

Using Cisco Service Assurance Agent and Internetwork Performance Monitor to Manage Quality of Service in Voice over IP Networks

Document ID: 13938

Introduction

Prerequisites

- Requirements

- Components Used

- Conventions

QoS Issues in a VoIP Network

Managing QoS with Cisco SAA and IPM

- Design

- Results

Related Information

Introduction

This document describes the use of Cisco Service Assurance Agent (SAA) and Internetwork Performance Monitor (IPM) to measure quality of service (QoS) in Voice over IP (VoIP) networks. This information is based on a real-world IP Telephony project. This document focuses on the application of the products, not on the products themselves. You should already be familiar with Cisco SAA and IPM and have access to the required product documentation. See Related Information for references to other documentation.

Note: The Cisco SAA functionality in Cisco IOS® software was formerly known as Response Time Reporter (RTR).

When you are managing a large-scale VoIP network, you must have the necessary tools to objectively monitor and report on the voice quality in the network. It is not feasible to rely on user feedback alone, because it is often subjective and incomplete. Voice quality problems typically stem from network QoS problems. So, when you identify voice quality issues, you need a second tool to manage and monitor the network QoS. The example in this document uses Cisco SAA and IPM for this purpose.

Cisco Voice Manager (CVM) is used with Telemate.net to manage voice quality. It reports on the voice quality of calls via the Impairment/Calculated Planning Impairment Factor (ICPIF) that is calculated by a Cisco IOS gateway for each call. This allows the network manager to identify sites that suffer from poor voice quality. Refer to Managing Voice Quality with Cisco Voice Manager (CVM) and Telemate for more information.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

This document is not restricted to specific software or hardware versions, but the examples in this document

use these software and hardware versions:

- Cisco IOS Software Release 12.1(4)
- IPM 2.5 for Windows NT
- Catalyst 4500 Series switch

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Conventions

For more information on document conventions, refer to the Cisco Technical Tips Conventions.

QoS Issues in a VoIP Network

Several factors can degrade voice quality in a packetized voice network:

- Packet loss
- Excessive delay
- Excessive jitter

It is particularly important that you monitor these figures on an ongoing basis, if packet switched services are used in the WAN (for example, ATM, Frame Relay, or IP Virtual Private Network). There are numerous scenarios where congestion in the carrier network, misconfigured traffic shaping on the edge devices, or misconfigured policing on the carrier side can cause packet loss or excessive buffering. When the carrier is dropping packets, there is no obvious evidence on the edge devices. Therefore, you need an end-to-end tool like Cisco SAA that can inject traffic on the ingress and validate its successful arrival at egress.

Managing QoS with Cisco SAA and IPM

There are three Cisco SAA and IPM components:

- RTR probe
- RTR responder
- IPM console

The RTR probe sends a burst of packets to the RTR responder. The RTR responder turns them around and sends them back to the probe. This simple operation allows the probe to measure packet loss and round trip delay. To measure jitter, the probe sends a control packet to the responder before it initiates the packet burst. The control packet informs the responder how many milliseconds (ms) to expect between each packet in the burst. The responder then measures the inter-packet delay during the burst, and any deviation from the expect interval is recorded as jitter.

The IPM console controls the QoS monitoring. It programs the RTR probes with the relevant information via Simple Network Management Protocol (SNMP). It also collects the results via SNMP. No command-line interface Cisco IOS configuration is required on the RTR probes.

Issue the **rtr responder** global configuration command, to manually configure the RTR responders.

The RTR probes and responders must run Cisco IOS Software Release 12.0(5)T or later. The latest maintenance release of 12.1 mainstream is recommended. The RTR probes and responders in the examples in this document are running release 12.1(4). The IPM version in use is IPM 2.5 for Windows NT. A patch is

available on Cisco.com for this version. This patch is important, as it fixes a problem where IPM configures the RTR probes with an incorrect IP Precedence setting (Cisco bug ID CSCds02402 (registered customers only)).

Design

Before you deploy a Cisco SAA and IPM solution, you must do some design work with these considerations in mind:

- Placement of RTR probes and responders
- Traffic type that is sent from probe to responder

There are a number of things to take into consideration when you decide about the placement of probes and responders. First, you want the QoS measurement to cover every site, not just problem sites. This is because the delay and jitter numbers that IPM reports for a given site are most useful when compared to other sites in the same network. Thus, you want to measure sites with good QoS *and* poor QoS. Also, a well-performing site may become a poor-performing site tomorrow, due to changes in traffic patterns or network changes. You will want to detect this before it affects voice quality and is reported by the users.

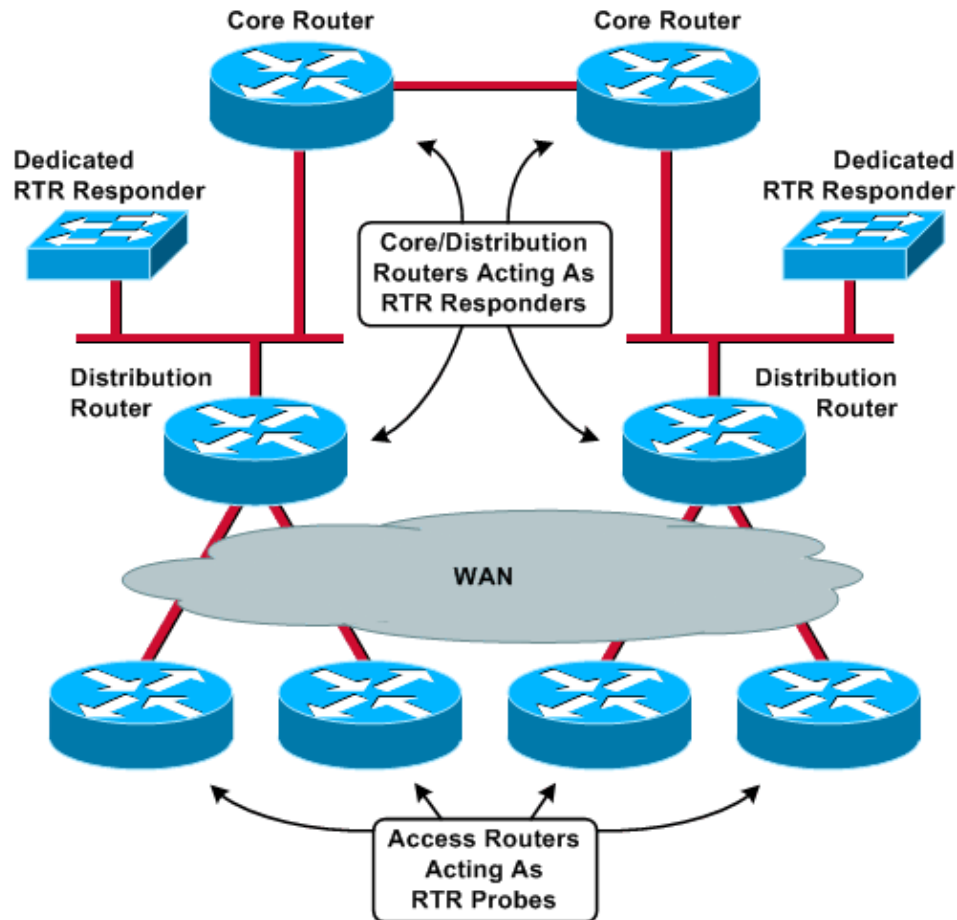
Second, CPU utilization is important. An already busy router may not be able to service the RTR component in a timely manner, and this may skew the results. Also, if you place too many probe instances on any single router, you might create high CPU utilization problems even though none existed before. The approach chosen for the example network in this document (and this should work in most networks) is to place the RTR probes on the remote/branch routers. These routers typically connect a single LAN to a relatively slow WAN service. Hence, branch routers often have very low CPU utilization and can easily cope with RTR. The other benefit of this design is that you distribute the load across as many routers as possible. Keep in mind that it is more work to be a probe than to be a responder, as probes take a certain amount of SNMP polling.

With this design, the RTR responders must be placed in the core. The responders will be busier than the probes, because they will be responding to many probes. Thus, a robust design deploys dedicated routers that act solely as responders. Most organizations have retired routers on the shelf that can perform this function. Any router with an Ethernet interface will suffice. Alternatively, the core/distribution routers can double as responders. The network diagram in this section depicts both scenarios.

Spread the load across as many routers as possible, and monitor the RTR CPU usage with this command:

```
Router# show processes cpu | i Rtt|PID

PID  Runtime(ms)  Invoked  uSecs  5Sec  1Min  5Min  TTY Process
67   0             7        0      0.00% 0.00% 0.00% 0 Rtt Responder
```



When you are matching probes with responders, it is recommended that you maintain a consistent topology between probe and responder. For example, all probes and responders should be separated by the same number of routers, switches, and WAN links. Only then can IPM results be directly compared among sites.

In this example, there are 200 remote sites and four core/distribution sites. A Catalyst 4500 at each distribution site acts as a dedicated RTR responder. Each of the 200 remote routers acts as a RTR probe. Each probe targets the responder that is located at the directly connected distribution site.

The bursts of traffic sent by the probes to the responders must be given the same QoS levels by the network as is given to voice. This may mean that you have to adjust the low latency queueing (LLQ) or Routing Table Protocol (RTP) Priority configurations on the router, so that traffic from the RTR probes is subject to strict priority queuing. When you are configuring the probe for RTP packets, only the destination User Datagram Protocol (UDP) port can be controlled and not the source port. A typical LLQ router configuration in this example network has access lists that specifically classify the RTR packets into the same queue as voice:

```
class-map Voicertp
  match access-group name IP-RTP

policy-map 192Kbps_site
  class Voicertp
    priority 110

ip access-list extended IP-RTP
  deny ip any any fragments
  permit udp 10.0.16.0 0.255.239.255
    range 16384 32768 10.0.16.0 0.255.239.255
    range 16384 32768 precedence critical
  permit udp any any eq 20000 precedence critical
  permit udp any eq 20000 any precedence critical
```

The IP–RTP access list has these classifying lines:

- deny ip any any fragments

Deny any IP fragment, as a layer 4 access list implicitly permits these.

- permit udp 10.0.16.0 0.255.239.255 range 16384 32768 10.0.16.0
0.255.239.255 range 16384 32768 precedence critical

Permit RTP packets from voice subnets with IP precedence set to 5.

- permit udp any any eq 20000 precedence critical

Permit RTP packets from RTR probe going to RTR responder.

- permit udp any eq 20000 any precedence critical

Permit RTP packets from RTR responder going back to RTR probe.

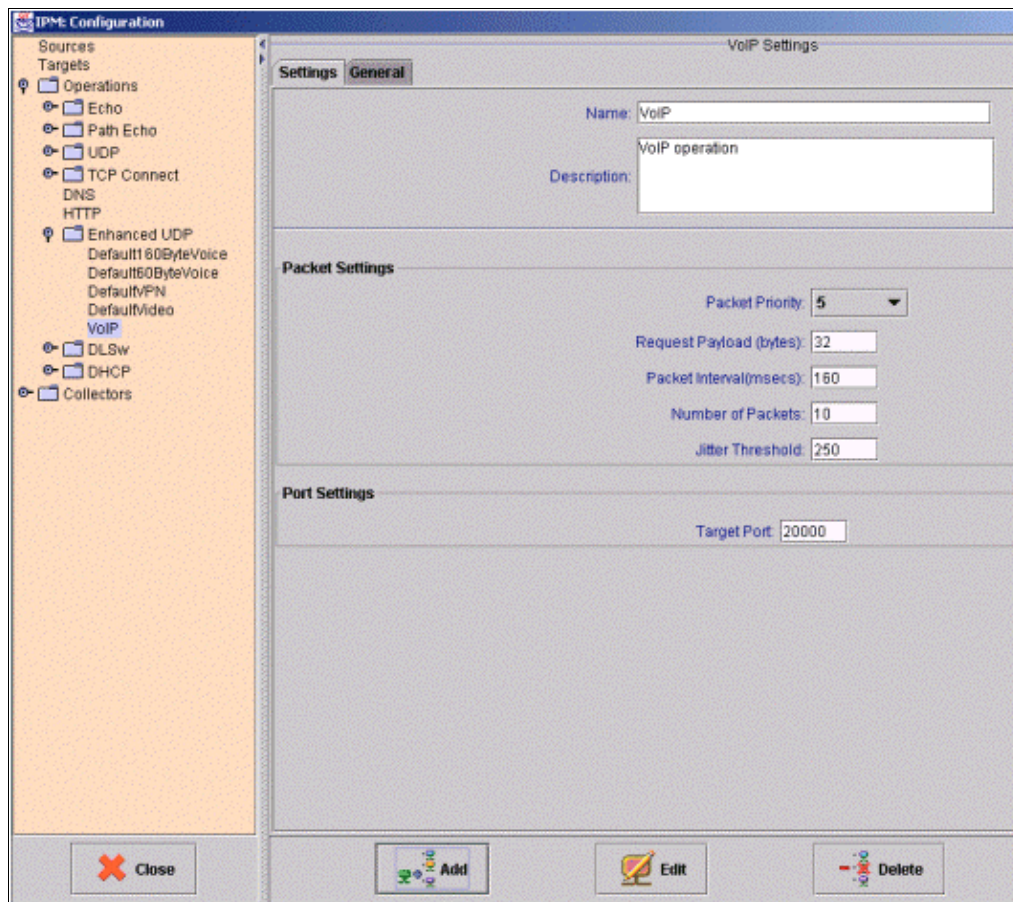
Be careful that the addition of RTR traffic does not cause the LLQ queues to be over–subscribed and cause real voice packets to be dropped. The standard **Default60ByteVoice** IPM operation sends bursts of RTP packets with these parameters:

- Request Payload: 60 bytes

Note: This is the RTP header and voice. Add 28 bytes (IP/UDP) to get the L3 datagram size.

- Interval: 20 ms
- Number of Packets: 10

This means that, during a burst, RTR consumes 35.2 Kbs of LLQ bandwidth. If there is not sufficient bandwidth for LLQ, then create a new IPM operation and increase the packet interval. With the parameters shown in this IPM configuration window, a burst consumes only 1 Kbps of bandwidth:



Results

The table in this section is an example of an IPM report. This report contains three RTR probe instances. Keep in mind that one physical probe may be configured with multiple RTR probe instances which target different responders or use different payloads.

Daily Jitter Summary Report										
11/15/2000										
Collector Info		Round Trip Latency		Src Dest Jitter		Dest Src Jitter		Completions		
Collector	Operation	Avg	Avg Max	Avg	Avg Max	Avg	Avg Max	Trys	Over %	Error %
haw-WN	VoIP	72.71	102.79	1.74	7.65	2.62	25.88	1440	0%	0%
	Last-Week	75.65	105.41	1.73	4.16	4.97	24.18	10113	0%	1%
	Last-Month	74.89	103.01	1.70	3.77	6.74	24.98	7822	0%	1%
wat-WN	VoIP	72.27	121.88	2.17	12.50	3.19	39.13	1447	0%	1%
	Last-Week	75.45	112.96	1.99	5.18	5.40	31.21	10127	0%	1%
	Last-Month	74.00	110.51	1.83	4.91	6.44	29.76	7826	0%	1%
sfd-WN	VoIP	70.43	114.13	1.80	8.08	2.68	32.08	1440	0%	0%
	Last-Week	73.92	112.17	1.75	4.68	4.94	30.19	10098	0%	1%
	Last-Month	72.90	104.13	1.79	4.82	6.41	27.30	7831	0%	1%

These are the meanings of each of the columns:

Avg:

IPM calculates an average for each hour of sampling. These hourly averages are then averaged across a longer period to get the daily, weekly, or monthly averages. In other words, for the daily report, IPM calculates the average for each hour for the past 24 hours. It then calculates the daily average as the average of these 24 averages.

Avg Max:

This value is the average of all hourly maximums for each day, week, and month in the chart. In other words, for the daily report, IPM takes the largest sample reported within each of the past 24 hours. It then calculates the daily maximum average as the average of these 24 samples.

Over %:

This is the percentage of samples that were over the configured threshold for the collector.

Error %:

This is the percentage of packets that encountered an error. A jitter probe reports several types of errors:

- SD Packet Loss Packets lost between source and destination
- DS Packet Loss Packets lost between destination and source
- Busies The number of occasions when a round-trip time (RTT) operation could not be initiated because a previous RTT operation had not been completed
- Sequence The number of RTT operation completions received with an unexpected sequence identifier. These are some possible reasons why this might occur:
 - ◆ A duplicate packet was received.
 - ◆ A response was received after it had timed-out.
 - ◆ A corrupted packet was received and was not detected.
- Drops The number of occasions when either of these occurred:
 - ◆ An RTT operation could not be initiated because some necessary internal resource was not available (for example, memory or the Systems Network Architecture [SNA] subsystem)
 - ◆ Operation completion could not be recognized.
- MIA (Missing in Action) The number of packets that are lost for which no direction can be determined.
- Late The number of packets that arrived after the timeout.

The question that arises from this information is what delay, jitter, and error values are acceptable in a VoIP network. Unfortunately, there is no simple answer to this question. Acceptable values depend on codec type, jitter buffer size, and other factors. In addition, there are interdependencies between these variables. A higher packet loss may mean that less jitter can be tolerated.

The best way to obtain workable delay and jitter figures is to compare similar sites in the same network. If all 192 Kbps-attached sites but one report jitter values around 50 ms, and the remaining site reports 100 ms jitter, then there is a problem, regardless of the nominal values. IPM can provide ongoing 24x7 delay and jitter measurement for the entire network, and it can provide a baseline to use as the benchmark for delay and jitter comparisons.

Errors are a different, however. In principle, any error percentage other than zero is a red flag. The RTR packets are given the same QoS treatment as voice packets. If the network QoS and call admission control is robust, no level of congestion should cause packet loss or excessive delay for voice or RTR packets.

Therefore, you can expect the IPM error counts to be zero. The only errors that could be considered normal are cyclic redundancy check (CRC) errors, but these should be rare in a quality infrastructure. If they are frequent, they constitute a risk to voice quality.

Related Information

- [Network Monitoring Using Cisco Service Assurance Agent \(SAA\)](#)
 - [IPM User Documentation](#)
 - [IPM Software Download Site](#)
 - [Voice Technology Support](#)
 - [Voice and IP Communications Product Support](#)
 - [Recommended Reading: Troubleshooting Cisco IP Telephony](#)
 - [Technical Support & Documentation – Cisco Systems](#)
-

[Contacts & Feedback](#) | [Help](#) | [Site Map](#)

© 2009 – 2010 Cisco Systems, Inc. All rights reserved. [Terms & Conditions](#) | [Privacy Statement](#) | [Cookie Policy](#) | [Trademarks of Cisco Systems, Inc.](#)

Updated: Jan 04, 2006

Document ID: 13938
