Introduction

Quality of Service (QoS) mechanisms prioritise critical traffic, curb congestion, and guarantee predictable performance. Among the formal QoS frameworks, **Integrated Services (IntServ)** and **Differentiated Services (DiffServ)** approach the problem from opposite ends of the scalability-versus-granularity spectrum. Understanding how they allocate resources and signal requirements helps administrators choose the model that best fits their workload profile.

IntServ: Per-Flow Guarantees

IntServ, standardised in RFC 1633, employs the Resource Reservation Protocol (RSVP) to establish explicit state for **every individual flow** along each hop. Routers track bandwidth, delay, and jitter commitments, enabling hard guarantees such as strict end-to-end latency for voice or telepresence sessions (Kurose & Ross, 2021). The strength of IntServ lies in its **fine-grained, deterministic** service level: once a reservation succeeds, the application knows capacity is locked in. However, per-flow state scales linearly with the number of sessions, imposing heavy memory and CPU burdens on core routers as the network grows.

DiffServ: Class-Based Prioritisation

DiffServ, defined in RFC 2475, sidesteps per-flow complexity by **aggregating traffic into classes**. Each packet carries a DSCP value in the IP header; edge devices mark these values, while core routers perform simple, stateless behaviours such as expedited forwarding (EF) or assured forwarding (AF) (White & Donohue, 2018). DiffServ therefore scales efficiently to carrier-grade backbones because routers only reference a small class table. The trade-off is **statistical assurance**: high-priority classes receive preferential treatment but no absolute guarantee if the class itself becomes congested.

Deployment Considerations

Criterion	IntServ	DiffServ
State in routers	Per-flow	Per-class
Scalability	Limited beyond thousands of flows	Handles millions of flows
Guarantee type	Hard, deterministic	Relative, probabilistic
Signalling	RSVP required end-to-end	Edge-only marking; no core signaling
Management overhead	High (flow setup/tear-down)	Moderate (policy definition)

Example: When to Choose Each Model

Video-conferencing inside a hospital.

Within a single campus hospital, latency-sensitive telemedicine feeds demand unwavering millisecond-level delay. With only a few hundred concurrent streams, the administrator can deploy IntServ/RSVP on internal switches to reserve bandwidth per session and pass annual compliance audits.

Cloud provider backbone.

Conversely, a public IaaS provider moves terabits of mixed tenant traffic across its backbone. Maintaining per-flow state is infeasible. DiffServ lets engineers mark storage replication as AF, real-time gaming as EF, and bulk backups as best-effort, ensuring high-value workloads experience minimal jitter without overloading router TCAM tables.

Conclusion

IntServ delivers airtight, flow-level guarantees at the cost of scalability and complexity, making it ideal for controlled domains with stringent service-level objectives. DiffServ trades deterministic precision for class-based simplicity, excelling in large-scale or multi-tenant networks where aggregate fairness and core router efficiency matter most. Selecting the right model hinges on flow count, administrative reach, and the criticality of absolute versus statistical performance assurances.

Word count: 436

References

Kurose, J. F., & Ross, K. W. (2021). *Computer networking: A top-down approach* (8th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/computer-networking/P200000003334/9780135928615?srsltid=AfmBOorih30FoFFXS7R1FwMnK dgtTIgdR8PgP1AJSzyOs6W3IE5plx54

White, C., & Donohue, G. (2018). *QoS quick start: A practical guide for Cisco networks* (2nd ed.). Cisco Press.