

## 10. QoS (Quality of service)

Prof.dr Igor Radusinović

igorr@ucg.ac.me

dr Slavica Tomović

slavicat@ucg.ac.me

## QoS

- Uvod
- Oblikovanje saobraćaja
- Scheduling
- WFQ

## QoS

### Uvod

- Internet nudi best-effort servis
- Internet aplikacije imaju različite zahtjeve u pogledu performansi poput minimalne propusnosti  $R_{min}$  i maksimalnog kašnjenja  $D_{max}$
- Velike razlike u pogledu zahtijevanih performansi sugeriše potrebu različitog tretmana paketa koji prenose saobraćaje različiti aplikacija

Aplikacija	$R_{min}(kb/s)$	$D_{max}(s)$
Video streaming	60	10
VoIP	10	0,2
Download	30	200
Web Browsing	50	2
Video Conference	80	0,2

Telekomunikacione mreže 10-3

3

## QoS

### Oblikovanje saobraćaja

- Zagušenje se javlja kada paketi dolaze brzinom koja je približna ili veća od brzine kojom se poslužuju
- Paketi na Internetu obično dolaze u velikim grupama (bursts)
- Mreža neprekidno nadzire saobraćajne tokove da bi obezbijedila poštovanje dogovora
- Kada neki paket nije poslat u skladu sa dogovorom, mreža ga može odbaciti ili mu dodijeliti niži prioritet što može dovesti do naknadnog odbacivanja ako se za to javi potreba
- Ako se zagušenje pojavi, paketi označeni sa nižim prioritetom se prvi odbacuju
- *Leaky Bucket Algoritam (Algoritam probušenog bureta)* je najčešće korišćeni mehanizam ove namjene
  - Bucket (Bure) ima brzinu oticanja koja odgovara srednjoj ugovorenoj brzini
  - Bucket ima određenu dubinu tako da može podržati varijacije dolazne brzine
  - Dolazni paketi su prihvatljivi ako ne izazivaju prekoračenje

Telekomunikacione mreže 10-4

4

Leaky Bucket algoritam se koristi za administrativni nadzor dolazne brzine paketa



Brzina oticanja odgovara nekoj brzini u ravnotežnom stanju

Dubina Bucket-a odgovara maksimalnom dozvoljenoj dolaznoj nivou bursta

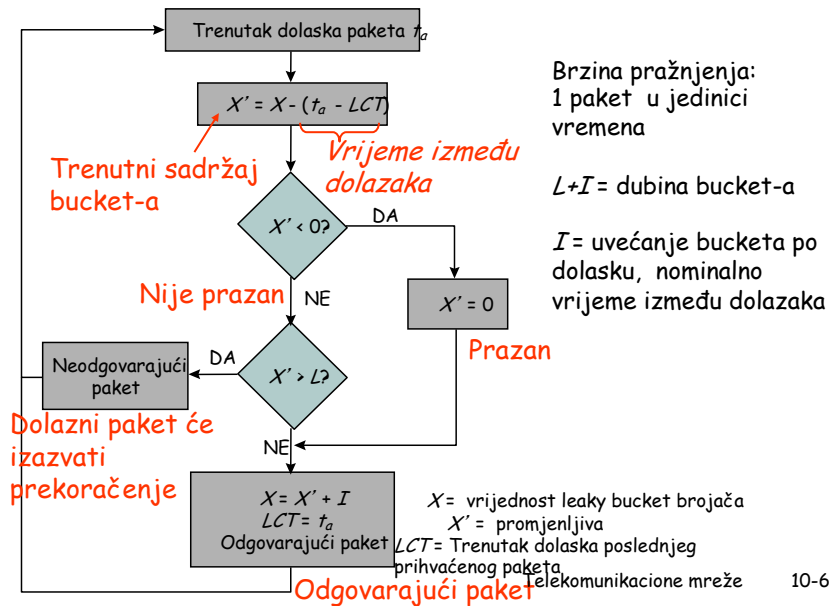
1 paket u jedinici vremena  
Neka su paketi konstantne veličine kao kod ATM

Neka je  $X$  = sadržaj bucket-a u momentu kada je stigao posljednji prihvatljivi paket  
Neka je  $t_a$  - trenutak dolaska paketa

Telekomunikacione mreže 10-5

5

## Leaky Bucket Algoritam

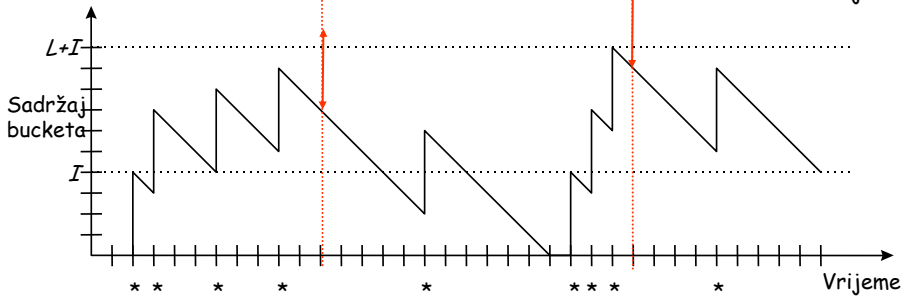
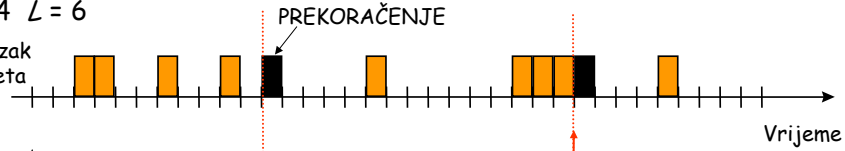


6

## Leaky Bucket primjer

$I = 4$   $L = 6$

Dolazak paketa



Paketi koji izazovu prekoračenje ne ulaze u bucket jer nisu uključeni u inicijalne kalkulacije

Telekomunikacione mreže 10-7

7

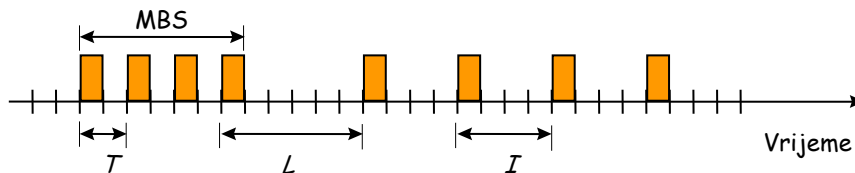
## Parametri algoritma

$T = 1 /$  maksimalna brzina

MBS = maksimalna veličina burst-a

$I =$  nominalno vrijeme između odlazaka =  $1 /$  prihvatljiva brzina

$$MBS = 1 + \left\lceil \frac{L}{I - T} \right\rceil$$

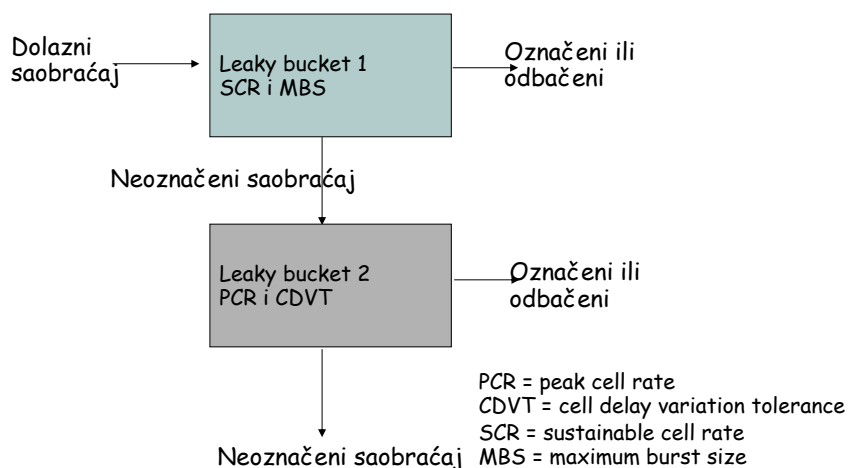


Telekomunikacione mreže 10-8

8

## Dvojni Leaky Bucket

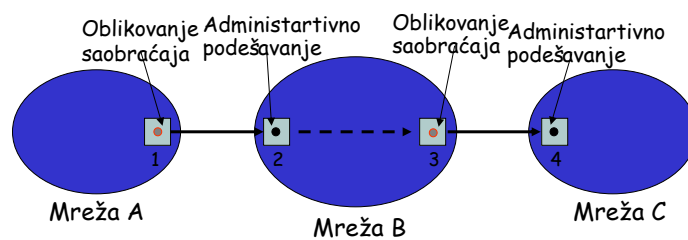
Dvojni leaky bucket za podešavanje PCR, SCR i MBS:



Telekomunikacione mreže 10-9

9

## Oblikovanje saobraćaja

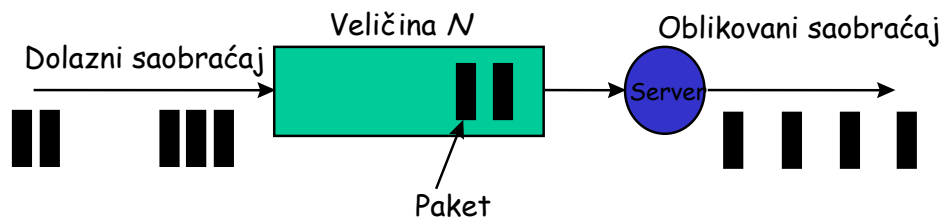


- ❑ Mreže administrativno podešavaju dolazne saobraćajne tokove
- ❑ *Oblikovanje saobraćaja* se koristi kada se želi postići da određeni saobraćajni tok odgovara određenim parametrima
- ❑ Mreže mogu oblikovati svoj saobraćaj prije predaje narednoj mreži

Telekomunikacione mreže 10-10

10

## Leaky Bucket oblikovanje saobraćaja



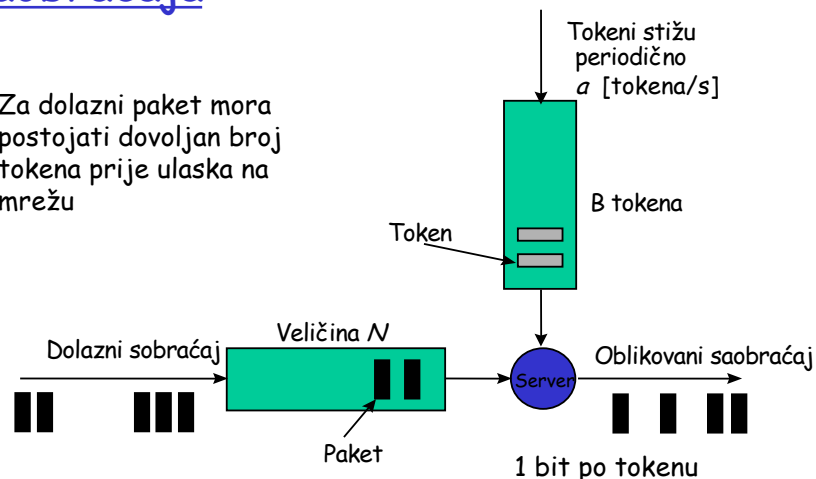
- ❑ Baferovanje dolaznih paketa
- ❑ Periodično se poslužuju saglasno odgovarajućim parametrima
- ❑ Skokovi saobraćaja se baferuju i poravnavaju
- ❑ Moguć je gubitak paketa zbog prekoračenja bafera
- ❑ Previše restriktivno podešavanja, pošto željeni saobraćaj ne treba da bude u potpunosti ravnomjeran

Telekomunikacione mreže 10-11

11

## Token Bucket oblikovanje saobraćaja

Za dolazni paket mora postojati dovoljan broj tokena prije ulaska na mrežu



- ❑ Brzina tokena reguliše prenos paketa
- ❑ Ako je dovoljan broj tokena na raspolaganju, paketi ulaze na mrežu bez kašnjenja
- ❑ K određuje kolika je grupisanost saobraćaja dozvoljena na mreži

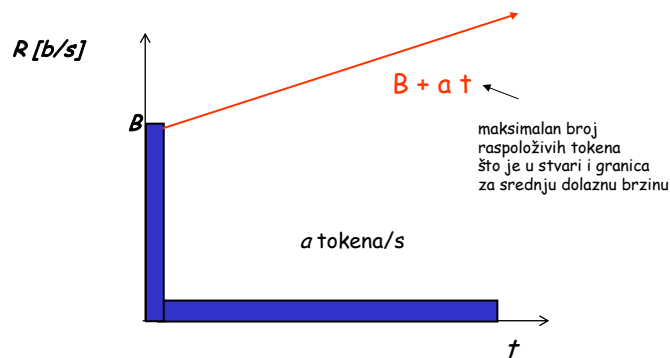
Telekomunikacione mreže 10-12

12

## Token Bucket efekat oblikovanja

Token bucket ograničava saobraćaj koji dolazi od izvora tako da je

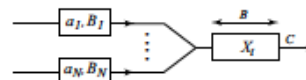
- srednja brzina  $R$  je manja od  $B + a \cdot t$  u intervalu trajanja  $t$
- Maksimalna veličina bursta je manja od  $B$



Telekomunikacione mreže 10-13

13

## QoS



### Granice kašnjenja

- Kao što se može ograničiti brzina toka i veličina burst-a može se ograničiti i kašnjenje u baferu
- Neka  $N$  saobraćajnih tokova oblikovanih token bucket sistemima parametara  $a_i$  i  $B_i$  dolazi na ulaz bafera koji se prazni brzinom  $C$
- $X_t$  - broj paketa u baferu u trenutku  $t$
- $B$  - veličina bafera izražena u bajtima
- Veličina paketa u toku  $i$  je ograničena sa  $B_i$
- Paket može imati maksimalnu veličinu  $P$ , dolazi na ulaz bafera bez prekida (ne miješaju se bajti različitih paketa)
- Neka je  $a_1 + a_2 + \dots + a_N \leq C$
- Može se pokazati da važi:

$$X_t \leq B_1 + B_2 + \dots + B_N + NP$$

- Odnosno da je kašnjenje bilo kojeg paketa ograničeno sa
- $$\frac{B_1 + B_2 + \dots + B_N + NP + P}{C}$$

Telekomunikacione mreže 10-14

14

## QoS

### Scheduling

- Scheduling: izbor narednog paketa za slanje
  - FIFO/Prioritet
  - Round Robin/ DRR
  - Weighted Fair Queuing
- Odbacivanje paketa:
  - Kada je bafer pun
  - Ne samo kada je bafer pun
    - Aktivni menadžment bafera
- Signalizacija zagušenja
  - Explicit Congestion Notification (ECN)

Telekomunikacione mreže 10-15

15

## QoS

### Scheduling

- Zašto ne koristiti beskonačno velike bafere?
  - Nema odbacivanja paketa!
- Mali baferi:
  - Često odbacuju pakete zbog pojave burstova
  - Unose malo kašnjenje
- Veliki baferi:
  - Smanjuju odbacivanje paketa
  - Povećavaju kašnjenje
- Kompromis je RED (Random Early Detection)

Telekomunikacione mreže 10-16

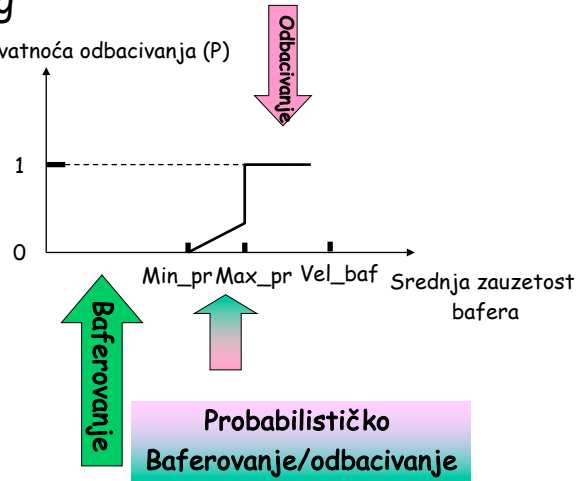
16



# QoS

## Scheduling

□ RED Vjerovatnoća odbacivanja (P)



17

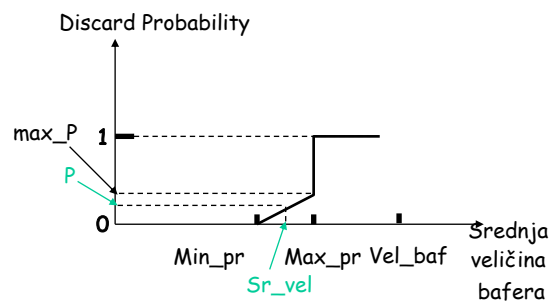
# QoS

## Scheduling

□ RED

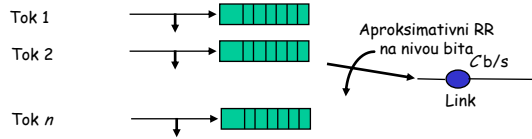
Podešavanje vjerovatnoće P:

$$P = \max_P \frac{Sr\_vel - Min\_pr}{Max\_pr - Min\_pr}$$



18

# QoS

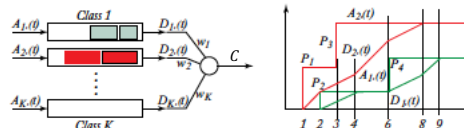


## Scheduling

### Fair Queueing/Generalized Processor Sharing (GPS)

- Svaki tok ima svoj logički bafer:
- $C$  b/s se alokira nepraznim baferima
  - Brzina prenosa =  $C / n(t)$ , gdje je  $n(t)$  broj nepraznih bafera
- Idealizovani sistem podrazumijeva fluidni tok u baferima
- Implementacija zahtijeva aproksimaciju:
  - simuliranje fluidnog sistema;
  - sortiranje paketa prema vremenu završetka idealnog sistema

# QoS

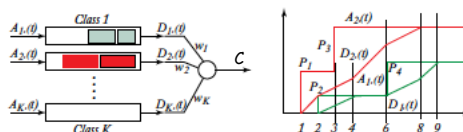


## Scheduling GPS

- Paketi se klasifikuju u K klasa saobraćaja i čekaju u odgovarajućem baferu dok ih scheduler ne odabere
- Svakoj klasi k odgovara težinski factor  $w_k$ .
- $C$  je brzina linka
- $W$  je zbir težinskih faktora bafera koji u trenutku  $t$  nijesu prazni
- Scheduler bira pakete iz bafera sa vjerovatnoćom proporcionalnom težinskom faktoru njihove klase.
- Trenutna brzina posluživanja klase k je  $\frac{w_k C}{W}$
- Ovaj model je samo matematički i nije primjenjiv jer da scheduler "miješa" bite iz različitih paketa i ne poštuje granice paketa.
- Može se pokazati da se toku k može obezbijediti minimalna brzina koja je jednaka  $\frac{w_k C}{\sum_{i=1}^K w_i}$
- Ovu brzinu tok dobija kada nijedan bafer nije prazan
- Ako na ulaz GPS schedulera dolaze tokovi saobraćaja oblikovani token bucket sistemima parametara  $(a_i, B_i)$ , kašnjenje paketa klase k je ograničeno vrijednošću  $\frac{B_k \sum_{i=1}^K w_i}{w_k C}$
- pri čemu je  $a_1 + a_2 + \dots + a_K \leq C$

## QoS

### Scheduling WFQ



- Predstavlja aprokcimaciju GPS koja se može realizovati
- Paketi su klasifikovani i stavljeni u redove kao kod GPS.
- Scheduler prenosi istovremeno samo jedan paket brzinom linka  $C$ .
- Kad završi prenos paketa scheduler započinje sa slanjem paketa kojeg bi GPS, ako bi bio primijenjen, prvo poslužio.
- Ako su  $F_n$  i  $G_n$  vremena odlaska paketa  $n$  ako se koristi WFQ i GPS, a  $T$  vrijeme prenosa najvećeg paketa onda važi

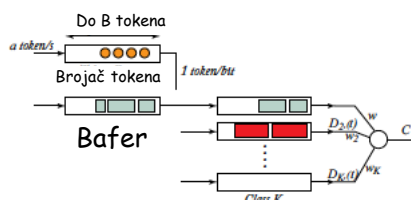
$$F_n \leq G_n + T$$

Telekomunikacione mreže 10-21

21

## QoS

### Scheduling WFQ



- Može se pokazati da ako je
- $$a < \frac{wC}{\sum_{i=1}^K w_i}$$
- Maksimalno kašnjenje paketa je
- $$\frac{B \sum_{i=1}^K w_i}{wC} + \frac{L}{C}$$
- Gdje je  $L$  veličina paketa

Telekomunikacione mreže 10-22

22

## Značaj WFQ

- ❑ WFQ može obezbijediti granice kašnjenja od kraja do kraja mreže
- ❑ WFQ nudi i ferrost i garancije performansi
- ❑ Granice veće bez obzira na ponašanje ostalog saobraćaja
- ❑ Može biti generalizovan za mreže gdje su scheduleri varijante WFQ i gdje brzine prenosa variraju u vremenu
- ❑ Ne ograničava jitter