APPLICATION PERFORMANCE MANAGEMENT FOR BUSINESS VPN SERVICES

USING ALCATEL-LUCENT 5670 REPORTING AND ANALYSIS MANAGER (RAM) TO IMPROVE APPLICATION PERFORMANCE AND CUSTOMER EXPERIENCE

APPLICATION NOTE

TABLE OF CONTENTS

ABSTRACT

Application performance management has become a significant area of focus not only for service providers looking to provide value-added WAN optimization services, but also for those service providers looking to provide application-based SLAs as part of their cloud strategy. Alcatel-Lucent 5670 Reporting and Analysis Manager (RAM) has the capability to collect, analyze and report on application performance and behavior, as well as end-user quality of experience.

This application note will describe the mechanisms by which application performance data is collected, how it is processed within the 5670 RAM, and implications for reporting. Additionally, this application note will describe configuration scenarios that have a significant impact on overall product performance and scalability.

OVERVIEW

Service providers today face significant marketing and operational challenges. A rapidly commoditizing market for traditional network connectivity and enterprise demand for new enterprise services centered around networked applications and cloud computing are putting pressure on their traditional role as network providers. Alcatel-Lucent's Application-Assured Virtual Private Network (AA-VPN) enables this transition by transforming a service provider's business service delivery infrastructure from being service-aware to being application-aware. In a service-aware network, packet loss, round-trip delay, jitter, QoS, etc., are all viewed from the context of the service being delivered.

Similarly management and control mechanisms, such as shaping and policing, are also applied at a service level — with all applications within each service class treated equally. In an application-aware network, leveraging application performance management, operators (and even the business customer directly) are able to view performance metrics and apply control on a per-application basis.

An Application-Assured VPN solution offers business service customers the ability to analyze and control application traffic without purchasing, deploying or managing complex and expensive CPE-based solutions. Application performance management, as part of an AA-VPN, extends this capability to include analysis of quality of experience and performance metrics on a per application basis.

The role of volume and performance monitoring in an AA solution

The value of data analysis within a WAN management can be characterized in terms of four broad categories:

- Network policy planning and validation
- Application performance analysis
- Application inventory
- Anomaly detection
- Network right-sizing
- Capacity planning and budget justification
- Service-level assurance
- Forensic analysis
- Fault isolation

While specifically mandated to manage and assure the availability and performance of critical applications, the vast majority of IT managers and directors have no clear view of exactly what traffic is leveraging the WAN. Application inventory analysis helps the IT manager understand the existing traffic, identify rogue traffic, categorize that traffic in terms of its value to the business, and ultimately use the information available in 5670 RAM to define application QoS policies to manage that same traffic.

Once the IT manager understands what applications are running on the WAN, and has an appropriate strategy for policing and managing traffic, it becomes very important to understand the evolution of application traffic over time. Different applications have different quality of service (QoS) requirements of the network. A view of how these

applications are growing helps the IT manager better plan what type of services (bandwidth, classes of service, etc.) will be required to support the business over time. Different views of that same data are also useful for the IT manager to justify and assure budget for WAN growth over time.

Application service-level assurance is about setting application Quality of Experience (QoE) targets that align with overall criticality of an application, and then continually measuring the actual application QoE and providing an assessment as compared against the previously defined targets. Application service-level assurance is an essential aspect of WAN management, as it helps business customers to i) quantify application performance in terms of their business needs, ii) define an investment level that aligns with their business needs, iii) set expectations that characterize intended application performance and iv) identify when application performance fails to satisfy the defined business need.

When the quality of experience fails to achieve the defined targets, anomaly detection and forensic analysis helps the enterprise IT organization answer the following questions:

- **What was the nature of the incident?** By correlating the underlying performance metrics to the quality of experience measurement, the WAN management system can define the problem in terms of availability, delay, jitter, or packet loss.
- **What was the impact of the incident?** Understanding the duration of the issue, as well as the number and type of related application transactions helps an IT organization to qualify the impact and urgency of the identified incident.
- **Where was the locality of the incident?** Application delivery of a WAN infrastructure can be a complicated transaction. Resolving related issues can be difficult without guidance. Providing direction in terms of where the issue originated (that is, server side, client side, or WAN) focuses attention on the root cause, and helps to speed resolution.

In deployment scenarios where the MS-ISA is deployed at the service edge, the operator can also use application assurance data to determine whether or not delay issues originate within the WAN, the client access network, the server access network, or even within the application infrastructure itself.

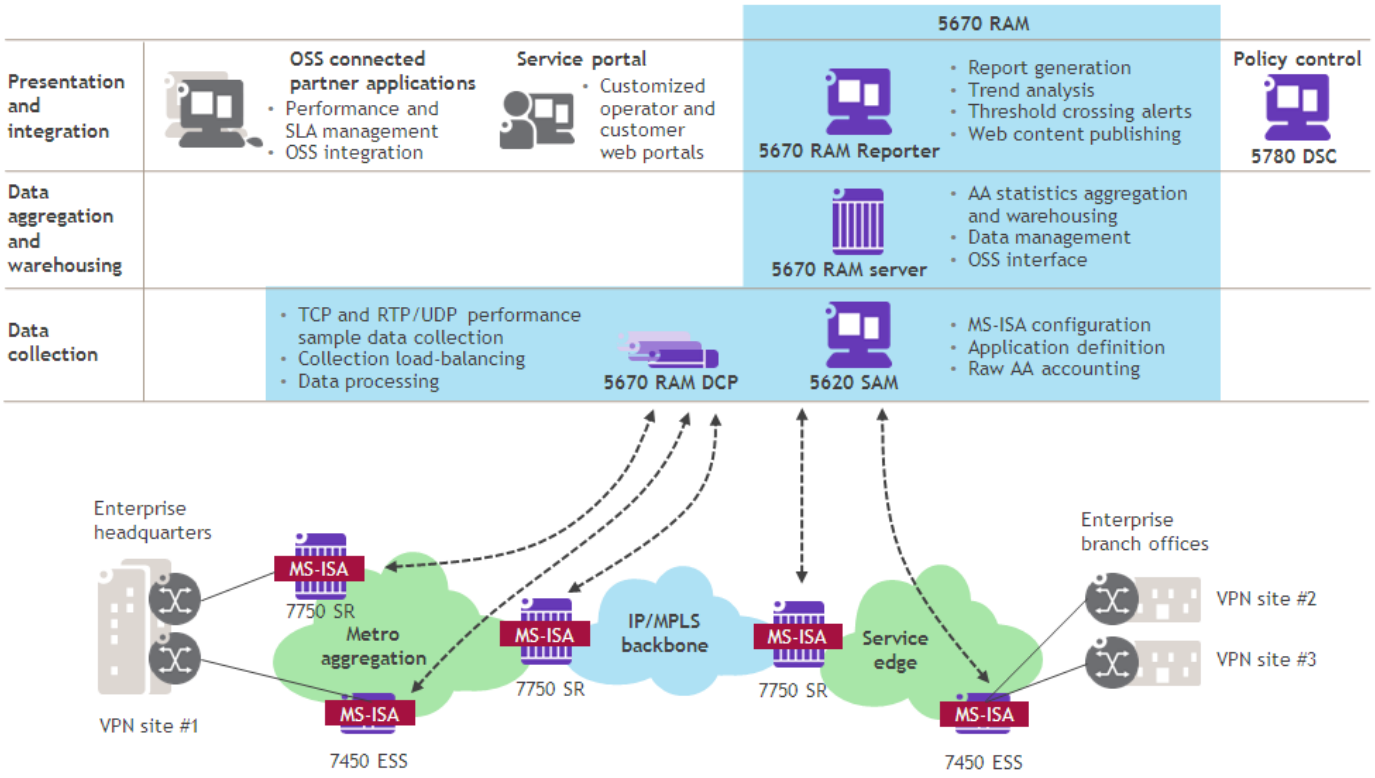
APPLICATION-ASSURED VPN SOLUTION OVERVIEW

The AA-VPN solution relies on the Application-Assurance (AA) feature set of the Alcatel-Lucent 7x50 routing and switching portfolio included in the Multi-Service Integrated Service Adapter (MS-ISA) blade. The MS-ISA is a blade for high-touch packet processing. In the AA mode, the MS-ISA provides stateful, pattern and string-based identification of applications to enable dynamic per-service, per-site and per-application reporting and QoS policy control — all at line speed.

The Alcatel-Lucent AA-VPN solution is enabled and operated by the service and application management capabilities of the Alcatel-Lucent 5620 Service Aware Manager (SAM) management suite, which includes the Alcatel-Lucent 5670 Reporting and Analysis Manager (RAM). Together, these products provide a comprehensive management solution that enables the operator to extend its existing service network to encompass AA-VPNs. Customer access to the application reports is provided through a Service Portal.

The MS-ISA modules in the 7x50 platform identify the application flows and generate aggregate or per-flow statistics, which are then sent to the Alcatel-Lucent 5620 SAM or 5670 RAM Data Collection and Processing (DCP) appliance, where they are processed for population into the 5670 RAM Data Warehouse. Through the 5670 RAM Reporter and Service Portal, these statistics can then be presented into graphical usage and trending reports, 'green wall' type performance reporting, and so on, as shown in Figure 1.

Figure 1.Application-Assured Business VPN management architecture



GENERATE FLOW OUTPUT

The AA mode-enabled MS-ISA (or AA-ISA) is able to report on the performance of TCP and RTP/UDP applications by leveraging cflowd v10 for per-flow meta data export, which is compliant with the IP Flow Information eXport (IPFIX) standard. While every flow is evaluated by the AA-ISA, cflowd output of the flow meta-data is sampled. A flow record is published upon termination of a tracked flow. The AA-ISA provides flow-based reporting for the following areas:

- TCP performance
- RTP/UDP voice performance
- RTP/UDP video conferencing performance
- RTP/UDP audio performance
- Flow-based volume

In addition to the standard 5-tuple information normally found in flow records (IP Source, IPdestination, source port, destination port and protocol), the flow record produced by the AA-ISA also contains a varying number of vendor-specific fields that positively identify

applications and application groups, as well as providing the raw data necessary to derive the application performance metrics.

Understanding the importance of the algorithms empowering Audio/Video performance metrics, Alcatel-Lucent has partnered with Telchemy Inc. to leverage Telchemy's decade-plus experience and leadership in performance management technology for VoIP and IP video conferencing. AA-ISA integrates Telchemy's VQmon Software agent — used by over 120 embedded and standalone test probes — to produce highly accurate estimates of perceptual quality for voice and video-conferencing sessions.

Each AA-ISA is capable of producing up to 10,000 flows per second (fps). A fully populated router chassis can produce up to 30,000 fps. As the AA-ISAs may be terminating millions of flows per second, not all application flows will typically generate a flow record as the AA-ISA samples the data for flow export. The sampled nature of the data precludes the use of cflowd for billing purposes in most circumstances, hence, cflowd data does not serve to replace the existing accounting data available through the 5670 RAM, but serves to augment it with performance behavior data. Performance reporting employs the flow sampling mechanism, while volume reporting uses the packet sampling technique. In both cases, the sample rate is configurable at the AA-Group[1] level. Such a configuration will affect all associated partitions[2] , and all applications within those partitions.

The AA-ISA does support cflowd publication to multiple destinations for redundancy, however, it is important to note that this does not change the overall maximum flow output rate. Sending cflowd records to two different destinations keeps the unique flow output rate at 10,000 flows per second per AA-ISA. In other words, the AA-ISA still sends 10,000 unique flows per second to each collector or 20,000 duplicated flows per second with one copy going to each collector.

[1] The AA-Group is an organization of physical AA-ISA cards within a router. The purpose of the AA-Group is to facilitate load balancing, failover, and hardware configuration.

[2] Partitions are the logical divisions of an AA Group. The purpose of a partition is to allow a single AA group to be used in multiple different ways, through the use of policies, each of which is restricted to its own partition.

5670 RAM DATA COLLECTION AND PROCESSING APPLIANCE

Based on the peak flow output of the AA-ISA, it is possible for a single service router to output more than 2.5 billion records each day. Clearly, a mechanism needs to be employed to identify the useful information within the individual flow records without actually keeping those individual records for an extended period of time.

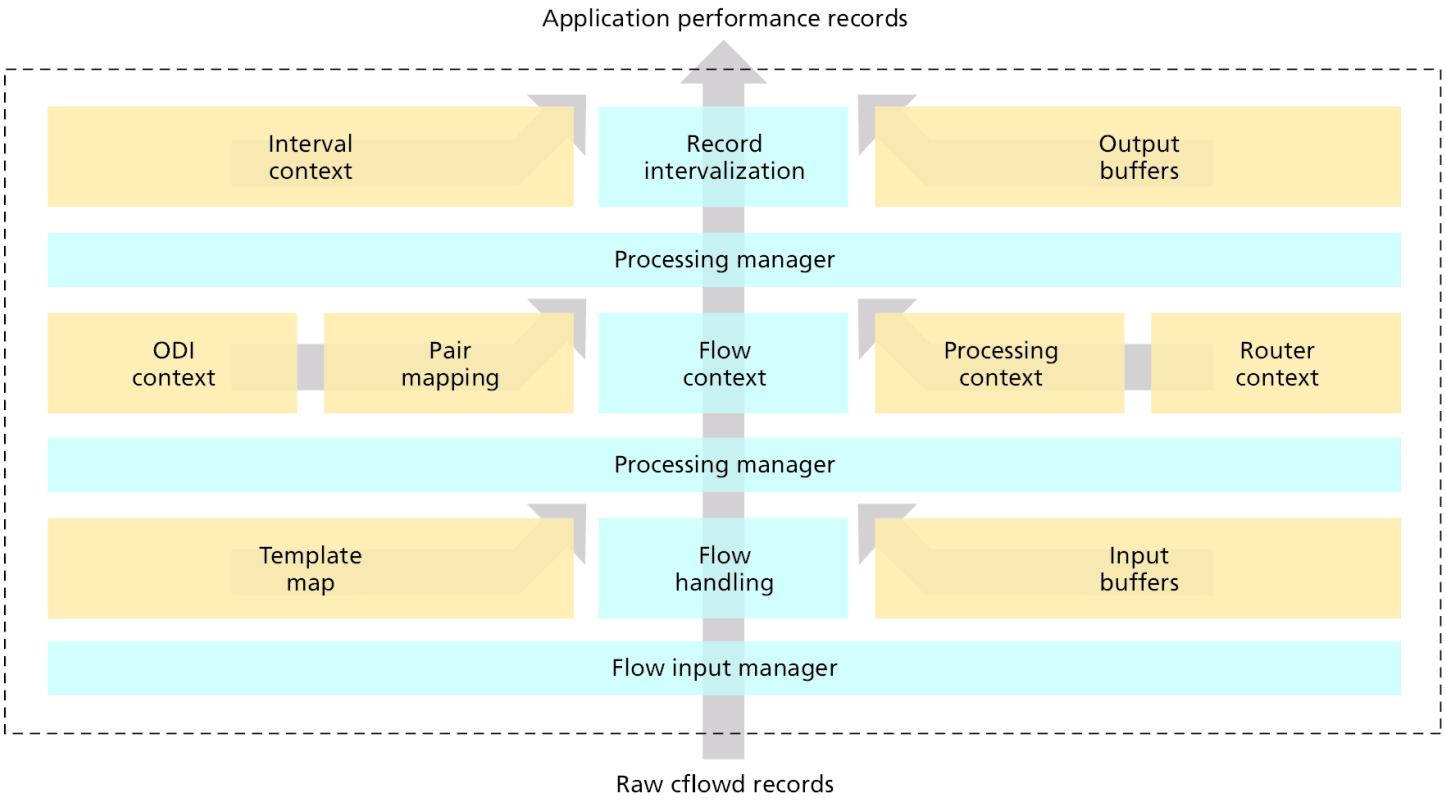
Introduced in Release 3.0R3, the 5670 RAM Data Collection and Processing (DCP) appliance is a distributable component of the 5670 RAM architecture. The purpose of the DCP is to provide scalable and intelligent flow collection, processing and analysis on behalf of 5670 RAM. The input to the DCPs is the individual flow records generated by the AA-ISAs from one or more routers. The output of the DCPs (to 5670 RAM) is a set of summarized records tied to dimensions of interest over a period of time.

The output of the DCP is a periodic set of records that represent a summary of application performance over the period. These records collapse the raw flow data along several dimensions of interest over a configurable time interval, as shown in Figure 2.

Each flow record received by the DCP is identified by a template ID. The DCP uses this template ID to properly parse the individual fields from the flow record, as well as to determine how to process the data contained within the record.

The processing manager will identify the observation domain of the flow based on the direction of the flow and the identified location of the observation point (AA-ISA generating the flow record) as it relates to the flow (client side of the network or server side of the network). In this way, the DCP is able to provide a much more useful context to the data (e.g., not just 'ingress delay', but 'server side delay') . The processing manager is also aware of the raw fields contained within the flow record, and how to combine these fields (across multiple records over time) to create a useful metric (i.e., using retransmitted packets to calculate packet loss). The output file period is configurable (5 min, 10 min, 15 min, 20 min, 30 min, 1 hr).

Figure 2.Application performance data processing flow



Dimensions of Interest

Dimensions of interest provide a framework along which the DCP will summarize data. These dimensions will be continued in the 5670 RAM server as the primary lines of analysis and reporting. DCP currently supports three dimensions of interest:

- Router level – applicable to TCP performance and volume reporting
- SAP / route of interest level – applicable to TCP performance and volume reporting
- SAP / route of interest / CODEC level – applicable to RTP/UDP performance records

Router level summarization creates a single record for each application and app group active on the router. The record will contain a set of application performance metrics for the given application (or app group) over the reporting interval. There is no configuration required for router level summarization.

As previously discussed, route of interest level summarization is a mechanism to, flexibly, collapse the source and destination addresses into a coarser and more meaningful context, thereby reducing the record count, without sacrificing useful information. This is accomplished through the use of DCP groups.

The primary objective of application performance reporting is to enable analysis of traffic travelling from one VPN site to another VPN site (site to site). The secondary objective is to enable analysis of traffic from a specific subnet of a VPN site to another VPN site (subnet to subnet). A third objective of application performance reporting is to enable analysis of traffic travelling between a VPN site and the Internet. Proper configuration of DCP groups support all three of these objectives.

Due to the relatively low volume of RTP/UDP records, RTP/UDP performance data is not summarized (and all flow detail is maintained, however, each flow record is enriched with the local and remote DCP group data.)

DCP Groups in detail

As previously noted the objectives of DCP groups are to collapse flow level detail into aggregate records that are more meaningful for the user, and more manageable for the RAM Server. In order to achieve the level of control and consistency required from the system, DCP Group configuration is done via 5620 SAM. Within 5620 SAM the operator has the ability to enable, create, and manage DCP groups on a per-service basis.

DCP Groups can exist at either of two levels - per service, or per AA-subscriber (VPN site) level. In general it is more useful to associate a DCP group to a specific VPN site, then to the service. The exceptions are two default DCP groups that are created when a service has AA performance reporting enabled: 'default Internet' and 'default Intranet'. Default Intranet includes IP membership rules associated to all private IP addresses and default Internet includes IP membership rules associated to all public IP addresses. Some enterprise customer's may leverage public IP addresses within their private networks. As default DCP groups are configurable on a per service basis, IP membership rules for default groups can be tailored for these scenarios.

Custom DCP Groups are defined by the operator through the 5620 SAM GUI, or via 5620 SAM-O. They can be associated to a service, but ideally should be associated to a specific AA-subscriber. Cflowd records from the AA-ISA include information regarding the local AA-subscriber. Association of the custom DCP group to AA-subscriber allows RAM to determine the remote AA-subscriber, which is key to providing VPN site-to-site reporting.

DCP IP membership rules may overlap between groups. As an example IP membership rules for one DCP group may specify an entire subnet associated to a given VPN site. IP membership rules for a second DCP group may specify a subset of IPs within the same subnet. DCP organizes rules such that smaller span rules are considered first.

TCP PERFORMANCE REPORTING

The AA-VPN solution can perform application performance reporting on TCP applications. This feature is a combination of functionality found within 5620 SAM, the AA-ISA and 5670 RAM. In this capacity, 5620 SAM provides end-to-end policy management for the AA-ISA. This facilitates a consistent definition of applications and app-groups across different routers in the network to ensure that accurate and consistent reporting is achieved.

Table 1 shows the supported performance metrics per dimension.

Table 1. Supported performance metrics per dimension

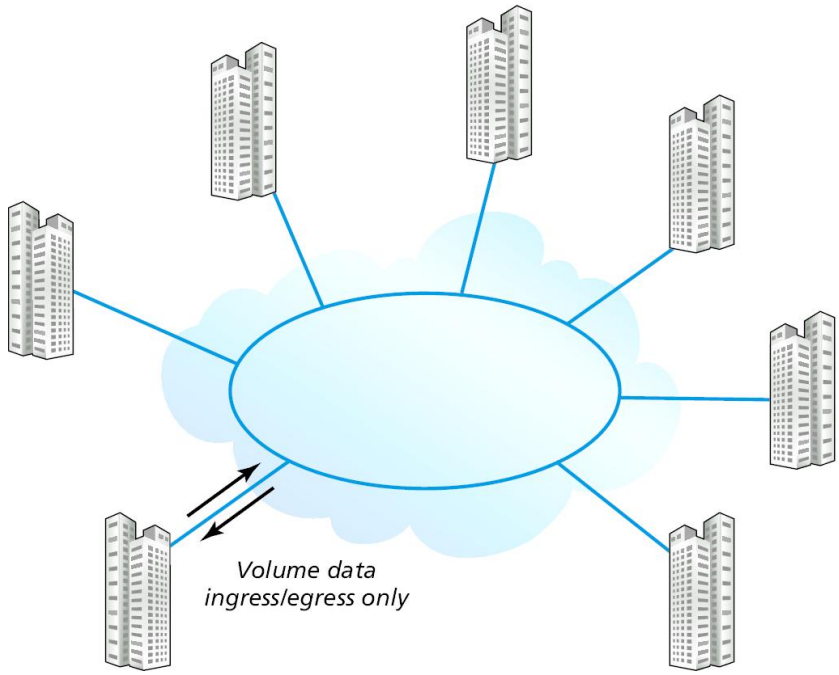
METRIC	ROUTER / APP	ROUTE / APP OF INTEREST	NOTES
Flow Count	X	X	Summation of all applicable flows in the record Available as of RAM3.0R3
Flow Duration	X	X	Average flow duration of all applicable flows in the record
Total Bytes	X	X	Directionally Min, Max, Avg Available as of RAM3.0R3
Total Packets	X	X	Directionally Min, Max, Avg Available as of RAM3.0R3
Client Trip Delay		X	Based on session establishment delay Only from observation points on client side of network Min, Max, Avg, Packet Delay Variation Available as of RAM3.0R3
Server Trip Delay		X	Based on session establishment delay Only from observation points on server side of network Min, Max, Avg, Packet Delay Variation Available as of RAM3.0R3
Round Trip Delay	X	X	Based on session establishment delay Available from either side of network Avg, Packet Delay Variation Available as of RAM3.0R3
Local Access Delay		X	Based on session establishment delay Min, Max, Avg, Packet Delay Variation Available as of RAM 5.0R1
WAN Delay		X	Based on session establishment delay Min, Max, Avg, Packet Delay Variation Available as of RAM 5.0R1
Packet Loss Ratio	X	X	Directionally. Available as of RAM3.0R3
Server Side Delay		X	Based on session TCP delay Average Available as of RAM 4.0R1
Server Side Server Delay	X	X	Based on session TCP delay Available on both sides of the network, but most trusted from the server side Weighted average based on sample count Excludes the Server Trip Delay (delay incurred by the server only) Available as of RAM 4.0R1
Total Delay	X	X	Based on session TCP delay Represents 1 full turn through the network Variance and weighted average based on sample count Available as of RAM 4.0R1
Total Samples Count	X	X	Sum of all samples, for all applicable flows. Available as of RAM 4.0R1

Apdex™ – Round Trip Time	X	Evaluating Round Trip Delay Available as of RAM 4.0R1
Apdex™ – Packet Loss Ratio	X	Evaluating packet loss ratio Available as of RAM 4.0R1
Apdex™ – Total Delay Average	X	Evaluating Total Delay Average Available as of RAM 4.0R1
Apdex™ Total Delay Variance	X	Evaluating Total Delay Variance Available as of RAM 4.0R1

FLOW-BASED VOLUMETRIC REPORTING

Traditionally, the AA-VPN solution has provided detailed volumetric analysis leveraging the filebased accounting statistic mechanism available on the service router. The accounting statistics data is complete in that every byte and every packet is counted, and not sampled. However, while the accounting data provides an accurate view of the volumetric activity into (and out of) a given SAP, there is no visibility on where that data is coming from (or going to) — in short, the volumetric data lacks a direction vector (as shown in Figure 3).

Figure 3. Volumetric reporting – Lacks directional component



The flow-based volumetric reporting is based on sampled data and is therefore unsuitable for billing purposes in of itself, however it does, through the inclusion of route-of-interest data, provide a direction vector and relative volumes — both of which can be used to enrich the accounting data in order to provide intelligence on the volumes, sources and destinations of application data, as shown in Figure 4.

Figure 4. Flow-based volumetric reporting – Volume data with source/destination context

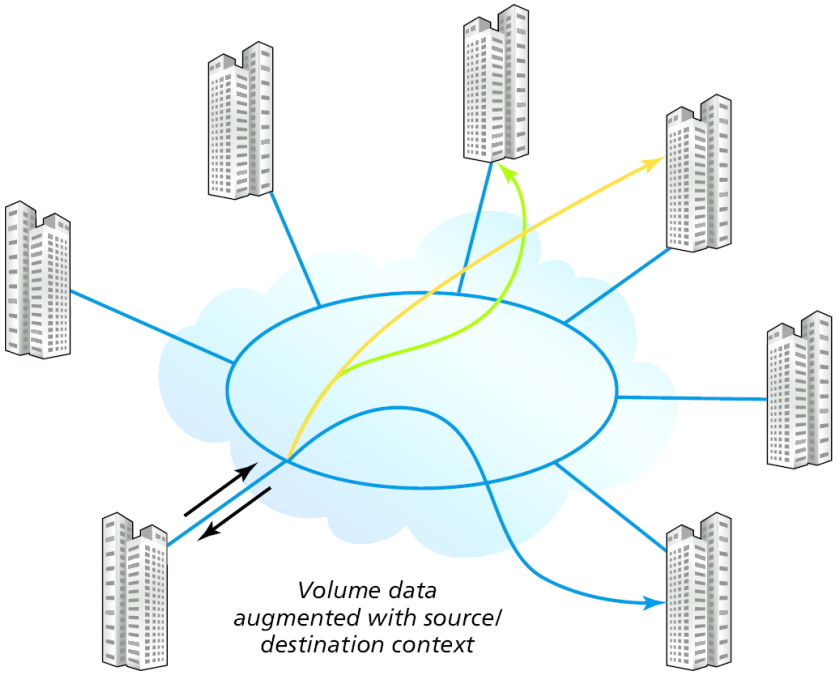


Table 2 shows the volumetric reporting metrics per dimension.

Table 2. The volumetric reporting metrics per dimension

METRIC	ROUTER / APP	ROUTE / APP OF INTEREST	NOTES
Total Bytes		X	The total number of bytes sampled and reported on over the period Directionally Available as of RAM3.0R3
Total Packets		X	The total number of packets sampled and reported on over the period Directionally Available as of RAM3.0R3
Flow count		X	Total number of applicable flows over the period Directionally Available as of RAM3.0R3
Duration		X	The average flow duration over the period Directionally Available as of RAM3.0R3
Averaged BPS		X	The average of each flow BPS over the period Directionally Available as of RAM 3.0R3
Averaged PPS		X	The average of each flow PPS over the period Directionally Available as of RAM 3.0R3

RTP/UDP APPLICATION PERFORMANCE REPORTING

Many business-critical real-time voice, video and audio applications are delivered over UDP and require a different mechanism for application performance management. The RTP/UDP performance management feature is consistent in the management and configuration elements of the solution.

There are two notable differences in the handling of cflowd records from RTP/UDP applications versus TCP applications:

- No DCP summarization: Given the expected volumes of RTP/UDP flow data, the DCP does not provide summarization services. Rather, the DCP organizes the raw flow records into the required dimensional categories and subsequently provides all the flow records to the 5670 RAM Server.
- The 5670 RAM Server will maintain all raw flow detail for a period of up to 60 days. Beyond that timeframe, only MOS, R-factor and degradation factor data are available in an aggregated format.

The organization of flow data will be similar to TCP data in that the DCP will leverage the IP Source and IP Destination addresses to develop a route of interest, however, data will be further organized by CODEC (SAP/Route of Interest / CODEC)

Table 1 describes the performance metrics for TCP applications. Table 2 describes the flow-based volumetric data available across all applications. Tables 3, 4 and 5 describe the supported performance metric fields for voice, video and audio applications.

Table 3. Voice application performance metrics

FIELD	DESCRIPTION
Forwarding Class	The forwarding class of the application
RTP Payload Type	The RTP payload type
Flow Duration	Measured in ms
Packets Received	A count of the number of packets received
Packets Lost	A count of the number of packets lost
Packets Discarded	A count of the number of packets discarded
Packets Out of Sequence	A count of the number of packets out of sequence
Burst Count	The total number of burst periods
Average Burst Length	The average length of a packet loss burst (in ms and packets)
Gap Count	The total number of gap periods
Average Gap Length	The average length of a gap period (in ms and packets)
Average Round Trip Delay	Latency
Round Trip Delay Source	The round trip delay value source (unknown = 0, default = 1, user config = 2, RTCP report = 3, RTCP XR report = 4, RTCP HR report = 5, SIP RTCP report = 6)
MAPDV	The running average mean-absolute packet delay variation in ms
MOS-LQ	The listening quality MOS
MOS-CQ	The conversational quality MOS
MOS-NOM	The nominal (generally accepted maximum attainable) MOS for the voice stream given a typical transmission system and voice CODEC selection for the call
R-factor LQ	The listening quality R-factor
R-factor CQ	The conversational quality R-factor
R-factor G107	The G107 R-factor
R-factor Nom	The nominal R-factor for the voice stream
R-burst	R-factor for burst conditions
R-gap	R-factor for gap conditions
Degradation Factors	Severity of call quality degradation due to the following factors: • Packet loss • Packet discard • Delay • Jitter • Codec type • Signal level • Noise level • Echo

Table 4. Video application performance metrics

FIELD	DESCRIPTION
RTP Payload Type	The RTP payload type
Flow Duration	Measured in ms
Packets Received	A count of the number of packets received
Packets Lost	A count of the number of packets lost
Packets Discarded	A count of the number of packets discarded
Packets Out of Sequence	A count of the number of packets out of sequence
Burst Count	The total number of burst periods
Average Burst Length	The average length of a packet loss burst (in ms and packets)
Gap Count	The total number of gap periods
Average Gap Length	The average length of a gap period (in ms and packets)
Average Round Trip Delay	Latency
Round Trip Delay Source	The round trip delay value source (unknown = 0, default = 1, user config = 2, RTCP report = 3, RTCP XR report = 4, RTCP HR report = 5, SIP RTCP report = 6)
Average Round Trip Delay Variance	Average packet to packet delay variance over the flow
VSTQ	A 0-50 CODEC independent score (50 being best) measuring the ability of the IP network to carry video reliably
MOS-V (absolute)	The absolute MOSV
MOS-V (relative)	The average MOSV
MOS-AV	The average audio/video MOS
Degradation Factors	Severity of call quality degradation due to the following factors: • Packet loss • Packet discard • Delay • Jitter • Codec quantization level • Codec encoding b/w constraints • Frame resolution • Frame presentation rate • GOP length • Network b/w constraints • Audio-video sync • Recency
GOP Type	Group of pictures type (i.e., IBBP)
GOP Length	The number of frames in a group of pictures
Image Width, Height	In pixels
Frame Rate	In frames per 1000 seconds
Slices per I-frame	The average number of slices contained in each I-frame
Reference Clock Rate	Video clock rate
Video Interlaced	Binary field indicating whether the video is interlaced
I, P, B, SI, SP Frames Received	The number of I, P, B, SI, SP frames received
I, P, B, SI, SP Frames Impaired	The number of I, P, B, SI, SP frames impaired
Average Video Bandwidth	The average video bandwidth in bits/second measured during one second window excluding transport packet header overhead and error correction/retransmission
Peak Video Bandwidth	The peak video bandwidth in bits/second measured during one second window excluding transport packet header overhead and error correction/retransmission
Frame Inter-arrival Jitter	The average frame inter-arrival jitter in ms – computed relative to the expected arrival time based on the frame rate
I-Frame Inter-arrival Jitter	The average I-frame inter-arrival jitter in ms – computed relative to the expected arrival time based on the frame rate
Average inter-arrival Delay	The average frame arrival delay in ms
Estimated PSNR	An estimate of the distortion that has occurred between the source video stream and the output video stream

Table 5. Audio application performance metrics

FIELD	DESCRIPTION
RTP Payload Type	The RTP payload type
Flow Duration	Measured in ms
Packets Received	A count of the number of packets received
Packets Lost	A count of the number of packets lost
Packets Discarded	A count of the number of packets discarded
Packets Out of Sequence	A count of the number of packets out of sequence
Burst Count	The total number of burst periods
Average Burst Length	The average length of a packet loss burst (in ms and packets)
Gap Count	The total number of gap periods
Average Gap Length	The average length of a gap period (in ms and packets)
Average Round Trip Delay	Latency
PPDVM	The instantaneous packet-to-packet delay variation measured at the end of the interval
MOS-A	Mean Opinion Score - Audio
Degradation Factors	Severity of call-quality degradation due to the following factors: packet loss, packet discard, codec type, recency
Number of audio channels	Measured number of audio channels
Rf Clock Rate	Clock rate
Avg Audio b/w	The average audio bandwidth in bits/second measured during one second window (excluding transport packet header overhead and error correction/retransmission)
Peak Audio b/w	The peak audio bandwidth in bits/second measured during one second window (excluding transport packet header overhead and error correction/retransmission)

QUALITY OF EXPERIENCE REPORTING

Quality of Experience (QoE) reporting allows for the interpretation of one or more performance metrics into a single score that represents the user's overall experience. A QoE score also allows a simple set of thresholds to convert the QoE score into an actionable QoE state. The 5670 RAM supports QoE reporting for both TCP and RTP/UDP applications.

Apdex™:QoE reporting for TCP application

Alcatel-Lucent is a contributing member of the Apdex Alliance™, and the 5670 RAM leverages a compliant approach to measuring and reporting Apdex™ scores. The Alcatel-Lucent approach to Apdex™ scoring allows for a highly personalized analysis of the performance data.

Apdex™ is based on a double-threshold system. The first set of thresholds evaluates the quality of each individual flow and ultimately results in a normalized Apdex™ score. The second set of thresholds serves to transform the Apdex™ score into an Apdex™ state that can easily be represented on a dashboard, or interpreted as an alarm.

The application flow thresholds are defined as part of the application object within the 5620 SAM GUI or through the 5620 SAM OSS Interface (SAM-O). Since the flow thresholds are definable on a per-application basis, and applications are defined per-AA partition, flow thresholds can be unique for each customer (in a non-shared partition deployment). The

5620 SAM supports the definition of flow thresholds for the following TCP application performance metrics:

- Round Trip Time (RTT) based on session establishment delay
- Packet loss
- Average Total Delay based on TCP Delay (bidirectional combined metric)
- Standard Deviation of the Total Delay based on TCP Delay (bidirectional combined metric)

Each flow is evaluated (per metric) by the DCP as it is received to determine whether the flow qualifies as 'satisfactory', 'tolerable' or 'failed'. This evaluation leads to a set of four Apdex™ scores — with each score a numeric representation between 0 and 1, where '1' is ideal — one for each of the above described performance metrics. The RAM Server also considers the overall Apdex™ score for the application to be the lowest of any per-metric scores.

The Apdex™ state is derived by evaluating the Apdex™ score against a second set of thresholds. These thresholds are also defined within 5620 SAM, however, the Apdex™ state thresholds are
defined on a per-customer basis. This allows for different service-level expectations even in a shared partition scenario. The 5670 RAM Server automatically calculates and maintains a set of per-metric Apdex™ states for each application on each SAP as well as single per-application, per-SAP state based on the worst case state. In keeping with the Apdex™ nomenclature, the possible Apdex™states are:

- Excellent
- Good
- Fair
- Poor
- Unacceptable

Mean Opinion Scores: QoE reporting for RTP/UDP applications

Mean Opinion Scores (MOSs) are an industry-standard means of examining multiple factors that contribute to an overall quality of experience (such as CODEC type, jitter, and delay) and of creating a single, normalized score that represents the quality of experience for that real-time voice, video, or audio session.

MOSs are always between 0 and 5, where 5 represents an ideal score. MOSs are calculated by the AA-ISA based on the capability of the CODEC, and other measured degradation factors. The AA-ISA supports per-customer Apdex™ thresholds to convert the Apdex™ score to an actionable state; the 5620 SAM also supports this set of per-customer MOS thresholds to convert the MOS
score to a MOS state. Similarly, the 5670 RAM Server will calculate and maintain a MOS state on a per-SAP, per-application basis. Some applications support multiple MOS variants. DCP leverages the per-customer MOS thresholds across all MOS variants, resulting in a MOS state per variant. The 5670 RAM Server will consider the overall MOS/state for the application to be the worst measured case. The possible MOS states are:

- Excellent
- Good
- Fair
- Poor
- Bad

DIMENSIONING CFLOWD COLLECTION FOR THE 5670 RAM

There are several factors to consider with respect to dimensioning cflowd collection for the 5670 RAM:

- Flow export rate from the network, which is heavily dependent on the flow sample rates
- The ratio of TCP Performance versus UDP/RTP performance versus cflowd volume records
- The number of subnet-to-subnet combinations (route of interest)
- The number of applications and application groups being tracked

Each DCP can collect flow records from numerous routers — the dimensioning factor is the flow collection rate. Each DCP can collect up to 100,000 flows per second. As a rule of thumb, each AA-ISA generates approximately 2,000 flows per second per GB of traffic (depending on the configured sampling rate).

The second important scaling factor is the volume of records produced by the deployed DCPs for consumption by the 5670 RAM Server. The DCPs compress TCP and flow-volume records, however, RTP/UDP records are passed directly to the 5670 RAM Server.

With respect to compression of TCP performance and cflowd-volume reports, the DCP supports configuration of DCP groups. These groups are configured within 5620 SAM on a per-service basis. A typical objective of DCP groups is to achieve VPN site-to-site level granularity. With this in mind, extremely large service sizes can have a significant impact on 5670 RAM scaling as the site-to-site combinations are N*(N-1) in a fully meshed service. Consider a 500 site service with 10 applications could generate 2.5 million records per interval to the 5670 RAM server (25% of the server capacity). These extreme service size cases must be identified and factored into the overall sizing activity as they may be the majority contributor to data growth if not carefully managed.

With this in mind, extremely large service sizes can have a significant impact on 5670 RAM scaling as the site-to-site combinations are N*(N-1) in a fully meshed service. Consider that a 500-site service with 10 applications could generate 2.5 million records per interval to the 5670 RAM Server (25% of the server capacity). These extreme service size cases must be identified and factored into the overall sizing activity, as they may be the majority contributor to data growth if not carefully managed.

Each 5670 RAM Server supports a maximum of 10 million records per interval in total. This is a combined metric that includes AA accounting files (from SAM), as well as records arriving from the DCP.

The number of applications (and associated app groups) for which performance reporting is enabled is also a key dimensioning factor that affects the degree of compression achieved for TCP performance reporting.

Voice, video and audio flow records are not summarized at a DCP level. Each application session results in two records to the 5670 RAM server.

CONCLUSION

Application performance management is a key enabler for service provider differentiation and offers high value for end customers. The Alcatel-Lucent approach — which leverages network-based application assurance capability and a pre-integrated software stack — ensures a cohesive end-to-end solution and carrier-class scalability with minimal incremental investment.