

Introduction to Quality of Service

Contents

Introduction to Quality of Service	1
SCOPEL Supported VoIP Protocols	1
SCOPEL Supported SIP Video CODECs	2
SCOPEL Supported Voice CODECs	2
Introduction to G.107 E-Model	2
Factors Affecting the R Value	3
Jitter	3
Network Transmission Delay	4
CODEC Vs. MOS	4
Jitter Buffer	4
Jitter Buffer and Delay	5
Bandwidth Usage per CODEC	5
CODEC MOS Vs. Random Packet Loss	5
Transcoding Delay	6
LAN Switching QoS Methods	6
LAN QoS Implementation Over Shared Data Cables	6
Scenario 1	6
Scenario 2	7
Scenario 3	7
LAN QoS Implementation Over Dedicated Data Cables	7
Scenario 1	7
Scenario 2	8
LAN QoS Implementation Over WANs	8
Recommendations	8
Other Factors to Consider	9

The use of IP as a foundation for converged networks has raised several issues for both enterprise IT departments and ISPs. IP and Ethernet are connectionless technologies and do not guarantee bandwidth. The protocol will not differentiate network traffic based on the type of flow to ensure that the proper amount of bandwidth and prioritization level are defined for a particular type of application. Since IP does not support the prioritization of network traffic network managers and service providers must make their network components aware of applications and their various performance requirements.

The E-Model (ITU-T Recommendation G.107) is a tool that can estimate the end-to-end voice quality, taking the IP Telephony parameters and impairments into account.

The Mean Opinion Score (MOS) is an industry standard model that uses the ITU-T Recommendation G.107 E-Model to calculate the performance of Voice over Internet Protocol (VoIP) over IP networks.

SCOPEL Supported VoIP Protocols

- SIP
- MGCP
- H.323

- SCCP/SKINNY
- IAX2
- Google Talk
- Jabber/XMPP
- ENUM
- DUNDi

SCOPTTEL Supported SIP Video CODECs

- H.261
- H.263
- H.263+
- H.264

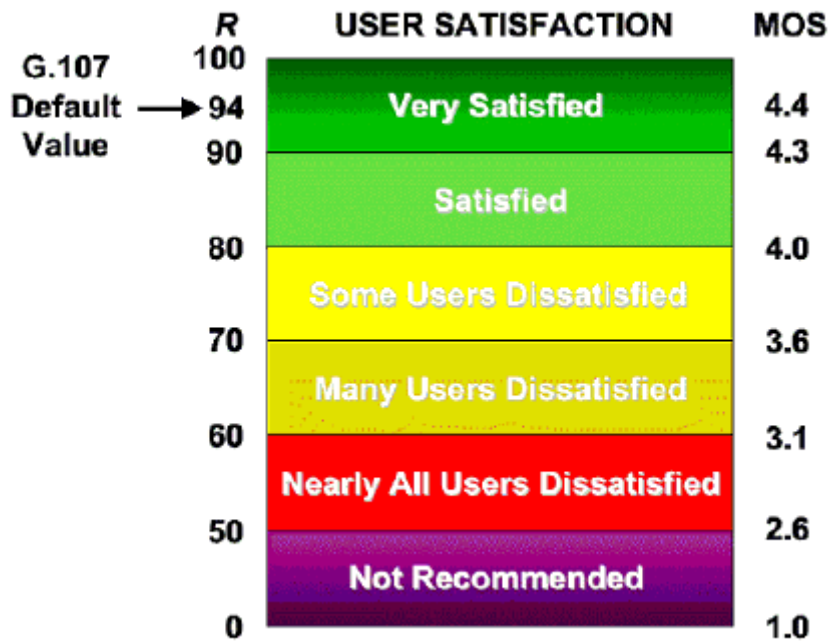
SCOPTTEL Supported Voice CODECs

- G.711 (ulaw)
- G.711 (alaw)
- G.722
- G.726
- G.729 (requires third party license or transcoding hardware)
- 16 bit Signed Linear PCM (slin)
- GSM
- LPC10
- Speex
- ADPCM

Introduction to G.107 E-Model

The output of the E-Model is a scalar called the “Rating Factor”, the “R-value”, or simply R. The scale is typically from 50 to 100, where everything below 50 is clearly unacceptable and everything above 94.15 (the maximum with the G.107 E-Model, version 19 default values) is unobtainable in narrowband (300 to 3400 Hz) telephony.

The R scale translates into a subjective model for QoS measurement known as the MOS (Mean Opinion Score).



Factors Affecting the R Value

Several factors within the VoIP application influence the output of the R Value

- Delay, including delay variation, transcoding and jitter buffers
- Echo
- Speech compression (CODEC)
- Packet loss

Echo is a result of too much delay.

Choppy speech is a result of too much packet loss.

Jitter

Jitter, also called delay variation, indicates the differences in arrival times among all datagrams sent during a VoIP call.

When a datagram is sent, the sender gives it a timestamp which is placed in the RTP header.

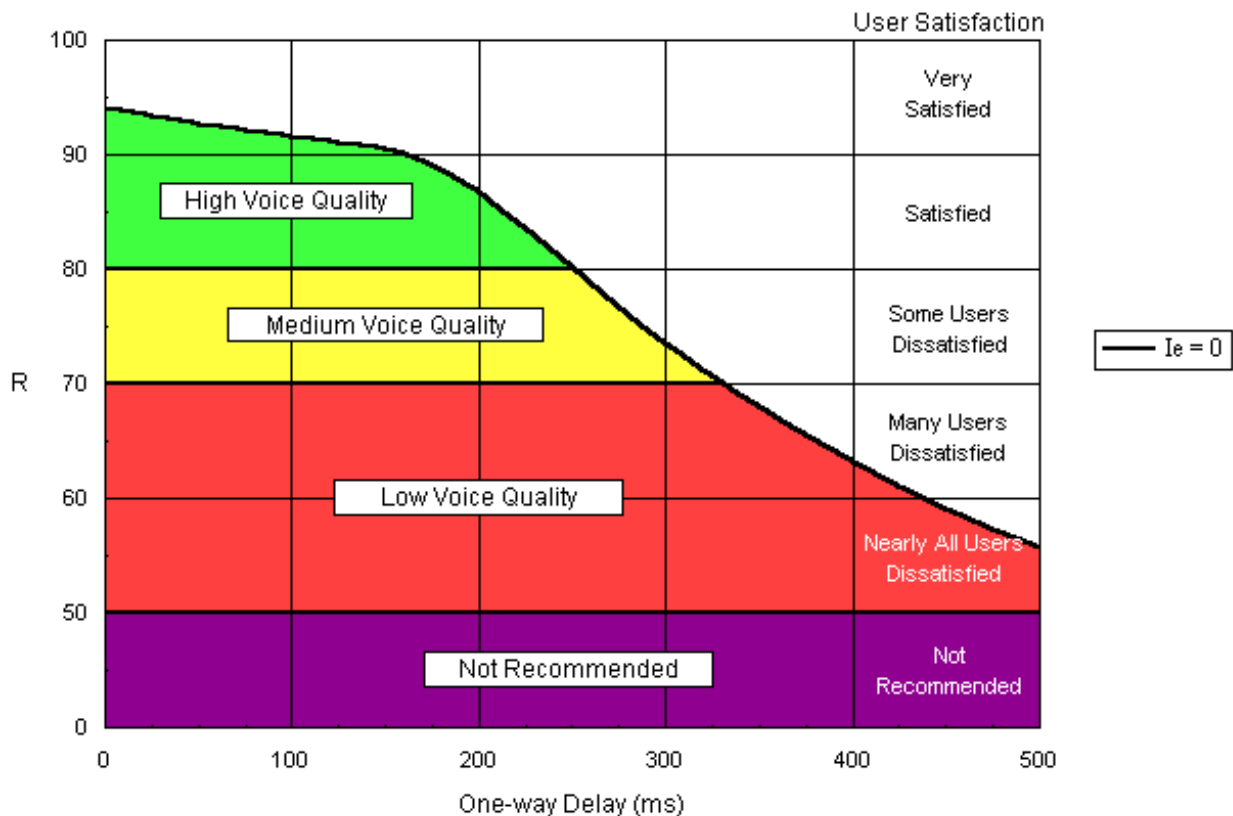
When it's received, the receiver adds another timestamp.

These two timestamps are used to calculate the packet's transit time.

If the transit times for datagrams within the same call are different, the call contains jitter.

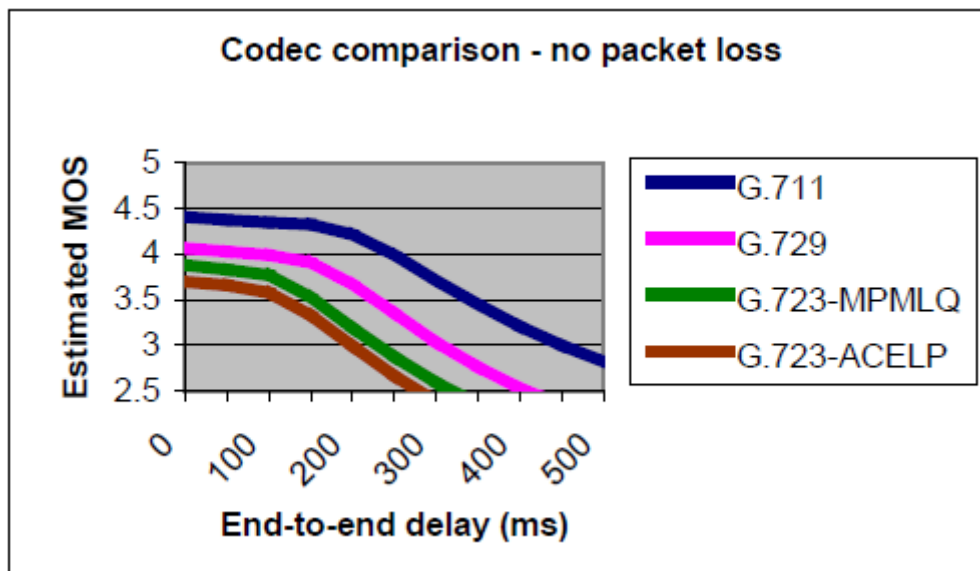
In a video application, jitter manifests itself as a flickering image, while in a telephone call, its effect may be similar to the effect of lost data: some words may be missing or garbled or delayed.

Network Transmission Delay



CODEC Vs. MOS

Delay affects some CODEC's more than others



Jitter Buffer

To lessen the impact of jitter, the jitter buffer may adjust itself dynamically based on the perceived jitter. As datagrams arrive, they are placed in the jitter buffer, which holds them long enough to supply them to the CODEC at a more constant rate.

If a datagram arrives too early or too late, it may not fit in the jitter buffer and is discarded.

You'd like to make the jitter buffer just large enough to handle any variation due to the data network.

Jitter Buffer and Delay

Jitter buffers solve the lost and late packets problem by holding datagrams to compensate for perceived jitter delay. As datagrams arrive, they are placed in the jitter buffer, which holds them long enough to supply them to the CODEC at a more constant rate.

Jitter buffers solve the lost and late packets problem by adding delay that reduces the available delay budget. Truly lost (discarded) packets will never show up. This translates into even lower voice quality for a given amount of propagation delay. The goal is to minimize jitter buffer delay.

Packetization delay is included in the MOS estimate, as is the "jitter buffer delay," the delay introduced by the effects of buffering to reduce inter-arrival delay variations.

Example: A fixed jitter buffer of 60ms will add 60ms of latency to each RTP packet.

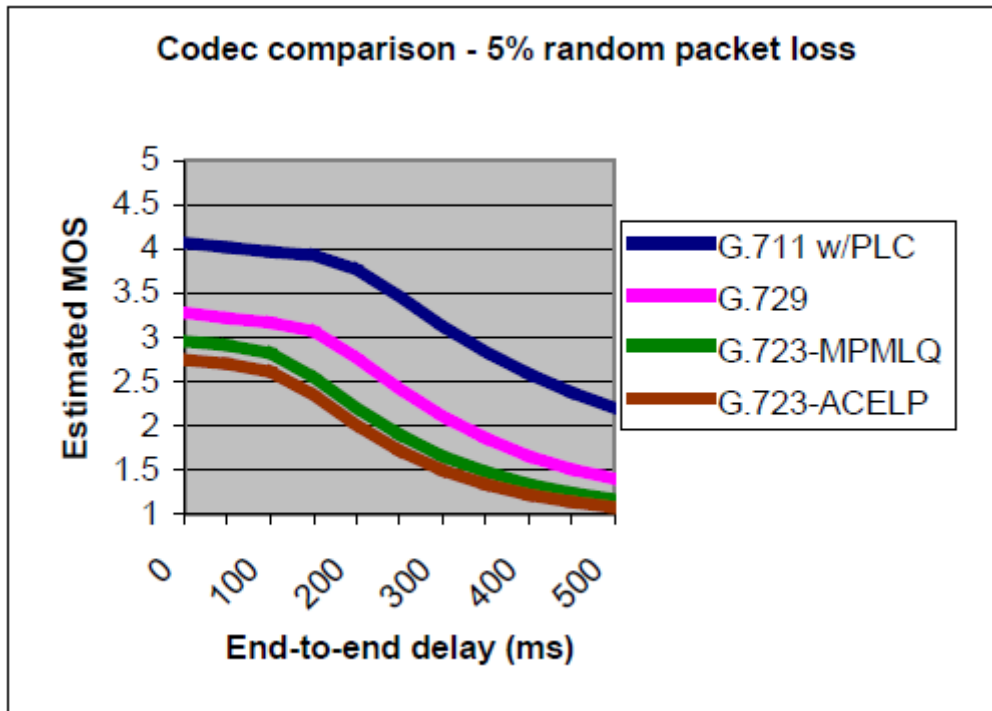
Bandwidth Usage per CODEC

Each CODEC requires a different amount of bandwidth for both the transmit and receive RTP streams. And each CODEC carries a maximum MOS score.

Codec	Data Rate	Typical Datagram Size	Packetization Delay	Combined Bandwidth for 2 Flows	Typical Jitter Buffer Delay	Theoretical Maximum MOS
G.711u	64.0 kbps	20 ms	1.0 ms	174.40 kbps	2 datagrams (40 ms)	4.40
G.711a	64.0 kbps	20 ms	1.0 ms	174.40 kbps	2 datagrams (40 ms)	4.40
G.726-32	32.0 kbps	20 ms	1.0 ms	110.40 kbps	2 datagrams (40 ms)	4.22
G.729	8.0 kbps	20 ms	25.0 ms	62.40 kbps	2 datagrams (40 ms)	4.07
G.723.1 MPMLQ	6.3 kbps	30 ms	67.5 ms	43.73 kbps	2 datagrams (60 ms)	3.87
G.723.1 ACELP	5.3 kbps	30 ms	67.5 ms	41.60 kbps	2 datagrams (60 ms)	3.69

CODEC MOS Vs. Random Packet Loss

Packet Loss affects some CODEC's more than others.



Transcoding Delay

Transcoding is defined as two or more encodings of a signal through different types of CODECS. Example: GSM EFR to G.711 to G.729

It can take considerable time in hardware or software to transcode CODECs and the time taken adds to overall end to end delay.

Since transcoding delays vary depending on the types of algorithms and hardware used the delay cannot be predicted but transcoding can add very significant end to end delay.

The transcoding delay is added to network delay + jitter buffer + other factors

LAN Switching QoS Methods

Layer 3 LAN switches supporting 802.1p 802.1q and/or DiffServ should always be used on the LAN.

These switches forward frames at Layer 2 speeds in order to reduce packet forwarding delays.

These switches typically prioritize traffic using Weighted Queuing hardware queues mapped to Layer 2 and Layer 3 priority markings.

These switches mark egress LAN traffic with priority markings that can be carried by MPLS networks in order to preserve QoS methods over WAN links.

LAN QoS Implementation Over Shared Data Cables

Scenario 1

When a phone and PC share a cable drop the phone is plugged into the data jack and the PC is plugged into the phone. Therefore the phone must support VLAN tagging and should also support VLAN priority tagging.

In one scenario the phone can be a member of the default VLAN but must have VLAN tagging enabled in order to send Layer 2 priority tags to the switch. The phone must have a higher layer 2 priority than the PC port. The phone priority should be set to a priority of 5 or 6 so the switch can prioritize traffic in its default priority queues. The PC port can be set to best effort or priority 3. The default VLAN is usually VLAN 1. So both the phone and the PC port are both members of VLAN 1 but voice frames are forwarded with a much higher priority by the switch.

Caveats:

- The DHCP server and subnet must be able to support enough IP addresses for the phones and the PC's.

- The VoIP server and any other required network device must be on a port where the priority is statically set to 5 or 6.

Scenario 2

When a phone and PC share a cable drop the phone is plugged into the data jack and the PC is plugged into the phone. Therefore the phone must support VLAN tagging and should also support VLAN priority tagging.

In this scenario the phone can be a member of the voice VLAN but must have its VLAN ID tagging enabled in order to send VLAN tags to the switch. The PC can have its VLAN tagging disabled since it will automatically be a member of the default VLAN. The major advantage of this approach is that a phone or PC can be plugged into any trunk or hybrid port and will automatically be in the correct VLAN due to the VLAN ID tag defined in the phone's configuration.

Caveats:

- Each switch port must support trunk or hybrid ports.
- Each trunk or hybrid port must allow untagged traffic from VLAN 1.
- Each trunk or hybrid port must allow tagged traffic from the configured voice VLAN.
- The VoIP server must be on a port with a PVID belonging to the voice VLAN.
- It is recommended that an IP address be assigned to each VLAN on the switch so the switch can route from one VLAN to other VLANs. Typically the VLAN IP address is used as the default gateway for each device on this VLAN and routes are configured on the switch.
- Any device that does not support VLAN tagging must be plugged into a port with a PVID belonging to the correct VLAN.

Scenario 3

When a phone and PC share a cable drop the phone is plugged into the data jack and the PC is plugged into the phone. Therefore the phone must support VLAN tagging and should also support VLAN priority tagging.

In this scenario the phone and Ethernet switch can support Cisco CDP in which case the phone will automatically be put into VLAN 100 when it boots and connects to the switch.

Caveats:

- The VoIP server must be on a port with a PVID belonging to the CDP voice VLAN 100.
- It is recommended that an IP address be assigned to each VLAN on the switch so the switch can route from one VLAN to other VLANs. Typically the VLAN IP address is used as the default gateway for each device on this VLAN and routes are configured on the switch.
- Any device that does not support VLAN tagging must be plugged into a port with a PVID belonging to the correct VLAN 100
- Any DHCP, DNS, TFTP server, or gateway etc... required for the voice VLAN must be plugged into a port with a PVID belonging to the correct CDP VLAN 100
- If your phone and switch natively support CDP but you do not wish to use CDP then manually disable CDP support on the phone.

LAN QoS Implementation Over Dedicated Data Cables

Scenario 1

When a phone and PC each have a dedicated cable plugged into dedicated switches for voice and data a QoS mechanism does not have to be used. This is because dedicated switches have the same effect of isolating traffic and broadcast storms as a configured VLAN.

Caveats:

- The voice and data networks cannot be members of the same subnet.
- Any DHCP, DNS, TFTP server, or gateway etc... required for the voice subnet must be plugged into the correct network.

- Any DHCP, DNS, TFTP server, or gateway etc... required for the data subnet must be plugged into the correct network.
- Routes must be configured for data and voice VLANs if traffic is to be routed between subnets.
- The SCOPEL server is by default bound to listen for voice and video traffic on any interface. This means that the WAN port of the SCOPEL server can listen even if it is configured in NAT/Gateway mode and the inbound ports are allowed by the SCOPEL firewall.

Scenario 2

When a phone and PC each have a dedicated cable plugged into shared switches for voice and data a QoS mechanism does not have to be used. This is because each configured VLAN isolates traffic and broadcast storms on each configured VLAN. The switch must be configured with one or more data VLANs. Each VLAN must have an IP address assigned. Routes must be built for each VLAN if packet forwarding is required. The switch VLAN IP address becomes the default gateway for each VLAN. A phone or other voice device must be plugged into a switch port configured for the VLAN PVID of the voice VLAN ID. A PC or other device must be plugged into a switch port configured for the VLAN PVID of the required VLAN ID. This method requires a dedicated switch port for each phone and network device.

Caveats:

- The voice and data networks cannot be members of the same subnet.
- Any DHCP, DNS, TFTP server, or gateway etc... required for the voice subnet must be plugged into the correct network.
- Any DHCP, DNS, TFTP server, or gateway etc... required for the data subnet must be plugged into the correct network.
- Routes must be configured for data and voice VLANs if traffic is to be routed between subnets.
- The SCOPEL server is by default bound to listen for voice and video traffic on any interface. This means that the WAN port of the SCOPEL server can listen even if it is configured in NAT/Gateway mode and the inbound ports are allowed by the SCOPEL firewall.

LAN QoS Implementation Over WANs

Multi-Protocol Label Switching [MPLS] is similar to DiffServ in some respects, as it also marks traffic at ingress boundaries in a network, and un-marks at egress points. But unlike DiffServ, which uses the marking to determine priority within a router, MPLS markings (20-bit labels) are primarily designed to determine the next router hop. MPLS is not application controlled (no MPLS APIs exist), nor does it have an end-host protocol component. Unlike any of the other QoS protocols we describe in this paper, MPLS resides only on routers. And MPLS is protocol-independent (i.e., "multi-protocol"), so it can be used with network protocols other than IP (like IPX, ATM, PPP or Frame-Relay) or directly over data-link layer as well [MPLS Framework, MPLS Architecture].

The purpose of MPLS networks is to extend an existing private network over the Internet but by using private network links created and sold by carriers. These networks are typically sold as secure layer 2 extensions of the Ethernet network.

MPLS can honour marked Layer 2 priority and VLAN tagging of frames originating from the LAN. Therefore MPLS is an ideal way of preserving QoS over Internet connections.

Recommendations

- Total end to end delay should never exceed 250ms.
- The highest quality CODEC should always be used in order to preserve the MOS score since network jitter effects higher quality CODECs less than compressed CODECs.
- Avoid transcoding whenever possible to preserve the network MOS.
- Avoid the use of large jitter buffers whenever possible.
- Use VLANs to separate voice and video traffic from data traffic.
- Take voice and video UDP traffic into consideration when sharing the network with TCP traffic.

- Use QoS switches on the LAN and MPLS networks on the WAN or dedicated voice/video WAN connections whenever possible.

Other Factors to Consider

Operating system. What version of operating system is running on the routers, switches, firewalls, and other devices? Is it a version that can support VoIP traffic? Does it have the proper functionality to support VoIP?

Memory. How much memory (RAM) is installed in the routers? Is there enough memory to support VoIP functions well? Is there enough memory to support the number of calls that will be added to the network?

QoS. Most vendors recommend some quality of service mechanism. Do the network devices support those QoS mechanisms? Is QoS already configured on the routers? What QoS mechanism is in use? How is VoIP traffic to be prioritized?

VLANs. A virtual LAN (VLAN) is used to group or segregate LAN traffic by users. VLANs allow for different data classes to be prioritized by the switches using the 802.1p/Q protocol. Do the switches support VLANs and 802.1p/Q? Do the switches have VLANs already configured?

Shared LAN hubs. Shared hubs offer no QoS guarantees. Any device attached to the hub, even an IP phone, end up competing with any other attached devices for bandwidth. Consider upgrading all shared hubs in the network to switches.

Interface speed. The interfaces in the routers operate at various speeds. Are the interfaces 56 kbps, 1.544 Mbps, 10 Mbps, 100 Mbps, or 1000 Mbps (gigabit)? Do the interfaces support full-duplex mode of operation? Do the interface speeds support the number of VoIP calls that will be added to the network?

Power to the phone. If you're about to upgrade your switches, ask your vendor if the specific platform supports.