

Flow control. Quality of services.

Ing. Yelena Trofimova

Department of Computer Systems
Faculty of Information Technology
Czech Technical University in Prague
©Yelena Trofimova, 2021

Computer networks, BIE-PSI
SS 2020/21, Lecture 8

<https://courses.fit.cvut.cz/BIE-PSI/>



Content

- QoS
- application requirements
- parameters of QoS
- technical solutions
 - ▶ SLA, SLS
 - ▶ queuing algorithms
 - ▶ RED, WRED
 - ▶ token bucket
 - ▶ leaky bucket

How to define QoS

- definition level:
 - ▶ network – throughput, delay, etc.
 - ▶ application – frames per second, transmission time, personal evaluation for voice services, availability, etc.
- type of definition:
 - ▶ quantitative
 - ▶ qualitative (relative)
- quality of service is the result of a joint activity of applications, middleware, OS and network.

QoS parameters I

- exact definitions:
 - ▶ IETF activities
 - IP Performance Metrics (IPPM) Working Group: defines metrics for performance measurements (delay, round-trip time, loss, etc.)
 - IP Flow Information eXport (IPFIX): defines protocol for export of flow data
 - ▶ ITU-T recommendations: I.350, I.381, Y.1541, Y.1545, Y.1561, etc.
 - QoS in specific networks (ISDN, ATM, MPLS, IP, Ethernet, etc.)
- parameters and metrics of network quality of service (QoS) could be used in agreements among network providers, and between end users and their network providers
- there is no uniform universal specification
 - ▶ several standardization organizations
 - ▶ individual projects
 - ▶ continuous attempts to better specify traffic classes

QoS parameters II

- quantitative parameters
 - ▶ delay
 - ▶ jitter
 - ▶ packet loss rate, bit error rate
 - ▶ throughput
 - ▶ availability
- qualitative parameters
 - ▶ classes

Quantitative parameters

description of the physical properties of the system

- immediately measured value
- relation to some time interval
- absolute limits ("never exceeds ...")
- quantile ("does not exceed 99% of the time ...")
- mean value and variance (σ) – for normal distribution

Delay

- RTT – Round-trip Time
 - ▶ both directions
 - ▶ easy to measure
- OWD – One-way Delay
 - ▶ may be different in both directions
 - ▶ difficult to measure – requires time synchronization

Delay: components

- transmission delay – time from first to last bit on medium:
 $D_T = N/S$, where N is the number of bits, S – transmission rate
- propagation delay – physical speed:
 - ▶ speed of light for vacuum or air
 - ▶ $\sim 2 \times 10^8$ m/s for electrical cable
 - ▶ $\sim 2 \times 10^8$ m/s for optical cable
- processing delay – header analysis, searching in routing tables, etc.
- queuing delay
 - ▶ output queue of the network interface
 - ▶ input queue of device processor (e.g. router)

Jitter

- one-way delay variation
- could be expressed as
 - ▶ absolute value: $D_{max} - D_{min}$
 - ▶ relative value: %
- important for real-time network communication
- compensated by the input buffer in the application
- critical for NTP

Data loss

- measured by number of lost transfer units
 - ▶ packets – packet loss rate (PLR)
 - ▶ bits – bit error rate (BER)
- depends on the medium
 - ▶ high for a wireless network (e.g. Wi-Fi)
 - ▶ very low for optical networks
- congestion of the line
 - ▶ dropping of packets \Rightarrow data loss

Throughput

- network capacity (bandwidth) – theoretical data transfer rate
 - ▶ limited by a link with the smallest capacity
- throughput – actual data transfer rate
 - ▶ influence of repaired losses
- goodput – the application layer throughput
- available bandwidth
 - ▶ free transmission capacity
 - ▶ is rapidly changing
 - ▶ difficult to evaluate

Availability

- probability that the network (service) will be available
 - ▶ expressed relatively (e.g. 0.99 of the year)
- includes
 - ▶ technical faults
 - ▶ administrative restrictions (planned maintenance, updates, etc.)

Qualitative parameters

Classes

- best-effort
 - ▶ network does not provide any guarantees
 - ▶ user obtains unspecified variable bit rate and delivery time, depending on the current traffic load
 - ▶ standard in the Internet
- other classes – depending on the technology, different specifications, less implementations
 - ▶ sensitive traffic – guaranteed bandwidth, low packet loss rate
 - ▶ LBE (Less Than Best Effort)
 - ▶ undesired –

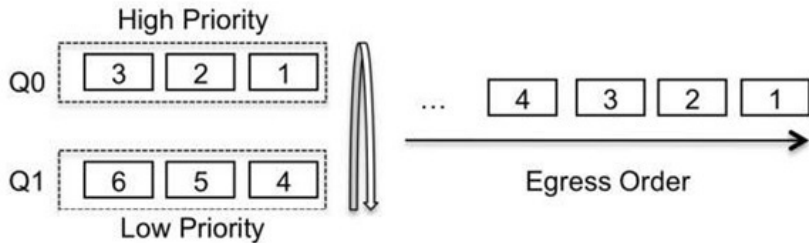
SLA/SLS

- SLA – Service Level Agreement
 - ▶ agreement on technical and non-technical service parameters between customer and provider
- SLS – Service Level Specification, specification of SLA technical parameters
 - ▶ Scope (ingress, egress, ...)
 - ▶ Flow description (IP addresses, ports, etc.)
 - ▶ Traffic envelope (peak rate p , token bucket r, b)
 - ▶ Excess treatment (drop, shape, remark)
 - ▶ Performance (loss, throughput, delay, jitter, ...)
 - ▶ Time
 - ▶ Reliability

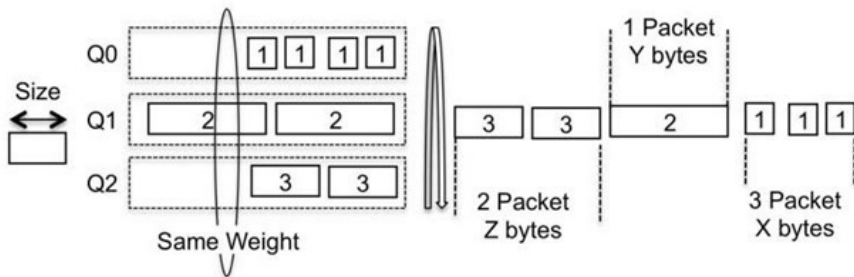
Queuing Algorithms

- based on arrival: FIFO (First In First Out)
- based on priority: PQ (Priority Queuing)
- sharing of the capacity:
 - ▶ WFQ (Weighted Fair Queuing)
 - ▶ WRR (Weighted Round Robin)
 - ▶ MDRR / DRR (Modified / Deficit Round Robin)
- prevention of congestion
 - ▶ WRED / RED ((Weighted) Random Early Detection)

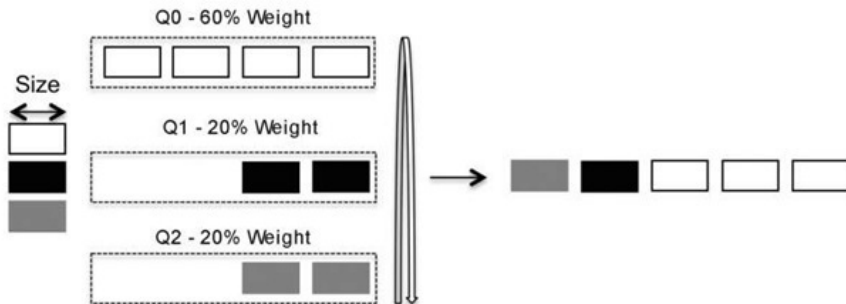
Priority Queuing



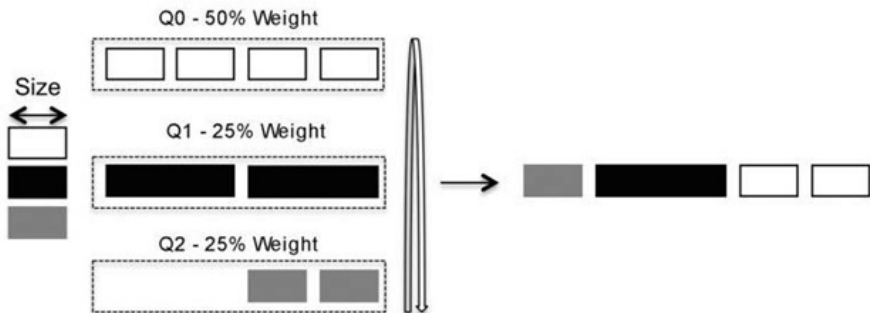
Weighted fair queuing



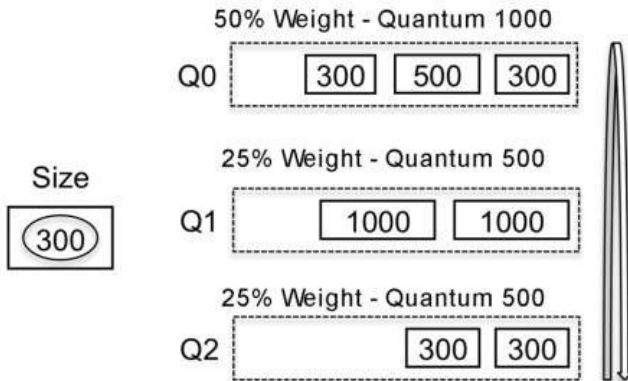
Weighted round robin I



Weighted round robin II



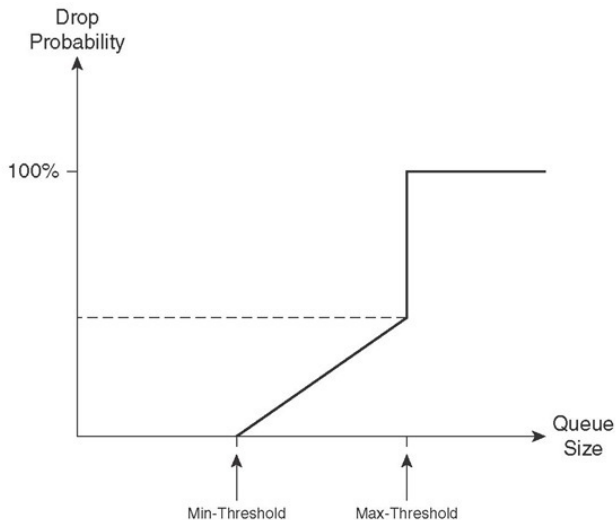
Deficit weighted round robin



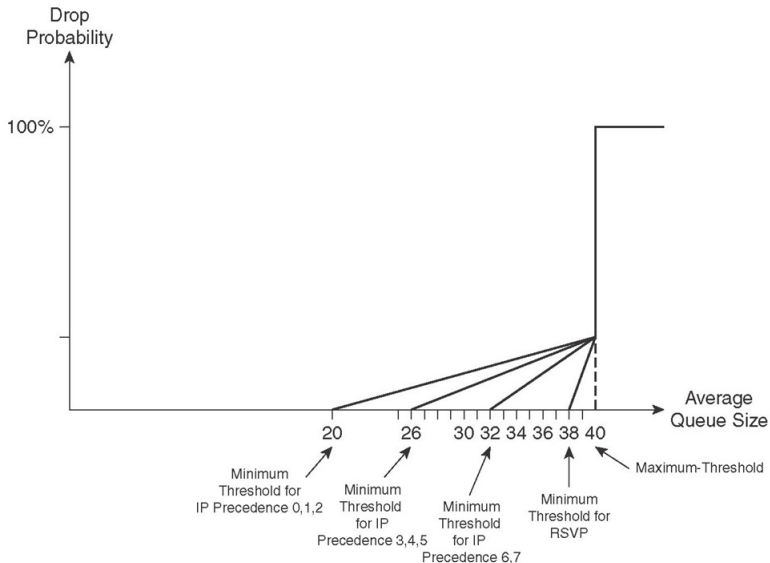
Network Congestion

- congestion
 - ▶ sharp increase in the loss rate
 - ▶ degradation of network parameters
- prevention (congestion avoidance)
 - ▶ bandwidth reservation for certain connections
 - ▶ reaction for signs of congestion – increased delay, packet dropping (TCP)

RED – random early detection



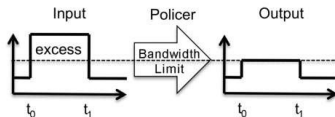
WRED – weighted random early detection



Policing & shaping: deal with burstiness and excess traffic

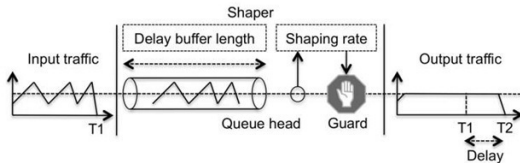
- policier, input parameters:

- ▶ bandwidth limit
- ▶ burst size limit

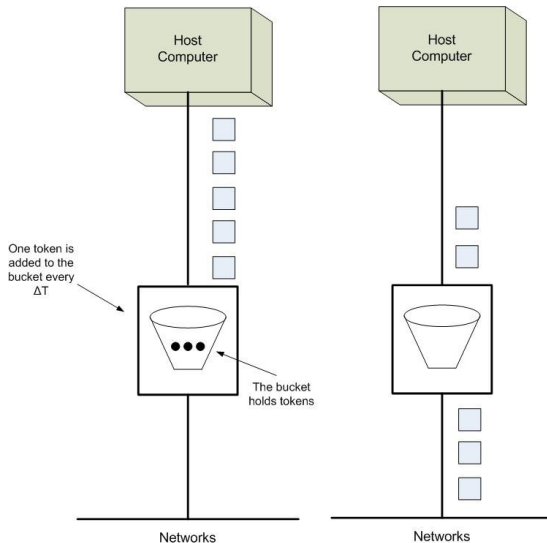


- sharper, input parameters:

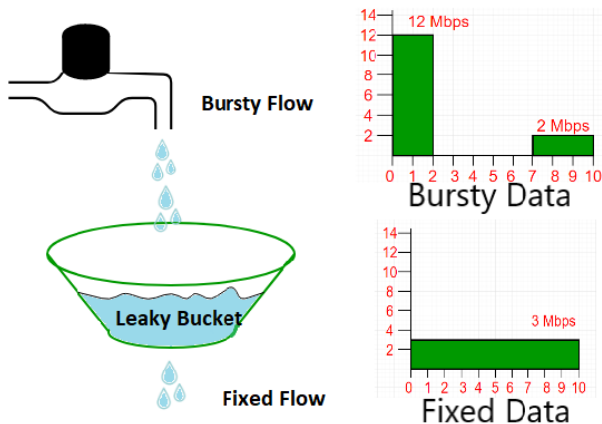
- ▶ shaping rate
- ▶ delay buffer length



Token bucket



Leaky bucket



Source: <https://www.geeksforgeeks.org>