Bandwidth—bits per second (bps) of a network link or usage. Voice (telephone) is a small and consistent amount of data (lots of tiny packets). Video varies widely based on how well it compresses at a given moment (how similar one frame is to the frame before).

QoS (Quality of Service)—prioritizes packets based on type to benefit time-sensitive applications like voice and video at the expense of things like file transfer where the difference won't be noticed. The goal is to control 3 factors:

	Goal For Voice	Goal For Video
Delay	≤150 ms	200-400 ms
Jitter—Delay Variability	≤ 30 ms	30 -50 ms
Loss	≤ 1%	0.1% to 1%

PACKET MARKING

Packets are inspected once to determine their type and header fields are set. From then on, other network devices in the enterprise use the headers to choose packets for priority handling. This may happen at the access switch, when the packet first enters the network. All L4 TCP and L3 IP fields are available to help make the choice.

NBAR (Network Based Application Recognition)—Cisco-proprietary QoS marking technology that analyses the type and source of packets in greater detail than just TCP port numbers.

Layer 3 IPv4 Marking

Preferable because the same header is used end-to-end. (L2 headers are replaced at each router).

TOS (Type of Service)—1-byte IPv4 header field. Its use has varied over time.

- Old Way—IPP (3 bits) + other (5 bits)
- New Way—DSCP (6 bits) + ECN (2bits)

DSCP (Differentiated Services Code Point)—Current way of assigning one of 64 classification markings to a packet.

IPP (IP Precedence)—Deprecated. First 3 bits of the TOS field. Provided 8 possible markings

Layer 3 IPv6 Marking

Traffic Class—one-byte field

Layer 2 802.1Q (VLAN) Ethernet Marking

802.1Q defines a 3-bit field for QoS marking. This only applies to trunks using 802.1Q VLAN encapsulation and disappears at the router, just like all L2 headers. The field has two interchangeable names:

- CoS (Class of Service)
- PCP (Priority Code Point)

TRUST BOUNDARIES

Since user devices can set DSCP and/or CoS marking, it's routine for admins to set a trust boundary such that only their own devices (switches & routers) can set QoS markings that are respected. For example, any QoS markings from a user would be rewritten by the access layer switch. The trust boundary would be in the middle of that switch.

IP phones are generally trusted and they rewrite the markings from any attached (daisy-chained) PC.

DIFFSERV

DiffServ standardizes some values for DSCP marking, with both values (shown in decimal), names, and uses.

Name	Value (decimal)	Uses
EF (Expedited Forwarding)	46	Voice audio payloads. Low latency, jitter, loss.
AF (Assured Forwarding)	(12 values)	
CS (Class Selector)	8 values, multiples of 8 (0x to 7x)	

AF (Assured Forwarding) defines 4 queues, each with 3 drop priorities, for a total of 12 values. This can be expressed as AFxy, where x is the queue (higher better) and y is the likelihood of drop (lower better). The following table shows that name and the decimal value for each. The highest priority queue is upper left, with the highest decimal value.

	Queue 4 (Best)	Queue 3	Queue 2	Queue 1 (Worst)
Drop = 1 (Unlikely)	AF41 DSCP = 34	AF31 DSCP = 26	AF21 DSCP = 18	AF11 DSCP = 10
Drop = 2 (Moderate)	AF42 DSCP = 36	AF32 DSCP = 28	AF22 DSCP = 20	AF12 DSCP = 12
Drop = 3 (Likely)	AF43 DSCP = 38	AF33 DSCP = 30	AF23 DSCP = 22	AF13 DSCP = 14

CS (Class Selector) uses only the bits of the old IPP field (the first three bits of the DSCP field). The lower 3 DSCP bits are set to 0. The result is 8 possible DSCP values, each a multiple of 8 (0, 8, 16, …, 56), corresponding to the old IPP values of 0 to 7.

QUEUING & SCHEDULING

A single output port can have more than one queue, each FIFO (First In First Out), to accommodate multiple classes of traffic. A scheduler can then implement prioritization as it chooses which queue to select a packet from for dispatch.

Round Robin—The queues take turns feeding packets to the output

Weighted Round Robin—The scheduler preferentially takes more or fewer packets from each queue

CBWFQ (Class-Based Weighted Fair Queuing)—Implements weighted round robin. Weights are configured as a percentage of total bandwidth. At worst, (congestion) each queue gets its share.

LLQ (Low Latency Queuing)—For voice & video. One or more high priority queues are exempted from round robin and always have their traffic immediately sent.

Starvation—If the traffic in an LLQ priority queue fills or exceeds the capacity of an interface, the other queues will never send anything.

Policing—Eliminate starvation by limiting entry of packets into the priority queue to a % of interface bandwidth, dropping the rest.

CAC (Call Admission Control)—Works as part of routing to prevent policing, starvation, etc. by not even trying to route more voice/video out a given interface than it can handle [out of scope].

SHAPING & POLICING

CIR (Committed Information Rate)—The amount of bandwidth that you have contracted for from a service provider or telco.

Policing—Keep bandwidth below a limit by chopping off the peaks and throwing them away. This is more of a tool for telcos and not something you would choose to do to yourself.

- Track bandwidth usage
- Keep usage below a limit by chopping off the peaks and throwing them away
- You can merely mark packets that are beyond the CIR (as "optional") so they can be dropped further downstream only if actual congestion requires it
- "Bursting" possible after inactivity (brief excess as long as the average is below CIR)

Shaping—Smooth bandwidth usage to keep below a limit by storing up packets and releasing them as capacity allows. This is a good way to make the most of a link if the telco is using policing.

Time Interval—How often a router sends the amount of data allowed in that time interval (at full line speed) before waiting until the next time interval. The result averages at a lower rate than line speed.

- A large interval can increase voice/video jitter by a matching amount—even the LLQ priority queue must wait for the next interval.
- Cisco recommends a 10 ms interval for video/voice (1/100 second)

CONGESTION AVOIDANCE

TCP Window Size—The number of bytes that can be sent before all of the segments get acknowledged at once. A field in the receiver's acknowledgement allows the receiver continual control over the window size in order to "throttle" the sender.

- If there were no transmission errors and no need to throttle, the receiver might double the window size with each acknowledgement, making large transmissions efficient
- If segments are lost, the receiver might shrink the window by half (for each missing segment) to limit the quantity of data that must be resent in response to future errors

Tail Drop—If a queue is full don't add a new packet; just drop it.

Congestion Avoidance—Monitor the filling of the queue. As it fills, drop more and more packets in the hopes that the transmission errors will cause TCP to shrink its window and slow down. The dropping can be sensitive to DSCP markings in order to slow down less sensitive flows.