VM.
- This tap is automatically removed if the VM is destroyed or moved to another compute node.

Figure 1: Integration with OpenStack inside the NFVI
The Solution

Enea’s Qosmos Probe enables deep and granular visibility into virtualized infrastructures. At the heart of this technology is Qosmos ixEngine®, the market-leading, deep packet inspection engine that provides detailed real-time IP traffic classification and metadata extraction.

The Qosmos Probe is built to support a high degree of end-user customization. Users have the freedom to define which metadata to extract and to specify the logic for their unique KPIs. In other words, the Qosmos Probe will focus on capturing and providing only the information that is relevant to a specific cloud or service provider.

For example, in the case of a web conferencing service, the provider might want to focus on SIP and RTP metrics for voice, file exchange and chat; RTSP metrics for the video portion and HTTP metrics for the control and web GUI portion.

Qosmos Probe Compared to Appliance-based Monitoring Probes

- Non-intrusive: No need to install agents on VMs, which translates into lower operational risk
- Requires no physical space in racks: No appliance to install, since the Qosmos Probe only uses existing CPU & RAM, and can be managed remotely
- Provides monitoring robustness and availability: The Qosmos Probe distributed monitoring architecture is less risky than a single point of failure with a dedicated physical probe
- No impact on bandwidth: Qosmos Probe does not need to send monitored packets to a dedicated probe and therefore does not generate additional bandwidth

Examples of Available Metadata

- Packet loss
- Round Trip Time (RTT)
- Latency
- Retransmission count
- Application error codes
- Jitter delay
- Mean Opinion Score (MOS)
- Individual call flow messages

Work Flow and Organizational Processes

The Qosmos Probe is a powerful tool to identify the root cause of quality issues or downtime both quickly and easily. This is especially valuable in the case of transient problems, which are hard to replicate. For example, in a recent situation, a virtual router became overloaded and sporadically lost packets. As the problem occurred only under certain conditions, it was particularly difficult to solve.

Based on precise packet loss information obtained from the Qosmos Probe it is possible to identify the impact of the losses per application and to pinpoint particular issues per flow (see figure 2 below).

It also enables differentiated SLAs for services (e.g. voice vs. email) even in a completely mixed workload situation in which it is not possible to physically differentiate between applications.

Examples of Available Metadata

- Packet loss
- Round Trip Time (RTT)
- Latency
- Retransmission count
- Application error codes
- Jitter delay
- Mean Opinion Score (MOS)
- Individual call flow messages

Figure 2: Using Qosmos Probe to identify impact per application