

★ 1. Classification

What it means:

Identifying and grouping traffic into categories.

Why it's needed:

Before giving priority to traffic, the network must know **what type of traffic it is**.

How it works:

Routers/switches look at packet details such as:

- Source/Destination IP
- TCP/UDP ports (e.g., port 5060 = SIP, RTP = voice packets)
- VLAN ID
- Protocol type
- MAC address
- Application ID (like Skype, Teams)

Simple Explanation:

“Classification is sorting packets before deciding how to treat them.”

★ 2. Marking

What it means:

Assigning a **priority label** to the packet.

Two types of marking:

- **CoS (Layer 2 marking)** → used by switches
- **DSCP (Layer 3 marking)** → used by routers

Example values:

- **Voice** → DSCP 46 (EF)
- **Video** → AF41
- **Normal Data** → 0

Simple Explanation:

“Marking puts a sticker on the packet telling the network how important it is.”

★ 3. Policing

What it means:

Controlling traffic rate by **dropping** or **remarking** packets when they exceed a limit.

Purpose:

- Prevents one type of traffic from using too much bandwidth
- Protects network from congestion

How it works:

If allowed bandwidth = 10 Mbps but traffic goes above that:

- Excess packets are **dropped** or
- Marked with **lower priority**

Simple Explanation:

“Policing is like a speed limit. If packets go over the limit, they get dropped.”

4. Shaping

What it means:

Controlling the **flow** of traffic by **buffering (holding)** and sending it at a smooth rate.

Difference from policing:

- **Policing drops extra traffic**
- **Shaping delays extra traffic** (holds it in queue)

Used in:

- WAN links
- Internet connections
- MPLS circuits

Simple Explanation:

“Shaping smooths out the traffic flow instead of dropping packets.”

5. Queuing

What it means:

Arranging packets in **different queues** (lines) based on priority.

Why it's needed:

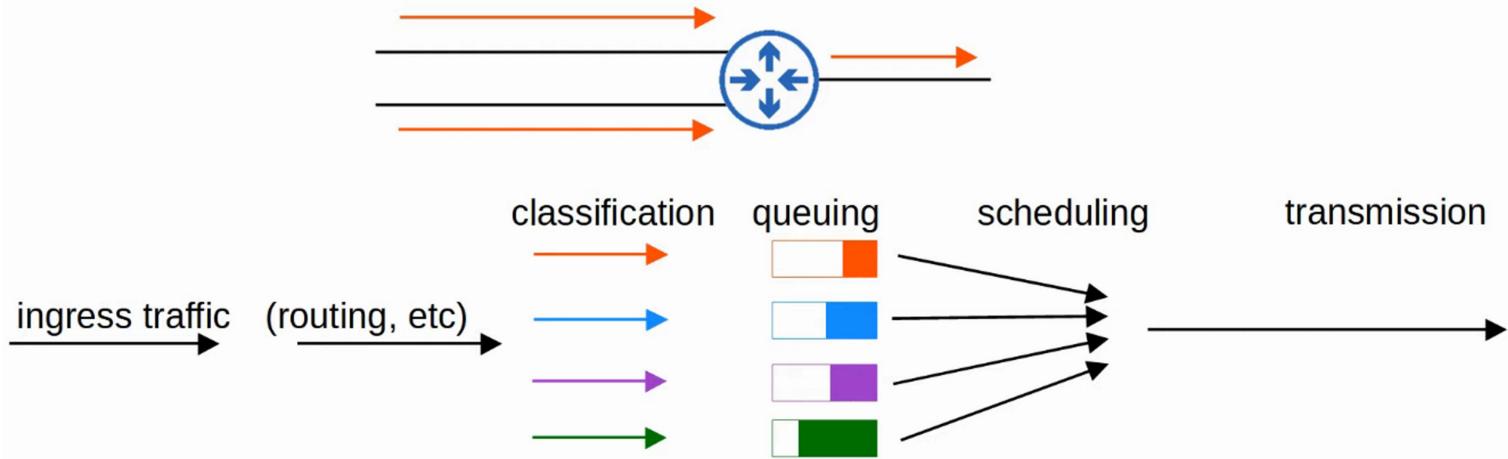
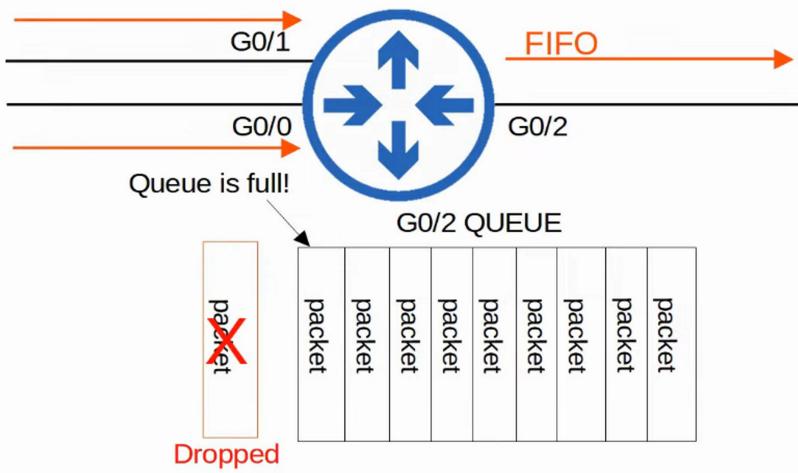
During congestion, all packets cannot be forwarded at the same time.

Types of queues:

- High priority queue (voice)
- Medium priority queue (video)
- Low priority queue (data)
- Background queue (updates/backups)

Simple Explanation:

“Queuing puts important packets in the front of the line.”



★ 1. FIFO (First In, First Out) Queuing

Meaning:

Packets are forwarded in the **exact order** they arrive. No priority.

Features:

- No classification
- No prioritization
- Simple, default method on many devices

Best for:

Low-traffic networks

Simple explanation:

“First come, first served — like a single line at a shop.”

★ 2. Priority Queuing (PQ)

Meaning:

Traffic is divided into queues with different priority levels (high, medium, low).

How it works:

- **High priority queue is always served first.**
- Lower priority queues are served only when higher queues are empty.

Problem:

Low priority traffic may starve (never get sent).

Best for:

Voice, real-time traffic

Simple explanation:

"VIP queue always goes first. Others wait."

★ 3. Weighted Fair Queuing (WFQ)

Meaning:

Provides **fair bandwidth sharing** among flows.

How it works:

- Automatically classifies traffic by flow
- Assigns each flow a "weight"
- Higher weight = more bandwidth

Best for:

Mixed traffic environments

Simple explanation:

"Everyone gets a fair share, but important flows get a bigger slice."

★ 4. Class-Based Weighted Fair Queuing (CBWFQ)

Meaning:

You create **custom classes** and assign **bandwidth or queue limits** to each.

How it works:

- Admin manually defines classes (e.g., voice, video, data)
- Each class gets a certain percentage of bandwidth

But:

No strict priority queue for voice (that's why LLQ was created).

Simple explanation:

"You build your own queues and decide how much bandwidth each gets."

★ 5. Low Latency Queuing (LLQ)

Meaning:

Adds a **strict priority queue** to CBWFQ specifically for real-time traffic.

Why:

Voice needs extremely low delay.

How it works:

- Voice traffic goes into Priority Queue (PQ)
- Other traffic uses CBWFQ queues

Best for:

VoIP, video conferencing

Simple explanation:

“Voice gets a special fast lane that always goes first.”

★ 6. Scheduling

What it means:

Deciding **which queue gets serviced first** and **how much bandwidth** each queue receives.

Scheduling algorithms include:

- **Priority Queuing (PQ)** → voice always first
- **Weighted Fair Queuing (WFQ)** → fair share for all traffic
- **Class-Based Weighted Fair Queuing (CBWFQ)** → customizable queues
- **LLQ (Low Latency Queue)** → adds strict priority for VoIP

Simple Explanation:

“Scheduling decides which queue gets priority and how fast packets are sent.”

★ Summary Table (Very Simple)

QoS Method	What It Does	Simple Meaning
Classification	Identify traffic	Sorting packets
Marking	Give priority label	Put sticker on importance
Policing	Drop extra traffic	Enforce speed limit
Shaping	Smooth flow by delaying	Make traffic slow but stable
Queuing	Put traffic in separate lines	VIP line for important packets
Scheduling	Decide order/priority of queues	Who goes first