



Guaranteed bandwidth for critical applications

Patrik Fältström

<paf@cisco.com>

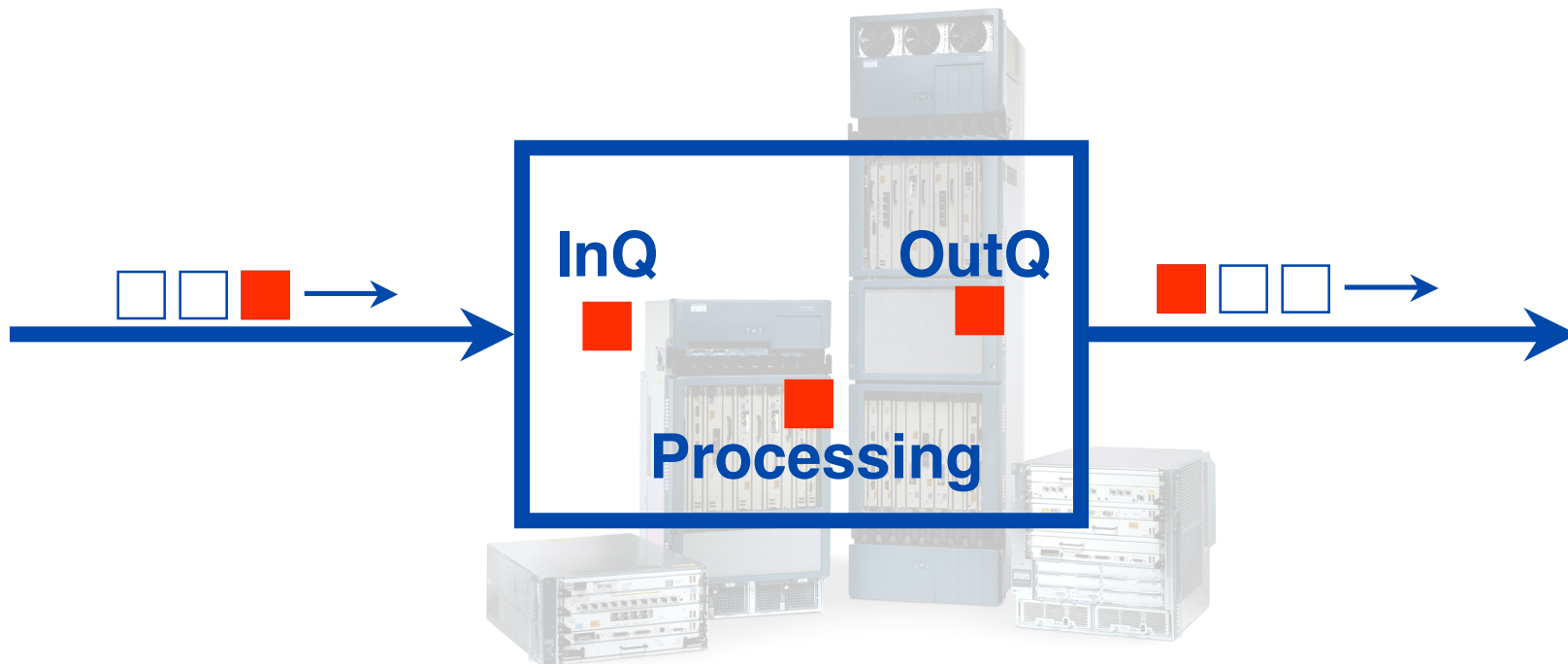
Area Director, Applications Area, IETF
Corporate Consulting Engineer, Cisco Systems

Congestion



- **Dropped packets**
 - If data is needed, it needs to be retransmitted
- **Delay**
 - A packet is stuck in queue, waiting to be sent
- **Variable delay (jitter)**
 - The delay might vary for different packets in the same stream
- **Different applications have different needs**

One network



What can happen?

- Packets arrive randomly
- Wait in InQ for processing
 - Overloaded processing unit
- Time for processing
 - Constant per packet
- Wait in OutQ for free slot in network
 - Depending on transmission what happens



Fastest path

- No packet in InQ, so processing start immediately
- When packet arrives to OutQ, it can be immediately sent onto network
- Only delay is processing time



What is QoS?

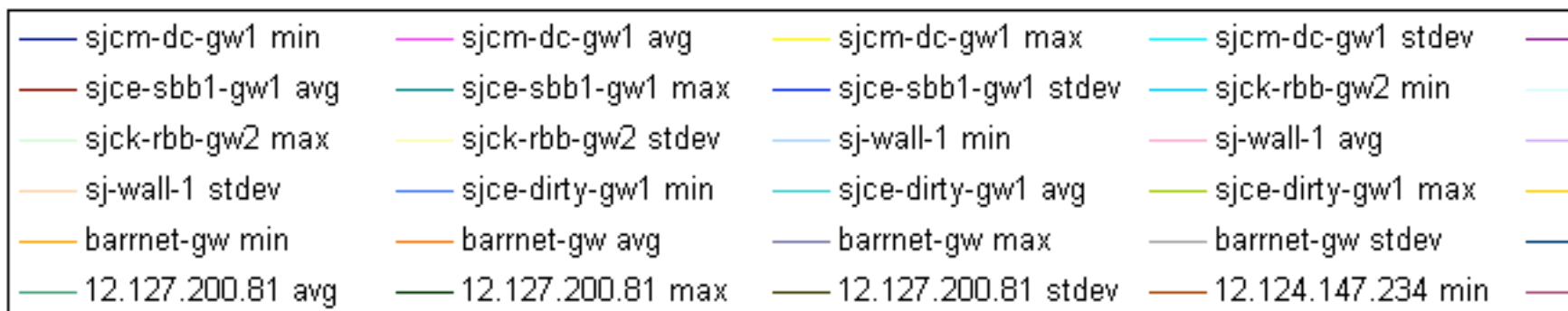
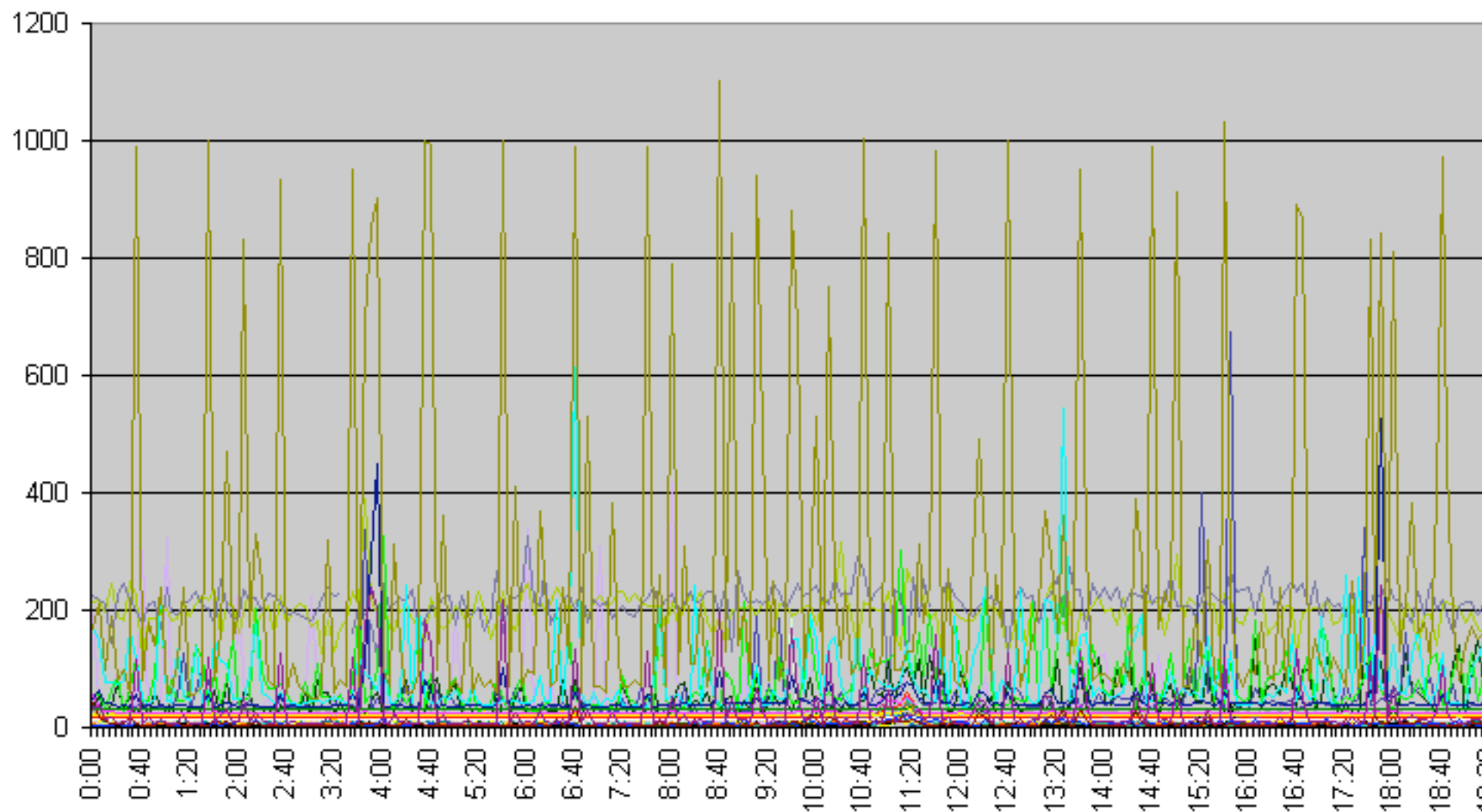
- If there are packets in InQ or OutQ, or possibly the queues are full so the packets have to be dropped, what to do?
- I.e. simplest thing is “priority” in the queues
 - Note, if there is no queue, priority doesn’t matter
 - Some transmission mechanisms have other implications



When do we need QoS?



- **We need QoS when either:**
 - **We get more packets per time unit in than processing engine can handle**
 - InQ is filled, and at one stage it will be full
 - **We have processed more packets per time unit than transmission can handle**
 - OutQ is filled, and at one stage it will be full
 - **While queues are filled, only delay will increase**
 - **If incoming traffic is bursty, queue size can handle temporary load without dropping packets**
 - Remember that different kind of traffic have different needs, for example VoIP can handle dropped packets better than delay or jitter



Patrik R 3:40 3:20 3:00 2:40 2:20 2:00 1:40 1:20 1:00 3:20 kholm

