

**Classification** entails using a traffic descriptor to categorize a packet within a specific group to define that packet and make it accessible for QoS handling on the network. For example, you can use classification to mark certain packets for IP Precedence. IP Precedence is usually deployed as close to the **edge of the network** or the administrative domain as possible.

**Queuing** is designed to accommodate temporary congestion on a network device's interface by storing excess packets in buffers until bandwidth becomes available.

When a queue is full, IOS has no place to put newly arriving packets, so it discards them. This phenomenon is called **tail drop**. Often, when a queue fills, several packets are tail dropped at a time, given the bursty nature of data packets.

**Marking** allows the QoS level of the packet to change based upon classification or policing. Tail drop is the default drop mechanism.

**Traffic shaping** prevents the bit rate of the packets exiting an interface from exceeding a configured shaping rate. To do so, the shaper monitors the bit rate at which data is being sent. If the configured rate is exceeded, the shaper delays packets, holding the packets in a shaping queue. The shaper then releases packets from the queue such that, over time, the overall bit rate does not exceed the shaping rate.

**Random Early Detection (RED)** monitors the average queue size and drops packets based on statistical probabilities. If the buffer is almost empty, all incoming packets are accepted. As the queue grows, the probability for dropping an incoming packet grows too. **When the buffer is full, the probability has reached 1 and all incoming packets are dropped.**

Shaping implies the existence of a queue and of sufficient memory to buffer delayed packets, while policing does not. Queuing is an outbound concept; packets going out an interface get queued and can be shaped. **Only policing can be applied to inbound traffic on an interface.** Ensure that you have sufficient memory when enabling shaping. In addition, shaping requires a scheduling function for later transmission of any delayed packets. This scheduling function allows you to organize the shaping queue into different queues. Examples of scheduling functions are Class Based Weighted Fair Queuing (CBWFQ) and Low Latency Queuing (LLQ).