Quality of Service (QoS) for Voice Traffic Prioritization - Official Technical Overview & Hardware Datasheet

SYSTEM OVERVIEW

This document provides a comprehensive technical overview of Quality of Service (QoS) mechanisms as implemented in carrier-grade network infrastructure, with specific focus on real-time voice traffic prioritization. The methodologies described herein ensure deterministic latency, jitter control, and lossless delivery for delay-sensitive voice streams across converged IP networks. Leveraging hardware-accelerated classification, marking, queueing, and shaping engines, our platform delivers sub-millisecond forwarding guarantees for voice traffic under full link saturation.

INTERNAL ASIC & FORWARDING PIPELINE DESIGN

The hardware-based QoS engine resides within the network processor ASIC, operating at line rate without CPU intervention. Key architectural elements include:

- Classification Engine: Performs deep packet inspection (DPI) up to Layer 7, identifying voice signaling (SIP, H.323, MGCP) and RTP media streams via protocol-specific signatures and port ranges (5060/5061 for SIP, 16384-32767 for RTP).
- Marking Module: Hardware-rewrites DSCP (Differentiated Services Code Point) bits within IPv4 ToS or IPv6 Traffic Class fields. Standard voice marking uses DSCP EF (Expedited Forwarding, 46) for RTP and DSCP AF31 (26) for SIP signaling.
- Policing & Shaping: Single-Rate Two-Color and Two-Rate Three-Color meters (RFC 2697, RFC 2698) enforce per-flow or per-class committed information rates (CIR). Excess voice traffic is dropped to preserve voice quality.
- Queueing Hierarchy: Low-Latency Queueing (LLQ) with Priority Queue (PQ) strict scheduling. Voice class receives absolute priority over all other traffic classes (data, video, best-effort).

PHYSICAL & ENVIRONMENTAL SPECIFICATIONS

Form Factor: 1RU Rack-Mountable Chassis

Switching Capacity: 336 Gbps (non-blocking)

Forwarding Rate: 250 Mpps (64-byte packets)

Voice-Focused Jitter Guarantee: <1ms under 80% load

Latency (Voice Class, 64B packets): <100 microseconds (store-and-forward)

Power Supply: 1+1 Redundant, Hot-swappable AC (100-240V) or DC (-48V)

Max Power Consumption: 85W (idle) - 220W (full load)

Operating Temperature: 0°C to 50°C (standard) / -40°C to +75°C (ruggedized option)

MTBF (Mean Time Between Failures): 350,000 hours (Telcordia SR-332)

Parameter	Voice Traffic Priority Specification
Classification Method	DPI (SIP/RTP), L3/L4 ACL, NBAR
DSCP Marking	EF (46) for RTP, AF31 (26) for SIP
Queueing Discipline	Low-Latency Queueing (LLQ) with Strict Priority
Policing Compliance	RFC 2697 (Single-Rate Three-Color Marker)
Shaping Granularity	64 kbps to 10 Gbps (1 kbps steps)
Jitter Buffer Support	Hardware-accelerated, 0-150ms adaptive
Max Concurrent Voice Flows	32,768 (G.711, G.729, Opus)

Drop Probability (Voice)	0% under CIR, tail drop for excess

QUALITY OF SERVICE PROTOCOLS & STANDARDS COMPLIANCE

The platform fully implements the following QoS frameworks and recommendations:

- IEEE 802.1p (CoS): 3-bit priority field within VLAN tag. Voice assigned CoS 5 (highest non-reserved).
- IETF RFC 2474 / 2475 (DiffServ): DSCP EF (46) for voice payload, DSCP AF31 (26) for voice control.
- IETF RFC 2597 (AF PHB): Assured Forwarding for signaling classes.
- IETF RFC 3246 (EF PHB): Expedited Forwarding per-hop behavior.
- ITU-T Y.1541: Network performance objectives for real-time voice (Class 0).
- MEF 23.1: Class of Service Phase 2 – Voice-specific performance attributes.

Hardware offload supports up to 8K unique traffic classes (flow-based) and 16 egress queues per port.