

Univerza v Ljubljani  
Fakulteta za elektrotehniko

Anton Kos

Zagotavljanje različnih stopenj kakovosti  
storitev v omrežjih s paketnim prenosom  
podatkov

DOKTORSKA DISERTACIJA

Mentor: prof. dr. Sašo Tomažič

Ljubljana, 2006



## Zahvala

Mentorju prof. dr. Sašu Tomažiču se zahvaljujem za strokovno vodstvo, koristne nasvete, konstruktivne predloge in komentarje. Sodelavcem v Laboratoriju za komunikacijske naprave pa se zahvaljujem za prijetno in ustvarjalno delovno vzdušje ter vso pomoč, ki so mi jo nudili.

Za koristne strokovne nasvete se zahvaljujem bratu Andreju. Za pomoč pri formalizaciji matematičnih zapisov se zahvaljujem Jeleni, za natančen pregled izpeljav in pravilnosti rezultatov pa tudi Jasni in Vesni.

Posebna zahvala velja mojima staršema, ki sta mi omogočila brezskrben študij in me pri tem ves čas podpirala in vzpodbujala.

Največja zahvala pa gre moji Mojci. Z veliko potrpežljivostjo, zaupanjem in pravšnjo mero vzpodbudnih besed mi je vedno znova uspela dati nov zagon.



## Povzetek

*Kakovost storitve* (QoS - Quality of Service) je vse pomembnejši dejavnik v sodobnih telekomunikacijskih omrežjih. Pomembna je tako za uporabnike kot za ponudnike storitev. Uporabnik s tem dobi storitev po svojih željah, ponudniku pa predstavlja dodano vrednost omrežja. To delo se ukvarja z možnostmi zagotavljanja različnih stopenj kakovosti storitve v omrežjih s paketnim prenosom podatkov.

Obstoječe definicije kakovosti storitve večinoma služijo ožjim področjem znotraj telekomunikacij. Iz podanih definicij izluščimo njihova glavna vodila in na njihovi osnovi predlagamo novo definicijo kakovosti storitve v telekomunikacijskih sistemih. Nas zanimajo predvsem telekomunikacijska omrežja in znotraj njih omrežja s paketnim prenosom podatkov. Dojemanje kakovosti storitve s strani uporabnika in ponudnika se zelo razlikuje. Uporabnik upošteva subjektivne kriterije, ki jih imenujemo tudi *kakovost izkušnje* (QoE - Quality of Experience), ponudnik pa objektivne, ki jih lahko meri in nadzoruje. Objektivni kriteriji temeljijo na parametrih kakovosti storitve, ki jih določajo lastnosti aplikacij, storitev in predvsem prenosnega omrežja. Ni nepomembno tudi dejstvo, da so ti parametri med seboj odvisni. Storitve s standardiziranimi parametri kakovosti storitve pa so zaenkrat še redke.

Nas še posebej zanima kakovost storitve v omrežjih z internetnim protokolom (IP - Internet Protocol), ki so trenutno prevladujoča paketna omrežja. Osnovne lastnosti omrežij IP kot so statistični multipleks, nezanesljivost, nepovezavnost in prenos po najboljših zmožnostih, ne govorijo v prid možnosti zagotavljanja različnih stopenj kakovosti storitve. Tudi široka paleta raznolikih prenosnih tehnologij, ki jih uporabljajo omrežja IP, ni temu v prid. Zato je pomembna analiza zmožnosti prenosnih tehnologij v smislu zagotavljanja različnih stopenj kakovosti storitve, ki jo podajamo v ta namen. Čeprav so omrežja IP že ob definiciji protokola IP vsebovala omejeno možnost zagotavljanja kakovosti storitve preko polja ToS (Type of Service - tip storitve), je sedaj to polje na novo definirano in se imenuje polje DS (Differentiated Services - diferencirane storitve). Preko slednjega lahko s pomočjo Diferenciranih storitev zagotavljamo relativne stopnje kakovosti storitve. Boljše, vendar zahtevnejše so Integrirane storitve, ki zagotovijo absolutne stopnje kakovosti storitve.

Predstavimo mehanizme za zagotavljanje kakovosti storitev v paketnih omrežjih in model omrežne naprave. Še posebej se posvetimo krmiljenju in konfiguraciji čakalnih vrst ter odpravnikom z izhodnim čakanjem. Obravnavamo glavne mere za oceno delovanja odpravnikov (pravičnost, algoritemska zakasnitev in zahtevnost) ter opišemo odpravnike, ki jih potrebujemo v nadaljevanju.

Jedro disertacije je predlog združitve Diferenciranih in Integriranih storitev v omrežju DIS (Differentiated and Integrated Services - diferencirane in integrirane storitve). Za združevanje smo se odločili, ker ne prva ne druga rešitev sama zase nista primerni za

vse vrste prometa in njihove zahteve po kakovosti storitve. V ta namen podajamo referenčno topologijo omrežja ter primerjavo funkcionalnosti in lastnosti Diferenciranih in Integriranih storitev. Izluščimo dobre in slabe lastnosti obeh. V podanem predlogu združevanja ohranimo dobre lastnosti obeh in izničimo večino slabih. Predlagano omrežje DIS deluje dvostopenjsko: na prvi stopnji deluje podobno kot omrežje z Diferenciranimi storitvami, na drugi stopnji pa kot omrežje z Integriranimi storitvami. Za omrežje DIS podamo predlog konfiguracije čakalnih vrst ter ga analiziramo teoretično in preko simulacij.

Za potrebe simulacij podamo model konfiguracije dvostopenjske čakalne vrste. Na osnovi tega modela in že prej podanega modela omrežne naprave razvijemo lastni simulator omrežne naprave v omrežju DIS. V preizkus pravilnega delovanja simulatorja najprej simuliramo vsako stopnjo posebej. Simulacijske rezultate primerjamo z analitičnimi, od katerih se praktično ne razlikujejo. Ti rezultati nam služijo tudi kot osnova za primerjavo s simulacijami dvostopenjskega delovanja omrežja DIS. Teoretična analiza in simulacije prve stopnje pokažejo, da z izbranim strogo prednostnim odpravljanjem lahko zagotovimo različne stopnje kakovosti storitve za pakete posameznih storitvenih razredov. Za glavni parameter pri tem izberemo zakasnitev. Na prvi stopnji omrežja DIS simulacije razširimo še na višje protokolne plasti, lastnosti posameznih aplikacij in lastnosti njihovih podatkovnih pretokov. Na drugi stopnji izberemo odpravnik DRR (Deficit Round Robin - deficitarno krožno odpravljanje), ki poskrbi za pravičnost med pretoki znotraj posameznega storitvenega razreda. To pomeni, da vsak pretok, glede na svojo rezervacijo, dobi pravični delež omrežnih virov. Simulacija dvostopenjskega delovanja pokaže, da dosežemo želene in pričakovane rezultate. Zato trdimo, da je omrežje DIS s predlagano konfiguracijo čakalnih vrst sposobno in primerno za zagotavljanje različnih stopenj kakovosti storitve.

Rezultati zgornjih simulacij nakazujejo, da lahko povprečno zakasnitev paketov še zmanjšamo, če znižamo vrednost kvanta pri odpravniku DRR pod njegovo kritično vrednost pri  $Q = L_{max}$ . Zato predlagamo nov odpravnik z oznako SCDRR (Sub-Critical Deficit Round Robin - podkritično deficitarno krožno odpravljanje), ki temelji na odpravniku DRR. Za odpravnik SCDRR podamo parametre in robne pogoje delovanja ter izpeljemo njegovo pravičnost, algoritemsko zakasnitev in računsko zahtevnost. Ugotovimo, da sta prvi dve meri ugodnejši kot pri odpravniku DRR, zadnja pa manj ugodna. Izvedemo tudi simulacije odpravnika SCDRR za primer samostojnega delovanja in delovanja na drugi stopnji omrežja DIS. Primerjava rezultatov samostojnega (enostopenjskega) delovanja z analitičnimi pokaže, da se ti med seboj praktično ne razlikujejo. Primerjava rezultatov dvostopenjskega delovanja odpravnikov SCDRR in DRR prav tako potrди pričakovanja izhajajoča iz analitičnih rezultatov. Pri uporabi odpravnika SCDRR se zmanjšajo povprečne zakasnitve in poveča pravičnost. Poveča se tudi računsko zahtevnost. Toda glede na faktorje izboljšanja in poslabšanja omenjenih mer ter njihove pomembnosti v smislu zagotavljanja različnih stopenj kakovosti storitev zaključimo, da je uvedba odpravnika SCDRR smiselna in koristna.

## Abstract

*Quality of service* (QoS) is ever more important factor in telecommunication networks. Quality of service is important for both, users and service providers. Through quality of service users get service to their requests, and providers added value of their networks. This work studies the possibilities of assuring different levels of quality of service in packet networks.

Existing definitions of quality of service are mainly used only within certain areas of telecommunications. The given definitions have presented us with the main guidance. Based on them we propose a new definition of quality of service for telecommunication systems. Our main interest is in telecommunication networks with emphasis on packet networks. The comprehension of quality of service from the side of users and from the side of service providers differs greatly. Users consider subjective criteria called *quality of experience* (QoE), but service providers consider objective criteria, which can be measured and controlled. Objective criteria are referred to as quality of service parameters. They are defined by the properties of applications, services, and most of all by the properties of transmission network. These parameters are interdependent. Services with standardised quality of service parameters are for now quite rare.

We take special interest in quality of service in IP (Internet Protocol) networks, which are currently predominant packet networks. The basic properties of IP networks such as statistical multiplex, unreliability, connectionless and best-effort data transfer, are not to the benefit of assuring different levels of quality of service. The wide choice of heterogeneous transmission networks that can be used by IP networks is also not an advantage in this case. Therefore we present the analysis of capabilities of transmission technologies in the sense of assuring different levels of quality of service. Although IP protocol has been offering a limited quality of service support through ToS (Type of Service) field from its definition on, this field is now redefined into a DS (Differentiated Services) field. Using DS and employing Differentiated Services relative levels of quality of service can be offered. More advanced but more complex are Integrated Services, which can assure absolute levels of quality of service.

Mechanisms for quality of service assurance in packet networks and a model of a network device are presented. We concentrate on managing and configuration of queues for schedulers with output queuing. We discuss the main measures for estimation of scheduler operation (fairness, latency and complexity) and describe the properties and operation of schedulers needed in further chapters.

The main part of this dissertation is the proposal of merging of Differentiated Services and Integrated Services into a DIS (Differentiated and Integrated Services) network. We decided for their merger because Integrated or Differentiated Services alone are not suitable for all kinds of traffic, applications and their quality of service

demands. For this purpose we give a reference network topology and a comparison of functionality and properties of Integrated and Differentiated Services. We list the most important good and bad properties of both. In the given proposal for a DIS network we preserve their good properties and eliminate most of their bad properties. The proposed DIS network operates in two stages. The first stage operates similar to a Differentiated Services network, while the second stage operates similar to an Integrated Services network. The configuration of queues is proposed and analysed theoretically and through simulations.

For the simulation purposes we give a model of a two-stage queue configuration. Based on this model and the previously given model of a network device we develop our own simulator for a DIS network device. To test the simulator, we first simulate each stage separately. Simulation results are compared with analytical results. The difference between them is practically negligible. These results also serve as the basis for comparison between one-stage and two-stage operation of DIS network. Theoretical analysis and simulations of the first stage show, that the chosen strict priority scheduling provides us with different levels of quality of service for packets belonging to different service classes. Delay is chosen as the main parameter. Simulations of the first stage of DIS network are extended to higher protocol layers, application properties and their traffic patterns. For the second stage DRR (Deficit Round Robin) scheduler is chosen. It takes care of the fairness among flows inside each service class. Meaning, that each flow gets its fair share of network resources according to its reservation. The simulation of a two-stage operation shows that we achieve desired and expected results. Therefore we claim that DIS network, with the proposed queue configuration, is suitable and capable of assuring different levels of quality of service.

Simulation results indicate that the average packet delay can be further reduced, if we decrease the quantum value of DRR scheduler under its critical value of  $Q = L_{max}$ . Therefore we propose a new SCDRR (Sub-Critical Deficit Round Robin) scheduler, which is based on the DRR scheduler. For SCDRR we give its parameters and boundary conditions. We also derive its fairness, latency, and computational complexity. We establish, that the derived measures for fairness and latency are better than the same measures for DRR, but computational complexity is worse. We carry out simulations of stand-alone SCDRR operation and its operation on the second stage of the DIS network. Comparison of stand-alone simulations with analytical results shows that they are practically identical. The comparison of simulation results of two-stage operation between SCDRR and DRR schedulers also confirm our expectations based on analytical results. Using SCDRR the average packet delay decreases and fairness increases. Unfortunately computational complexity also increases. Taking into consideration the factors of improvement and deterioration of the mentioned scheduler measures and their importance in the sense of quality of service assurance, we conclude, that the introduction of SCDRR scheduler is reasonable and useful.



## Prispevki k znanosti

- Nova definicija kakovosti storitve,
- predlog združevanja diferenciranih in integriranih storitev (DIS) v dvostopenjskem omrežju IP, ki vsebuje način združevanja obeh vrst storitev, izdelavo simulatorja omrežne naprave omrežja DIS in simulacijo omrežja ter ovrednotenje sposobnosti omrežja DIS za zagotavljanje različnih stopenj kakovosti storitve,
- predlog izboljšane DRR odpravnik s podkritičnim delovanjem (SCDRR), njegova analiza in simulacija,
- odkritje in odprava napak pri izpeljavi analitične mere pravičnosti in analitične mere algoritemske zakasnitve v referenčnih člankih,
- izpeljava analitične mere zahtevnosti za SCDRR odpravnik.



# Kazalo

<b>1</b>	<b>Uvod</b>	<b>5</b>
<b>2</b>	<b>Kakovost storitve</b>	<b>9</b>
2.1	Definicija kakovosti storitve . . . . .	9
2.2	Kakovost storitve v telekomunikacijskih omrežjih . . . . .	10
2.3	Kakovost storitve z vidika ponudnika in uporabnika . . . . .	10
2.4	Parametri kakovosti storitve . . . . .	11
2.4.1	Medsebojna odvisnost parametrov kakovosti storitve . . . . .	12
2.5	Standardizirane vrednosti parametrov kakovosti storitve . . . . .	12
2.6	Parametri kakovosti storitve za zahtevne aplikacije . . . . .	13
2.7	Psihološki in sociološki vidiki uvajanja različnih stopenj kakovosti storitev . . . . .	14
<b>3</b>	<b>Kakovost storitve v omrežjih IP</b>	<b>15</b>
3.1	Osnovne lastnosti paketnih omrežij . . . . .	15
3.1.1	Statistični multipleks . . . . .	15
3.1.2	Razpoložljiva kapaciteta prenosne poti, zakasnitev in njena spre- menljivost . . . . .	16
3.1.3	Kakovost storitve v paketnih omrežjih . . . . .	17
3.2	Omrežja IP . . . . .	18
3.2.1	Omrežja IP in prenosne tehnologije . . . . .	18
3.2.2	Nadgradnja omrežij IP za zagotavljanje kakovosti storitev . . . . .	21
3.2.3	Diferencirane in integrirane storitve . . . . .	24
3.2.4	Osnovne lastnosti in delovanje omrežij z IS . . . . .	24
3.2.5	Osnovne lastnosti in delovanje omrežij z DS . . . . .	25
3.3	Kakovost storitve med končnima točkama komunikacije . . . . .	25
<b>4</b>	<b>Mehanizmi zagotavljanja kakovosti storitve v paketnih omrežjih</b>	<b>27</b>
4.1	Model omrežne naprave . . . . .	28
4.2	Konfiguracija in krmiljenje čakalnih vrst . . . . .	30
4.2.1	Konfiguracija čakalnih vrst . . . . .	30
4.2.2	Krmiljenje čakalnih vrst . . . . .	31
4.3	Odpravljanje v paketnih omrežjih . . . . .	33
4.3.1	Zelene lastnosti odpravnikov . . . . .	33
4.3.2	Osnovne možnosti izbire glede lastnosti odpravnikov . . . . .	34
4.4	Glavne mere za oceno delovanja odpravnikov . . . . .	35
4.4.1	Pravičnost . . . . .	36
4.4.2	Zakasnitev algoritma . . . . .	37

4.4.3	Zahtevnost . . . . .	38
4.5	Odpravniki z izhodnim čakanjem . . . . .	39
4.5.1	FCFS . . . . .	40
4.5.2	Strogo prednostno odpravljanje . . . . .	40
4.5.3	Pravično odpravljanje . . . . .	42
4.5.4	Souporaba procesnih zmogljivosti . . . . .	43
4.5.5	Uteženo pravično odpravljanje . . . . .	43
4.5.6	Deficitarno krožno dodeljevanje . . . . .	44
<b>5</b>	<b>Združitev Diferenciranih in Integriranih storitev</b>	<b>47</b>
5.1	Referenčna topologija prenosnega omrežja . . . . .	47
5.2	Primerjava Diferenciranih in Integriranih storitev . . . . .	49
5.3	Zakaj in na kakšen način združiti omrežja z IS in DS . . . . .	50
5.4	Predlog združitve Diferenciranih in Integriranih storitev . . . . .	51
5.4.1	Možne težave in rešitve pri združevanju IS in DS . . . . .	52
5.4.2	Drugi predlogi združevanja IS in DS . . . . .	53
5.4.3	Predlog konfiguracije in krmiljenja omrežnih naprav . . . . .	54
5.5	Analiza delovanja prve stopnje omrežja DIS . . . . .	56
5.5.1	Razlikovanje prometa glede na zakasnitve . . . . .	57
5.5.2	Teoretična analiza diferenciacije prometa . . . . .	58
5.5.2.1	Čakalna vrsta M/G/1 s prednostjo . . . . .	58
5.5.2.2	Ohranitveni zakon . . . . .	62
5.5.2.3	Omrežje čakalnih vrst . . . . .	63
5.5.3	Kratek opis simulatorja omrežne naprave v omrežju DIS . . . . .	63
5.5.4	Simulacija diferenciacije prometa ter primerjava rezultatov z analitičnimi . . . . .	63
5.5.5	Simulacija prve stopnje omrežja DIS . . . . .	66
5.5.5.1	Simulacijska shema in podatkovni pretoki . . . . .	66
5.5.5.2	Simulacijski scenariji in parametri . . . . .	67
5.5.5.3	Simulacija osnovnega stanja omrežij IP . . . . .	68
5.5.5.4	Uvedba različnih stopenj kakovosti storitve . . . . .	70
5.5.6	Komentar rezultatov . . . . .	77
5.6	Analiza delovanja druge stopnje omrežja DIS . . . . .	78
5.6.1	Simulacija samostojnega delovanja odpravnika DRR . . . . .	78
5.6.2	Simulacija dvostopenjskega delovanja odpravnika v omrežju DIS . . . . .	84
5.6.2.1	Rezultati simulacij dvostopenjskega odpravljanja in primerjava z enostopenjskim . . . . .	84
5.6.3	Komentar simulacijskih rezultatov . . . . .	94
<b>6</b>	<b>Odpravnik SCDRR</b>	<b>95</b>
6.1	Analiza lastnosti in delovanja odpravnika SCDRR . . . . .	96
6.2	Parametri in robni pogoji delovanja SCDRR . . . . .	96
6.3	Pravičnost algoritma SCDRR . . . . .	98
6.3.1	Relativna mera pravičnosti . . . . .	98
6.3.2	Absolutna mera pravičnosti . . . . .	100
6.3.3	Primerjava pravičnosti nekaterih pomembnejših odpravnikov . . . . .	101

6.4	Zakasnitev algoritma SCDRR . . . . .	103
6.5	Analiza računske zahtevnosti SCDRR . . . . .	105
6.5.1	Definicije in seznam spremenljivk . . . . .	105
6.5.2	Robni pogoji . . . . .	107
6.5.3	Izračun števila potrebnih operacij . . . . .	107
6.5.4	Zahtevnost pri eksponentni porazdelitvi dolžin paketov . . . . .	113
6.5.5	Zahtevnost pri enakomerni porazdelitvi dolžin paketov . . . . .	114
6.6	Simulacija samostojnega delovanja SCDRR . . . . .	117
6.7	Simulacija SCDRR na drugi stopnji omrežja DIS . . . . .	118
6.8	Komentar simulacijskih rezultatov za SCDRR . . . . .	122
<b>7</b>	<b>Zaključek</b>	<b>123</b>
	<b>Dodatek</b>	<b>125</b>
<b>A</b>	<b>Vedro z žetoni</b>	<b>127</b>
<b>B</b>	<b>Osnove teorije čakalnih vrst</b>	<b>129</b>
B.1	Poissonov proces . . . . .	130
B.2	Markovove verige ter proces rojstev in smrti . . . . .	131
B.3	Najpomembnejše lastnosti in parametri čakalnih vrst . . . . .	134
B.3.1	Čakalna vrsta M/M/1 . . . . .	135
B.3.2	Čakalna vrsta M/M/1/K . . . . .	138
B.3.3	Čakalna vrsta M/G/1 . . . . .	138
<b>C</b>	<b>Analiza lastnosti odpravnika SCDRR</b>	<b>141</b>
C.1	Parametri SCDRR . . . . .	141
C.2	Pravičnost algoritma SCDRR . . . . .	141
C.2.1	Relativna mera pravičnosti . . . . .	142
C.3	Zakasnitev algoritma SCDRR . . . . .	147
C.4	Zahtevnost SCDRR pri eksponentni porazdelitvi dolžin paketov . . . . .	159
C.4.1	Definicije . . . . .	159
C.4.2	Izračun števila potrebnih operacij . . . . .	160
C.4.3	Zbir rezultatov za eksponentno porazdelitev dolžin paketov . . . . .	165
C.4.4	Primer . . . . .	165
C.4.5	Poenostavljen izračun $DC_m$ . . . . .	168
C.5	Zahtevnost SCDRR pri enakomerni porazdelitvi dolžin paketov . . . . .	171
C.5.1	Definicije . . . . .	171
C.5.2	Izračun števila potrebnih operacij . . . . .	171
C.5.2.1	Izračun pri izpolnjenem pogoju $DC_{m-1} < L_{min}$ . . . . .	174
C.5.2.2	Izračun pri izpolnjenem pogoju $DC_{m-1} \geq L_{min}$ . . . . .	176
C.5.3	Zbir rezultatov za enakomerno porazdelitev dolžin paketov . . . . .	179
C.5.4	Primer 1 . . . . .	180
C.5.5	Primer 2 . . . . .	181
C.5.6	Poenostavljen izračun $DC_m$ . . . . .	184
C.5.7	Poenostavljen izračun $O_m$ . . . . .	187

<b>Literatura</b>	<b>189</b>
<b>Seznam Kratic</b>	<b>195</b>

# 1 Uvod

V zadnjem desetletju smo priča izredno hitremu razvoju in širjenju telekomunikacijskih omrežij in v zvezi z njimi je vse pogostejše omenjan tudi pojem *kakovost storitve*. Dostop do enega izmed mnogih telekomunikacijskih omrežij danes ni več problem. Uporabnik navadno lahko izbira med več različnimi omrežji in/ali ponudniki. Izbira je odvisna od želja in potreb uporabnika predvsem pa od nabora storitev, ki jih omogočajo ponudniki. Uvajanje različnih stopenj kakovosti storitve v omrežje je v interesu ponudnikov in uporabnikov. Prvi na ta način lahko nudijo širši nabor storitev in optimalneje izkoriščajo omrežje, drugim pa je na voljo večja in pestrejša izbira, ki se lahko prilagodi njihovim potrebam.

V okviru telekomunikacijskih omrežij nas najbolj zanimajo podatkovna omrežja. Zmožnosti zagotavljanja različnih stopenj kakovosti storitve se med njimi precej razlikujejo. Nekatera izmed njih ne nudijo prav nobenih možnosti za kakovost storitve, druga pa ponujajo omejene možnosti. Še najboljše se odreže omrežje ATM (Asynchronous Transfer Mode - asinhroni način prenosa), ki je sposobno nuditi različne stopnje kakovosti storitve ob zagotavljanju dogovorjenih prenosnih parametrov.

Od vseh podatkovnih omrežij so najhitrejši in najburnejši razvoj dosegla omrežja IP (Internet Protocol - internetni protokol). Lahko bi rekli, da so se omrežja IP (Internet) pojavila ob pravem času s pravo aplikacijo. V času, ko smo čakali na širokopasovna podatkovna omrežja, temelječa na tehnologiji ATM, so se omrežja IP prebila praktično do vsakega končnega uporabnika; tako doma kot v podjetjih in raziskovalnih ustanovah. Uporabnike je najprej navdušila elektronska pošta, piko na i pa so pritisnili brskalniki in spletne strani. Te so omrežje naredile prijazno tudi manj večšim uporabnikom in jim hkrati ponudile ogromen, skoraj že preobsežen vir informacij.

Po bliskovitem prodoru Interneta so omrežja IP postala najpomembnejša, najbolj razširjena in najlažje dostopna podatkovna omrežja, ki pa v osnovi žal ponujajo le storitev prenosa po najboljših zmožnostih (best-effort). Ta je sicer povsem primerna za aplikacije, ki so nastale za omrežja IP, veliko manj pa za aplikacije, ki prihajajo iz drugih okolij in omrežij. Kot primer naj omenimo le sproten prenos govora in videa, katerih prenosni parametri se precej razlikujejo od prenosnih parametrov izvirnih IP aplikacij. Omrežja IP za take storitve niso bila načrtovana in zanje v osnovi niso primerna, saj ne podpirajo različnih stopenj kakovosti storitve (QoS - Quality of Service) in ne vsebujejo mehanizmov, potrebnih za uspešen sproten prenos podatkov ali podatkov, ki zahtevajo storitev boljšo od storitve po najboljših zmožnostih.

Ideje o nadgradnji omrežij IP v smislu zagotavljanja različnih stopenj kakovosti storitve so se realizirale v obliki specifikacij Integriranih storitev (IS - Integrated Services) in Diferenciranih storitev (DS - Differentiated Services). Oba pristopa se že v osnovi precej razlikujeta.

Integrirane storitve preko signalizacijskega in rezervacijskega protokola vpeljejo povezavno načelo v sicer nepovezavno orientirano omrežje. S tem uvedejo nadzor nad omrežnimi viri in vzpostavijo pogoje, potrebne za prenos podatkov s točno določenimi zahtevami po kakovosti storitve. Ta pristop zahteva nadgradnjo omrežnih naprav in aplikacij, ki uporabljajo Integrirane storitve. Te ponujajo tri razrede storitev: storitev po najboljših zmožnostih (best-effort), ki predstavlja današnje stanje v omrežjih IP, storitev nadzorovanega bremena (controlled load), ki posnema stanje nizko obremenjenega omrežja in zagotovljene storitve (guaranteed service), pri katerih omrežje uporabniku zagotovi prenos podatkov z dogovorjenimi parametri. Integrirane storitve so nekateri označili za preveč zapletene, razsipne z omrežnimi viri in slabo razširljive.

Zaradi pomislekov glede Integriranih storitev so bile kasneje definirane še Diferencirane storitve (DS), ki uvajajo različne stopnje kakovosti storitve na osnovi omejenega števila prednostnih razredov. Diferencirane storitve ohranjajo nepovezavno orientiranost omrežja. Podatke o pripadnosti storitvenemu razredu in s tem željeni stopnji kakovosti storitve nosi vsak paket sam. Procesiranje v omrežnih napravah je tako mnogo preprostejše in razširljivost na višje prenosne hitrosti, za razliko od IS, ni več tako pereč problem. Največji kompromis pa naredimo pri stopnjah kakovosti storitve, ki niso več absolutne (kot pri IS) ampak relativne. To pomeni, da omrežje uporabniku ne zagotavlja prenosnih parametrov v absolutno določenih okvirih ampak le glede na parametre prometa z višjo ali nižjo stopnjo kakovosti storitve.

Jedro disertacije je nov predlog združevanja Diferenciranih in Integriranih storitev, ki ohrani njune dobre lastnosti in izniči slabe. Po našem predlogu omrežje deluje dvostopenjsko. Na prvi stopnji razločuje le posamezne storitvene razrede, na drugi pa tudi posamezne pretoke. Dvostopenjsko delovanje se uporabi le za določene storitvene razrede; tiste v katerih posamezni pretoki zahtevajo zagotovljeno kakovost storitve. Tako naredimo dober kompromis med zahtevnostjo in možnostmi nujenja različnih stopenj kakovosti storitve.

Delo je razdeljeno na 8 poglavij. V 2. poglavju najprej predstavimo pojem kakovosti storitve, podamo njegove definicije in predlagamo novo definicijo. Osvetlimo tudi kakovost storitve v telekomunikacijskih omrežjih ter poglede uporabnikov in ponudnikov nanjo. Nadaljujemo z navedbo parametrov kakovosti storitve in njihovimi standardiziranimi vrednostmi. Na koncu omenimo še psihološke in sociološke vidike uvažanja različnih stopenj kakovosti storitve v omrežja.

V 3. poglavju se ukvarjamo s kakovostjo storitve v omrežjih IP. Začnemo z opisom splošnih lastnosti paketnih omrežij in njihovimi zmožnostmi zagotavljanja različnih stopenj kakovosti storitve. Nato se osredotočimo na omrežja IP in prenosne tehnologije, ki jih le ta uporabljajo za prenos podatkov. Opišemo zmožnosti osnovnega delovanja omrežij IP in njihove nadgradnje v smislu kakovosti storitve. Na koncu podamo še pogoje za zagotavljanje kakovosti storitve na celotni prenosni poti.

V 4. poglavju opišemo mehanizme za zagotavljanje kakovosti storitev v paketnih omrežjih. Podamo model omrežne naprave ter možnosti za konfiguracijo in krmiljenje čakalnih vrst. Nadaljujemo z opisom odpravljanja in odpravnikov. Navedemo zelene lastnosti odpravnikov ter glavne mere za njihovo ocenjevanje. Na koncu opišemo odpravnike z izhodnim čakanjem, ki jih potrebujemo v nadaljevanju.

V 5. poglavju podamo nov predlog združevanja Diferenciranih in Integriranih storitev. Začnemo z opisom referenčne topologije omrežja ter opisom in primerjavo



---

lastnosti Diferenciranih in Integriranih storitev. Nato podamo razloge za združevanje in sam predlog združitve. Novo omrežje imenujemo DIS - omrežje z Diferenciranimi in Integriranimi storitvami (Differentiated and Integrated Services). Navedemo tudi možne težave, ki lahko nastanejo zaradi združevanja. Nadaljujemo z opisom drugih predlogov združevanja ter podamo predlog konfiguracije in krmiljenja omrežnih naprav, ki je primeren za naš predlog združevanja.

Nato naredimo analizo delovanja prve stopnje DIS omrežja. Podamo teoretično analizo diferenciacije prometa na osnovi prednostnih razredov in izvedemo simulacije. V ta namen smo razvili lasten simulator, ki omogoči simulacijo dvostopenjskega delovanja omrežja. Na prvi stopnji omrežja uporabimo strogo prednostno odpravljanje. Teoretične rezultate primerjamo s simulacijskimi in ugotovimo, da so razlike zanemarljive. Nato izvedemo še simulacijo prve stopnje omrežja z upoštevanjem protokolov na višjih plasteh protokolnega sklada. To nam da rezultate, ki so še bližje realnim razmeram na omrežju.

Na koncu naredimo analizo delovanja druge stopnje omrežja DIS. Izvedemo simulacije in podamo primerjavo med enostopenjskim in dvostopenjskim delovanjem omrežja. Na drugi stopnji omrežja uporabimo DRR (Deficit Round Robin - deficitarno krožno odpravljanje) odpravnik. Pri primerjavi simulacijskih rezultatov ugotovimo, da se pri nestandardnih, podkritičnih parametrih odpravnik DRR zmanjšajo povprečne zakasnitve paketov.

V 6. poglavju predlagamo nov odpravnik. Imenujemo ga *DRR s podkritičnim delovanjem* ali SCDRR (Sub-Critical Deficit Round Robin). Analiziramo njegove lastnosti in delovanje ter podamo mere za pravičnost, algoritemsko zakasnitev in zahtevnost. Izvedemo simulacije samostojnega delovanja odpravnik SCDRR ter simulacije odpravnik SCDRR na drugi stopnji omrežja DIS. Ugotovimo, da je odpravnik SCDRR v vseh lastnostih, razen pri zahtevnosti, boljši od odpravnik DRR.



## 2 Kakovost storitve

*Kakovost storitve* je vse pogosteje uporabljan in mnogokrat tudi zlorabljan pojem. Naj ga zato na tem mestu najprej osvetlimo z več zornih kotov, podamo njegove definicije, njegovo razumevanje s strani uporabnikov in ponudnikov storitve, predstavimo njegove tehnične, psihološke in sociološke vidike ter možnost njegove implementacije v sodobnih telekomunikacijskih omrežjih.

Razvoj telekomunikacijskih omrežij je v zadnjem desetletju napredoval z velikimi koraki in biti povezan ali "on-line" ni več le domena posvečenih. Na začetku je uporabnikom vsekakor najbolj pomembna sama povezljivost v omrežje in obstoj določenih aplikacij, ki jim olajšajo delo. Ko pa uporaba omrežij postane rutina, celo nuja, zgolj osnovne lastnosti ne zadoščajo več. Tedaj se uporabniki naenkrat zavedo potrebe po kakovosti - najsi bo to za aplikacije, povezave v omrežje, prenos podatkov ali kak drug gradnik sodobnih telekomunikacijskih omrežij.

### 2.1 Definicija kakovosti storitve

Splošno veljavne in enoznačne definicije kakovosti storitve ni. Obstaja pa množica definicij, ki služijo širšim ali ožjim področjem znotraj sodobnih telekomunikacij.

Tako je ITU (International Telecommunication Union - mednarodna telekomunikacijska unija) kakovost storitve za potrebe odprtega porazdeljenega procesiranja (Open Distributed Processing) definirala kot: "Niz zahtev za kakovost kolektivnega obnašanja enega ali več objektov." [21] Kakovost storitve lahko opišemo z več parametri, ki opisujejo delovanje in obnašanje sistema.

Bolj splošna pa je definicija v priporočilu E.800 [18], ki pravi: "Kakovost storitve je skupni učinek lastnosti storitve, ki določajo zadovoljstvo uporabnika."

V ATM (Asynchronous Transfer Mode) leksikonu [103] najdemo naslednjo definicijo kakovosti storitve: "Kakovost storitve se nanaša na niz performančnih parametrov, s katerimi opišemo promet preko dane navidezne povezave."

IETF (Internet Engineering Task Force - skupina za razvoj Interneta) se je v svojem dokumentu RFC<sup>1</sup> 1946 [64] kakovosti storitve dotaknila z naslednjimi besedami: "Z naraščanjem povpraševanja po sočasnih omrežnih storitvah narašča tudi potreba po določljivih storitvah prenosa podatkov v teh omrežjih. Določljivost pa od aplikacij in omrežne infrastrukture zahteva sposobnost zahtevati, vzpostaviti in uveljaviti tako določen prenos podatkov. Skupaj lahko tem storitvam rečemo rezervacija pasovne širine in kakovost storitev." V dokumentu RFC 1932 [63] pa takole: "Parametri kakovosti storitve za sočasne aplikacije so taki, da imajo ti podatki prednost pred podatki RSVP

---

<sup>1</sup>Request For Comments - poziv za komentarje

(Resource reSerVation Protocol - protokol za rezervacijo virov) pretokov ali pa se v določeni obliki prenašajo znotraj njih.”

V članku *Distributed Multimedia and Quality of Service: A Survey* [109] je predlagana splošna definicija kakovosti storitve za aplikacije s potrebo po sočasni komunikaciji: ”Kakovost storitve določa niz kvantitativnih in kvalitativnih lastnosti porazdeljenega večpredstavnega sistema, ki so potrebni za doseganje zahtevane funkcionalnosti aplikacije.”

Iz zgornjih odstavkov lahko izluščimo dve glavni vodili za definicijo kakovosti storitve: zadovoljstvo uporabnika in doseganje ustreznih parametrov prenosnega sistema. Zato predlagamo novo definicijo kakovosti storitve, ki se glasi: ”**Kakovost storitve je zagotavljanje vnaprej dogovorjenih parametrov prenosnega sistema, ki jamčijo zadovoljstvo uporabnika.**”

## 2.2 Kakovost storitve v telekomunikacijskih omrežjih

Predvsem nas zanima kakovost storitve v telekomunikacijskih omrežjih. V nadaljevanju vso pozornost namenjamo kakovosti storitve v paketnih omrežjih, še posebej pa obravnavi kakovosti storitve v omrežjih IP.

Tudi nekatere definicije iz prejšnjega razdelka se nanašajo zgolj na telekomunikacijska omrežja ali določeno tehnologijo prenosa podatkov. Če se omejimo zgolj na telekomunikacijska omrežja in prenos podatkov, potem lahko kakovost storitve predstavimo s parametri prenosnega sistema, kot so: zakasnitev, bitna hitrost, zanesljivost prenosa podatkov, propustnost, delež napak in drugimi. Dosežene vrednosti teh parametrov pa morajo jamčiti zadovoljstvo uporabnika. Vrednosti parametrov so med uporabniki in ponudniki storitev dogovorjene vnaprej ali pa se dogovarjajo sproti. Pomemben element zagotavljanja kakovosti storitve, ki je implicitno zaobjet tudi v naši definiciji kakovosti storitve, je namreč dogovor med ponudnikom in uporabnikom. Le tako lahko jamčimo, da so usklajene možnosti in želje obeh strani.

Za primer vzemimo omrežja ATM, kjer imamo definirane naslednje parametre kakovosti storitve: zakasnitev celic, število izgubljenih celic, število okvarjenih celic, število napačno vstavljenih celic, spremenljivost zakasnitve celic in povprečno zakasnitev celic. S temi parametri je definiranih pet razredov kakovosti storitve: od razreda storitev po najboljših možnostih brez zagotovil, do razreda, ki omogoča sočasni prenos podatkov. Ti razredi so definirani za prenos podatkov na nižjih protokolnih plasteh in aplikacijam niso neposredno dostopni.

## 2.3 Kakovost storitve z vidika ponudnika in uporabnika

Pojem *kakovost storitve* je zelo širok, a se večinoma uporablja zgolj za objektivne mere kakovosti povezav in prenosa podatkov v omrežju. Glede na uporabnika ter njegovo dožemanje dela in storitev pa kakovost postaja vse bolj subjektivna mera, ki jo določa on sam. Zato lahko rečemo, da je bistven vidik določanja primerne kakovosti storitve mnenje uporabnikov, ki s svojo oceno izrazijo stopnjo zadovoljstva s storitvijo.

Večinoma uporabnik dojema kakovost storitve precej drugače kot njen ponudnik. Med najpomembnejše razloge za to sodi način njenega dožemanja. Medtem ko ponud-

nik za to uporablja objektivne mere, uporabnik kakovost storitve ocenjuje subjektivno. S stališča prvega so pomembne mere za kakovost storitve zakasnitev, bitna hitrost, propustnost, stopnja napak pri prenosu in podobno. Drugi pa storitev večinoma ocenjuje na podlagi svojih izkušenj pri njeni uporabi. Te niso odvisne zgolj od zmogljivosti omrežja ali opreme ponudnika storitve, temveč tudi od terminalske opreme uporabnika, prijaznosti njegovih aplikacij, njegove usposobljenosti in podobno. V zadnjem času se vse bolj uveljavlja pojem *kakovost izkušnje* (QoE - Quality of Experience), ki predstavlja celovito uporabnikovo zaznavanje storitve. V nadaljevanju za objektivne mere kakovosti storitve uporabljamo pojem kakovost storitve, za subjektivne mere zaznavanja kakovosti storitve s strani uporabnikov pa pojem kakovost izkušnje.

Ker so storitve namenjene njenim uporabnikom, bi se lahko strinjali naj v končni fazi njihovo kakovost določajo in ocenjujejo sami. Žal pa imajo uporabniki lahko zelo različne predstave o kakovosti storitve in svoje želje večinoma izražajo v netehničnem jeziku. Tega morajo ponudniki storitve ovrednotiti s parametri kakovosti, ki jih lahko zagotovi prenosno omrežje. Mnogokrat uporabniki zahtevajo "kakovost storitve", čeprav ne vedo točno kaj to je.

Problem lahko rešimo s sklenitvijo *sporazuma o ravni storitve* (SLA - Service Level Agreement) med ponudnikom in uporabnikom. Pri tem je potrebno uskladiti želje uporabnikov in zmožnosti omrežja. Seveda je pomembno, da ponudnik in uporabnik najdeta skupni jezik. Prvi mora razumeti, kako kakovost storitve dojema uporabnik, ta pa, kaj lahko od ponudnika zahteva in pričakuje. S tem sporazumom se navadno določi tudi cena, ki je odvisna od dogovorjene stopnje kakovosti storitve.

## 2.4 Parametri kakovosti storitve

Splošno sprejetih subjektivnih mer za kakovost storitve ali izkušnje nimamo. Subjektivne mere so navadno povprečne ocene velikega števila ocenjevalcev, ki temeljijo na vnaprej dogovorjeni in pripravljene lestvici ocen. Ponavadi subjektivne mere obstajajo za točno določene storitve, na primer za sproti prenos govora. Na tem mestu zato podajamo parametre, s pomočjo katerih lahko določimo objektivne mere in ocene za kakovost storitev. Najpomembnejši parametri so:

- bitna hitrost (pasovna širina),
- zakasnitev,
- spremenljivost zakasnitve,
- verjetnost napak pri prenosu,
- rezervacija omrežnih virov,
- zasedenost omrežja,
- razpoložljivost omrežja ter
- varnost in zasebnost podatkov.

Glede na nabor parametrov kakovosti storitve, ki jih ponuja omrežje, in potrebe uporabnika, se ta dva dogovorita o vrednostih parametrov prenosa, ki lahko veljajo celoten čas prenosa ali pa se dinamično spreminjajo glede na trenutne potrebe.

Najprimernejši način za doseganje zadovoljivih rezultatov s stališča operaterjev omrežij in ponudnikov storitev je zagotavljanje kakovosti storitev preko omejenega

nabora vrednosti parametrov prenosa, ki jih operaterji omrežja in ponudniki storitev lahko spremljajo in nadzorujejo. Da bi lahko s spreminjanjem parametrov prenosa vplival na delovanje ponujene storitve, mora operater ali ponudnik podrobno poznati njeno obnašanje. Ob tem pa mora zagotoviti primerno (želeno) kakovost storitve, s katero zadovolji uporabnika ter hkrati čimbolj optimalno in ekonomično izkoristiti prenosno omrežje.

#### 2.4.1 Medsebojna odvisnost parametrov kakovosti storitve

Zgoraj naštetih parametri kakovosti storitve so med seboj odvisni in jih zato večinoma ne moremo obravnavati ločeno. Sprememba nekaterih izmed njih lahko odločilno vpliva na druge, odvisno od stopnje njihove soodvisnosti.

Zelo lep primer take soodvisnosti sta bitna hitrost povezave in zakasnitve prenosa podatkov preko nje. Če povečamo prvo, se druga zmanjša. Zakasnitev prenosa  $\tau$  po fizični povezavi med dvema sosednjima vozliščema v omrežju je sestavljena iz zakasnitve razširjanja (propagacije) signala in zakasnitve predaje paketa (čas, ki ga oddajnik porabi, da preda celoten paket na linijo). Zakasnitev razširjanja signala med sosednjima vozliščema je konstantna, zakasnitev predaje paketa pa je odvisna od njegove dolžine. Na primer, če 1000 oktetov dolg paket prenašamo v 100 km oddaljeno napravo preko povezave s hitrostjo 64 kbit/s, ima zakasnitev 125,5 ms, če pa isti paket prenašamo preko povezave s hitrostjo 128 kbit/s, ima zakasnitev 63 ms.

Podobna soodvisnost obstaja med zakasnitvijo in zasedenostjo omrežja, med razpoložljivostjo in zasedenostjo omrežja, med spremenljivostjo zakasnitve in zasedenostjo omrežja, itd. Zagotavljanje kakovosti storitev preko nadzora omenjenih parametrov zatorej ni preprosta naloga. Pri spremembi enega izmed njih moramo vedno upoštevati tudi njegov vpliv na druge parametre.

## 2.5 Standardizirane vrednosti parametrov kakovosti storitve

Glede na zelo splošne in ohlapne definicije (razdelek 2.1), velikega nabora parametrov ter njihove soodvisnosti, ne moremo pričakovati natančno določenih in standardiziranih vrednosti parametrov, ki bi zagotavljali točno določeno ali standardizirano kakovost storitve.

Ena redkih standardiziranih storitev je sprotni prenos govora (interaktivni pogovor). ITU priporočilo G.114 [19] določa maksimalne dovoljene zakasnitve na celotni prenosni poti ob predpostavki, da je ustrezno poskrbljeno za kontrolo odboja<sup>2</sup>.

Zakasnitev na celotni prenosni poti je določena kot zakasnitev zvoka od ust govorca do ušes poslušalca. Sestavljena je iz stalnega in spremenljivega dela. Prvi predstavlja zakasnitev obdelave govornega signala, ki je za dani sistem stalna, drugi pa zakasnitev prenosa okvirov govora, ki je odvisna od prenosnega omrežja in se navadno spreminja

---

<sup>2</sup>Vzrokov za odboj na prenosni poti je več. Najpomembnejša sta odboj zaradi neprilagojenih linijskih impedanc in odboj v prostoru. Za govorca odboj sam po sebi ni nezaželen, ko pa nastopa z večjo zakasnitvijo, postane skrajno neprijeten in moteč. Proti njemu se borimo z zmanjševalniki (Echo Suppressor) in izničevalniki odboja (Echo Celler), ki naj bi se uporabili že pri zakasnitvah nad 25 ms [20].

s časom. Na prvi del zakasnitve omrežje nima vpliva. V disertaciji se zato ukvarjamo s spremenljivim delom zakasnitve in z drugimi parametri kakovosti storitve, ki so neposredno odvisni od delovanja in zmogljivosti prenosnega omrežja.

Za sprotni prenos govora veljajo naslednje standardizirane vrednosti zakasnitve na celotni prenosni poti [19]:

- |              |  |
|--------------|--|
| < 150 ms     | primerno za večino aplikacij,  |
| 150 - 400 ms | sprejemljivo za nekatere aplikacije, če se uporabnik zaveda vpliva zakasnitve na delovanje aplikacije, |
| > 400 ms     | neprimerno za večino aplikacij.  |

Zakasnitve okoli 150 ms so zadovoljive za večino interaktivnih aplikacij, vendar pa nekatere visoko interaktivne govorne in podatkovne aplikacije za dobro delovanje vseeno zahtevajo nižje vrednosti. Podobno velja za zakasnitve nad 400 ms, ki so v osnovi neprimerne za večino interaktivnih aplikacij in se jih moramo izogibati, razen v primerih, ko je dostopnost do storitve (aplikacije) pomembnejša od njene kakovosti. Podrobnejše rezultate meritev vpliva večjih zakasnitev na kakovost interaktivnega pogovora najdemo v dodatkih A in B priporočila ITU G.114 [19], v [23] pa so podrobneje obdelane možnosti prenosa govora preko podatkovnih omrežij, tudi s stališča kakovosti storitve.

## 2.6 Parametri kakovosti storitve za zahtevne aplikacije

Uveljavljene aplikacije<sup>3</sup>, ki izhajajo iz podatkovnega sveta, so večinoma nezahtevne glede kakovosti storitve pri prenosu podatkov. Večini je dovolj že povezljivost in razumne vrednosti ostalih parametrov kakovosti storitve, podanih v razdelku 2.4. Z napredkom tehnologije, večanjem zahtev uporabnikov, konkurence in zlivanjem omrežij pa so se na podatkovnih omrežjih pojavile aplikacije, ki jim zgolj povezljivost ne zadošča več. Za ustrezno delovanje od omrežja pričakujejo izpolnjevanje določenih, večkrat zelo strogih zahtev glede prenosnih parametrov. Kot najzahtevnejše se izkažejo aplikacije, ki delujejo v realnem času. Kot primer zopet uporabimo sprotni prenos govora, ki od omrežja zahteva naslednje:

- zadostna razpoložljiva bitna hitrost,
- rezervacija pasovne širine,
- nizka zakasnitev (prenos v realnem času),
- čimmanjša spremenljivost zakasnitve,
- nizek nivo napak na prenosni poti.

Vse našteje zahteve, razen zadnje, je v klasičnih paketnih omrežjih težko izpolniti. Zadostna bitna hitrost za prenos govora načeloma ni problematična, težave pa nastanejo, ker mora biti na voljo ves čas aktivnosti aplikacije. V zvezi s tem je tudi želja po rezervaciji pasovne širine, ki aplikacijam zagotovi pasovno širino ves čas njenega trajanja. Nizka zakasnitev se navezuje na vrednosti podane v razdelku 2.5 in je tesno povezana s svojo spremenljivostjo, ki mora biti v sprejemniku kompenzirana z dodatno

<sup>3</sup>Pri tem imamo v mislih predvsem aplikacije za prenos elektronske pošte, datotek in brskanja po svetovnem spletu.

zakasnitvijo, ki se prišteva skupni zakasnitvi na prenosni poti. Še večje težave nastopijo pri sprotnem prenosu videa, ki zahteva poleg tega tudi dokaj veliko pasovno širino.

## 2.7 Psihološki in sociološki vidiki uvajanja različnih stopenj kakovosti storitev

Uvajanje različnih stopenj kakovosti storitev v paketnih omrežjih narekuje več dejavnikov. Čeprav ponavadi najprej pomislimo na velike zahteve aplikacij, uvajanje novih storitev in želje uporabnikov pa smo verjetno bliže resnici, če rečemo, da bo gonilna sila uvajanja kakovosti storitev ekonomija. Ponudniki storitev in operaterji omrežij bodo v želji pridobivanja konkurenčne prednosti ponudili različne stopnje kakovosti storitve. Glede na to, da je začetna osnova večinoma zelo nizka<sup>4</sup>, naj bi vsaka ponudba boljše kakovosti storitve pomenila tudi višji dohodek. Za boljšo kakovost storitve bo seveda potrebno plačati več.

Vprašanje je, kako bodo to sprejeli uporabniki. Če se osredotočimo na omrežje Internet, ki ga uporablja daleč največ ljudi, opazimo, da je le to izredno demokratično. Vsi uporabniki, ki v istem trenutku zasedajo določeno povezavo, imajo enake možnosti za prenos podatkov in zanje dosegajo enake prenosne parametre. Z uvedbo različnih stopenj kakovosti storitve to demokratičnost porušimo in razslojimo trenutno enakopravne uporabnike omrežja na "reveže" in "bogataše". Zgolj ugibamo lahko, kako bo to vplivalo na njih ter posledično tudi na operaterje omrežij in ponudnike storitev. Vsekakor je to snov nadaljnjih raziskav.

---

<sup>4</sup>V mislih imamo predvsem javna omrežja IP, ki trenutno podpirajo le prenos po najboljših zmožnostih.



## 3 Kakovost storitve v omrežjih IP

V tem poglavju obravnavamo možnosti zagotavljanja različnih stopenj kakovosti storitve v omrežjih IP. Izbrali smo jih zaradi njihove vsesplošne prevlade med trenutno delujočimi paketnimi omrežji. Zato najprej predstavimo osnovne lastnosti paketnih omrežij in njihovo zmožnost zagotavljanja različnih stopenj kakovosti storitve. Nadaljujemo s kratko navedbo zmožnosti zagotavljanja različnih stopenj kakovosti storitve za najpogostejše tehnologije povezavne plasti protokolnega sklada, ki jih za prenos podatkov uporabljajo omrežja IP. Zaključimo z obravnavo omrežij IP, njihovih nadgradenj v smislu kakovosti storitve in s pogoji, ki so potrebni za zagotavljanje kakovosti storitve na celotni prenosni poti.

### 3.1 Osnovne lastnosti paketnih omrežij

V paketnih omrežjih prenašamo podatke s pomočjo paketov različnih dolžin, ki se na prenosni poti statistično multipleksirajo s podatki drugih povezav. Statistični multipleks ima kar nekaj prednosti pred časovnim ali frekvenčnim multipleksom, ima pa tudi svoje slabosti.

#### 3.1.1 Statistični multipleks

Pri statističnem multipleksu podatke posredujemo, ko se ti pojavljajo na vhodu multipleksorja. Celoten paket pošljemo naenkrat in v tem času morajo paketi ostalih povezav (kanalov) čakati na sprostitev prenosne poti. Zato imamo v takih multipleksorjih čakalne vrste, ki začasno hranijo pakete, ki še niso prišli na vrsto za posredovanje in jih nato po določenem algoritmu posredujejo na izhod multipleksorja.

Tako kot pri drugih tehnikah multipleksiranja je tudi pri statističnem bistvo združevanje kanalov. Pri večini multipleksirnih tehnik lahko vzpostavljamo nove povezave dokler ima multipleksiran kanal proste kapacitete. V nasprotnem primeru je vzpostavitev povezave zavrnjena. Vsaka vzpostavljena povezava ima zagotovljeno kapaciteto ves čas trajanja povezave. Pri statističnem multipleksu število povezav načeloma ni omejeno, kapaciteta za posamezne vzpostavljene povezave pa ni več zagotovljena. Posledici te odločitve sta boljša izkoriščenost prenosne poti in stalna dosegljivost povezave (stalna povezljivost). Oboje velja za dobre lastnosti statističnega multipleksa. Druge multipleksirne tehnike (kot sta časovni in frekvenčni multipleks) imajo slabšo izkoriščenost kapacitet in možnost zavrnitve povezave ob polni zasedenosti multipleksiranih kanalov. Slaba stran statističnega multipleksa je možnost zasičenja omrežja in s tem izguba paketov ter nedeterministične razmere na prenosni poti. Druge multipleksirne tehnike, na primer časovni in frekvenčni multipleks, imajo rezervirane kapacitete

za vsako povezavo in s tem točno določene parametre prenosa.

Večina podatkovnih aplikacij tvori neenakomeren tok podatkov. Prenosna pot je zato s paketi teh aplikacij zasedena le v času njihovega pošiljanja in ne ves čas trajanja njihove povezave. Vstopni kanali multipleksorja niso ves čas enako aktivni ampak se količina podatkov na njih spreminja s časom. Kapaciteta prenosne poti na izhodu multipleksorja je zato lahko manjša od vsote maksimalnih pretokov vhodnih kanalov. Dolgoročno zadošča že, če je njena kapaciteta vsota povprečnih vrednosti pretokov na vhodnih kanalih. Seveda se zaradi tega zgodi, da je v nekem času pritek paketov v multipleksor prevelik. Čakalna vrsta, ki zaradi omejene dolžine lahko kompenzira le kratkotrajne presežke pritoka, se začne polniti in ob daljšem obdobju presežnega prometa prične zavračati novoprispele pakete. Takemu stanju rečemo zasičenje vozlišča ali omrežja. Zaradi tega je v omrežjih s statističnim multipleksom pomembno uporabljati mehanizme, ki zmanjšajo možnost zasičenja in po možnosti omogočajo ponudbo različnih stopenj kakovosti storitve.

Okoli statističnega multipleksa, čakalnih vrst, algoritmov njihove strežbe, izločanju odvečnih paketov, itd, se je razvila obsežna teorija, ki se je dotaknemo v naslednjih poglavjih.

### 3.1.2 Razpoložljiva kapaciteta prenosne poti, zakasnitev in njena spremenljivost

Statistično multipleksiranje dokaj učinkovito izrablja kapacitete prenosnih poti, vendar ima lahko manj ugodne vplive na prenos podatkov. Ker ob času odpošiljanja paketov ne vemo, na kakšne razmere bodo ti naleteli v omrežju, tudi ne vemo kakšni bodo njihovi prenosni parametri.

Zakasnitev prenosa podatkov v omrežju je vsota časov, potrebnih za prenos paketa med vozlišči in je odvisna od njihove medsebojne oddaljenosti ter razmer v omrežju. Čas prenosa paketa med vozlišči je sestavljen iz fizične zakasnitve paketa in njegove zakasnitve v vozlišču. Fizična zakasnitev je odvisna le od razdalje, hitrosti razširjanja signala in kapacitete povezave, zakasnitev v vozlišču pa od stanja vozlišča in je sestavljena iz zakasnitve obdelave paketa ter zakasnitve zaradi uvrstitve paketa v čakalno vrsto<sup>1</sup>. Zakasnitev obdelave paketa je ponavadi zanemarljiva, zakasnitev zaradi uvrstitve paketa v čakalno vrsto pa se spreminja v odvisnosti od zasedenosti čakalne vrste in je lahko precejšnja. Zasedenost čakalnih vrst je odvisna od kratkoročne vsote prometa na njenem vhodu. Podatki posameznih povezav jo občutijo kot spremenljivo razpoložljivo kapaciteto prenosne poti in s tem spremenljivo zakasnitev.

Vidimo, da statistično multipleksiranje, ob dobri izrabi omrežnih kapacitet, v prenos podatkov vnese negotovost v smislu razpoložljivih prenosnih kapacitet, zakasnitve in njene spremenljivosti.

---

<sup>1</sup>Uvrstitev paketa v čakalno vrsto ima za posledico njegovo čakanje na vrsto za oddajo in s tem dodatno zakasnitev. Kdaj paket pride na vrsto, pa je odvisno od načina strežbe čakalne vrste in njene prednosti glede na možne druge vzporedne čakalne vrste na istem izhodu multipleksorja.

### 3.1.3 Kakovost storitve v paketnih omrežjih

Dejansko lahko kakovost storitve ocenimo glede na dosežene vrednosti parametrov kakovosti storitve, ki so podani v razdelku 2.4. Najpomembnejše lastnosti paketnih omrežij, ki so povezane s kakovostjo storitve, so:

- statistično multipleksiranje,
- spremenljiva razpoložljiva prenosna kapaciteta,
- nevarnost zasičenja omrežja,
- srednja do velika in spremenljiva zakasnitev.

Zato lahko rečemo, da paketna omrežja niso najbolj primerna za prenos podatkov, ki bi zahtevali stalno in zagotovljeno kakovost storitve. Kot primer vzemimo prenos G.729 kodiranega interaktivnega govora, ki ima do prenosnega omrežja naslednje zahteve:

- bitna hitrost 8 kbit/s,
- zagotovljena stalna prenosna kapaciteta,
- zakasnitev na celotni prenosni poti manjša od 150 ms,
- čim manjša spremenljivost zakasnitve<sup>2</sup>,
- izguba okvirov govora manjša od 5%.

Ugotovimo, da prva zahteva običajno ni problematična, saj imajo vsa paketna omrežja večje prenosne kapacitete kot jih zahteva govor (tipično 5.3 do 64 kbit/s). Zaplete pa se že pri drugi zahtevi, ki je paketna omrežja s svojim statističnim multipleksiranjem ne morejo v celoti zagotoviti. Pri nekaterih paketnih omrežjih skušamo to rešiti z različnimi rezervacijskimi protokoli, ki rezervirajo prenosno kapacitete za govor na celotni prenosni poti, pri tem pa seveda izgubimo nekatere prednosti statističnega multipleksa.

Zagotavljanje primerno nizke in čimbolj enakomerne zakasnitve pri paketnih omrežjih postane problem, ko se obremenitev omrežja veča in približuje meji zasičenja. Zakasnitev se tedaj veča in postaja vse bolj neenakomerna do te mere, da postane pogovor močno moten ali celo nemogoč. Na sprejemni strani moramo spremenljivost zakasnitve uravnati z dodatno zakasnitvijo, kar nas kaj hitro privede na mejo še sprejemljive zakasnitve in s tem kakovosti govora.

V povezavi z zasičenjem omrežja je tudi izguba podatkovnih paketov (v našem primeru govornih), kar povzroči izpad krajših ali daljših odsekov govora tako, da sporazumevanje postane moteno ali onemogočeno.

Če strnemo zgornje ugotovitve, pridemo do spoznanja, da so paketna omrežja zaradi svojih lastnosti manj primerna za prenos govora, saj ne morejo zadostiti vsem zahtevam govornih aplikacij. Enako ugotovimo za nekatere druge aplikacije s podobnimi ali celo višjimi zahtevami po kakovosti storitve, še posebej, če zahtevajo interaktivnost ali prenos v realnem času.

V nadaljevanju se osredotočimo na omrežja IP in njihovo zmožnost zagotavljanja različnih stopenj kakovosti storitve.

---

<sup>2</sup>Dodatna zakasnitev, ki jo zaradi spremenljivosti zakasnitve vnese sprejemni medpomnilnik (jitter buffer), in zakasnitev na prenosni poti skupaj ne smeta preseči 150 ms.

## 3.2 Omrežja IP

Prvi zametki omrežij IP segajo že v daljna šestdeseta leta prejšnjega stoletja. Sedanjo obliko so dobila ob koncu sedemdesetih let z definicijo para protokolov TCP (Transmission Control Protocol - protokol za nadzor prenosa) in IP. V začetku je bilo to vojaško omrežje, kasneje pa se je razširilo tudi v akademske kroge. Načrtovano je bilo za robusten in zanesljiv prenos podatkov, ki naj bi deloval tudi ob hkratnem izpadu več vozlišč v omrežju. Poudarek je bil torej na povezljivosti in ne na prenosnih zmogljivostih, kar se v veliki meri pozna še danes. V devetdesetih letih pa so omrežja IP doživela komercialen razmah in postala najpomembnejša in najbolj razširjena podatkovna omrežja.

Omrežja IP uporabljajo Internetni protokol (IP), ki je najpogosteje uporabljan omrežni protokol za prenos podatkov znotraj posameznih omrežij in med omrežji, ki temeljijo na različnih prenosnih tehnologijah. Njegova uspešnost temelji predvsem na njegovi robustnosti in preprostosti.

V osnovi so omrežja IP nezanesljiva, nepovezavno orientirana paketna omrežja, ki prenašajo podatke po najboljših zmožnostih.

**Nezanesljivost** pomeni, da se paketi na poti lahko izgubijo, imajo nepredvidljivo zakasnitev, se lahko podvojijo, so dostavljeni v napačnem vrstnem redu ali na cilj pridejo okvarjeni. Protokol IP na omrežni plasti nudi nezanesljivo storitev prenosa podatkov protokolom transportne plasti. Najpomembnejši in najpogosteje uporabljan transportni protokol je TCP, ki končnima točkama na prenosni poti zagotavlja zanesljiv, povezavno orientiran prenos podatkov na osnovi potrjevanja in morebitnega ponovnega pošiljanja okvarjenih ali izgubljenih podatkovnih segmentov. TCP tako nadgradi protokol IP z zanesljivim prenosom podatkov.

**Nepovezavnost** pomeni, da se pred prenosom podatkov ne vzpostavi povezava med končnima točkama. Podatki se preprosto predajo omrežju, ki poskrbi za njihovo dostavo na cilj. To pomeni, da se vsak paket na omrežju obravnava neodvisno od drugih in po njem tako tudi potuje. Vzpostavlanje in vzdrževanje povezav je v omrežjih IP stvar višjih protokolnih plasti (TCP).

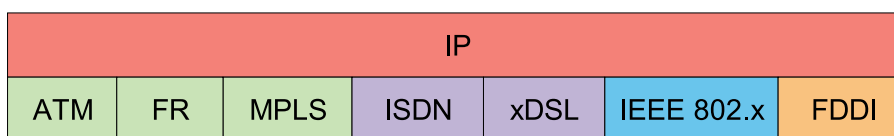
**Prenos po najboljših zmožnostih** ni nujno nekaj slabega, čeprav se v smislu zagotavljanja različnih stopenj kakovosti storitve redno znajde na dnu lestvice. Omrežje paketov ne zavrže samovoljno, brez pravega vzroka, ampak se to najverjetneje zgodi zaradi zasičenosti omrežja ali izpada povezave. Če so razmere v omrežju dobre in to deluje z nizko obremenitvijo, potem tudi prenos po najboljših zmožnostih zadošča večini prenosnih zahtev.

### 3.2.1 Omrežja IP in prenosne tehnologije

Omrežja IP za prenos podatkov uporabljajo različne prenosne tehnologije in protokole (slika 3.1). Če želimo, da omrežja IP zagotovijo različne stopnje kakovosti storitve, morajo to zagotoviti tudi tehnologije in protokoli, ki jih omrežja IP uporabijo za prenos. Zato v tem razdelku za pomembnejše prenosne tehnologije in protokole navajamo njihove zmožnosti zagotavljanja kakovosti storitve. Več o delovanju in lastnostih navedenih tehnologij in protokolov najdemo v [54], [6] in [23]. Osredotočili smo se na dva kriterija:

- zmožnost nudenja različnih stopenj kakovosti storitve za različne podatkovne pretoke in
- zmožnost zagotavljanja dogovorjenih prenosnih parametrov.

Pri tem s  $\checkmark$  označimo, da je omrežje tega sposobno, z  $\times$ , da tega ni sposobno in s  $\bigcirc$ , da je tega sposobno le pod določenimi pogoji.



Slika 3.1 – Nekatere tehnologije in protokoli, ki jih za prenos podatkov uporabljajo omrežja IP

V tabelah 3.1 do 3.3 razdelimo obravnavane tehnologije na tiste, ki se uporabljajo v krajevnih (LAN - Local Area Network) in mestnih omrežjih (MAN - Metropolitan Area Network), na tiste, ki se uporabljajo v dostopovnih omrežjih in na tiste, ki se uporabljajo v prostranih omrežjih (WAN - Wide Area Network).

Tehnologija	Zagotavljanje QoS		Komentar
	stopnje	parametri	
Ethernet	$\checkmark$	$\times$	relativne stopnje QoS ob uporabi IEEE 802.3p
IzoEthernet	$\checkmark$	$\checkmark$	možnost rezervacije izohronih kanalov
Token Ring	$\checkmark$	$\bigcirc$	relativne stopnje QoS, vrednost prenosnih parametrov za posamezni razred je omejena in odvisna od števila aktivnih delovnih postaj
Token Bus	$\checkmark$	$\bigcirc$	
ARCnet	$\checkmark$	$\bigcirc$	
Local Talk	$\times$	$\times$	ni QoS mehanizmov
Wireless LAN	$\checkmark$	$\bigcirc$	ob uporabi IEEE 802.11e
FDDI/CDDI	$\times$	$\bigcirc$	vrednost parametrov odvisna od števila postaj
FDDI II	$\times$	$\checkmark$	možnost rezervacije izohronih kanalov
DQDB	$\checkmark$	$\checkmark$	relativne stopnje QoS, izohroni kanali

Tabela 3.1 – Kakovost storitve v lokalnih in mestnih omrežjih

V omrežjih Ethernet lahko z uporabo prednostne sheme IEEE 802.3p zagotovimo relativne stopnje kakovosti storitve, ne moremo pa zagotavljati prenosnih parametrov. Pri Izohronem Ethernetu lahko z ustrezno rezervacijo izohronih kanalov<sup>3</sup> jamčimo za bitno hitrost, zakasnitev in druge prenosne parametre. Omrežja Token Ring, Token Bus in ARCnet delujejo na skoraj enakem principu podajanja žetona<sup>4</sup>. Vsa tri omrežja imajo možnost relativnega razlikovanja prometa. Vsa tri omrežja zagotavljajo tudi neko minimalno bitno hitrost, ki pa je odvisna od trenutnega števila aktivnih delovnih postaj v omrežju. Ob uporabi več prednostnih razredov je minimalna bitna hitrost zagotovljena le za razred z najvišjo prednostjo. V brezžičnih krajevnih omrežjih je ob uporabi

<sup>3</sup>Beseda izohrono pomeni istočasno, sočasno. Izohroni prenos podatkov pomeni, da ti prihajajo v rednih, enakomernih in vnaprej definiranih intervalih. Zaradi tega so izohrona omrežja še posebej primerna za aplikacije, ki imajo stalen in enakomeren pretok podatkov ter zahtevajo prenos v realnem času.

<sup>4</sup>Pri Token Ringu je obroč izveden fizično, pri ostalih dveh pa logično.

priporočila IEEE 802.11e možno zagotavljati različne stopnje kakovosti storitve. Osnovne funkcije priporočila podpirajo štiri relativne stopnje kakovosti storitve, napredne funkcije pa omogočijo determinističen dostop do prenosnega medija in s tem možnost zagotavljanja prenosnih parametrov. FDDI (Fiber Distributed Data Interface - optični porazdeljeni podatkovni vmesnik) in CDDI (Copper Distributed Data Interface - žično porazdeljeni podatkovni vmesnik) sta mestni omrežji, ki ne podpirata različnih stopenj kakovosti storitve, zagotavljata pa pravičen in determinističen dostop do prenosnega medija. Vrednosti prenosnih parametrov pa so spet odvisne od trenutnega števila aktivnih delovnih postaj. Opcijska izvedba omrežij FDDI in CDDI podpira tudi sinhrono delovanje, ki vsaki delovni postaji na omrežju določi časovni interval, v katerem zagotovo lahko oddaja podatke. FDDI II je nadgradnja omrežij FDDI, ki podpira tudi izohrone kanale. DQDB (Distributed Queue Dual Bus - dvojno vodilo s porazdeljeno čakalno vrsto) je tehnologija za mestna omrežja, ki podpira relativne stopnje kakovosti storitve in tudi izohrone kanale ter jo lahko primerjamo z Izohronim Ethernetom. Obe omrežji sta sposobni nuditi različne stopnje kakovosti storitve in jamčiti za prenosne parametre.

Tehnologija	Zagotavljanje QoS		Komentar
	stopnje	parametri	
xDSL	×	✓	za vse podatkovne pretoke skupaj
Kabelska	×	○	odvisno od uporabljene tehnologije in izvedbe
PSTN	×	✓	za vse podatkovne pretoke skupaj
ISDN	×	✓	za vse podatkovne pretoke skupaj

Tabela 3.2 – Kakovost storitve in dostopovne tehnologije

Za dostopovne tehnologije (tabela 3.2) je značilno, da enemu ali več uporabnikom omogočajo dostop do nekega omrežja. Večinoma nudijo le podatkovni kanal, preko katerega, najpogosteje s pomočjo PPP (Point-to-Point Protocol - protokol točka točka), prenašamo podatke omrežne plasti (IP). PPP je protokol povezavne plasti, ki ne podpira različnih stopenj QoS in ne zagotavlja vrednosti prenosnih parametrov. Značilno za omenjene dostopovne tehnologije je, da za vse podatkovne pretoke skupaj nudijo nespremenljive prenosne parametre. PSTN (Public Switched Telephone Network -javno komutirano telefonsko omrežje) in ISDN (Integrated Services Digital Network - digitalno omrežje z integriranimi storitvami) bi lahko obravnavali tudi kot omrežje velikega dosega, vendar je taka raba zanje zelo netipična. Glede na to, da dostopovne tehnologije nudijo nespremenljive prenosne parametre, lahko z ustreznimi mehanizmi na višjih protokolnih plasteh zadovoljivo rešimo problem zagotavljanja kakovosti storitve.

Poglejmo si še najpomembnejša omrežja velikega dosega. V omrežjih Frame Relay (FR - blokovno posredovanje) ima posamezna povezava skozi omrežje zagotovljeno minimalno bitno hitrost ali CIR (Committed Information Rate - zjamčena informacijska hitrost). Ta velja za vse podatkovne pretoke na tej povezavi skupaj. Zagotavljanje različnih stopenj storitve med pretoki je zato odvisno od realizacije FR omrežnih naprav. Omrežja ATM so bila že od vsega začetka načrtovana za storitve z različnimi zahtevami po kakovosti storitve. Če želimo na omrežju ponuditi in zagotavljati različne stopnje kakovosti storitve, so od vseh omrežij najprimernejša prav omrežja ATM saj omogočajo

Tehnologija	Zagotavljanje QoS		Komentar
	stopnje	parametri	
X.25	×	×	ni QoS mehanizmov
Frame Relay	○	✓	odvisno od naprav, zagotovljena bitna hitrost
ATM	✓	✓	absolutne stopnje QoS
MPLS	✓	○	relativne stopnje QoS, prometni inženiring
SDH	×	✓	možnost rezervacije izohronih kanalov

Tabela 3.3 – Kakovost storitve v omrežjih velikega dosega

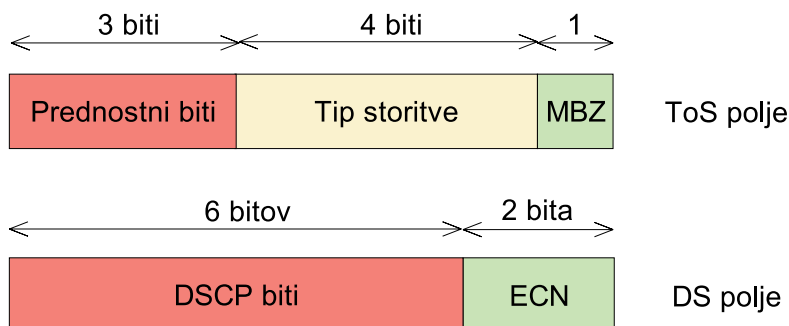
zagotavljanje absolutne stopnje kakovosti storitve. Omrežja MPLS (Multiprotocol Label Switching - Multiprotokolna komutacija na osnovi label) podpirajo relativne stopnje kakovosti storitve, prenosne parametre pa pod določenimi pogoji lahko zagotovimo s prometnim inženiringom, ki ga dovoljuje MPLS. Omrežja SDH (Synchronous Digital Hierarchy - Sinhrona digitalna hierarhija) delujejo na fizični plasti protokolnega sklada. Za prenos podatkov IP moramo uporabiti še protokol povezavne plasti, največkrat kar PPP. Tu jih omenjamo zato, ker omogočajo rezervacijo izohronih kanalov, ki so potrebni za prenos podatkov aplikacij z visokimi zahtevami po kakovosti storitve.

### 3.2.2 Nadgradnja omrežij IP za zagotavljanje kakovosti storitev

V osnovi omrežja IP uporabnikom večinoma ponujajo zgolj povezljivost. Prenos podatkov se vrši po načelu "najboljših zmožnosti" (best effort), kar z drugimi besedami pomeni - nobenih zagotovil. Razlogi za tako stanje so zgodovinski, saj so bila omrežja IP načrtovana zgolj za prenos podatkov aplikacij, ki od omrežja ne zahtevajo nikakršnih zagotovil o kakovosti storitve. Problem je nastal z uvajanjem novih storitev, ki pa v omrežju z zgolj "najboljšimi zmožnostmi" ne delujejo dobro (ali pa sploh ne). Omrežja IP ob vsakem času na vsakem mestu vsem aplikacijam ponujajo enake pogoje. To pomeni, da ne razlikujejo med podatki različnih aplikacij. Taka demokratičnost ne škodi in lepo deluje v nizko obremenjenih omrežjih, ko vsaka aplikacija uspe dobiti svoj delež omrežnih virov. Težave se začnejo z višanjem obremenjenosti omrežja, ko se razmere (zopet demokratično) slabšajo za vse aplikacije enako. Aplikacije z večjimi zahtevami kmalu naletijo na razmere, ki niso več zadovoljive in močno poslabšajo ali celo onemogočijo njihovo delovanje.

Lepa predstavnik aplikacij primernih in neprimernih za omrežja IP sta elektronska pošta in prenos interaktivnega govora. Prva je tipična aplikacija načrtovana za omrežja IP, ki od omrežja ne zahteva veliko. Potrebuje le povezljivost med pošiljateljem in naslovnikom ter ne izkazuje nobenih posebnih zahtev glede zakasnitve ali drugih parametrov kakovosti storitve. Prenos interaktivnega govora zahteva precej več: stalno in rezervirano bitno hitrost ter nizko in čim manj spremenljivo zakasnitev. To pa so za omrežja IP že precej hude zahteve, ki jih ne morejo vedno zagotoviti. Omrežja IP je za zagotavljanje različnih stopenj kakovosti storitve potrebno nadgraditi.

Napisano označuje zgolj trenutno stanje omrežij IP in ne pomeni, da omrežja IP že v osnovi niso zmožna nuditi različnih stopenj kakovosti storitve. Glava paketa IP namreč vsebuje polje velikosti enega okteta, ki je rezervirano prav v ta namen in je bilo določeno že ob sami definiciji protokola IP v standardu RFC 791 [86] ter podrobneje



Slika 3.2 – Zgradba polja ToS in DSCP

definirano v standardu RFC 1349 [61]. Polje se imenuje ToS (Type of Service - tip storitve) in je prikazano na sliki 3.2. Vključuje 3 prednostne bite, 4 bite za tip storitve ter en neuporabljen bit (MBZ - Must Be Zero), ki ima določeno vrednost nič. S tremi prednostnimi biti je mogoče določiti do 8 različnih prednostnih razredov, ki lahko ustrezajo do 8 različnim stopnjam kakovosti storitve. Vrednosti in pomen prednostnih bitov so navedene v tabeli 3.4. Poleg tega lahko s 4 biti za tip storitve opišemo do 16 različnih storitev omrežja. Zaenkrat je definiranih le pet tipov storitev, ki so navedeni v tabeli 3.5. Čeprav vmesne kombinacije bitov niso določene, to ne pomeni, da so prepovedane, le njihova uporaba ni standardizirana. Podrobnejšo razlago najdemo v [61].

Prednost	Prednostni biti	Storitev
0	000	normalno posredovanje
1	001	prednostno posredovanje
2	010	takojšnje posredovanje
3	011	krizno posredovanje
4	100	prednostno krizno posredovanje
5	101	kritično posredovanje
6	110	nadzor medmrežja
7	111	nadzor omrežja

Tabela 3.4 – Pomen prednostnih bitov v polju ToS paketa IP

Oznaka	Tip storitve
1000	minimiziraj zakasnitev
0100	maksimiziraj propustnost
0010	maksimiziraj zanesljivost
0001	minimiziraj ceno
0000	normalna storitev

Tabela 3.5 – Pomen bitov za tip storitve v polju ToS paketa IP



Vendar pa se polje ToS, kljub za tiste čase dobro zamišljenemu konceptu, v omrežjih IP ni uporabljalo. Za potrebe Diferenciranih storitev (glej tudi razdelek 3.2.5) je bilo zato kasneje s standardoma RFC 2474 [72] in RFC 3168 [83] določeno polje DS, ki v glavah paketov IPv4 in IPv6 nadomesti polje ToS. Zgradba polja DS je prikazana na sliki 3.2. Polje je razdeljeno na dva dela. Prvih 6 bitov predstavlja kodno oznako diferencirane storitve ali DSCP (Differentiated Services Codepoint) in dopušča do 64 različnih storitvenih razredov (stopenj kakovosti storitve), zadnja dva bita (ECN - Explicit Congestion Notification) pa omogočata neposredno obveščanje o zamašitvi (zasičenju) v omrežju [83]. V tabeli 3.6 so navedene DSCP vrednosti za že definirane storitvene razrede, ki jih v terminologiji omrežij DS označujemo s PHB (Per-Hop Behaviour - obnašanje vedenjske skupine paketov). PHB tako označuje skupino paketov, ki so v omrežju deležni enake stopnje kakovosti storitve.

Storitveni razred (PHB)	DSCP	Storitev
Privzeti	000000	po najboljših zmožnostih
Izbirni	xxx000	xxx = enako bitom polja ToS
Ekspeditivno posredovanje	101110	nizka zakasnitev, nizke izgube, nizka spremenljivost zakasnitve...
Zagotovljeno posredovanje	aaabbb	definiranih dvanajst podrazredov

Tabela 3.6 – Storitveni razredi za nekatere priporočene vrednosti DSCP

*Privzeti PHB* določa, da je paket z vrednostjo polja DSCP "000000" obravnavan po najboljših zmožnostih. *Z izbirnim PHB* se paketi obravnavajo na enak način kot pri prednostni obravnavi, določeni s poljem ToS (tabela 3.5). Pri *ekspeditivnem posredovanju* (EF - Expedited Forwarding) so paketi deležni ugodnih prenosnih parametrov: nizka zakasnitev, nizka spremenljivost zakasnitve, nizke izgube, rezervacija bitne hitrosti in podobno. Pri *zagotovljenem posredovanju* (AF - Assured Forwarding) pa je definiranih dvanajst podrazredov, ki so navedeni v tabeli 3.7. Podrobnejšo razlago definiranih DSCP vrednosti najdemo v [72]-[75] in [84]-[85].

Razred	AFy1	AFy2	AFy3
AF1x	001010	001100	001110
AF2x	010010	010100	010110
AF3x	011010	011100	011110
AF4x	100010	100100	100110

Tabela 3.7 – Podrazredi zagotovljenega posredovanja (AF)

Pri AF imamo definirane štiri storitvene razrede AF1x do AF4x. Pri tem x pomeni prednost odmetavanja paketov v posameznem razredu. Uporaba prednosti odmetavanja znotraj posameznega storitvenega razreda je koristna, saj se na ta način kaznuje promet, ki prekorači enega izmed dogovorjenih parametrov (na primer bitno hitrost). Kaznovane pakete se ponovno označi, tokrat z višjo prednostjo odmetavanja.

Za preprosto razlikovanje stopenj kakovosti storitve v omrežjih IP imamo torej na voljo dve možnosti: uporabo polja ToS ali polja DS. Če na prenosni poti različni

deli omrežja uporabljajo različni možnosti, mora omrežje pri prehodu paketov iz dela omrežja, ki za razlikovanje stopnje kakovosti storitve uporablja polje ToS, v del omrežja, ki za to uporablja polje DS, zagotoviti čimbolj ustrezno preslikavo med vrednostmi prednostnih bitov polja ToS in DSCP vrednostmi polja DS ter obratno.

Preslikava iz ToS v DSCP vrednost poteka brez problemov, saj porabimo zgolj eno osmino možnih vrednosti polja DSCP. Polje ToS namreč preslikamo v izbirni DSCP razred. V obratni smeri pa gre težje, saj moramo 64 vrednosti polja DSCP preslikati v le 8 vrednosti polja ToS, kar lahko vodi v okrnjeno razločevanje različnih stopenj kakovosti storitve.

### 3.2.3 Diferencirane in integrirane storitve

Najpomembnejši nadgradnji za zagotavljanje različnih stopenj kakovosti storitve v omrežjih IP sta uvedba Integriranih storitev (IS - Integrated Services) in Diferenciranih storitev (DS - Differentiated Services). Pri IETF (Internet Engineering Task Force) so najprej definirali priporočila s skupnim imenom Integrirane storitve, katerih izvedba se je pokazala za zapleteno in dokaj zahtevno. Zato so kasneje definirali še Diferencirane storitve, ki so za ceno nekoliko slabših lastnosti manj zahtevne za izvedbo.

### 3.2.4 Osnovne lastnosti in delovanje omrežij z IS

Koncept IS se je začel razvijati v začetku devetdesetih let prejšnjega stoletja in se uveljavil z IETF standardom RFC 1633 [62]. IS omogočajo rezervacijo virov omrežja na osnovi podatkovnih pretokov<sup>5</sup> na celotni prenosni poti. To pomeni, da se za vsak podatkovni pretok rezervirajo omrežni viri v vsaki omrežni napravi na poti. IS v omrežja IP uvedejo tri razrede storitev. Poleg razreda najboljših zmožnosti (best effort), ki predstavlja osnovno stanje omrežja, sta uvedena še razreda zajamčenih storitev (Guaranteed Services) in storitev nadzorovanega bremena (Controlled Load Services).

Razred zajamčene storitve uporabniku nudi trdna zagotovila glede prenosnih parametrov, kot sta bitna hitrost in zakasnitev, zato je zelo primeren za zahtevne aplikacije, na primer prenos podatkov v realnem času. Razred storitve nadzorovanega bremena ne nudi trdnih zagotovil glede zakasnitev. Zakasnitev storitve nadzorovanega bremena je v predpisanih mejah, če se vsi podatkovni pretoki, ki ji pripadajo, držijo dogovorjenih prenosnih parametrov. Lahko rečemo, da aplikacijam ponuja storitev, ki je približno enaka storitvi po najboljših zmožnostih na nizko obremenjenem omrežju. Ta razred je primeren predvsem za aplikacije, ki zahtevajo boljše lastnosti prenosa podatkov kot jih lahko nudi prenos po najboljših zmožnostih. Več o omenjenih razredih storitev najdemo v [70]-[71].

Za vzpostavitev rezervacij v omrežju potrebujemo nek kontrolni ali signalizacijski protokol. V RFC priporočilih tak protokol ni specificiran, vendar se v omrežjih z IS skoraj izključno uporablja protokol RSVP, ki je podrobno opisan v [66]-[71], [77], [80] in [110].

Potrebno je poudariti, da sta IS in RSVP komplementarni, a neodvisni tehnologiji. IS namreč privzamejo, da je omrežje sposobno zagotoviti zahtevano stopnjo kakovosti

---

<sup>5</sup>Primer podatkovnega pretoka so, na primer, vsi podatki z enakim izvornim in ponornim IP naslovom

storitve. Z uporabo RSVP pa to isto omrežje pripravimo za prenos podatkov preko zahtevanega razreda IS storitve. Namesto RSVP bi lahko uporabili tudi kakšen drug rezervacijski protokol.

### 3.2.5 Osnovne lastnosti in delovanje omrežij z DS

DS so nastale nekoliko kasneje kot IS, ko se je izkazalo, da slednje sicer zadovoljivo rešujejo diferenciacijo storitev na Internetu, vendar pa so (pre)zahtevne za izvedbo. Vzdrževanje stanj za večtisoč do večstotisoč hkratnih pretokov v hrbteničnih usmerjevalnikih namreč ni preprosta naloga.

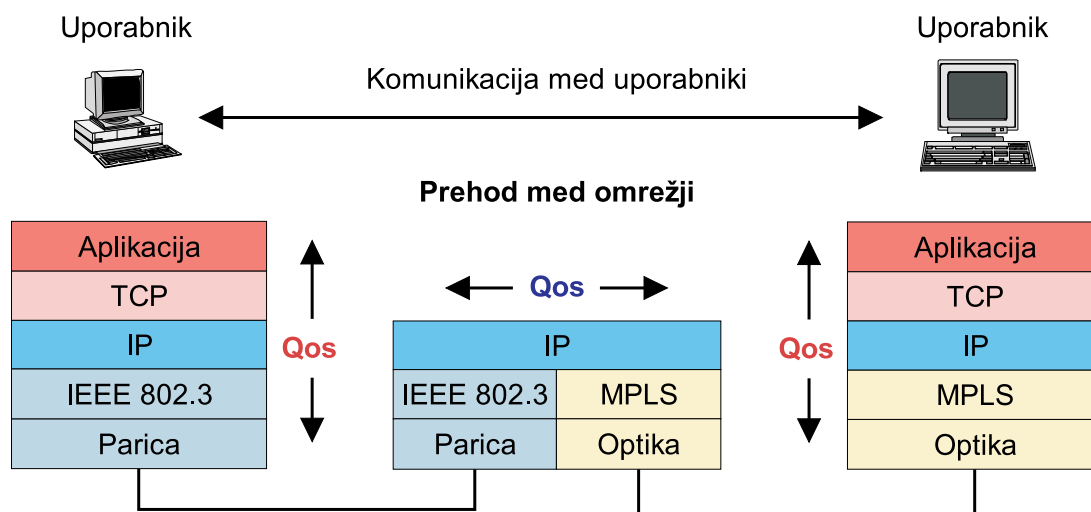
DS se problema lotijo preko diferenciacije na osnovi storitvenih razredov. Tu informacijo o zahtevani stopnji kakovosti storitve nosi vsak posamezen paket v svoji glavi. Za to uporabimo že obstoječa polja v glavah protokolov IPv4 in IPv6, ki se jim spremeni pomen, tako da določajo stopnjo kakovosti storitve (glej razdelek 3.2.2). Na ta način sicer izgubimo možnost zagotavljanja trdnih prenosnih parametrov, kot je na primer zakasnitev, in na plano pridejo vse prednosti in slabosti diferenciacije na osnovi storitvenih razredov. Pridobimo pa na enostavnosti izvedbe, saj usmerjevalnikom na poti ni več potrebno vzdrževati podatkov mnogoštevilnih pretokov, odpade pa tudi potreba po zapleteni signalizaciji za vzpostavitev in vzdrževanje povezave. Usmerjevalniki prispele pakete, glede na informacije zapisane v njihovih glavah, razporedijo v ustrezen razred, ki je deležen vnaprej določene stopnje kakovosti storitve. Trdnih prenosnih zagotovil pri DS ni. Stopnja kakovosti storitve posameznega storitvenega razreda pa je relativna glede na ostale razrede. Več o DS najdemo v [72]-[75] in [84]-[85].

## 3.3 Kakovost storitve med končnima točkama komunikacije

Doslej smo spoznali možnosti nudenja različnih stopenj kakovosti storitve v posameznih paketnih omrežjih. Navedene stopnje kakovosti storitve veljajo le, če celotna prenosna pot med končnima točkama komunikacije poteka v istem omrežju. To je pogosto res, v splošnem pa komunikacija poteka preko več omrežij, ki za prenos podatkov uporabljajo različne tehnologije in protokole.

Poudariti moramo, da je za zagotavljanje določene stopnje kakovosti storitve na celotni prenosni poti potrebno zagotoviti sodelovanje prav vseh elementov na njej. To med drugim vključuje ustrezno programsko opremo v končnih napravah, ki prepozna zahtevo po kakovosti storitve in jo ustrezno preda protokolnemu skladu ter protokole, prenosne tehnologije in omrežne naprave, ki omogočajo označevanje, ločevanje in prenos podatkov z različnimi stopnjami kakovosti storitve.

Stopnja kakovosti storitve je navadno določena z naborom parametrov kakovosti storitve in njihovimi vrednostmi. Če želimo na celotni prenosni poti zadržati enako stopnjo kakovosti storitve, potem morajo vsi protokoli in tehnologije na tej prenosni poti parametre kakovosti storitve obravnavati in izpolnjevati na enak način. Če je nabor protokolov in tehnologij enak na celotni prenosni poti, potem se morajo parametri kakovosti storitve pravilno preslikati samo med protokolnimi plastmi, če pa je prenosna pot sestavljena iz več omrežij z različnimi prenosnimi tehnologijami, pa se morajo parametri kakovosti storitve pravilno preslikati tudi med temi omrežji.



Slika 3.3 – Prenos parametrov kakovosti storitve med končnima točkama komunikacije

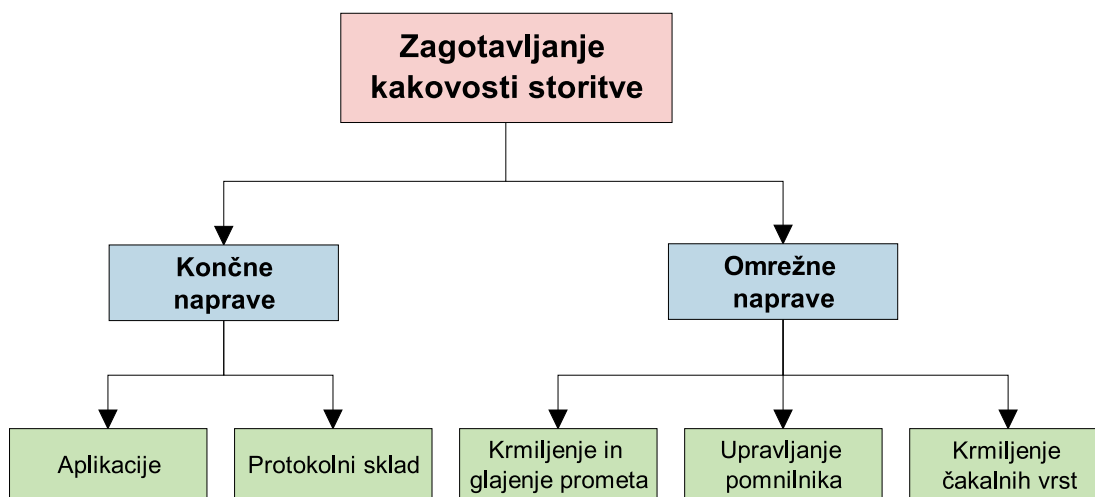
Preslikavo parametrov kakovosti storitve na celotni prenosni poti najlažje ponazorimo s preprostim primerom na sliki 3.3. Komunikacija med dvema uporabnikoma (končnima točkama) poteka po prenosni poti, ki na omrežni plasti uporablja protokol IP, na njej pa podatke prenašamo s pomočjo dveh različnih prenosnih tehnologij. V našem primeru uporabniška aplikacija zahteva določeno stopnjo kakovosti storitve, ki se na IP plasti preslika v ustrezno vrednost polja ToS ali DSCP v glavi paketa IP. To polje se ob predaji paketa IEEE 802.3 plasti preslika v ustrezno kombinacijo 802.1p bitov, ki določajo prednostno obravnavo Ethernet okvirov. Po prenosu paketa IP do roba omrežja IEEE 802.3 se izvrši obratna preslikava, nato pa se mora vrednost iz polja ToS ali DSCP v glavi paketa IP preslikati v ustrezno kombinacijo bitov eksperimentalnega polja MPLS okvira, ki določa prednostno obravnavo MPLS okvirov. Na sprejemni strani se izvede obratna preslikava v polje ToS ali DSCP. Uporabniška aplikacija na drugi strani mora upoštevati vrednost polja ToS ali DSCP paketa IP in ustrezno ukrepati. Poudarimo naj, da je to zelo poenostavljen primer, s katerim želimo predstaviti načelo preslikave določene stopnje kakovosti storitve skozi protokolne sklade in različna prenosna omrežja na celotni prenosni poti. V resnici so te preslikave lahko precej bolj zapletene in na žalost niso enoumne. Različne prenosne tehnologije namreč uporabljajo različno število stopenj kakovosti storitve in tudi te največkrat med seboj niso primerljive. Na primer, če želimo preslikati kakovost storitve določeno z eno izmed 64 možnih vrednosti polja DSCP paketa IP, v eksperimentalno polje MPLS okvira z 8 možnimi vrednostmi, moramo skleniti kompromis in združevati stopnje kakovosti storitve določene v polju DSCP.

## 4 Mehanizmi zagotavljanja kakovosti storitve v paketnih omrežjih

Z razvojem paketnih omrežij, novih aplikacij, prehajanjem nekaterih storitev iz tradicionalnih na paketna omrežja, se porajajo nove potrebe in zahteve, ki jih morajo izpolnjevati sodobna paketna omrežja. Ena najpomembnejših zahtev je zagotavljanje različnih stopenj kakovosti storitve.

Če želimo doseči zadovoljstvo uporabnika po predlagani definiciji kakovosti storitve (glej razdelek 2.1), moramo zagotoviti ustrezne prenosne parametre na celotni prenosni poti (glej razdelek 3.3). Glede na sliko 3.3 lahko zagotavljanje kakovosti storitve razdelimo v dva sklopa, kar prikazuje slika 4.1.

Prvi sklop je zagotavljanje kakovosti storitve v končnih napravah. Pri tem je pomembno predvsem, da zagotovimo ustrezne končne naprave, pravilno delovanje aplikacij in ustrezen protokolni sklad. Slednji mora vsebovati ustrezne protokole, podpirati zagotavljanje zahtevanih prenosnih parametrov skozi celoten protokolni sklad, njihove preslikave med protokoli in podpirati tudi vse druge funkcije, potrebne za zagotavljanje ustrezne stopnje kakovosti storitve.



Slika 4.1 – Model zagotavljanja kakovosti storitve

Drugi sklop zadeva omrežne naprave, ki obvladujejo manjši del protokolnega sklada in imajo le na prvi pogled lažjo nalogo kot končne naprave. Medtem ko morajo končne naprave hkrati obvladovati le prometni pretok ene ali kvečjemu nekaj aplikacij in/ali povezav, se omrežne naprave spopadajo z nekaj sto do nekaj milijoni hkratnih povezav in njihovimi prometnimi pretoki. Zato je v omrežnih napravah obvladovanje in

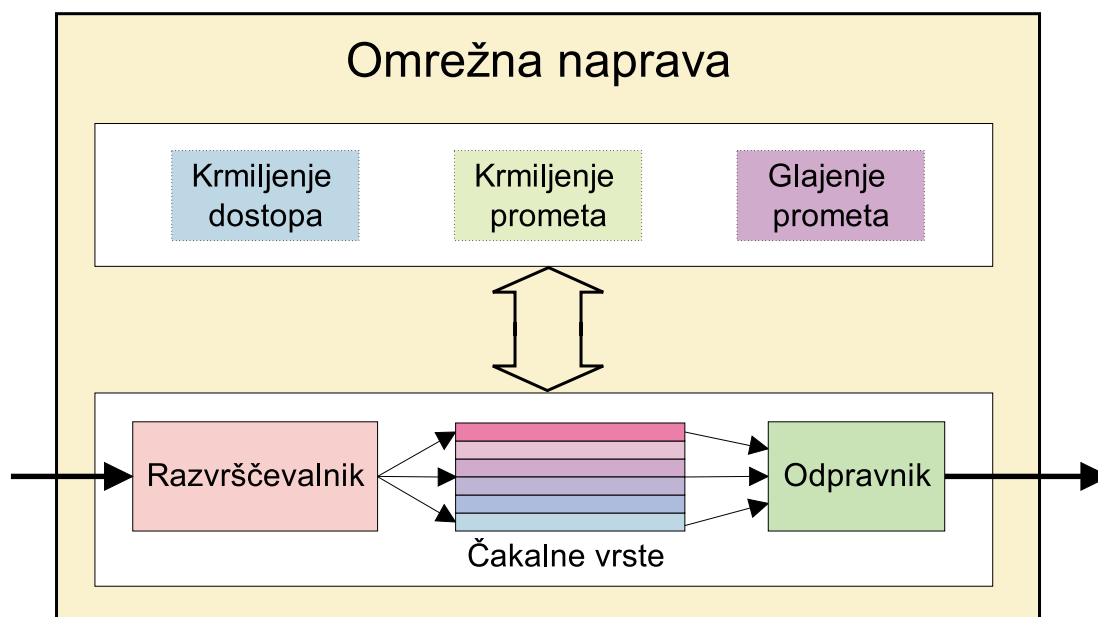
delitev omrežnih virov izredno pomembna naloga. Za zagotavljanje kakovosti storitev je pomembno: krmiljenje in glajenje prometa, krmiljenje čakalnih vrst ter upravljanje pomnilnika. Te naloge podrobneje predstavimo v nadaljevanju.

Končne naprave so izredno raznolike in navadno pod nadzorom uporabnika, omrežne naprave pa so pod nadzorom operaterjev omrežij in precej manj raznolike. Ker za večino končnih naprav velja, da nimajo težav z razpoložljivimi viri, zagotavljanje kakovosti storitve s tega stališča ne bi smelo biti problematično. Zato se v nadaljevanju posvečamo predvsem omrežnim napravam ter si ogledamo osnovna načela in mehanizme deljenja omrežnih virov za zagotavljanje različnih stopenj kakovosti storitve.

#### 4.1 Model omrežne naprave

Na sliki 4.2 je predstavljen poenostavljen model omrežne naprave z elementi, ki so pomembni pri zagotavljanju različnih stopenj kakovosti storitve.

Ko paketi prispejo v omrežno napravo, jih razvrščevalnik razporedi v ustrezne čakalne vrste, kjer počakajo, da bodo odpravljeni. Razvrščanje, čakanje in odpravljanje se izvajajo v sodelovanju z elementi za krmiljenje dostopa (ob vzpostavljanju povezave ali spreminjanju njenih lastnosti) ter krmiljenje in glajenje prometa (med prenosom).



Slika 4.2 – Poenostavljen model omrežne naprave

Preden se lotimo podrobnejše razlage delovanja omrežne naprave, definirajmo nekaj osnovnih pojmov.

**Omrežna naprava** (network device) je vsaka naprava na prenosni poti, ki posreduje pakete in po možnosti vrši funkcije zagotavljanja kakovosti storitve.

**Podatkovni pretok** (flow) je skupina paketov z eno ali več enakimi lastnostmi. Podatkovni pretok največkrat prestavlja skupina paketov, ki imajo isto izvorno in ponorno točko (pripadajo isti povezavi, vzpostavljeni med dvema končnima

napravama v omrežju). Podatkovni pretok so lahko tudi paketi, ki imajo enako vrednost nekega drugega polja v svojih glavah (npr. enako DSCP vrednost) in zato pripadajo istemu storitvenemu razredu.

**Krmiljenje dostopa** (admission control) je funkcija omrežne naprave, ki novemu podatkovnemu pretoku dovoli ali zavrne dostop do zahtevanih omrežnih virov. Enako velja pri spreminjanju lastnosti pretoka. Krmiljenje dostopa je funkcija, ki se izvaja nad celotnim podatkovnim pretokom.

**Krmiljenje prometa** (traffic policing) je funkcija omrežne naprave, ki se izvaja nad posameznimi paketi in skrbi za to, da se promet podatkovnih pretokov zadržuje v dogovorjenih mejah (je skladen z dogovorjenimi prometnimi specifikacijami). Na primer, če podatkovni pretok preseže maksimalno dogovorjeno bitno hitrost, lahko omrežna naprava presežne pakete zavrže ali jim dodeli slabšo kakovost storitve od dogovorjene. Funkcija krmiljenja prometa lahko opravlja tudi funkcijo označevanja paketov (marking) in funkcijo merjenja (metering). Navadno se izvaja le na vstopnih točkah v omrežje, redkeje pa tudi v vsaki omrežni napravi na prenosni poti.

**Glajenje prometa** (traffic shaping) je funkcija, ki je povezana s funkcijo krmiljenja prometa in poskrbi, da podatkovni pretoki ostajajo v dogovorjenih mejah. Na primer, če podatkovni pretok trenutno preseže maksimalno dogovorjeno bitno hitrost, se presežni paketi začasno shranijo v medpomnilniku gladilnika, iz katerega se odpošiljajo s hitrostjo, ki je še v dogovorjenih mejah. Na ta način se seveda lahko odpravijo le kratkotrajne kršitve prometnega pretoka. Tudi glajenje prometa se, zaradi velikih potreb po omrežnih virih (predvsem pomnilniku omrežne naprave) navadno izvaja le na vstopnih točkah v omrežje, lahko pa tudi v vsaki omrežni napravi na prenosni poti. Funkcija glajenja prometa lahko opravlja tudi funkcijo odmetavanja presežnih paketov (dropping).

**Razvrščevalnik** (classifier) skrbi za pravilno razvrščanje pripuščenih<sup>1</sup> paketov v ustrežne čakalne vrste. Razvrščanje je v večini primerov preprosta in z viri nepotratna operacija. Razvrščanje se navadno vrši na podlagi informacij zapisanih v glavi paketa: polje DSCP, izvorni ali ponorni naslov in podobno.

**Čakalne vrste** (queues) v omrežnih napravah hranijo pakete, ki čakajo, da pridejo na vrsto za posredovanje. Za zagotavljanje več stopenj kakovosti storitve potrebujemo več (logičnih) čakalnih vrst. V vsaki izmed njih se, na primer, nahajajo paketi, ki pripadajo istemu storitvenemu razredu. V skrajnem primeru ima lahko vsak posamezen podatkovni pretok svojo čakalno vrsto.

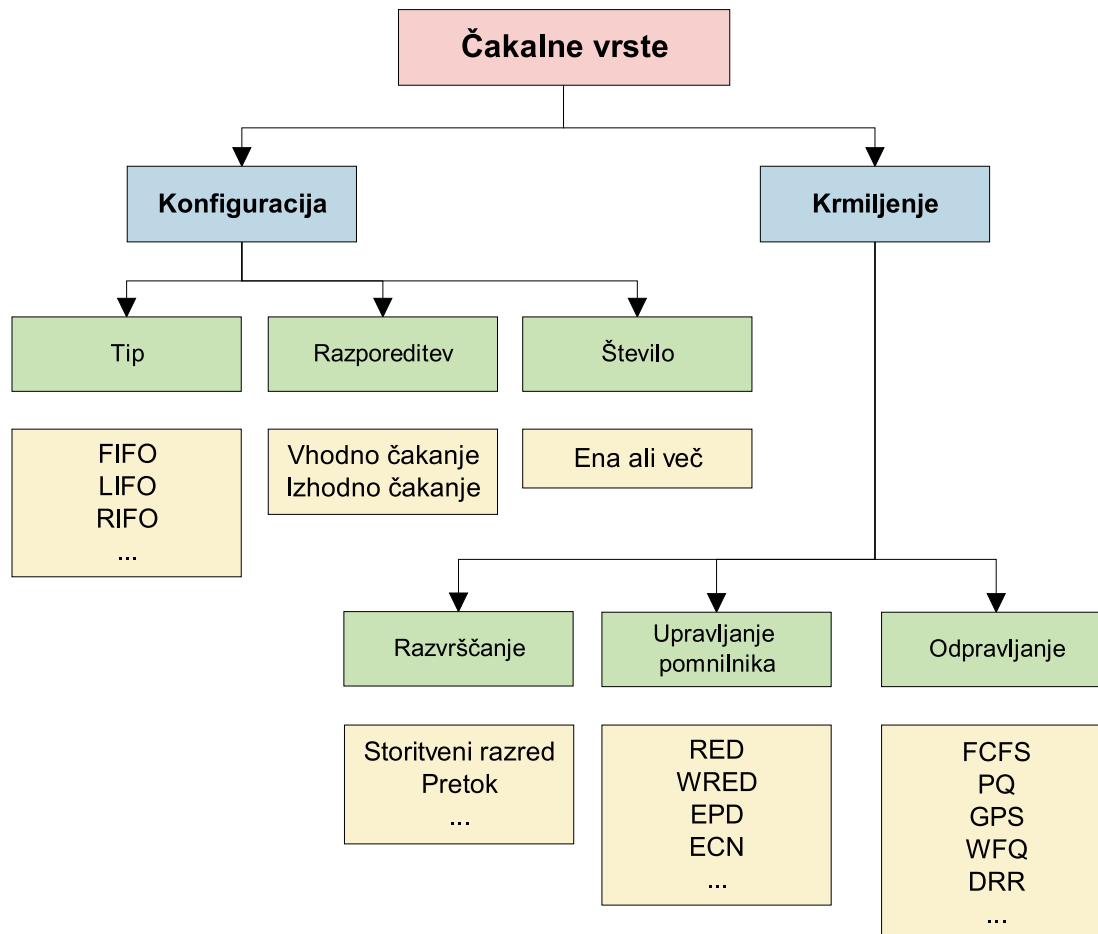
**Odpravnik** (packet scheduler) skrbi za pravilen vrstni red posredovanja čakajočih paketov. Delovanje odpravnika (odpravljanje paketov) in algoritem izbire čakalne vrste (strežna disciplina), iz katere bo posredovan naslednji paket, sta ključna za zagotavljanje različnih stopenj kakovosti storitve.

Funkcije krmiljenja dostopa ter krmiljenja in glajenja pretokov poskrbijo, da so podatkovni pretoki skladni z dogovorjenimi prometnimi specifikacijami. To omrežni napravi omogoča, da s pravilno uporabo razvrščanja in odpravljanja paketov zagotovi dogovorjeno kakovost storitve za vsak storitveni razred ali pretok. Zato si podrobneje pogledjmo možnosti, ki jih imamo pri konfiguraciji in krmiljenju čakalnih vrst.

<sup>1</sup>Paketov, ki so bili znotraj dogovorjenih mej ali paketov, ki so prišli iz gladilnika.

## 4.2 Konfiguracija in krmiljenje čakalnih vrst

Za zagotavljanje različnih stopenj kakovosti storitve v omrežnih napravah sta ključnega pomena konfiguracija in krmiljenje čakalnih vrst. Konfiguracija predstavlja tip, razporeditev in število čakalnih vrst v omrežni napravi. Krmiljenje pa predstavlja razvrščanje, upravljanje pomnilnika in način odpravljanja. Oboje je prikazano na sliki 4.3.



Slika 4.3 – Možnosti konfiguracije in krmiljenja čakalnih vrst

### 4.2.1 Konfiguracija čakalnih vrst

Najbolj pogosta je čakalna vrsta **tipa** FIFO (First In First Out - prvi noter prvi ven), pri kateri se novoprispeli paketi vstavijo na konec čakalne vrste, odpravljajo pa se z njenega začetka. Tako pri odpravljanju zagotovimo vrstni red paketov glede na čas njihovega prihoda v omrežno napravo. Obstajajo pa tudi čakalne vrste tipov LIFO (Last In First Out - zadnji noter prvi ven), LIRO (Last In Random Out - zadnji noter naključno ven), RIRO (Random In Random Out - naključno noter naključno ven) in druge [16]. V tem delu obravnavamo izključno čakalne vrste tipa FIFO.

Pri **razporeditvi** čakalnih vrst imamo dve osnovni možnosti. Čakalne vrste so lahko na vhodu v omrežno napravo ali na izhodu iz nje. Prvim rečemo vhodne čakalne



vrste, drugim pa izhodne čakalne vrste. Vsaka izmed razporeditev ima svoje prednosti in slabosti, ki jih tu ne naštevamo posebej, opisane pa so v [16]. Po analizi obeh možnosti se izkaže, da so za zagotavljanje različnih stopenj kakovosti storitve bolj primerne izhodne čakalne vrste [16]. V tem delu obravnavamo izključno izhodne čakalne vrste.

Z eno samo čakalno vrsto za vse pakete ne moremo zagotoviti različnih stopenj kakovosti storitve za različne skupine paketov. Zato se v omrežnih napravah uporablja **več čakalnih vrst**, v katere razvrščamo novoprispele pakete. Te so večinoma realizirane kot logične čakalne vrste. To pomeni, da vse čakalne vrste uporabljajo isti fizični pomnilnik, ki pa je razdeljen na logične enote, ki so na voljo posameznim čakalnim vrstam. Paketi se iz čakalnih vrst odpravljajo na način, ki za vsako izmed njih zagotavlja različno stopnjo kakovosti storitve, kar pa je odvisno od načina krmiljenja teh čakalnih vrst.

#### 4.2.2 Krmiljenje čakalnih vrst

Naloga razvrščevalnika je, da pripuščene pakete razporedi v ustrezne čakalne vrste. **Razvrščanje** poteka na več načinov, odvisno od načina zagotavljanja kakovosti storitve, števila čakalnih vrst in podobno.

Pakete se navadno razvršča v čakalne vrste glede na storitveni razred, ki mu pripadajo paketi. Recimo, da je v omrežni napravi  $N$  čakalnih vrst, po ena za vsakega izmed  $N$  storitvenih razredov. Ko razvrščevalnik prevzame paket, najprej ugotovi, kateremu izmed storitvenih razredov pripada in nato paket razvrsti v ustrezno čakalno vrsto. Storitveni razred je največkrat določen z vrednostjo določenega polja v glavi paketa (npr. DSCP pri paketih IP) ali pripadnostjo določenemu podatkovnemu pretoku. *Storitveni razred* je v tem kontekstu splošen pojem, ki ga uporabljamo za razlikovanje skupin paketov, ki so deležni enake stopnje kakovosti storitve. Stopnja kakovosti storitve se načeloma razlikuje za vsakega izmed  $N$  storitvenih razredov.

Kljub možni uporabi krmiljenja dostopa in prometa pa zaradi statističnega multipleksiranja v paketnih omrežjih lahko pride do zasičenja. To se zgodi, ko se zaradi presežnega prometa napolnijo čakalne vrste<sup>2</sup>. To novoprispelim paketom onemogoča uvrstitev vanje in so zato zavrženi. Zasičenje ima na delovanje omrežja več negativnih vplivov [90]:

- odmetavanje (zavračanje) paketov je tratenje omrežnih virov na prenosni poti do njihove zavrnitve (točke zasičenja),
- ob zasičenju se zaradi podaljšanja povprečne zasedenosti čakalnih vrst poveča tudi povprečna zakasnitev paketov,
- dalj časa trajajoče zasičenje lahko povzroči prekinitve povezave ali delovanja aplikacij,
- zasičenje v omrežjih IP lahko povzroči fenomen poznan pod imenom "globalna TCP sinhronizacija", ko vsi TCP izvori, ki jih prizadene zasičenje, ob približno istem času zmanjšajo oddajno hitrost in jo potem vsi naenkrat enakomerno

---

<sup>2</sup>Čakalne vrste se začnejo polniti, ko paketi v omrežno napravo pritekajo hitreje kot iz nje odtekajo. Če take razmere trajajo dalj časa, se čakalne vrste zapolnijo in ne morejo sprejemati novih paketov - tako stanje označujemo kot zasičenje.

povečujejo. To lahko vodi do nezaželenih oscilacij v zasedenosti omrežja, ko se izmenjujejo obdobja zasičenja in obdobja nizke zasedenosti omrežja.

Osnovnemu načinu **upravljanja pomnilnika** (čakalnih vrst) rečemo *neposredno zavračanje* (tail drop). To je v bistvu pasivni način upravljanja pomnilnika, saj omrežna naprava začne zavračati novoprispele pakete šele, ko se čakalne vrste napolnijo - se že nahajajo v stanju zasičenja. Zato ima ta način vse zgoraj naštetih slabosti.

Če se želimo zasičenju izogniti ali ga vsaj omiliti, moramo uporabiti aktivne načine upravljanja pomnilnika. Ti temeljijo na dveh osnovnih tehnikah: odmetavanju paketov iz čakalne vrste še preden se ta napolni in označevanju paketov.

Najpogostejša tehnika odmetavanja paketov v omrežjih IP je Random Early Detection (RED - naključno zgodnje odmetavanje paketov). Ta pakete zavrača z določeno stopnjo verjetnosti, ki je odvisna od zapoljenosti čakalne vrste. Tako naključno zavrača pakete in na ta način posredno sporoča podatkovnim izvorom, naj zmanjšajo hitrost oddajanja. Z naključnim odmetavanjem paketov se, med drugim, izognemo globalni TCP sinhronizaciji. Ta tehnika deluje samo za TCP in SCTP (Stream Control Transmission Protocol - protokol za krmiljenje prenosa pretokov) promet, na ostale vrste prometa, kot je na primer UDP (User Datagram Protocol - uporabniški datagramski protokol), nima vpliva. Naprednejše različice opisanega mehanizma so FRED (Fair Random Early Detection - pravično naključno zgodnje odmetavanje paketov) in WRED (Weighted Random Early Detection - uteženo naključno zgodnje odmetavanje paketov) [12]. Za druge prenosne tehnologije obstajajo druge tehnike. V omrežjih ATM se tako uporabljata Partial Packet Discard (PPD - delno odmetavanje paketov) in Early Packet Discard (EPD - zgodnje odmetavanje paketov) [12].

Pri odmetavanju paketov podatkovnim izvorom posredno (implicitno) sporočamo, naj zmanjšajo hitrost oddajanja. Pri tem žal tratimo omrežne vire. Prikladnejši način je neposredno (eksplicitno) obveščanje z označevanjem paketov. Pri tem omrežna naprava nastavi določena polja v glavah paketov in na ta način drugim omrežnim ali končnim napravam na prenosni poti sporoča nevarnost zasičenja. V omrežjih Frame Relay se tako uporabljata bita FECN (Forward Explicit Congestion Notification - neposredno vnaprejšnje obvestilo o zamašitvi) in BECN (Backward Explicit Congestion Notification - neposredno povratno obvestilo o zamašitvi), v omrežjih IP pa je z dokumentom RFC 2481 definirano polje ECN (Explicit Congestion Notification).

Prednosti in slabosti posameznih načinov upravljanja pomnilnika so podrobneje opisane v [90] in [12].

**Odpravljanje** paketov je najpomembnejša funkcija omrežne naprave, s katero zagotovimo različne stopnje kakovosti storitve. Način in vrstni red odpravljanja paketov namreč določa stopnjo kakovosti storitve, ki jo prejmejo paketi uvrščeni v posamezne čakalne vrste. Odpravljanje je tesno povezano z ostalimi, že opisanimi funkcijami omrežne naprave.

Ustrezne pogoje za uspešno delovanje odpravnikova namreč zagotavljajo funkcije krmiljenja dostopa in prometa ter glajenja prometa, ki skrbijo, da so podatkovni pretoki, ki dosežejo čakalne vrste, skladni z dogovorom. Funkcija razvrščanja poskrbi, da se prispeli paketi uvrstijo v ustrezne čakalne vrste, funkcija upravljanja pomnilnika pa poskuša poskrbeti, da omrežna naprava ne pade v stanje zasičenja. Posredno za to skrbita tudi funkciji krmiljenja dostopa in prometa, ki načeloma v omrežno napravo

pripuščata samo toliko prometa, kot ga ta lahko posreduje. Ker se ti dve funkciji večinoma izvajata le ob vstopu prometa v omrežje (v robnih omrežnih napravah), znotraj omrežja (v jedrnih omrežnih napravah) še vedno lahko pride do zasičenja. Poleg tega pa navadno ves promet na omrežju ni podvržen tema funkcijama, kar zopet povečuje možnost zasičenja.

Naloga odpravnik je, da določi vrstni red odpošiljanja paketov iz čakalnih vrst. Tako ob koncu odpošiljanja vsakega paketa izbere aktivno čakalno vrsto<sup>3</sup>, iz katere bo odposlan naslednji paket. Ker obravnavamo FIFO čakalne vrste, bo to vedno tisti paket, ki se nahaja na začetku (head of the queue) izbrane čakalne vrste in je izmed vseh paketov v tej čakalni vrsti prvi vstopil vanjo.

Izbira naslednje čakalne vrste je odvisna od načina odpravljanja. V literaturi je predlaganih in opisanih veliko število odpravnikov. Ker je način odpravljanja tako pomemben za zagotavljanje različnih stopenj kakovosti storitve, mu namenimo podrobnejšo obravnavo.

### 4.3 Odpravljanje v paketnih omrežjih

V tem razdelku obravnavamo različne načine odpravljanja paketov v omrežnih napravah z izhodnim čakanjem. Posvečamo se predvsem njihovim možnostim zagotavljanja ustreznih prenosnih parametrov, ki jih zahteva prenos podatkov z različnimi stopnjami kakovosti storitve. Pri tem predpostavimo, da je omrežna naprava dovolj zmogljiva in prenos paketa preko stikalne matrike<sup>4</sup> ne vnaša dodatne zakasnitve.

#### 4.3.1 Želene lastnosti odpravnikov

Odpravniki naj bi zadovoljevali naslednje, včasih nasprotujoče si zahteve [47]:

- preprosta implementacija,
- pravičnost in zaščita pretokov,
- delovanje znotraj postavljenih okvirov,
- preprosto in učinkovito krmiljenje dostopa.

V omrežjih z visoko prenosno hitrostjo je pomembna **preprosta implementacija** odpravnik, da ga lahko realiziramo kar v strojni opremi naprave. Pomembno pa je tudi, da je število operacij za odpravo paketa čim bolj neodvisno od števila čakalnih vrst (pretokov). Zaželeno je uporaba odpravnik, ki ima zahtevnost  $O(1)$ . Tako je zagotovljena tudi skalabilnost, ki je pri odpravnikih z zahtevnostjo  $O(N)$  ali  $O(\log(N))$  vprašljiva. Več o zahtevnosti odpravnikov pa v razdelku 4.4.3.

**Pravičnost** je zaželeno predvsem med pretoki z enako prednostjo (kakovostjo storitve). Odpravnik je pravičen, če zadosti *max-min kriteriju*, ki je podrobno opisan v [47]. V grobem lahko rečemo, da po max-min kriteriju, pretoki z "nizkimi" zahtevami dobijo zelen delež omrežnih virov (npr. bitne hitrosti), medtem ko si pretoki z "visokimi zahtevami" enakomerno porazdelijo preostanek virov. **Pravičen odpravnik**

<sup>3</sup>Čakalna vrsta je aktivna, kadar so v njej paketi, ki čakajo na odpošiljanje.

<sup>4</sup>Stikalne matrike v naš poenostavljen model omrežne naprave nismo vključili. Stikalna matrika skrbi za prenos paketov iz vhodnih na izhodna vrata omrežne naprave [16].

zagotavlja tudi **zaščito pretokov**, kar pomeni, da neobvladan pretok<sup>5</sup> ne vpliva na storitve, ki jih prejema ostali (obvladani) pretoki. Več o zahtevnosti odpravnikov pa v razdelku 4.4.1.

Za **delovanje v mejah postavljenih okvirov** velja, da mora odpravnik operaterju omrežja omogočiti, da lahko le ta jamči za prenosne parametre posameznih pretokov. Omejitev postavlja ohranitveni zakon (glej 5.5.2.2), ki pravi, da posameznemu pretoku lahko izboljšamo prenosne parametre le na račun preostalih pretokov. Tako, na primer, ne moremo vsem pretokom zagotoviti zakasnitve, ki je nižja od zakasnitve FCFS (First Come First Serve - prvi prispe prvi postrežen) odpravnika. Najpogosteje uporabljeni prenosni parametri za razlikovanje kakovosti storitve so: bitna hitrost, zakasnitev, spremenljivost zakasnitve in izgube. Meje teh parametrov pa so lahko deterministične (veljajo za vsak paket posebej) ali statistične (veljajo za vse pakete skupaj).

Pri **krmiljenju dostopa** moramo paziti predvsem na to, da nov pretok, ki mu omogočimo dostop, ne ogrozi storitev, ki jo prejema že pripuščeni pretoki. Poleg tega pa naj to ne vodi v zasičenje ali prenizko izrabljenost omrežja.

#### 4.3.2 Osnovne možnosti izbire glede lastnosti odpravnikov

Pri izbiri in načrtovanju lastnosti odpravnika imamo več stopenj prostosti:

- število prednostnih razredov,
- delovno intenzivni in delovno umirjeni odpravniki,
- odpravniki z urejenimi prednostmi in časovnimi okviri,
- stopnja združevanja pretokov znotraj prednostnega razreda,
- razpored strežbe paketov znotraj prednostnega razreda.

S **prednostnimi razredi** na enostaven način diferenciramo promet na omrežju. Pretoki, ki pripadajo višjim prednostnim razredom, imajo na voljo boljše prenosne pogoje kot tisti iz nižjih prednostnih razredov, na primer nižjo povprečno zakasnitev. Število prednostnih razredov je poljubno, navadno pa jih je le nekaj. Že s tremi prednostnimi razredi lahko precej dobro diferenciramo storitve omrežja: najvišji razred za nujna sporočila (nadzor omrežja), srednji razred za promet z zagotovljenimi storitvami in nižji razred za navaden promet. Za implementacijo prednostnih razredov je najpreprostejša uporaba strogo prednostnega odpravljanja in njegovih variant.

**Delovno intenzivni** (work-conserving) odpravnik miruje samo takrat, kadar so vse čakalne vrste prazne. Nasprotno pa **delovno umirjeni** (non-work-conserving) odpravnik lahko miruje tudi, ko so v čakalnih vrstah paketi, ki čakajo na odpravo. Na prvi pogled se zdi, da delovno umirjeni odpravniki razmetavajo z omrežnimi viri<sup>6</sup>, saj kljub čakajočim paketom puščajo linijo prazno. Prednosti delovno umirjenih odpravnikov so v tem, da s svojim mirovanjem dosežejo bolj predvidljive prometne porazdelitve ter tako omogočijo zmanjšanje velikosti čakalnih vrst v omrežnih napravah in znižajo spremenljivost zakasnitve. Uporaba primerne vrste odpravnika je odvisna od omrežja in storitev, ki jih nudi. Več o prednostih in slabostih obeh vrst odpravnikov najdemo

<sup>5</sup>Pretok, ki pošilja pakete s hitrostjo, ki je večja kot njegov pravični delež

<sup>6</sup>To je v mnogih primerih res, lahko pa tak odpravnik v času "mirovanja" na linijo predaja pakete, ki imajo določeno storitev po najboljših zmožnostih (best-effort).

v [47]. V nadaljevanju se posvečamo predvsem delovno intenzivnim odpravnikom, saj se v obstoječih omrežnih napravah uporabljajo skoraj izključno te vrste odpravnikov.

Pri odpravnikih z **urejenimi prednostmi** (sorted-priority) je določena globalna spremenljivka, navadno se imenuje *navidezni čas* (virtual time). Njena vrednost se osveži ob vsakem prihodu ali odpravi paketa iz čakalne vrste. Vsakemu prispelemu paketu se določi časovni žig, ki je funkcija te globalne spremenljivke. Paketi v čakalni vrsti so urejeni po časovnih žigih in se v tem vrstnem redu tudi odpravljajo. V odpravnikih s **časovnimi okviri** (frame-based) je čas razdeljen na okvire stalne ali spremenljive dolžine. Pretoki si znotraj enega časovnega okvira rezervirajo časovno rezino za prenos svojih paketov. Pri odpravnikih s stalno velikostjo časovnega okvira se v primeru, da pretoki ne izrabijo svojih časovnih rezin v celoti, zgodi da odpravnik preostanek časa do izteka časovnega okvira miruje. Ti odpravniki so torej delovno umirjeni. Pri odpravnikih s spremenljivo dolžino časovnega okvira se to ne more zgoditi, saj se njegova dolžina sproti prilagaja prenosnim potrebam pretokov. Ti odpravniki so zatorej delovno intenzivni.

Četrta stopnja prostosti nam določa **stopnjo združevanja pretokov**. V eni skrajnosti vse pretoke združimo v en storitveni razred (service class) z eno spremenljivko stanja. Posledično prejemaajo vsi pretoki v omrežju enako stopnjo kakovosti storitve. V drugi skrajnosti vsak pretok uvrstimo v svoj storitveni razred in načeloma lahko vsak posamezen pretok prejema različno stopnjo kakovosti storitve. V vmesnem primeru združujemo pretoke v več storitvenih razredov, ki prejemaajo različne stopnje kakovosti storitve. Pri tem vsi pretoki združeni v posamezen storitveni razred prejemaajo enako stopnjo kakovosti storitve. Pri združevanju pretokov delamo kompromis med številom spremenljivk stanja in številom stopenj kakovosti storitve odpravnika.

Z **razporedom strežbe** določamo vrstni red odpravljanja paketov posameznih pretokov (združenih pretokov) znotraj prednostnega razreda. Na voljo imamo dve osnovni izbiri: strežba po vrstnem redu njihovih prihodov (FCFS) ali glede na njihovo storitveno oznako. Prva možnost ne omogoča diferenciacije storitve za pretoke (združene pretoke) znotraj prednostnega razreda in ni pravična, saj dodelitev deleža bitne hitrosti ne sledi max-min kriteriju. Druga možnost pa vključuje zahtevno izračunavanje storitvene oznake za vsak posamezen paket in njihovo uvrščanje na pravilno mesto v čakalne vrste. Vse to pa močno poveča zahtevnost implementacije čakalne vrste in poveča zahteve po računski moči omrežne naprave [47].

## 4.4 Glavne mere za oceno delovanja odpravnikov

Glede na različne zahteve za prenos podatkov lahko za ocenjevanje primernosti odpravnikov uporabimo različne mere. Najpomembnejše izmed njih so:

- pravičnost,
- zakasnitev algoritma,
- zahtevnost.

Dober odpravnik je pravičen, ima nizko algoritemsko zakasnitev in je nezahteven za izvedbo.

#### 4.4.1 Pravičnost

Zahteva po pravičnem dodeljevanju virov se pojavi, ko si več podatkovnih pretokov deli iste omrežne vire. V kratkih časovnih intervalih lahko med različnimi odpravniki nastanejo precejšnja odstopanja v stopnji storitve, ki jo nudijo posameznim podatkovnim pretokom. Na primer, nekateri odpravniki lahko kaznujejo podatkovne pretoke, ki so predhodno prejeli storitev iznad svojih rezervacij. Zato tem pretokom, do izenačitve njihove normalizirane storitve z normalizirano storitvijo ostalih pretokov, storitev odrečejo. To pa vodi v kratkoročno nepravičnost. Dva odpravnika imata tako lahko enake lastnosti glede povprečne zakasnitve, a zelo različne lastnosti glede pravičnosti.

Splošno sprejete metode za oceno pravičnosti odpravnikov ni. Definiramo jo na neformalen, intuitiven način. V splošnem želimo, da odpravnik streže podatkovne pretoke glede na njihovo rezervirano raven storitve (rezervirano bitno hitrost). Neizkoriščen del virov (bitne hitrosti), ki nastane zaradi trenutne neaktivnosti nekaterih izmed podatkovnih pretokov, pa sorazmerno razdeli med aktivne pretoke. Poleg tega naj podatkovni pretoki ne bi bili kaznovani za storitev prejeto iznad svojih rezervacij, če so bili v tem času nekateri izmed pretokov neaktivni. Vse druge formalnejše in strožje definicije odsevajo to intuitivno definicijo.

Glede na podano intuitivno definicijo pravičnosti je idealen kar odpravnik s *souporabo procesnih zmogljivosti*, opisan v razdelku 4.5.4. Delež bitne hitrosti, ki je rezervirana s strani posameznega podatkovnega pretoka  $i$ , je določen z realnim številom  $r_i$ . Predpostavimo, da je v časovnem intervalu  $(t_1, t_2)$  v odpravniku  $N$  aktivnih pretokov. S funkcijo  $SENT_i(t_1, t_2)$  podamo količino podatkov pretoka  $i$ , ki so odposlani v časovnem intervalu  $(t_1, t_2)$ . Glede na [99] lahko za časovni interval  $(t_1, t_2)$  zapišemo naslednjo neenačbo:

$$SENT_i(t_1, t_2) \geq \frac{r_i}{\sum_{j=1}^N r_j} (t_2 - t_1) R \quad (4.1)$$

kjer je  $R$  bitna hitrost izhodne linije,  $N$  število aktivnih pretokov,  $r_i$  in  $r_j$  pa rezervirani bitni hitrosti pretokov  $i$  in  $j$ . Najmanjša količina podatkov pretoka  $i$ , ki je odposlana v kateremkoli časovnem intervalu, pa je:

$$\frac{r_i}{\sum_{j=1}^V r_j} (t_2 - t_1) R \quad (4.2)$$

kjer je  $V$  največje število podatkovnih pretokov. Pri tem velja  $\sum_{j=1}^V r_j \leq R$ .

Odpravnik s *souporabo procesnih zmogljivosti* vsakemu aktivnemu podatkovnemu pretoku vedno zagotavlja prenos podatkov, ki je vsaj enak njegovi rezervirani bitni hitrosti. Na ta način je poskrbljeno tako za zaščito pretokov kot za popolno pravičnost.

Pravičnost odpravnikov v telekomunikacijskih omrežjih se navadno meri z eno izmed dveh mer. Prva mera za pravičnost temelji na maksimalni razliki med storitvijo, ki jo pretok prejme od merjenega odpravnika, in med storitvijo, ki bi jo prejel od idealnega odpravnika s *souporabo procesnih zmogljivosti*. Tej meri pravimo *absolutna mera pravičnosti* ali AFM (Absolute Fairness Measure). Čeprav štejemo, da je to

najprimernejša mera za pravičnost, pa jo je največkrat zelo težko določiti. Zato se uporablja alternativna mera, ki temelji na največji razliki med storitvama, ki ju prejmeta dva različna pretoka v merjenem odpravniku. Tej meri pravimo *relativna mera pravičnosti* ali RFM (Relative Fairness Measure). Natančnejše definicije obeh mer za pravičnost so podane v razdelku C.2 in v [115].

#### 4.4.2 Zakasnitev algoritma

Pomembna lastnost odpravnika je tudi *zakasnitev algoritma* (latency) pri odpravi prvega paketa določenega podatkovnega pretoka, ki prispe v prazno čakalno vrsto.

Za zagotavljanje različnih stopenj kakovosti storitve je med drugim potrebno zagotoviti tudi določljivo in navzgor omejeno zakasnitev na celotni prenosni poti. Zato morajo biti omejene tudi zakasnitve paketov v vseh omrežnih napravah na tej prenosni poti. Za doseg tega cilja pa morata biti izpolnjena dva ključna pogoja: omejena izbruhnost podatkovnih pretokov ter določljiva in navzgor omejena zakasnitev algoritma odpravnika.

Za omejitev izbruhnosti podatkovnih pretokov se najpogosteje uporablja oblikovanje prometa po principu *vedra z žetoni* (token bucket), ki je opisan v dodatku A, redko pa tudi po principu *puščajočega vedra* (leaky bucket). Pri tem upoštevamo, da je promet podatkovnega pretoka  $i$  zglajen s parametri  $(\sigma_i, \lambda_i)$ , kjer:

$\sigma_i$         označuje maksimalno velikost izbruha pretoka  $i$ ,  
 $\lambda_i$         označuje povprečno bitno hitrost pretoka  $i$ .

Zakasnitev algoritma odpravnika je del celotne zakasnitve v omrežni napravi. Intuitivno jo definiramo kot najdaljši možen čas, ki preteče od trenutka, ko prvi paket prispe v prazno čakalno vrsto, do trenutka, ko jo v celoti zapusti, vendar pa je primernejša definicija podana v [98]. V tem delu sta Stiliadis in Varma definirala splošno skupino odpravnikov, imenovanih Latency-Rate ( $\mathcal{LR}$ ) odpravniki. Njihovo delovanje s stališča določene čakalne vrste je odvisno le od dveh parametrov: algoritemske zakasnitve (latency) in dodeljene bitne hitrosti (rate). Pri tem je algoritemska zakasnitev definirana kot najdaljši možen čas, ki preteče od trenutka, ko prvi paket prispe v prazno čakalno vrsto, do trenutka, ko postane storitev te čakalne vrste enaka njeni dodeljeni bitni hitrosti.

Algoritemska zakasnitev posameznega odpravnika je odvisna od njegovih internih parametrov in delovanja, bitne hitrosti izhodne linije, števila čakalnih vrst (podatkovnih pretokov) ter njihovih rezerviranih bitnih hitrosti. Glede na to, da je algoritemska zakasnitev  $\mathcal{LR}$  odpravnika čas do trenutka, ko čakalna vrsta začne prejemati storitev, vsaj enako rezervirani, je ta čas neposredno povezan s potrebno dolžino čakalnih vrst in velikostjo sprejemnega medpomnilnika v napravah, ki predvajajo podatkovne pretoke v realnem času (real-time streaming). V najslabšem primeru morajo omrežne naprave hraniti pakete posameznega pretoka, dokler ne začne prejemati storitev enako rezervirani. Potrebno velikost čakalnih vrst ali sprejemnega medpomnilnika tako lahko izračunamo iz algoritemske zakasnitve in rezervirane bitne hitrosti pretoka.

Definicija  $\mathcal{LR}$  odpravnikov velja tako za zvezne odpravnike kot za paketne odpravnike. Glede na algoritmsko zakasnitev je zopet najprimernejši odpravnik s *souporabo*

procesnih zmogljivosti, opisan v razdelku 4.5.4. Njegova algoritemska zakasnitev je enaka 0. V nadaljevanju navedene rezultate povzemamo po [98] in [99].

Zanima nas tudi celotna zakasnitev v omrežni napravi. Če upoštevamo, da je promet pretoka  $i$  zglajen s pomočjo vedra z žetoni s parametri  $(\sigma_i, \lambda_i)$  in da je  $\lambda_i$  manjša od rezervirane bitne hitrosti  $r_i$ , potem je za  $\mathcal{LR}$  odpravnikove zakasnitev paketov pretoka  $i$ , ki jo označimo z  $d_i$ , omejena z izrazom [98]:

$$d_i \leq \frac{\sigma_i}{r_i} + \theta_i \quad (4.3)$$

kjer  $\theta_i$  označuje algoritemsko zakasnitev odpravnikovega podatkovnega pretoka  $i$ .

Ker definicija  $\mathcal{LR}$  odpravnikov ne temelji na kakršnikoli predpostavki glede vhodnega prometa, lahko izpeljemo tudi meje za vhodni promet, ki se ne pokorava modelu vedra z žetoni. Na primer, če poznamo povprečno in vršno bitno hitrost podatkovnega pretoka, lahko zapišemo izraz [98]:

$$d_i \leq \left( \frac{P_i - r_i}{r_i} \right) \left( \frac{\sigma_i}{P_i - \rho_i} \right) + \theta_i \quad (4.4)$$

kjer predstavlja  $\rho_i$  povprečno bitno hitrost pretoka in  $P_i$  njegovo vršno hitrost.

Za omrežje  $\mathcal{LR}$  odpravnikov in vedro z žetoni s parametri  $(\sigma_i, \lambda_i)$  velja izraz [98]:

$$D_i \leq \frac{\sigma_i}{r_i} + \sum_{j=1}^k \theta_i^j \quad (4.5)$$

kjer je  $k$  število zaporednih  $\mathcal{LR}$  odpravnikov (strežnikov) na prenosni poti pretoka  $i$  in  $\theta_i^j$  algoritemska zakasnitev pretoka  $i$  v  $j$ -tem zaporednem odpravniku. Za poznano povprečno in vršno bitno hitrost podatkovnega pretoka pa velja izraz [98]:

$$D_i \leq \left( \frac{P_i - r_i}{r_i} \right) \left( \frac{\sigma_i}{P_i - \rho_i} \right) + \sum_{j=1}^k \theta_i^j \quad (4.6)$$

Za zmanjšanje zakasnitve na celotni prenosni poti lahko naredimo dvoje: zmanjšamo izbruhnost pretoka  $\sigma_i$  ali algoritemsko zakasnitev odpravnikov  $\theta_i^j$  (z uporabo odpravnikov z manjšo algoritemsko zakasnitvijo). Slednji pristop je boljši, saj zagotovi zmanjšanje zakasnitve v vsakem odpravniku na prenosni poti.

Nekoliko natančnejši opis  $\mathcal{LR}$  odpravnikov je podan v razdelku C.3, podrobnosti in primerjave med različnimi odpravniki pa so navedene v [98] in tabeli 6.2.

### 4.4.3 Zahtevnost

Pod zahtevnost odpravnikovega štejemo: računsko zahtevnost, zahtevnost izvedbe in potrebo po virih (pomnilnik). Dober odpravnik je preprosto izvedljiv v strojni opremi (ne le programsko), za svoje delovanje ne potrebuje velike količine pomnilnika in je računsko nezahteven.



Za samo implementacijo odpravnika je najpomembnejša njegova računska zahtevnost. Ta je definirana kot maksimalni napor pri vstavljanju paketa v čakalno vrsto ali pri njegovem odpravljanju. Zaželeno je, da je računska zahtevnost odpravnika čim manjša in da je napor v čim manjši meri odvisen od števila čakalnih vrst ali števila hkratnih podatkovnih pretokov. To še posebej velja za hrbtencična omrežja, kjer je hitrost prenosa velika in ima odpravnik le malo časa za svoje odločitve.

Računsko zahtevnost označimo z  $O(x)$ , kjer je z  $x$  podana odvisnost števila operacij glede na število čakalnih vrst ali pretokov. Predpostavimo, da odpravnik streže  $N$  čakalnih vrst. Pri odpravniku z odvisnostjo  $O(N)$ , je število operacij premo sorazmerno s številom čakalnih vrst, pri odvisnosti  $O(\log(N))$  je ta odvisnost logaritemska, pri  $O(1)$  pa je število operacij konstantno ne glede na  $N$ . Za omrežne naprave manjših omrežij so primerni tudi odpravniki z  $O(N)$ , za omrežne naprave hrbtencičnih omrežij pa je zaželeno, da ima odpravnik konstantno računsko zahtevnost  $O(1)$ .

## 4.5 Odpravniki z izhodnim čakanjem

Kot smo že ugotovili, so za zagotavljanje različnih stopenj kakovosti storitve najbolj primerne naprave z izhodnim čakanjem. Zato se v nadaljevanju posvečamo odpravnikom z izhodnim čakanjem.

V literaturi je predlaganih in opisanih cela vrsta odpravnikov [1]-[4], [7]-[8], [11], [14], [15], [8], [42]-[46], [51]-[53], [55], [58]-[60], [88], [93], [96]-[102], [106], [111]-[115]. Navedeni pa še zdaleč ne predstavljajo celotne zbirke. Kot smo navedli že v razdelku 4.3.2, odpravnike delimo po več kriterijih, med drugim tudi glede na njihovo delavnost in notranjo strukturo [99]. Glede na delavnost jih delimo na: delovno intenzivne (work-conserving) in delovno umirjene (non-work-conserving). Glede na notranjo strukturo pa na: odpravnike z urejenimi prednostmi (sorted-priority) in odpravnike s časovnimi okviri (frame-based). Najbolj pomembni in za naše delo najrelevantnejši odpravniki so po svoji pripadnosti obema skupinama navedeni v tabeli 4.1.

Najbolj znani odpravniki z urejenimi prednostmi so: FCFS [89], Strict Priority [107], Generalized Processor Sharing (GPS - posplošena souporaba procesnih zmogljivosti) [58], Weighted Fair Queueing (WFQ - uteženo pravično odpravljanje) [112], Worst Case Fair Weighted Fair Queueing (WF<sup>2</sup>Q - uteženo pravično odpravljanje za najslabši primer) [112], Self-Clocked Fair Queueing (SCFQ - Pravično odpravljanje z lastnim proženjem) [11] in Virtual Clock (VC - navidezna ura) [112]. V skupini odpravnikov s stalno velikostjo časovnega okvira so med najbolj poznanimi Stop-and-Go, Jitter Earliest Due Date, Hierarchical Round Robin (HRR - hierarhično krožno odpravljanje) [112]. Pri odpravnikih s spremenljivo dolžino časovnega okvira pa so najbolj poznani odpravniki Round Robin<sup>7</sup> (RR - krožno odpravljanje), Weighted Bit-by-bit Round Robin (WBRR - uteženo bitno krožno odpravljanje) [89], Weighted Round Robin (WRR - uteženo krožno odpravljanje) [89] in Deficit Round Robin (DRR) [93].

V nadaljevanju opišemo tipične predstavnike določenih kombinacij lastnosti odpravnikov iz tabele 4.1. Še posebej se posvetimo opisu odpravnikov, ki jih v nadaljevanju uporabljamo pri simulacijah in izpeljavah (FCFS, Strict Priority, DRR, GPS). Podajamo tudi opise odpravnikov, s katerimi jih kasneje primerjamo. Delovanje in lastnosti

<sup>7</sup>Ponekod v literaturi ga imenujejo Fair Queueing

Odpravnik	Delavnost		Notranja urejenost	
	intenziven	umirjen	urejene prednosti	časovni okviri
FCFS	✓		✓	
Strict Priority	✓		✓	
GPS	✓			
WBRR	✓			✓
Virtual Clock	✓		✓	
WFQ	✓		✓	
SCFQ	✓		✓	
RR (Fair Queueing)	✓			✓
WRR	✓			✓
DRR	✓			✓
Stop-and-Go		✓		✓
J-EDD		✓	✓	
HRR		✓		✓

Tabela 4.1 – Uvrstitev odpravnikov glede na delavnost in notranjo urejenost

preostalih naštetih ali omenjenih odpravnikov pa so opisane v navedeni literaturi [1]-[4], [7]-[8], [11], [14], [15], [8], [42]-[46], [51]-[53], [55], [58]-[60], [88], [93], [96]-[102], [106], [111]-[115].

#### 4.5.1 FCFS

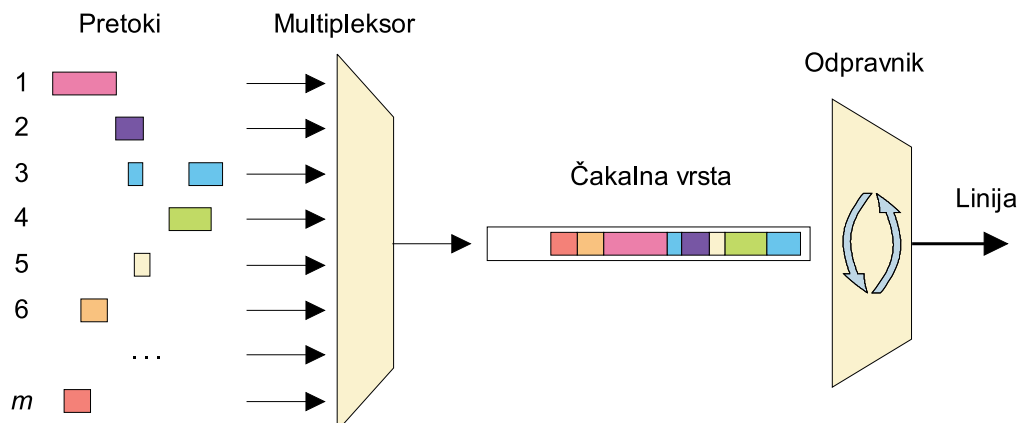
Odpravniki tipa FCFS<sup>8</sup> so najpreprostejši in najpogosteje uporabljeni odpravniki. So delovno intenzivni in delujejo po načelu urejenih prednosti. Vsak novoprispeli paket dobi časovni žig, ki ustreza dejanskemu času njegovega prihoda. Paketi se torej odpravljajo po enakem vrstnem redu, kot so prihajali v čakalno vrsto. So zelo preprosti in nezahtevni za implementacijo, njihova računsko zahtevnost je  $O(1)$ , imajo pa celo vrsto pomanjkljivosti: so nepravilni, ne nudijo zaščite pretokov<sup>9</sup>, ne ponujajo možnosti določitve (omejitve) največjih vrednosti zakasnitve in njene spremenljivosti. S stališča zagotavljanja različnih stopenj kakovosti storitve je ta odpravnik večinoma neuporaben, njegova velika prednost pa je seveda preprostost in nezahtevnost. Delovanje odpravnika FCFS je prikazano na sliki 4.4.

#### 4.5.2 Strogo prednostno odpravljanje

Strogo prednostno odpravljanje (Strict Priority Scheduling) pakete razvrsti v več prednostnih čakalnih vrst, znotraj katerih velja FIFO načelo. Na sliki 4.5 je prikazan primer s štirimi prednostnimi čakalnimi vrstami. Razvrščevalnik pakete posameznih pretokov razvrsti v ustrezne prednostne čakalne vrste. V našem primeru pretoka 3 in 4 pripadata najvišjemu prednostnemu razredu (4). Razvrščevalnik njune pakete uvršča v čakalno vrsto po časih njihovega prihoda. Enako velja za pretoka 1 in 5, ki pripadata

<sup>8</sup>Mnogokrat imenovani tudi FIFO.

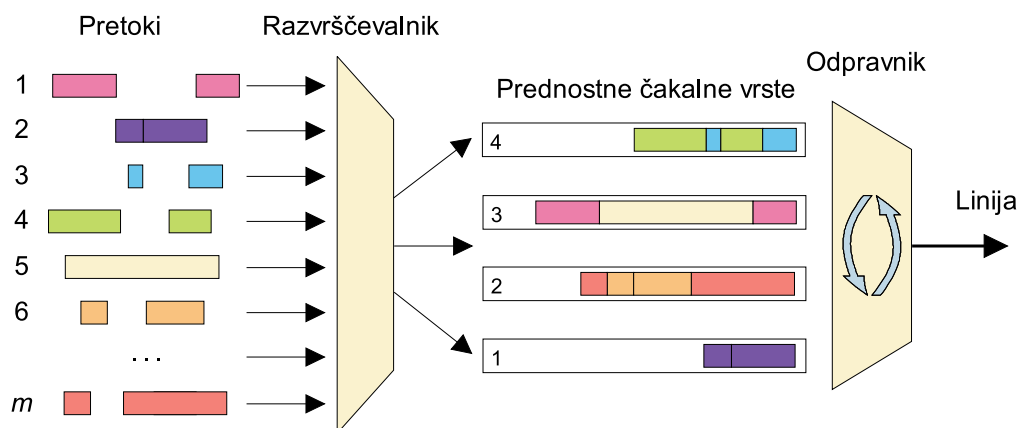
<sup>9</sup>En sam neobvladan pretok škodljivo vpliva na vse ostale pretoke.



Slika 4.4 – Odpravnik FCFS

prednostnemu razredu 3, pretokoma 6 in  $m$ , ki pripadata prednostnemu razredu 2 ter pretoku 2, ki pripada najnižjemu prednostnemu razredu (1).

Odpravnik streže najprej pakete iz neprazne čakalne vrste z najvišjo prednostjo, na primer  $p$ , ko je ta prazna, začne streči pakete v čakalni vrsti s prednostjo  $p - 1$ . Če v vmesnem času prispe paket v čakalno vrsto s prednostjo višjo kot  $p - 1$ , se strežba nadaljuje tam. Paketi v čakalni vrsti z določeno prednostjo so zato lahko postreženi šele, ko so vse čakalne vrste z višjimi prednostmi prazne.



Slika 4.5 – Odpravnik s strogimi prednostmi

Tako kot FCFS je tudi odpravnik s strogimi prednostmi enostaven in nezahteven za implementacijo. Njegova prednost je zmožnost zagotavljanja relativne stopnje kakovosti storitve. Višji prednostni razredi imajo boljše pogoje kot nižji, na primer nižjo zakasnitev. Ostajajo pa vse glavne pomanjkljivosti FCFS odpravnika. Ne nudi zaščite pretokov in je nepravilčen tako med prednostnimi razredi kot med pretoki znotraj prednostnega razreda. Neobvladan pretok znotraj enega izmed prednostnih razredov škodljivo vpliva tako na pretoke znotraj istega prednostnega razreda kot tudi na vse pretoke v nižjih prednostnih razredih. Prav tako pa zanj ne moremo določiti največjih vrednosti zakasnitve in njene spremenljivosti. V primeru, da je prometni pretok višjih

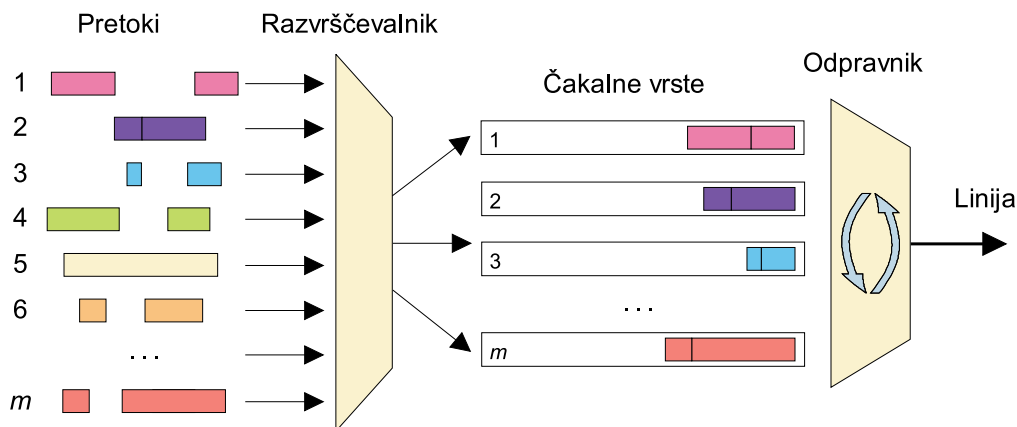
prednostnih razredov enak ali večji kot je kapaciteta povezave, pa nižji prednostni razredi lahko ostanejo nepostreženi.

### 4.5.3 Pravično odpravljanje

Pravično odpravljanje (FQ - Fair Queueing) ali krožno dodeljevanje (RR - Round Robin) [55] je osnova za vse odpravnike, ki želijo zagotoviti pravičnost med pretoki. To je delovno intenziven odpravnik, ki deluje po načelu časovnih okvirov. Pri pravičnem odpravljanju se vsakemu identificiranemu pretoku dodeli svoja čakalna vrsta. Paketi se krožno odpravljajo iz aktivnih čakalnih vrst. V vsakem obhodu<sup>10</sup> odpravnik odpošlje po en paket iz vsake aktivne čakalne vrste. Delovanje odpravnika je prikazano na sliki 4.6.

Ta odpravnik zagotavlja zaščito pretokov in pod pogojem, da so povprečne dolžine paketov vseh pretokov enake, tudi pravičnost. V primeru, ko določen pretok pošilja samo dolge in drug pretok samo kratke pakete, o pravičnosti ne moremo govoriti, saj pretok z dolgimi paketi zaseže precej večji delež bitne hitrosti povezave kot pretok s kratkimi paketi.

Slabosti odpravnika pa so predvsem te, da pretokom ni mogoče dodeliti različnih stopenj kakovosti storitve ali deležev prenosne hitrosti povezave. Zaradi tega ni preveč primeren za zagotavljanje različnih stopenj kakovosti storitve.



Slika 4.6 – Pravično odpravljanje

Različica pravičnega odpravljanja je uteženo krožno dodeljevanje (Weighted Round Robin - WRR), pri katerem se paketi iz čakalnih vrst odpravljajo glede na utež, ki pripada posamezni čakalni vrsti. Na primer, če imamo tri čakalne vrste z utežmi 1, 2 in 4, potem se v vsakem obhodu iz prve odpravi en paket, iz druge dva in iz tretje 4 paketi. WRR ima enake slabosti kot FQ ali RR. Ob določenih pogojih (vnaprej znane povprečne dolžine paketov) pa lahko na osnovi uteži pretokom zagotovi različne deleže bitne hitrosti in s tem pod določenimi pogoji tudi stopnje kakovosti storitve.

<sup>10</sup>Obhod je čas enega časovnega okvira in pomeni čas, potreben za odpravo po enega paketa iz vsake aktivne čakalne vrste.

#### 4.5.4 Souporaba procesnih zmogljivosti

Pri souporabi procesnih zmogljivosti (PS - Processor Sharing) odpravnik oddaja pakete iz vseh aktivnih čakalnih vrst naenkrat, vsakega z deležem hitrosti  $1/N$ , pri čemer  $N$  predstavlja število aktivnih čakalnih vrst. Posplošena souporaba procesnih zmogljivosti (GPS) je nadgradnja PS in omogoča uporabo uteženih deležev prenosne hitrosti za vsako izmed čakalnih vrst. Odpravnika sta zvezne narave, pri čemer se predpostavlja, da lahko odpravljamo neskončno majhne delčke paketov. Delež bitne hitrosti, ki je rezervirana s strani posameznega podatkovnega pretoka  $i$ , je tako določen z realnim številom  $r_i$ . Odpravnika lahko predstavimo tudi kot zvezno različico pravičnega odpravljanja in uteženega pravičnega odpravljanja, pri čemer se v vsakem obhodu odpravi neskončno majhen delček paketa. PS in GPS glede na notranjo urejenost ne moremo uvrstiti v nobeno izmed navedenih skupin, saj istočasno odpravljata pakete iz vseh aktivnih čakalnih vrst.

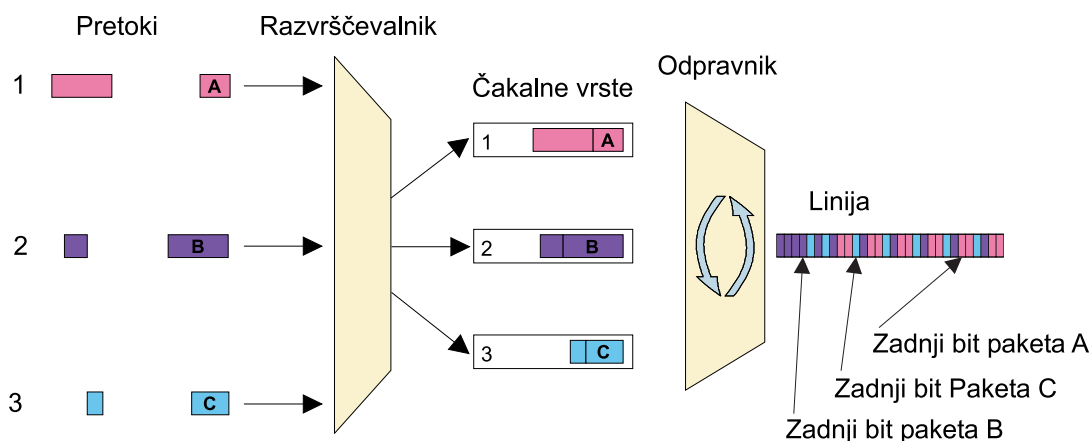
Prednost PS in GPS je v absolutni pravičnosti ter algoritemski zakasnitvi, ki je enaka nič. Odpravnika vsakemu aktivnemu podatkovnemu pretoku vedno zagotavljata prenos podatkov, ki je vsaj enak njegovi rezervirani bitni hitrosti. Poleg tega pa vso neizkoriščeno bitno hitrost izhodne linije porazdelita med vse aktivne pretoke v deležih, ki so sorazmerni njihovim rezerviranim vrednostim bitne hitrosti. Na ta način je poskrbljeno tako za zaščito pretokov kot za popolno pravičnost. Algoritemska zakasnitev obeh odpravnikov je enaka 0. Ker po definiciji v vsakem času strežeta vse čakalne vrste hkrati, se novoprispeli pretok začne streči v trenutku, ko prispe v omrežno napravo.

Žal pa sta ta dva odpravnika praktično neizvedljiva, saj paketov ne moremo deliti na neskončno majhne delčke. Podrobnejšo razlago tega odpravnika najdemo v [58], [59] in [112].

#### 4.5.5 Uteženo pravično odpravljanje

PS in GPS imata zelo privlačne lastnosti, žal pa nista izvedljiva v praksi. Odpravnik namreč lahko istočasno oddaja le en paket iz ene čakalne vrste. Najbližja aproksimacija PS in GPS sta BRR (Bit-by-bit Round Robin - bitno krožno odpravljanje) in WBRR, ki krožno strežeta aktivne čakalne vrste in iz njih odpravljata po en bit naenkrat (BRR) ali več bitov naenkrat (WBRR). BRR aproksimira PS, WBRR pa GPS. Primer delovanja WBRR odpravnika je prikazan na sliki 4.7. Pri tem ima pretok 1 utež 2, pretoka 2 in 3 pa utež 1. To pomeni, da pretok 1 lahko ob vsakem obhodu odda 2 bita, pretoka 2 in 3 pa po en bit. Odpravnika BRR in WBRR sta delovno intenzivna in delujeta po načelu časovnih okvirov.

Pri odpravnikih BRR in WBRR moramo na drugem koncu linije sprejete bite zopet sestaviti v pakete, kar vnaša dodatno delo in poveča kompleksnost odpravnika. Temu se izognemo z odpravnikoma WFQ (Weighted Fair Queueing) [8] in PGPS (Packet-by-packet GPS - paketna posplošena souporaba procesnih zmogljivosti) [58], ki sta bila razvita neodvisno, a sta po delovanju identična, zato v nadaljevanju omenjamo le prvega. WFQ odpravlja celotne pakete. Vrstni red odpravljanja paketov se določi glede na njihove izhodne čase, če bi le ti uporabljali GPS. Izhodni čas paketa predstavlja čas, ob katerem paket popolnoma zapusti čakalno vrsto. To lahko predstavimo na primeru s slike 4.7, kjer vidimo kdaj se odpošlje zadnji bit paketov A, B in C. Če



Slika 4.7 – Odpravnik WBRR

te pakete razporedimo po njihovih izhodnih časih, potem bi bil vrstni red oddajanja odpravnik WFQ A, C in B. Podrobnejša analiza izračuna izhodnih časov in vrstnega reda odpravljanja paketov po WFQ je navedena v [8], [58] in [47].

WFQ je odpravnik, ki je delovno intenziven in deluje po načelu urejenih prednosti, pri čemer so paketi urejeni glede na njihov izračunani izhodni čas. WFQ zagotavlja pravičnost in zaščito pretokov. Ob predpostavki, da so pretoki nadzorovani z mehanizmom *puščajočega vedra* (leaky bucket), pa imajo pretoki zagotovljeno tudi minimalno bitno hitrost in omejeno vrednost zakasnitve [58]-[59].

Slabost WFQ pa je predvsem v računski kompleksnosti, ki je potrebna za izračun izhodnih časov, obnavljanje navideznega časa ter za urejeno razvrščanje paketov po njihovih izhodnih časih. Iz WFQ je bila izpeljana cela vrsta odpravnikov. Med njimi sta najbolj znana SCFQ (Self Clocked Fair Queueing) [11] in WF<sup>2</sup>Q (Worst-case Fair WFQ) [4].

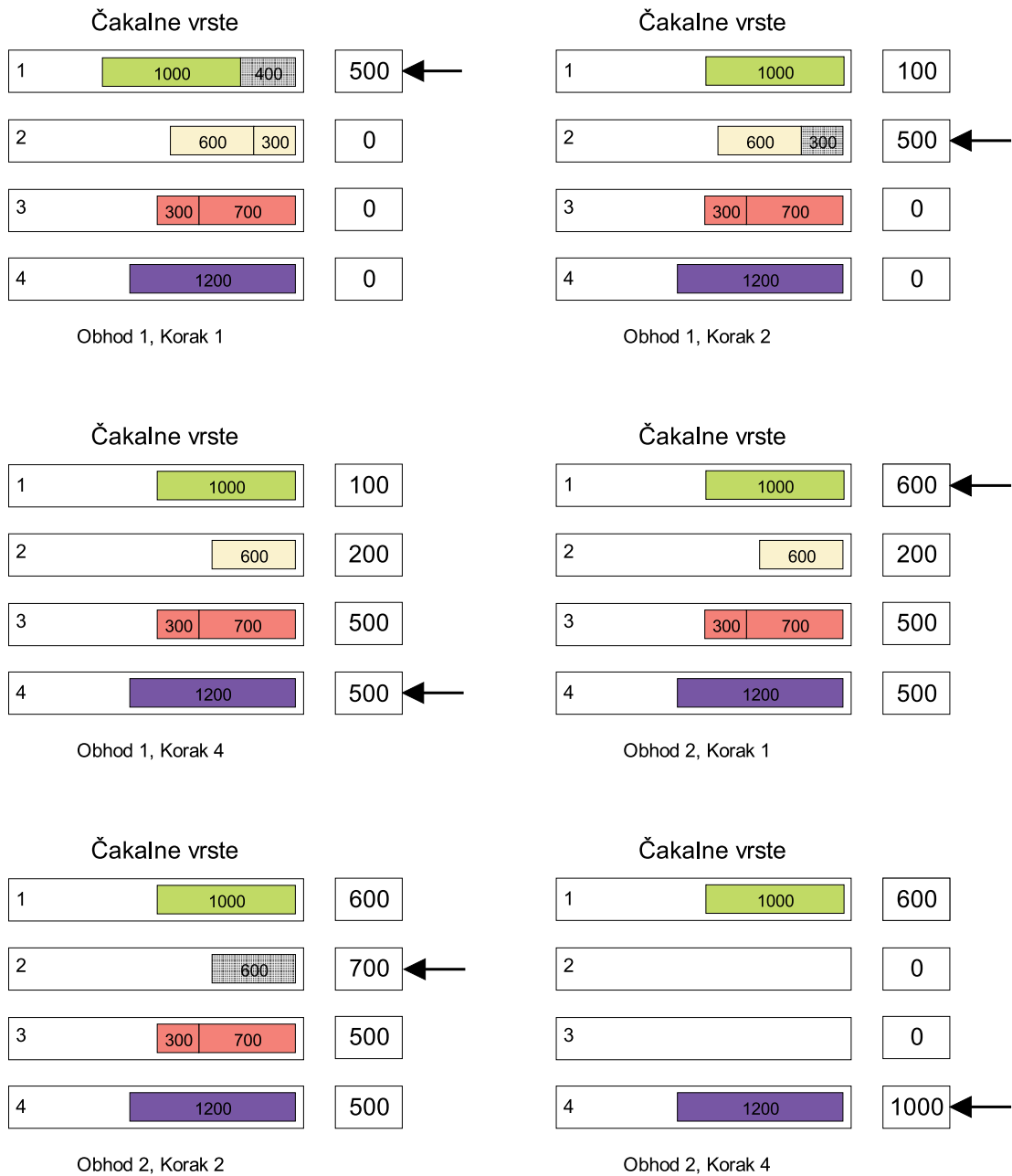
#### 4.5.6 Deficitarno krožno dodeljevanje

Odpravnik z deficitarnim krožnim dodeljevanjem (Deficit Round Robin - DRR) [93] je izpeljanka pravičnega odpravljanja. Odpravnik DRR je delovno intenziven in deluje po načelu časovnih okvirov. Odpravlja nekatere pomanjkljivosti RR in WRR odpravnikov, ki sta nepravična, če paketi niso enakih dolžin, ter pomanjkljivosti odpravnik WFQ, ki je računsko (pre)zahteven.

Najpomembnejša ideja DRR je, da v vsakem obhodu vsaki aktivni čakalni vrsti<sup>11</sup> dodeli določeno število bitov ali kvant. Kvant, ki pripada čakalni vrsti  $i$ , označimo s  $Q_i$ . Za spremljanje števila oddanih bitov vsaki čakalni vrsti priredimo spremenljivko stanja ali *števec primanjkljaja*, ki hrani število bitov, ki jih ima čakalna vrsta "v dobro". Delovanje odpravnik DRR najlaže predstavimo na primeru.

Na sliki 4.8 imamo štiri aktivne čakalne vrste. Ob začetku delovanja odpravnik DRR se v njih že nahajajo paketi, števci primanjkljaja pa so postavljeni na vrednost 0.

<sup>11</sup>Čas aktivnosti čakalne vrste je interval  $(t_0, t_1)$ , v katerem se v čakalni vrsti stalno nahaja vsaj en paket (backlogged queue). Čakalna vrsta preide iz pasivnega v aktivno stanje, ko ob času  $t_0$  prazna čakalna vrsta sprejme prvi paket. Čakalna vrsta preide iz aktivnega v pasivno stanje, ko jo ob času  $t_1$  v celoti zapusti zadnji čakajoči paket.



Slika 4.8 – Delovanje odpravnika DRR

V našem primeru so kvanti za vse čakalne vrste enaki in znašajo 500 bitov. Odpravnik začne svoj prvi obhod z obiskom čakalne vrste 1, kar označimo za korak 1. Najprej poveča njen števec primanjkljaja za  $Q$  ali 500 bitov. Če je prvi paket v čakalni vrsti manjši ali enak vrednosti števca, se ga odpravi in vrednost števca zmanjša za njegovo dolžino. Ta postopek se ponavlja, dokler so dolžine paketov na začetku čakalne vrste manjše ali enake preostali vrednosti števca. V našem primeru je bil iz prve čakalne vrste odpravljen samo en paket, vrednost števca primanjkljaja pa se je zmanjšala na 100. To se vidi pri drugem koraku na sliki 4.8. Tudi pri tem in vseh nadaljnjih korakih izvedemo enak postopek: povečamo števec primanjkljaja za  $Q$  in iz čakalne

vrste toliko časa odpravljamo pakete in zmanjšujemo vrednost števca, da je paket na začetku čakalne vrste daljši od preostale vrednosti števca. Tako smo v koraku 2 odpravili paket dolžine 300 bitov, preostala vrednost števca pa je 200 bitov. Pri koraku 3 (ni narisano) in koraku 4 prvega obhoda vidimo, da je paket na začetku čakalne vrste prevelik in mora zato počakati na zadostno povečanje števca primanjkljaja v naslednjih obhodih. Enako velja za prvo čakalno vrsto v prvem koraku drugega obhoda. V drugem koraku drugega obhoda odpravimo zadnji paket iz druge čakalne vrste. Čeprav je v števcu ostalo še 100 bitov, ga postavimo na 0, saj čakalna vrsta po izpraznitvi ni več aktivna in je odpravnik pri naslednjih obhodih ne bo več obiskal. Enako velja za tretjo čakalno vrsto v koraku 3 (ni na sliki). V koraku 4 dobi števec četrte čakalne vrste vrednost 1000, kar pa je še vedno premalo za odpravo paketa. Opisani postopek se sedaj ponovi v tretjem in nadaljnjih obhodih, kjer pa sta aktivni samo še prva in četrta čakalna vrsta.

Glavna odlika DRR je v računski nezahtevnosti. Ohrani in nadgradi vse odlike krožnega dodeljevanja (Round Robin - RR) in WRR. V primerjavi z WFQ in podobnimi odpravniki pa je nekoliko slabši glede pravičnosti in najvišjih vrednosti zakasnitve. Obstaja cela vrsta njegovih izpeljank, med katerimi so: Pre-order Deficit Round Robin [106], Carry-Over Round Robin [88], Nested Deficit Round Robin [43], Elastic Round Robin [44] in drugi.



## 5 Združitev Diferenciranih in Integriranih storitev

V nadaljevanju se ukvarjamo predvsem s sposobnostmi omrežij IP za zagotavljanje različnih stopenj kakovosti storitev. Kot smo ugotovili že v razdelku 3.2, osnovne funkcije omrežij IP ne zadoščajo za uvedbo različnih stopenj kakovosti storitve. V razdelkih 3.2.4 in 3.2.5 smo predstavili osnovne koncepte in način delovanja omrežij z Integriranimi storitvami (IS) in Diferenciranimi storitvami (DS), ki so nadgradnja omrežij IP. Že bežna primerjava obeh konceptov nam pokaže, da sta si precej različna. Vsak ima svoje dobre in slabe lastnosti. To poglavje zato posvečamo podrobnejši primerjavi IS in DS. Podamo tudi predlog združevanja obeh v smislu izrabe njihovih dobrih in izničenju njihovih slabih lastnosti.

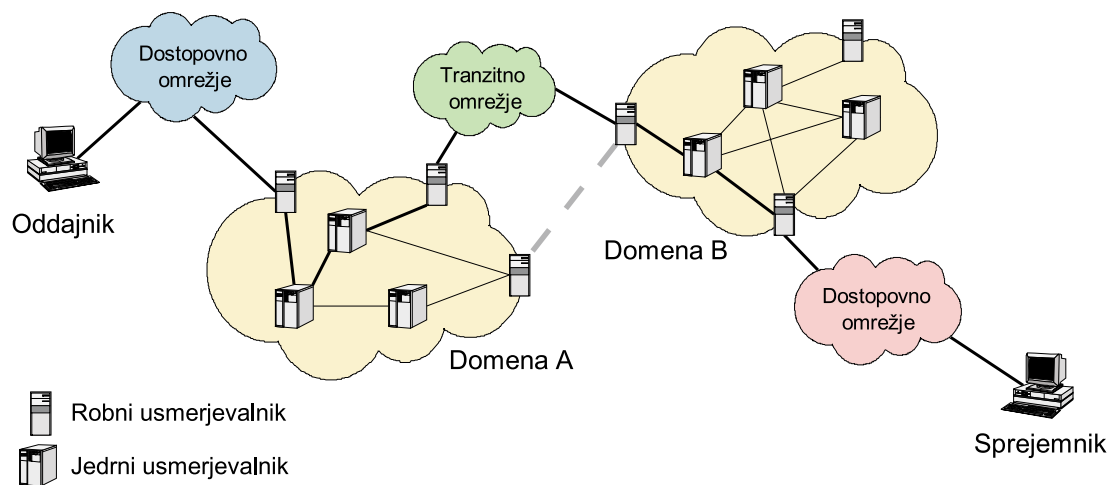
Najprej si oglejmo topologijo prenosnega omrežja, na kateri temeljita naša analiza in predlog združitve.

### 5.1 Referenčna topologija prenosnega omrežja

Referenčna topologija prenosnega omrežja, prikazana na sliki 5.1, je podlaga za podrobnejšo razlago našega predloga združevanja Diferenciranih in Integriranih storitev. Elementi tega omrežja so:

**Izvorna in ponorna točka komunikacije** sta končna uporabnika (napravi), ki si med seboj izmenjujeta podatke. Taki povezavi dveh uporabnikov rečemo povezava **od konca do konca** (end-to-end) in lahko poteka preko več logičnih omrežij. Splošen primer je prikazan na sliki 5.1, kjer sta končna uporabnika (oddajnik in sprejemnik) povezana vsak na svoje dostopovno omrežje in si podatke izmenjujeta preko domen A in B ter tranzitnega omrežja. V nadaljnjih razlagah se sklicujemo na primer prenosa podatkov, ki je prikazan na sliki 5.1.

**Domena** je del omrežja, ki za prenos podatkov uporablja določeno tehnologijo in je pod nadzorom enega ponudnika storitev (operaterja). Tako na primer domeni A in B lahko uporabljata isto tehnologijo, a sta v lasti dveh različnih operaterjev, ali pa sta v lasti istega operaterja, vendar imata različni tehnologiji prenosa podatkov (npr. IP/ATM in IP/SDH), lahko pa sta v lasti različnih operaterjev in temeljita na različnih tehnologijah. Domena je torej območje, na katerem veljajo enaka pravila, se ponujajo enake storitve in se nudijo enake stopnje kakovosti storitve. Pri prehodu iz ene domene v drugo se lahko pravila, storitve in stopnje kakovosti storitve, ki so na voljo, spreminjajo. Ker preslikave med različnimi stopnjami storitve na prehodu med domenami največkrat niso nedvoumne, to predstavlja precejšen problem zlasti za uporabnike (podatke), ki zahtevajo določeno stopnjo kakovosti storitve na povezavi od konca do konca.



Slika 5.1 – Referenčna topologija omrežja in komunikacija med končnima točkama preko IS/DS domen

**Dostopno omrežje** je del prenosne poti, po kateri se končni uporabnik poveže v omrežje (domeno). Navadno je to neka razširjena dostopovna tehnologija kot so: Ethernet, ADSL (Asymmetric Digital Subscriber Line - Asimetrična digitalna naročniška linija), klicna povezava, kabelsko omrežje in druge. V redkih primerih je lahko končni uporabnik povezan tudi neposredno v domeno. V našem predlogu od dostopnega omrežja zahtevamo polno IS/RSVP propustnost, kar pomeni, da dostopno omrežje ne sme biti vzrok za zavrnitev RSVP povezave in da mora zagotavljati želeno stopnjo kakovosti storitve. Predpostavljamo tudi, da dostopno omrežje nima problemov s skalabilnostjo, saj se preko njega ne povezuje veliko število uporabnikov.

**Tranzitno omrežje** je del prenosne poti, o katerem načeloma ne vemo nič in zaradi vnosa neznanj ni zaželen del prenosne poti. Na sliki je dodan zaradi splošnosti. V našem predlogu ga ne razdelamo podrobneje. Domeni A in B sta lahko povezani tudi neposredno. To je na sliki 5.1 prikazano s prekinjeno sivo črto.

**Robni usmerjevalnik** je omrežna naprava, ki se nahaja na robu posamezne domene in jo povezuje z drugimi deli omrežja. Na primer, robni usmerjevalnik povezuje domeno z dostopovnim ali tranzitnim omrežjem, ali pa dve domeni med seboj. V slednjem primeru ima na prehodu vsaka domena svoj robni usmerjevalnik. Robni usmerjevalnik skrbi za pravilno obravnavo podatkov, ki pritekajo v domeno ali pa jo zapuščajo. V kontekstu združitve Diferenciranih in Integriranih storitev mora robni usmerjevalnik skrbeti za celo vrsto opravil. Naj naštejemo le nekatere: skrb za pravilno vzpostavitev povezave in rezervacije preko domene, preverjanje upravičenosti podatkov do zahtevane kakovosti storitve, uravnavanje in oblikovanje prometa posameznih podatkovnih pretokov, preslikava parametrov in zahtevane stopnje kakovosti storitve med domenami, in drugo. Robni usmerjevalnik mora imeti popoln nadzor nad pretokom podatkov. Za robne usmerjevalnike na robu domene, ki se povezujejo na dostopovna omrežja, to ni problem, saj je količina podatkov navadno majhna ali srednje velika. Robni usmerjevalniki na robu domene, ki se povezujejo na tranzitna omrežja ali druge domene, pa morajo biti zmogljivejši.

**Jedrni usmerjevalnik** je omrežna naprava, ki se nahaja znotraj domene in se povezuje z drugimi jedrnimi in robnimi usmerjevalniki iste domene. Skrbi samo za posredovanje podatkov na njihovi prenosni poti. Ker opravlja zgolj posredovalno funkcijo, lahko obdela precej večjo količino podatkov kot enako zmogljiv robni usmerjevalnik. Jedrnim usmerjevalnikom tako ni potrebno skrbeti za preslikavo parametrov in stopenj kakovosti storitve ter druge funkcije, ki jih imajo robni usmerjevalniki. Njihov namen je hiter in učinkovit prenos podatkov med dvema robnima usmerjevalnikoma iste domene.

## 5.2 Primerjava Diferenciranih in Integriranih storitev

Kot smo že spoznali, sta si koncepta Diferenciranih in Integriranih storitev precej različna. Za primerjavo njune najpomembnejše lastnosti navajamo v tabeli 5.1 [9].

	Integrirane storitve	Diferencirane storitve
Diferenciacija storitev	na osnovi pretokov	na osnovi storitvenih razredov
Potrebe po omrežnih virih	za vsak pretok	za posamezen razred
Osnova razvrščanja prometa	kombinacija polj v glavi paketa IP	polje DS v glavi paketa IP
Trdnost prenosnih zagotovil	deterministična ali statistična	relativna
Nadzor dostopa do storitve	zahtevan	ni potreben
Signalizacijski protokol	zahtevan (RSVP)	ni potreben
Usklajenost diferenciacije	od konca do konca	znotraj DS domene
Skalabilnost	omejena s številom pretokov	omejena s številom razredov
Obračunavanje storitev	glede na lastnosti pretoka in QoS zahtev	glede na uporabljen razred
Upravljanje omrežja	podobno kot pri povezavnih omrežjih	podobno kot pri omrežjih IP

Tabela 5.1 – Primerjava glavnih lastnosti Diferenciranih in Integriranih storitev

Opazimo, da so si lastnosti obeh konceptov praktično nasprotne. Zato posebej navajamo prednosti in slabosti posameznega koncepta.

Prednosti koncepta IS so:

- zagotovljena kakovost storitev, ki je deterministična (zajamčena storitev) ali statistična (nadzorovano breme), ter velja na celotni prenosni poti,
- ločevanje posameznih pretokov, ki ne morejo vplivati drug na drugega,
- lažji nadzor dostopa do storitev in možnost podrobnejše razdelitve obračunavanja storitev.

Slabosti koncepta IS pa so:

- velike potrebe po omrežnih virih (za vsak posamezen pretok),

- obvezna uporaba signalizacijskega protokola, ki pomeni dodatno breme glede prenosa in procesnih zmogljivosti omrežnih naprav,
- omejena skalabilnost, saj potrebe po omrežnih virih naraščajo sorazmerno z večanjem prometa, ki je sorazmeren številu pretokov.

Prednosti koncepta DS so:

- majhne potrebe po omrežnih virih (za vsak storitveni razred),
- uporaba signalizacijskega protokola ni potrebna,
- skalabilnost ni problem, saj potrebe po omrežnih virih naraščajo s številom razredov, ki niso odvisni od količine prometa.

Slabosti koncepta DS pa so:

- kakovost storitve je zgolj relativna in velja samo znotraj posamezne DS domene,
- ločevanje posameznih pretokov ni mogoče in ti lahko vplivajo eden na drugega,
- težji nadzor dostopa do storitev in obračunavanja le teh.

Zaradi večinoma diametralno nasprotnih si lastnosti IS in DS popolna izraba dobrih in izničenje slabih lastnosti obeh konceptov ni mogoča, vsaj ne za ves promet na omrežju. Zato je naša ideja uporaba posameznih dobrih lastnosti obeh konceptov na način, da te zajamejo čim večji delež prometa na omrežju. Ker želimo uporabiti dobre lastnosti tako omrežij z IS kot z DS, mora naše omrežje podpirati oba koncepta.

### 5.3 Zakaj in na kakšen način združiti omrežja z IS in DS

Na sliki 5.1 vidimo, da je na celotni prenosni poti lahko več različnih domen. Če so to domene, ki uporabljajo koncept DS z relativno stopnjo kakovosti storitve, je zagotavljanje ustrezne stopnje kakovosti storitve na celotni poti zelo težavna naloga. Te domene so lahko v lasti različnih operaterjev, lahko temeljijo na različnih prenosnih tehnologijah, zelo verjetno ponujajo nezdružljive ali težko združljive stopnje kakovosti storitve in preslikava parametrov kakovosti storitve navadno temelji na kompromisih. Od takega omrežja je zato težko pričakovati, da bo uporabnik, ki pri svojem domenskemu operaterju zahteva in dobi določeno stopnjo kakovosti storitve (na primer v domeni A), enako stopnjo kakovosti storitve dobil tudi v ostalih domenah, skozi katere poteka prenosna pot. Tu lahko negativno vlogo odigra tudi poslovna tekmovalnost in nelojalna konkurenca med ponudniki storitev in/ali operaterji omrežij.

Zgornji problem lahko elegantno rešimo z uporabo IS. Tu uporabnik ob vzpostavitvi povezave zahteva določeno stopnjo kakovosti storitve in če je omrežje ne more zagotoviti, je zahteva zavrnjena. V nasprotnem primeru pa uporabnik dobi zahtevano stopnjo kakovosti storitve na celotni prenosni poti in preko vseh domen.

V naslednjem razdelku zato podajamo nov predlog združevanja omrežij z DS in IS, za katerega trdimo, da na preprost in obvladljiv način zagotovi njuno združitev. Pri tem ohrani njune dobre lastnosti in izniči večino njunih slabih lastnosti.

Iz zgoraj zapisanega in iz primerjave DS in IS v razdelku 5.2 zaključimo, da moramo ohraniti funkcionalnost obeh konceptov. Za to moramo zagotoviti nekatere minimalne

pogoje. Omrežje mora nuditi deterministično raven kakovosti storitve in možnost zaščite posameznih pretokov pred potencialno kvarnimi vplivi drugih pretokov na isti prenosni poti ali delu te poti. To zahteva funkcionalnost omrežij z IS in ohranitev signalizacijskega protokola (npr. RSVP). Na ta način ohranimo najpomembnejše dobre lastnosti IS. Po drugi strani pa je potrebno zagotoviti skalabilnost in zmanjšati potrebe po omrežnih virih. To zahteva funkcionalnost omrežij z DS. Na ta način ohranimo dobre lastnosti DS in izničimo slabe lastnosti IS.

## 5.4 Predlog združitve Diferenciranih in Integriranih storitev

V tem razdelku podajamo nov predlog združevanja omrežij z IS in DS. Novo, združeno omrežje, imenujemo omrežje z Diferenciranimi in integriranimi storitvami (DIS) [38].

Po našem predlogu omrežje DIS v osnovi in na prvi pogled deluje po načelu DS. To pomeni, da diferenciacija prometa poteka na osnovi pripadnosti storitvenemu razredu. Omrežje posamezne razrede obravnava glede na vrednost polja DS v njihovih glavah in jim tako omogoča boljše ali slabše prenosne pogoje. Ob predpostavki, da večina prometa v omrežju zahteva zgolj relativno stopnjo kakovosti storitve, lahko na tak način izkoristimo večino prednosti koncepta DS, ki so našteje v razdelku 5.2. Omrežje DIS mora prepoznati vrsto prometa in ga ustrezno obravnavati. Nekatere pakete samo posreduje, druge pred posredovanjem pregleda in po potrebi tudi spreminja.

Za manjši del prometa z višjimi ali bolj kompleksnimi zahtevami glede stopnje kakovosti storitve je zato definiranih nekaj storitvenih razredov, ki jih omrežje obravnava drugače. Kot primer prometa z višjimi/strožjimi zahtevami glede parametrov kakovosti storitve lahko navedemo prenos podatkov v realnem času, kot primer prometa z bolj kompleksnimi zahtevami pa podatke RSVP signalizacijskega protokola. Poleg tega, da tak promet od omrežja zahteva določena zagotovila ali okvire, v katerih se gibljejo parametri kakovosti storitve, so njegovi paketi podrobneje obdelani in zahtevajo več omrežnih virov. Za te storitvene razrede lahko rečemo, da sledijo konceptu IS in tako izkoriščajo njegove dobre lastnosti. Glede na predpostavko, da le manjši delež prometa od omrežja zahteva trdna zagotovila o stopnji kakovosti storitve, slabe lastnosti IS ne pridejo do izraza. Hkrati pa s tem za ta del prometa izničimo pomanjkljivosti koncepta DS (za vse našteje glej 5.2). Za preostali promet pa privzamemo, da mu zadošča funkcionalnost omrežij z DS.

Naj poudarimo, da to ni nekakšen "IS preko DS" ali nekaj podobnega, ampak združitev obeh konceptov. Naše omrežje deluje dvostopenjsko: na prvi stopnji omrežje deluje kot omrežje z DS, na drugi stopnji (samo za določene pretoke) pa kot omrežje z IS!

Delovanje omrežja DIS najlažje predstavimo na primeru. Definirajmo 6 storitvenih razredov z oznakami A do F z lastnostmi, navedenimi v tabeli 5.2. V omrežju z DS vsaki od teh oznak ustreza določena vrednost DSCP v glavi paketa IP (glej razdelek 3.2.2).

Ko posamezen paket potuje skozi omrežje, vsaka omrežna naprava na poti najprej preveri kateremu storitvenemu razredu pripada in ga glede na to ustrezno obravnava. Tako so paketi iz storitvenih razredov D do F razvrščeni v ustrezno čakalno vrsto<sup>1</sup>

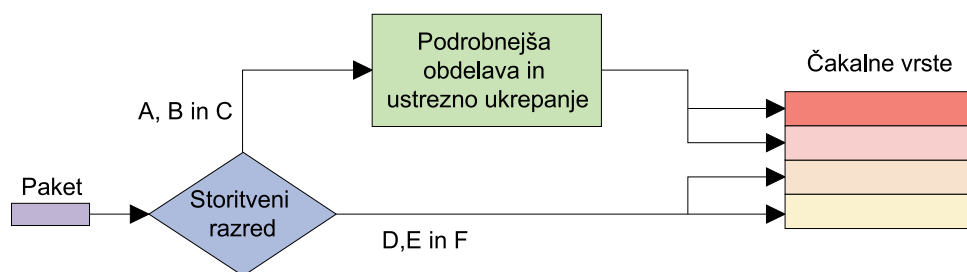
---

<sup>1</sup>Razvrščanje paketov v čakalne vrste smo obravnavali v poglavju 4

Razred	Ime	Komentar	Omrežje
A	zajamčene storitve	trdni parametri QoS	IS
B	nadzorovano breme	statistični parametri QoS	IS
C	signalizacijski protokol	npr. RSVP	IS/DS
D	prednostni promet	promet s prednostjo	DS
E	po najboljših zmožnostih	navaden promet	DS
F	neprednostni promet	promet v ozadju	DS

Tabela 5.2 – Primer razdelitve prometa v storitvene razrede

neposredno na podlagi vrednosti DSCP v glavi paketa IP in niso podrobneje obdelani. Ti paketi so torej deležni le enostopenjske obdelave.



Slika 5.2 – Poenostavljen primer poteka obdelave paketov v omrežjih DS

Paketi iz storitvenega razreda C pripadajo signalizacijskemu protokolu (v našem primeru je to kar RSVP) in jih morajo zato omrežne naprave podrobneje obdelati. To se zopet ugotovi na podlagi vrednosti DSCP v glavi paketa IP in tak paket se preda ustreznemu procesu, na primer, procesu za krmiljenje dostopa. Za promet razredov A in B se na podlagi vrednosti DSCP v glavi paketa IP ugotovi, da pripadata pretokom, ki imajo zajamčeno določeno stopnjo kakovosti storitve. Ti paketi se najprej predajo ustreznim procesom, na primer, procesu za krmiljenje pretoka in nato razporedijo v ustrezno čakalno vrsto. Paketi pretokov A do C so torej deležni dvostopenjske obdelave.

Tako je promet storitvenih razredov D do F deležen le preverjanja pripadnosti storitvenemu razredu preko vrednosti polja DS - enostopenjska obdelava, promet razredov A do C pa je poleg preverjanja vrednosti polja DS deležen še podrobnejše obdelave v vsaki omrežni napravi - dvostopenjska obdelava. Poenostavljen potek obdelave paketov posameznih storitvenih razredov je viden na sliki 5.2.

V našem predlogu združevanja IS in DS promet storitvenih razredov A do C zahteva funkcionalnost omrežij z IS, preostali trije razredi pa funkcionalnost omrežij z DS. Ob predpostavki, da je delež prometa razredov A in B ter posledično tudi C majhen v primerjavi s prometom razredov D do F, je združevanje obeh konceptov prav gotovo smiselno in koristno.

#### 5.4.1 Možne težave in rešitve pri združevanju IS in DS

Zgornja izvajanja temeljijo na predpostavki, da je delež prometa z višjimi in kompleksnejšimi zahtevami glede kakovosti storitve, ki uporablja IS funkcionalnost omrežja,

majhen v primerjavi s prometom, ki zahteva zgolj DS funkcionalnost. Na ta način se izognemo težavi, ki nastane zaradi neskalabilnosti omrežij z IS, ki je posledica višanja potreb po omrežnih virih sorazmerno z večanjem števila pretokov. Pri nekaj stotisoč ali milijonih hkratnih pretokov je to lahko prehudo breme tudi za najzmogljivejše omrežne naprave.

Zato moramo predvideti tudi situacije, ko je delež zahtevnega prometa višji, ko je omrežje sestavljeno iz manj zmogljivih vozlišč ali ko tudi najzmogljivejše omrežne naprave ne zmorejo postreči vsega prihajajočega QoS prometa.

Rešitev je združevanje RSVP pretokov in s tem zmanjšanje potreb po omrežnih virih, saj združen pretok zahteva manj omrežnih virov kot je vsota uporabljenih omrežnih virov s strani nezdruženih pretokov. V obstoječi literaturi [5], [95], [81], [13] lahko najdemo kar nekaj predlogov združevanja, ki imajo vsak svoje dobre in slabe strani in so na kratko predstavljeni v nadaljevanju. V [40] pa smo predlagali tudi nov način združevanja RSVP pretokov, ki je prilagojen omrežju DIS. Ker izkorišča možnosti, ki jih ponuja dvostopenjska obdelava, je za omrežje DIS optimalnejši kot drugi omenjeni predlogi.

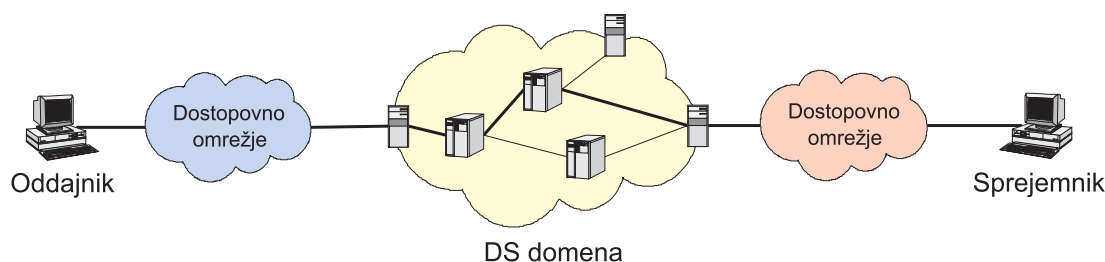
#### 5.4.2 Drugi predlogi združevanja IS in DS

V [5] je po spoznanju, da omrežja z DS, če niso predimenzionirana, sama po sebi ne morejo vedno zagotoviti ustrezne stopnje kakovosti storitve za vse aplikacije, RSVP zopet uveden kot pomemben element pri zagotavljanju kakovosti storitve v omrežju. Omrežja z DS imajo namreč težave s predvidenimi deleži prometa, ki pripadajo posameznemu storitvenemu razredu, in zato ni mogoča optimalna fiksna nastavitvev omrežnih naprav. Če to težavo poskusimo reševati z dinamičnim dodeljevanjem virov, to v omrežje prinese dodatno kompleksnost in stroške. S tem pa v veliki meri izničimo prednosti omrežij z DS. Zato je v [5] predlagana kombinacija DS/RSVP, kjer je RSVP v omrežju z DS uporabljen predvsem za dinamično kontrolo prometa posameznih storitvenih razredov in ne za vzpostavljanje rezervacij posameznih pretokov. Tak primer je, ko se RSVP uporabi kot *pripustitveni mehanizem* (admission mechanism) v omrežje z DS (glej sliko 5.3). Posamezni pretoki na dostopovnem omrežju tako še vedno uporabljajo RSVP, ki pa ga vhodni robni usmerjevalnik DS domene uporabi zgolj za pripustitev ali zavrnitev pretoka in njegovo uvrstitev v ustrezni storitveni razred<sup>2</sup>. Slabost pristopa v [5] je predvsem ta, da še vedno ni možna rezervacija virov za posamezne pretoke in s tem možnost njihove izolacije od potencialno škodljivih vplivov drugih pretokov. Ravno tako ni možno podati trdnih zagotovil o stopnji kakovosti storitve, ki je je deležen posamezen storitveni razred, saj jedrno omrežje še vedno deluje le po načelu DS.

V [95] in [81] je ideja iz [5] bolj podrobno razvita, načelo pa ostaja isto. Prenosna pot je razdeljena na tri dele. Prvi in zadnji del potekata po omrežju, ki je vsaj delno IS/RSVP sposobno. Na sliki 5.3 sta ta dva dela omrežja predstavljena kot dostopovno omrežje. Predlog navaja dve možni konfiguraciji DS domene. Prva predpostavlja, da DS domena ne vsebuje RSVP sposobnih omrežnih naprav in zato so znotraj nje viri statično določeni (neoptimalno). Robni usmerjevalniki tako izvajajo le funkcijo

---

<sup>2</sup>Če je pretok pripuščen v omrežje z DS, se ga uvrsti v zeleni storitveni razred, drugače pa se ga zavrne ali uvrsti v kak drug storitveni razred z nižjo stopnjo kakovosti storitve.



Slika 5.3 – Omrežje z DS domeno

pripravitve ali zavrnitve RSVP pretoka, znotraj domene pa vse poteka po načelu DS. Druga konfiguracija pa predpostavlja, da so znotraj domene vsaj nekatere omrežne naprave RSVP sposobne. To pomeni, da lahko sodelujejo v RSVP signalizaciji, promet pa še vedno posredujejo zgolj po načelu DS. Prednost tega pristopa je v možnosti dinamičnega dodeljevanja omrežnih virov v RSVP sposobnih napravah, ki poteka s pomočjo RSVP rezervacij (združenih ali posameznih) ali preko katerega drugega mehanizma za dinamično dodeljevanje omrežnih virov. Čeprav ta predlog odpravi nekatere slabosti iz [5], pa še vedno ne more podati trdnih zagotovil o stopnji kakovosti storitve za posamezen pretok, saj prenos podatkov preko DS domene še vedno poteka le po načelu DS, izboljša se le nadzor nad omrežnimi viri.

Naš predlog gre en korak dlje, saj predpostavi, da so omrežne naprave v domenah RSVP sposobne in da podpirajo IS. V našem predlogu tako omrežje v osnovi deluje po načelu DS, ko pa gre za prometne pretoke z višjimi zahtevami po kakovosti storitve pa lahko deluje tudi po načelu IS s podporo RSVP in po potrebi tudi s podporo združevanju RSVP pretokov.

Glavna prednost našega predloga omrežja DIS je možnost nudenja različnih stopenj kakovosti storitve, vključno z rezervacijo omrežnih virov in zagotovljenimi prenosnimi parametri na celotni prenosni poti, ne da bi pri tem imeli težave s skalabilnostjo. S tem izkoriščamo dobre lastnosti omrežij z DS in IS.

### 5.4.3 Predlog konfiguracije in krmiljenja omrežnih naprav

Podajmo še predlog konfiguracije in krmiljenja omrežnih naprav, ki po našem mnenju zadosti potrebam in delovanju predlaganega omrežja DIS. Podrobnosti glede konfiguracije in krmiljenja omrežnih naprav so podane v razdelku 4.2.

Pri krmiljenju omrežnih naprav le te razdelimo na: končne naprave, robne usmerjevalnike in jedrne usmerjevalnike. **Končne naprave** na oddajni strani morajo izvajati funkcije krmiljenja dostopa ter krmiljenja in glajenja prometa. Predpostavljamo, da odpravljanje v smislu kakovosti storitve v končnih napravah ni problematično. **Robni usmerjevalniki** imajo od vseh naprav v našem omrežju najtežje delo, saj morajo opravljati vse funkcije. Tako skrbijo za krmiljenja dostopa ter krmiljenja in glajenja prometa, razvrščanje prometa v čakalne vrste (glede na storitveni razred in v drugi stopnji tudi glede na pretok), upravljajo s pomnilnikom v smislu odmetavanja odvečnih paketov v času zasičenja (ali v izogib temu) ter pravilno odpravljanje paketov iz čakalnih vrst. Glavna naloga **jedrnih usmerjevalnikov** pa je pravilno krmiljenje čakalnih vrst (razvrščanje, upravljanje pomnilnika in odpravljanje), ki zagotavlja do-



govorjeno stopnjo kakovosti storitve.

O večini funkcij, ki jih omenjamo zgoraj, smo že govorili. Nismo pa še navedli katere odpravnikove uporabimo za posredovanje paketov. V standardih, ki določajo delovanje omrežij z DS in IS ter protokola RSVP, ni določeno katere mehanizme se uporabi za doseg cilja. Zato imamo pri izbiri odpravnikov proste roke, če le ti zadostijo vsem pogojem. S stališča omrežnih naprav je najbolje uporabiti čimbolj nezahtevne odpravnikove, ki so nepotrati z njihovimi viri. Problem takih odpravnikov je, da so navadno nepravilni ali imajo neugodno algoritemsko zakasnitev.

Ker naš predlog omrežja DIS uvede dvostopenjsko obravnavo paketov v omrežnih napravah, tako konfiguriramo tudi čakalne vrste v njih. Ugotovili smo že, da je za zagotavljanje različnih stopenj kakovosti storitve primernejše izhodno čakanje. Zato vzamemo omrežno napravo z izhodnim čakanjem, ki ima več čakalnih vrst tipa FIFO in na vsaki izmed obeh stopenj uporabimo svoj odpravnik, ki zadosti potrebam posamezne stopnje. Prva stopnja mora zagotoviti različne stopnje kakovosti storitve za obstoječe storitvene razrede, druga stopnja pa, kadar je to potrebno, še pravičnost med posameznimi pretoki znotraj istega storitvenega razreda. Za naše omrežje DIS zato predlagamo uporabo strogo prednostnega odpravljanja (Strict Priority Queueing) na prvi stopnji in deficitarnega krožnega dodeljevanja (DRR) na drugi stopnji. Utemeljitev:

- Oba predlagana odpravnikova imata zahtevnost  $O(1)$ .
- S **strogo prednostnim odpravljanjem** (glej razdelek 4.5.2) zagotovimo relativno stopnjo kakovosti storitve med posameznimi storitvenimi razredi. Če zopet vzamemo primer v razdelku 5.4, potem ima storitveni razred A najvišjo prednost in storitveni razred F najnižjo. Postavi se vprašanje, ali je to za naše omrežje zadosti dobro. Glede na lastnosti strogo prednostnega odpravljanja se paketi storitvenega (prednostnega) razreda A obnašajo kot da je celotna pasovna širina povezave njihova, saj jih pri prenosu paketi nižjih prednostnih razredov ne motijo. Ravno tako se paketi storitvenega (prednostnega) razreda B obnašajo kot da si pasovno širino delijo samo s paketi prednostnega razreda A. Podobno velja za pakete vse do prednostnega razreda F. Glede na to, da sta število in bitne hitrosti pretokov razreda A in B omejeni, se ne more zgoditi, da bi ta dva prednostna razreda porabila vso pasovno širino povezave in bi ostali storitveni razredi zaradi tega stradali. Spomnimo se, da je storitveni razred A namenjen *zajamčenim storitvam*, storitveni razred B pa *storitvi nadzorovanega bremena*. Ker je prvi zahtevnejši kot drugi, je tudi ta prednostni red pravilen.
- Z **DRR** (glej razdelek 4.5.6) pa zagotovimo, da so pretoki znotraj prednostnega razreda med seboj ločeni in pravično obravnavani. To pomeni, da dolgoročno noben izmed pretokov ne dobi pasovne širine, ki je večja kot mu pripada glede na rezervacije. To drugostopenjsko odpravljanje je smiselno le za storitveni razred A (zajamčene storitve) in po potrebi tudi za storitveni razred B (nadzorovano breme).

V naslednjih poglavjih je predstavljena analiza delovanja obeh zgoraj predlaganih odpravnikov v omrežju DIS. Najprej si ogledamo kako in na kakšen način strogo prednostno odpravljanje zagotovi različne stopnje kakovosti storitve na prvi stopnji omrežja DIS, potem pa še kako deluje in kakšne lastnosti ima odpravnik DRR. V obeh primerih

naredimo teoretično analizo, ki jo podpremo s simulacijskimi rezultati. Nato s pomočjo simulacij analiziramo še dvostopenjsko delovanje omrežja DIS z obema odpravnikoma. Na koncu uvedemo še izboljšavo odpravnika DRR. Zanj naredimo teoretično analizo in simulacije. Teoretične in simulacijske rezultate primerjamo med seboj in z rezultati delovanja odpravnika DRR.

## 5.5 Analiza delovanja prve stopnje omrežja DIS

V prejšnjih razdelkih smo podali predlog združevanja omrežij z DS in IS v omrežje DIS. Predlagali smo tudi možno konfiguracijo omrežnih naprav, ki zadovoljivo reši prenos podatkov v takem omrežju. Predlagano omrežje DIS na prvi stopnji uporablja *strogo prednostno odpravljanje*, ki zagotovi različne stopnje kakovosti storitve. Prva stopnja omrežja DIS v bistvu deluje po principu omrežij z DS. Zato v tem poglavju s teoretično analizo in simulacijami podajamo razmere v omrežju IP z diferenciranimi storitvami, ki temeljijo na prednostnem odpravljanju. Pri tem se z izbranimi scenariji poskušamo kar najbolj približati realnim razmeram. Osredotočimo se predvsem na jedrne usmerjevalnike, saj so ti najbolj obremenjeni in posredujejo največje količine prometa.

Promet na nekem omrežju lahko razlikujemo glede na več kriterijev: zakasnitve, propustnost, spremenljivost zakasnitve, izgube in drugo. Najprimernejša in najpogosteje uporabljena mera je zakasnitev, saj so s stališča zagotavljanja kakovosti storitve najzahtevnejše tiste aplikacije, ki zahtevajo prenos podatkov v realnem času (npr. IP telefonija). Pri teh aplikacijah je najpomembnejša mera kakovosti storitve prav zakasnitev.

Zakasnitev prenosa podatkov v omrežju je vsota zakasnitev prenosa podatkov med vozlišči na prenosni poti. Zakasnitev prenosa podatkov med vozlišči sestavljajo:

**Zakasnitev obdelave** (forwarding delay) je čas, ki ga omrežna naprava porabi, da paket sprejme, ga obdela, se odloči o načinu njegovega posredovanja in ga začne oddajati na prosti in nezasedeni izhodni liniji (če se ne oddaja ali na oddajo ne čaka noben drug paket). To je najkrajši čas, ki ga omrežna naprava porabi za posredovanje posameznega paketa. Te zakasnitve so navadno reda mikrosekund.

**Čakalni čas** (queuing delay) je čas, ki ga paket prebije v čakalni vrsti, ko čaka, da se odpravijo paketi, ki so prispeli pred njim ali imajo pred njim prednost. Čakalni čas v omrežni napravi je odvisen od mnogih dejavnikov in se s časom spreminja. Lahko je enak nič (nezasedena izhodna linija) ali pa je vsota časov, potrebnih za oddajo vseh paketov, ki so na vrsti pred čakajočim paketom.

**Zakasnitev predaje** (serialization delay) je čas, ki ga omrežna naprava porabi, da na izhodno linijo preda celoten paket. Zakasnitev predaje  $\tau$  je odvisna od dolžine paketa  $L$  in bitne hitrosti izhodne linije  $R$  ter se izračuna s pomočjo izraza  $\tau = L/R$ .

**Zakasnitev razširjanja signala** (propagation delay) je čas, ki ga signal porabi za pot med sosednjima vozliščema. ITU priporočilo G.114 [19] navaja vrednosti  $4 \mu\text{s}/\text{km}$  za zemeljske koaksialne vodnike,  $5 \mu\text{s}/\text{km}$  za optične vodnike in  $6 \mu\text{s}/\text{km}$  za podmorske kable.

V nadaljevanju analiziramo zakasnitve v omrežju z DS, kar ustreza analizi delovanja prve stopnje omrežja DIS. Uporabimo strogo prednostno odpravljanje. Najprej naredimo teoretično analizo zakasnitev v omrežni napravi in jo dopolnimo s simulacijami pod enakimi pogoji ter nato rezultate med seboj primerjamo. Tu se osredotočimo le na posredovalno funkcijo in ne upoštevamo vrste prometa in podrobnosti delovanja prometnih izvorov. Nadaljujemo s simulacijami, ki vključujejo tudi vplive višjih protokolnih plasti (predvsem protokola TCP), vrsto prometa ter lastnosti in delovanje prometnih izvorov. Rezultate teoretične analize in simulacij diferenciranih storitev s strogo prednostnim odpravljanjem primerjamo z rezultati delovanja omrežja po principu FCFS.

### 5.5.1 Razlikovanje prometa glede na zakasnitve

Kot glavni parameter kakovosti storitve smo si izbrali zakasnitev, zato promet na omrežju razlikujemo glede na zakasnitev, ki naj jo občuti. Bolj zahtevne aplikacije imajo višjo stopnjo kakovosti storitve in občutijo nižje zakasnitve, manj zahtevne aplikacije pa imajo nižjo stopnjo kakovosti storitve in zato občutijo višje zakasnitve. Izkaže se, da je za omrežja IP tako razlikovanje prometa zelo primerno, saj zlahka identificiramo aplikacije, ki zahtevajo nizke vrednosti zakasnitve in aplikacije, za katere zakasnitev ni najpomembnejši parameter kakovosti storitve.

Promet razdelimo v štiri storitvene razrede, ki jim priredimo štiri kakovostne stopnje. V vsakem razredu identificiramo najznačilnejšo aplikacijo in želeno vrednost zakasnitve, kot je navedeno v tabeli 5.3. Zakasnitve navedene v tabeli služijo le kot primer in so glede na zahteve lahko tudi drugačne.

Aplikacija	Stopnja QoS	Zakasnitev
Govor	4	$\leq 150$ ms
Video	3	$\leq 250$ ms
WWW	2	$\leq 1$ s
FTP	1	do nekaj sekund

Tabela 5.3 – Stopnje kakovosti storitve za značilne aplikacije

**Prenos interaktivnega govora** je izredno zahtevna aplikacija, ki tvori stalen pretok podatkov, za katerega mora biti ob vsakem času na voljo zadosti prenosnih zmogljivosti (bitne hitrosti). Poleg tega zahteva tudi zelo nizke zakasnitve in njihovo čim manjšo spremenljivost.

**Prenos videa** je precej podoben prenosu govora, le dopustne zakasnitve so nekoliko višje. Pretok podatkov je zaradi lastnosti kodirnikov slike spremenljiv in količinsko precej večji kot pri govoru.

**Brskanje po svetovnem spletu (WWW<sup>3</sup>)** je načeloma zelo nezahtevna aplikacija, ki pa zaradi občasne uporabnikove zahteve po "interaktivnosti" vseeno potrebuje nekoliko boljše kakovost storitve.

**Prenos datotek (FTP<sup>4</sup>)** ponavadi poteka v ozadju in od omrežja v smislu zakasnitev ne zahteva veliko. Potrebuje le primerno propustnost.

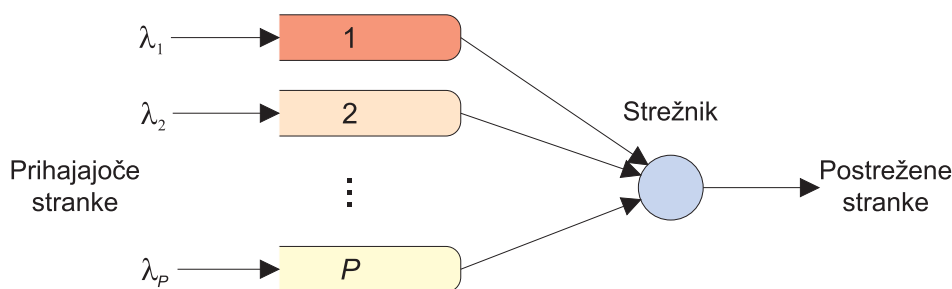
### 5.5.2 Teoretična analiza diferenciacije prometa

V tem razdelku podajamo teoretično analizo diferenciacije prometa paketnega omrežja na osnovi zakasnitev. Že v prejšnjih razdelkih smo se odločili za uporabo strogo prednostnega odpravljanja, čemur v teoriji čakalnih vrst ustreza sistem čakalne vrste M/G/1 s prednostjo.

Osnove teorije čakalnih vrst, definicijo spremenljivk in drugih pojmov povezanih z analizo čakalnih vrst, ki jih uporabljamo v nadaljevanju, najdemo v dodatku B. Rezultati so povzeti po [48], [49] in [50]. V naslednjih razdelkih uporabljamo terminologijo iz teorije čakalnih sistemov. Ustrezne termine paketnih omrežij navajamo v oklepajih.

#### 5.5.2.1 Čakalna vrsta M/G/1 s prednostjo

V čakalni vrsti M/G/1 s prednostjo stranke (pakete), ki prihajajo v sistem (omrežno napravo) razdelimo v storitvene razrede, ki jih obravnavamo prednostno. Znotraj posameznih razredov se stranke strežejo po vrstnem redu prihodov (FIFO). V nadaljevanju obravnavamo *strogo prednostno strežbo brez prekinitev*<sup>5</sup> (Non-preemptive Strict Priority Queueing) za čakalno vrsto M/G/1. Model čakalne vrste s prednostmi je podan na sliki 5.4.



Slika 5.4 – Model čakalne vrste s prednostmi za paketna omrežja

Strogo prednostna strežba čakalne vrste streže najprej stranke iz neprazne čakalnice z najvišjo prednostjo, na primer  $p$ , ko je ta prazna, začne streči stranke v čakalnici s prednostjo  $p - 1$ . Če v vmesnem času prispe stranka v čakalnico s prednostjo višjo kot  $p - 1$ , se strežba nadaljuje tam. Stranke v čakalnici z določeno prednostjo so zato lahko postrežene šele, ko so vse čakalnice z višjimi prednostmi prazne.

Predpostavimo, da prihajajoče stranke pripadajo enemu izmed  $P$  razredov, ki jih označimo z indeksom  $p$  ( $p = 1, 2, \dots, P$ ), kjer višja vrednost indeksa pomeni višjo prednost. Stranke iz razreda  $p$  so postrežene "bolje" kot stranke iz razreda  $p - 1$ . Za analizo definirajmo naslednje parametre in zveze med njimi [49]:

$$\lambda = \sum_{p=1}^P \lambda_p \quad (5.1)$$

<sup>5</sup>To pomeni, da prispela stranka iz razreda z višjo prednostjo ne more prekiniti potekajoče strežbe stranke iz razreda z nižjo prednostjo

$$\bar{x} = \sum_{p=1}^P \frac{\lambda_p}{\lambda} \bar{x}_p \quad (5.2)$$

$$\rho_p = \lambda_p \bar{x}_p \quad (5.3)$$

$$\rho = \lambda \bar{x} = \sum_{p=1}^P \rho_p \quad (5.4)$$

kjer  $\lambda_p$  pomeni hitrost prihodov strank iz razreda  $p$ ,  $\rho_p$  pomeni delež časa, ki ga strežnik porabi za obdelavo strank razreda  $p$  in  $\bar{x}_p$  povprečen čas, potreben za strežbo stranke iz razreda  $p$ . Analize se lotimo s stališča prihajajočih strank. Izberimo si eno opazovano stranko iz razreda  $p$ . Najprej lahko zapišemo povprečno zakasnitev  $W_0$ , ki jo opazovana stranka, vstopajoča v sistem, utрпи zaradi časa, ki je preostal do konca strežbe trenutno strežene stranke.

$$W_0 = \sum_{i=1}^P \rho_i \frac{\bar{x}_i^2}{2\bar{x}_i} = \sum_{i=1}^P \frac{\lambda_i \bar{x}_i^2}{2} \quad (5.5)$$

Povprečen čas prebit v čakalnici  $W_p$  za opazovano stranko iz razreda  $p$  pa lahko zapišemo kot

$$W_p = W_0 + \sum_{i=1}^P \bar{x}_i \bar{N}_{ip} + \sum_{i=1}^P \bar{x}_i \bar{M}_{ip} \quad (5.6)$$

kjer je  $\bar{x}_i$  povprečen čas strežbe stranke iz razreda  $i$ .  $\bar{N}_{ip}$  je poprečno število strank razreda  $i$  v sistemu ob prihodu opazovane stranke iz razreda  $p$ , ki bodo postrežene pred njo. To so stranke iz višjih prednostnih razredov ( $i > p$ ) in stranke istega prednostnega razreda, ki so že v sistemu.  $\bar{M}_{ip}$  pa je povprečno število strank iz višjih prednostnih razredov ( $i > p$ ), ki bodo prispele v sistem v času, ki ga bo opazovana stranka prebila v čakalnici in bodo postrežene pred njo.

Za strogo prednostno strežbo čakalnih vrst lahko zapišemo naslednje trditve:

$$\bar{N}_{ip} = 0 \quad i = 1, 2, \dots, p-1 \quad (5.7)$$

$$\bar{M}_{ip} = 0 \quad i = 1, 2, \dots, p \quad (5.8)$$

ki govori o tem, da nobena izmed strank iz nižjih prednostnih razredov ( $i < p$ ), ki so že v sistemu in nobena stranka iz prednostnega razreda  $i \leq p$ , ki bo v sistem prišla po opazovani stranki, ne bo postrežena pred njo. Sedaj pa s pomočjo Littlovega teorema določimo še povprečno število strank, ki bodo postrežene pred njo.

$$\bar{N}_{ip} = \lambda_i W_i \quad i = p, p+1, \dots, P \quad (5.9)$$

$$\bar{M}_{ip} = \lambda_i W_p \quad i = p+1, p+2, \dots, P \quad (5.10)$$

Prvi izraz predstavlja zatečeno stanje v prednostnih razredih ( $i \geq p$ ), katerih stranke bodo vsekakor postrežene pred opazovano, drugi pa določa povprečno število strank iz

prednostnih razredov ( $i > p$ ), ki bodo prispele v sistem v času  $W_p$  po opazovanju stranki ter bodo postrežene pred njo.  $W_p$  predstavlja povprečni čakalni čas opazovane stranke. Enačbo (5.6) sedaj lahko zapišemo kot

$$W_p = W_0 + \sum_{i=p}^P \bar{x}_i \lambda_i W_i + \sum_{i=p+1}^P \bar{x}_i \lambda_i W_p \quad p = 1, 2, \dots, P \quad (5.11)$$

iz katere lahko ob upoštevanju  $\rho_i = \bar{x}_i \lambda_i$  izrazimo  $W_p$

$$W_p = \frac{W_0 + \sum_{i=p+1}^P \rho_i W_i}{1 - \sum_{i=p}^P \rho_i} \quad p = 1, 2, \dots, P \quad (5.12)$$

Ta sistem enačb lahko rešimo rekurzivno in dobimo

$$W_p = \frac{\sum_{i=1}^P \frac{\lambda_i \bar{x}_i^2}{2}}{(1 - \sum_{i=p}^P \rho_i)(1 - \sum_{i=p+1}^P \rho_i)} \quad p = 1, 2, \dots, P \quad (5.13)$$

Ob upoštevanju relacije  $\sigma_p = \sum_{i=p}^P \rho_i$  pa lahko zapišemo še

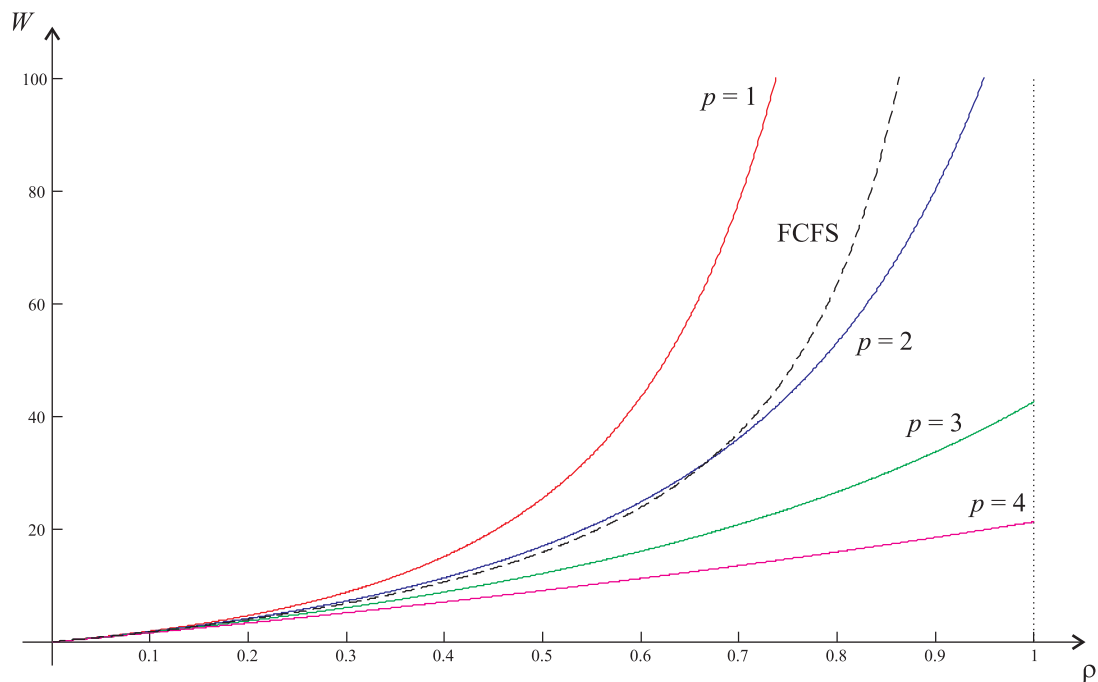
$$W_p = \frac{W_0}{(1 - \sigma_p)(1 - \sigma_{p+1})} \quad p = 1, 2, \dots, P \quad (5.14)$$

Povprečen čas prebit v sistemu pa izračunamo preprosto po enačbi  $T_p = W_p + \bar{x}_p$ , saj povprečnemu času prebitemu v čakalnici  $W_p$  dodamo le še povprečni strežni čas  $\bar{x}_p$ .

Grafični prikaz enačbe (5.13) je na sliki 5.5. Parametri za izračun so bili naslednji: povprečni čas strežbe vseh strank  $\bar{x} = 0.25$  s, povprečni čas strežbe strank posameznega prednostnega razreda  $\bar{x}_p = \bar{x}$ , število prednostnih razredov  $P = 4$ , povprečna hitrost prihodov v sistem  $0 \leq \lambda < 4$  s<sup>-1</sup>, povprečna hitrost prihodov v posamezni prednostni razred  $\lambda_p = \lambda/P$ , izkoriščenost sistema  $\rho = \lambda \bar{x}$  je bila  $0 \leq \rho < 1$ . Komentar: za primer paketnih omrežij pogoja  $\bar{x}_p = \bar{x}$  in  $\lambda_p = \lambda/P$  pomenita, da so imeli vsi prednostni razredi v povprečju enak promet. Iz slike vidimo, da so povprečni časi prebiti v čakalnici močno odvisni od prednostnega razreda. Podana je tudi primerjava s sistemom FCFS. V primerjavi z njim se ob nizko izkoriščenem sistemu prednostnima razredoma 3 in 4 čakalni čas zmanjša na račun razredov 1 in 2, ob visoko izkoriščenem sistemu pa tudi razred 2 doživi boljše razmere, seveda na račun razreda 1.

V sistemu M/G/1 s prednostjo dosežemo stacionarno stanje za vse čakalne vrste le, če velja, da v sistem prihaja manj strank kot jih ta lahko postreže. To zapišemo s pogojem:

$$\rho = \sum_{i=1}^P \rho_i < 1$$

Slika 5.5 – Povprečen čas prebit v čakalnici za razred  $p$  ob pogoju  $\rho < 1$ 

Zanimiva je še primerjava čakalnih časov za pakete v FCFS sistemu in obravnavanem sistemu. Opazimo, da je ob naraščajoči zasedenosti sistema le promet najnižjega prednostnega razreda na slabšem kot v FCFS sistemu, promet vseh drugih razredov pa pridobi. Če torej ustrezno razporedimo promet po prednostnih razredih, lahko precej izboljšamo razmere za določene vrste prometa, ki to potrebujejo. Pri tem pa najnižji prednostni razred napolnimo s prometom, ki ni občutljiv na večje zakasnitve. Seveda pa moramo paziti, da skupni promet ne preseže zmogljivosti sistema in je  $\rho < 1$ .

Oglejmo pa si še razmere za primer, ko je prihajajočih strank več kot jih je sistem zmožen obdelati ( $\rho \geq 1$ ), in je prikazan na sliki 5.6. Sistem tu ne postane nestabilen v celoti, ampak samo za prednostne razrede z nižjimi prednostmi.

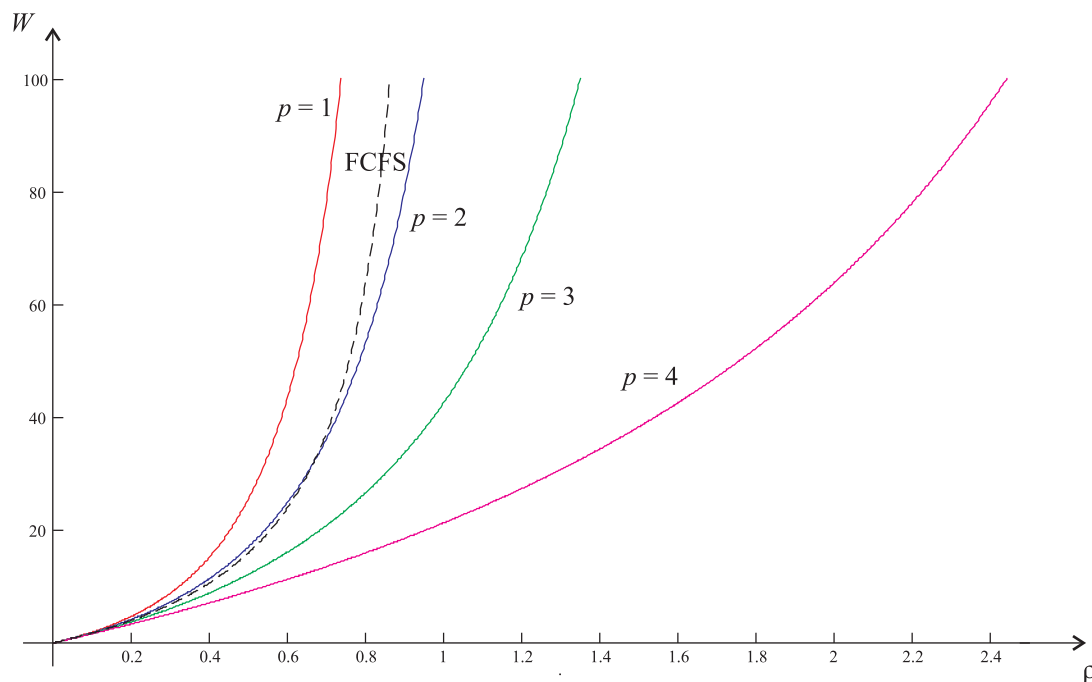
Če predpostavimo, da je  $\rho \geq 1$ , potem prednostni razred  $p$  lahko doseže stacionarno stanje pod pogojem, da velja  $\rho_p < 1$  in

$$\sum_{i=p}^P \rho_i < 1$$

V takem sistemu je  $P_s$  stacionarnih vrst.  $P_s$  je pozitivno celo število, za katero velja:

$$\sum_{i=P-P_s}^P \rho_i < 1 \leq \sum_{i=P-P_s-1}^P \rho_i \quad (5.15)$$

Če zopet pogledamo sliko 5.6, vidimo, da ko začne  $\rho$  naraščati preko svoje mejne vrednosti, začne naraščati tudi čakalni čas v čakalnih vrstah. V tistih z nižjo prednostjo ta čakalni čas hitro naraste v neskončnost. To pomeni, da paketi, ki pridejo v tako

Slika 5.6 – Povprečen čas prebit v čakalnici za razred  $p$ 

čakalno vrsto, nikoli ne bodo postreženi. Enako bi se zgodilo tudi s prometom v čakalni vrsti z najvišjo prioriteto, ko bi njen promet presegel zmognosti sistema pri  $\rho_P \geq 1$ .

### 5.5.2.2 Ohranitveni zakon

Po analizi čakalne vrste s prednostjo je to primerno mesto za kratko diskusijo o *ohranitvenem zakonu* (Conservation Law). V večini sistemov velja, da se nečesa ne more dobiti iz nič ali brez posledic za preostale elemente sistema. Nič drugače ni pri čakalnih vrstah. Če določeni stranki ali skupini strank podelimo prednostno obravnavo, lahko to, ob nespremenjeni kapaciteti sistema, naredimo samo na račun preostalih strank ali skupin v sistemu.

Ob predpostavki, da je sistem delovno intenziven, lahko za čakalno vrsto M/G/1 s prednostjo in brez prekinitev zapišemo naslednji izraz [50]

$$\sum_{p=1}^P \rho_p W_p = \begin{cases} \frac{\rho W_0}{1-\rho} & \rho < 1 \\ \infty & \rho \geq 1 \end{cases} \quad (5.16)$$

ki pravi, da utežena vsota povprečnih čakalnih časov  $W_p$  posameznih razredov ostaja konstantna. To pomeni, da če se čakalni čas za eno izmed skupin zmanjša, se morajo čakalni časi ostalih skupin ustrezno povečati. To pravilo velja neodvisno od načina strežbe čakalnih vrst, le funkcije zmanjšanja/povečanja čakalnih časov so za različne načine strežbe različne.

Pravkar zapisano dobro vidimo na slikah 5.5 in 5.6. Če za osnovo vzamemo sistem FCFS in predpostavimo, da so bile stranke tudi tam porazdeljene v razrede, ki pa so vsi imeli enako prednost, potem ob uvedbi različnih prednosti nekateri razredi pridobijo, drugi pa izgubijo, odvisno od izkoriščenosti sistema  $\rho$ .



### 5.5.2.3 Omrežje čakalnih vrst

Za omrežje čakalnih vrst (Network of queues) je v [49] dokazano, da ob predpostavki eksponentne porazdelitve med časi prihodov lahko obravnavamo vsak strežni sistem posebej. To je pomembna ugotovitev, saj tako lahko obravnavamo preproste posamezne sisteme in rezultate prenesemo na strežnike, ki so med seboj povezani v omrežje.

### 5.5.3 Kratak opis simulatorja omrežne naprave v omrežju DIS

Za simulacije zakasnitev smo razvili simulator omrežne naprave v omrežju DIS. Simulator je napisan v modularnem objektnem programskem jeziku MODSIM III podjetja CACI. Po sintaksi in nadzornih mehanizmih je podoben programskim jezikom Pascal, Modula-2 in Ada. MODSIM III je v osnovi objektni programski jezik, ki pa podpira tudi simulacije na osnovi diskretnih dogodkov. To pomeni, da ima vgrajene mehanizme, ki preko vgrajenega koledarja dogodkov (event calendar) nadzorujejo izvajanje določenih delov kode. Ob določenih dogodkih tako lahko izvajanje delov kode zaustavijo, poženejo ali za določen čas prekinejo. Za pravilen vrstni red pa poskrbi koledar dogodkov.

Naš simulator vključuje večino funkcionalnosti modela omrežne naprave, prikazane na sliki 4.2. Realizirani so: razvrščevalnik, krmiljenje prometa, čakalne vrste (dvostopenjske) in odpravnik. Poleg samega modela omrežne naprave pa programska koda vsebuje tudi module za generacijo prometa, module za zbiranje statističnih podatkov prometa in stanja omrežne naprave ter module za zapis rezultatov v datoteke s standardnim formatom. Za prometne izvore lahko izbiramo verjetnostno porazdelitev časov generacije in dolžin paketov. Pri izbiri odpravnika pa lahko izbiramo med več razvitimi odpravniki na prvi in drugi stopnji omrežja DIS. Zaenkrat so razviti odpravniki: FCFS, Strict Priority, RR, DRR in SCDRR.

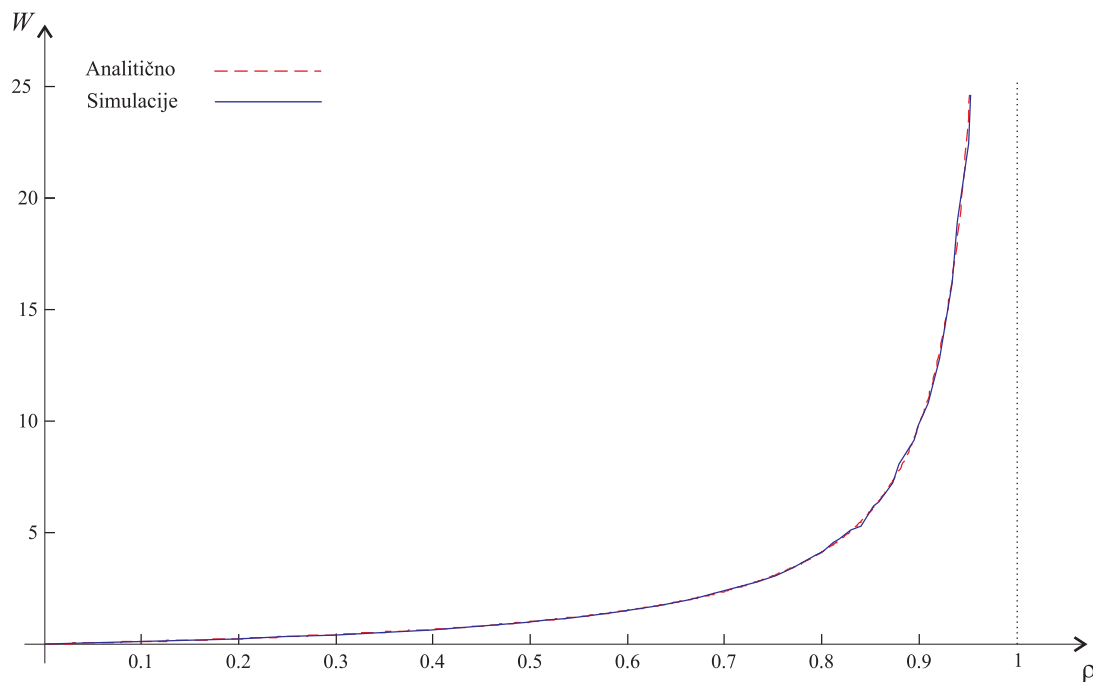
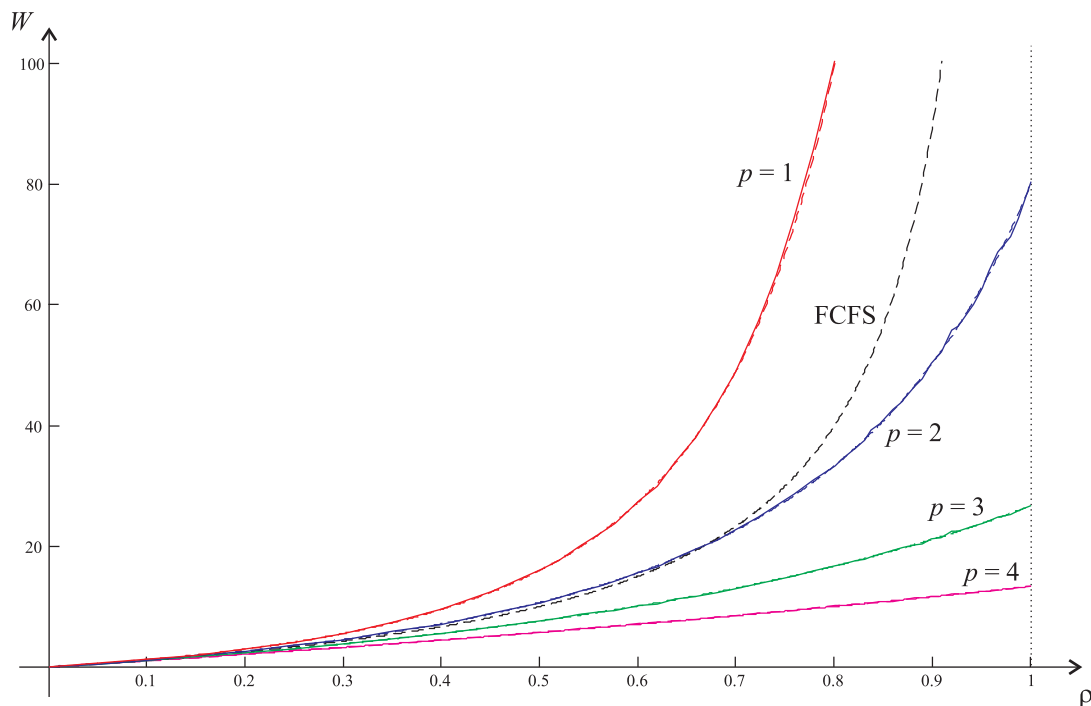
Simulator konfiguriramo preko vhodnih datotek, v katerih so zapisani podatki o prometnih izvori, konfiguraciji čakalnih vrst, načinu odpravljanja, vrsti in dolžini in parametrih simulacije ter načinu zbiranja in zapisa rezultatov. Simulacijski rezultati se preko makrojev avtomatsko prenesejo v pred-definiran format Excel datotek in grafov.

### 5.5.4 Simulacija diferenciacije prometa ter primerjava rezultatov z analitičnimi

Teoretične rezultate smo iz dveh razlogov preverili tudi s simulacijami. Prvi razlog je bil neposredna primerjava teoretičnih in simulacijskih rezultatov, drugi razlog pa je bil preizkus našega, zgoraj opisanega simulatorja čakalnega sistema za paketna omrežja na dobro poznanem in analitično obvladljivem čakalnem sistemu.

Najprej si pogledajmo najpreprostejši čakalni sistemu  $M/M/1$ , ki je opisan v dodatku B. Na sliki 5.7 je prikazana primerjava rezultatov za povprečen čas  $W$  prebit v čakalnici. Rezultati teoretične analize in simulacij se lepo ujemajo, odstopanja so praktično zanemarljiva. S tem smo potrdili predvsem pravilno delovanje prometnih izvorov (generatorjev paketov), pravilno delovanje samega simulatorja ter pravilnost postopkov izbire in izračuna simulacijskih parametrov.

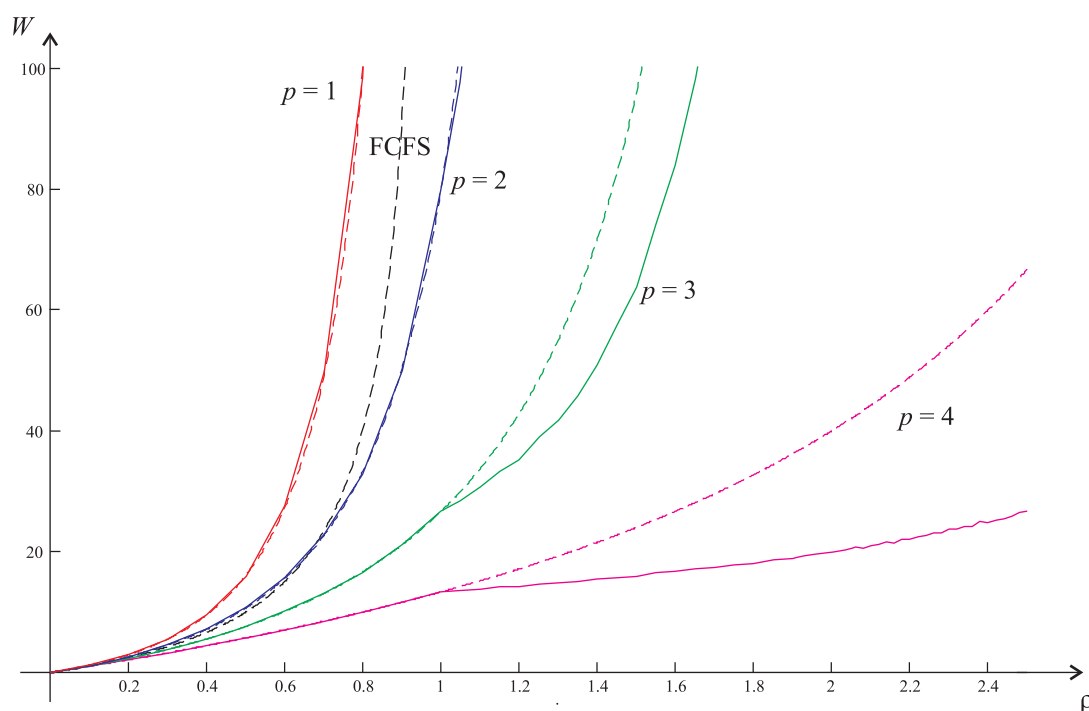
Sedaj se lahko posvetimo primerjavi čakalnega sistema s prednostjo. V razdelku 5.5.2.1 smo podali analitične rezultate za poljubno porazdelitev strežnih časov  $x$ . Iz

Slika 5.7 – Primerjava teoretičnih in simulacijskih rezultatov za povprečen čas  $W$  prebit v čakalniciSlika 5.8 – Primerjava teoretičnih in simulacijskih rezultatov za povprečen čakalni čas  $W$  pri  $0 < \rho < 1$ 

enačb (5.5) in (5.13) vidimo, da je ob izbrani izkoriščenosti sistema  $\rho = \sum_p \rho_p$  povprečen čas  $W_p$  prebit v čakalnici za posamezen prednostni razred  $p$  linearno odvisen le od povprečja kvadratov strežnih časov  $\bar{x}_p^2$  posameznih prednostnih razredov. Če zaradi enostavnejše izvedbe simulacij predpostavimo še, da je verjetnostna porazdelitev časov

strežbe enaka za vse prednostne razrede<sup>6</sup>, potem lahko zapišemo  $\bar{x} = \bar{x}_p$  in  $\overline{x^2} = \overline{x_p^2}$ . Povprečen čas strežbe in povprečje kvadratov strežnih časov sta za vse prednostne razrede enaka. Sedaj lahko enačbo (5.13) zapišemo v obliki

$$W_p = \frac{\sum_{i=1}^P \frac{\lambda_i \overline{x^2}}{2}}{(1 - \sum_{i=p}^P \rho_i)(1 - \sum_{i=p+1}^P \rho_i)} \quad p = 1, 2, \dots, P \quad (5.17)$$



Slika 5.9 – Primerjava teoretičnih in simulacijskih rezultatov za povprečen čakalni čas  $W$  pri  $0 < \rho < 2.5$

Če naredimo še korak dlje in predpostavimo, da je porazdelitev časov strežbe eksponentna<sup>7</sup>, potem veljajo naslednje zveze:  $\bar{x} = 1/\mu$  in  $\overline{x^2} = 2/\mu^2$ , pri čemer  $\mu$  predstavlja povprečno število postreženih strank v časovni enoti. Sliki 5.8 in 5.9 prikazujeta primerjavo med analitično izračunanimi in s simulacijami pridobljenimi čakalnimi časi za čakalni sistem z naslednjimi parametri: eksponentna porazdelitev časov strežbe, število prednostnih razredov  $P = 4$ , povprečno število postreženih strank v časovni enoti  $\mu = 0.1 \text{ s}^{-1}$ , povprečen čas strežbe  $\bar{x} = 1/\mu = 10 \text{ s}$  in povprečje kvadratov strežnih časov  $\overline{x^2} = 2/\mu^2 = 200 \text{ s}^2$ .

Izkoriščenost sistema  $\rho$  smo spreminjali z nastavitvijo skupne hitrosti prihodov v sistem  $\lambda$ , pri čemer so bile hitrosti prihodov za vse prednostne razrede enake  $\lambda_p = \lambda/P$ , kar nam da naslednjo zvezo  $\rho = \sum_p \rho_p = \sum_p \lambda_p \bar{x} = \lambda \bar{x}$ . Na sliki 5.8 so prikazane razmere

<sup>6</sup>S tem ne prizadenemo splošnosti rezultata, saj ničesar ne zanemarimo. Vzamemo le poseben primer, ki je nekoliko enostavnejši za simulacijo in lažje razumljiv.

<sup>7</sup>To pomeni, da je porazdelitev dolžin paketov eksponentna.

za  $0 < \rho < 1$ , pri katerih je sistem v celoti stabilen. Na sliki 5.9 pa so prikazane razmere za  $0 < \rho < 2.5$ , pri katerih je sistem za  $\rho > 1$  stabilen le za prednostne razrede z višjo prednostjo, prednostni razredi z nižjo prednostjo pa z višanjem  $\rho$  eden za drugim padajo v nestabilnost, pri čemer njihove čakalne vrste in zakasnitve rastejo v neskončnost. Delovanje sistema pri  $\rho > 1$  na daljši časovni rok ni smiselno, na sliki pa lepo vidimo, da so v tem primeru prizadeti zgolj nižji prednostni razredi. To pa smo s podeljevanjem prednosti na nek način tudi želeli doseči. Pri nizkih obremenitvah sistema razlike med prednostnimi razredi niso velike, z njegovo obremenjenostjo pa se te razlike hitro večajo.

### 5.5.5 Simulacija prve stopnje omrežja DIS

Doslej smo se, tako pri teoretični analizi kot pri simulacijah, ukvarjali s posameznimi strežnimi sistemi. Pri tem smo na vhodu sistema vzeli nek matematično obvladljiv pretok podatkov, npr. eksponentno gostoto porazdelitve verjetnosti časov med prihodi strank v sistem. Teoretični rezultati in simulacija prenosa podatkov brez upoštevanja vpliva višjih protokolnih plasti in upoštevanja drugih lastnosti omrežij dajo le približne rezultate. V primeru omrežij IP predvsem neupoštevanje lastnosti in delovanja protokola TCP, ki pomembno vpliva na pretok podatkov. Bolj popolno sliko nam dajo simulacije, pri katerih upoštevamo tudi prej zanemarjene lastnosti. Z njimi zajamemo več parametrov omrežja in se veliko bolj približamo realnim razmeram in rezultatom na konkretnem omrežju.

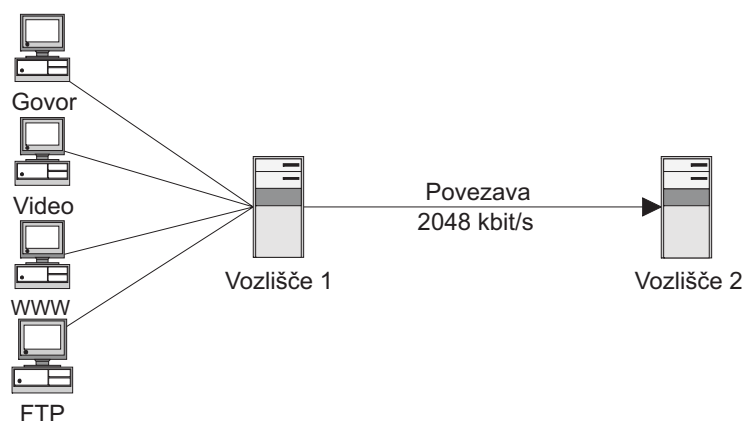
Zaradi obvladljivosti simulacije smo simulirali preprosto omrežje IP z nekaj končnimi napravami, ki tvorijo različne podatkovne pretoke, ter nekaj omrežnimi napravami. Pri tem smo upoštevali tudi vse protokolne plasti v končnih in omrežnih napravah. Iz istega razloga smo si za opazovanje izbrali le najpomembnejše simulacijske parametre. Simulirali smo po dveh osnovnih scenarijih. V prvem scenariju smo vzeli osnovno stanje omrežij IP s kakovostjo storitve po najboljših zmožnostih in FCFS strežbo. Te rezultate smo primerjali s simulacijami omrežja IP z diferenciranimi storitvami po drugem scenariju, pri čemer smo uporabili strogo prednostno strežbo čakalnih vrst. To ustreza simulaciji prve stopnje predlaganega omrežja DIS.

#### 5.5.5.1 Simulacijska shema in podatkovni pretoki

Za simulacijo smo si izbrali preprosto omrežno topologijo, ki je prikazana na sliki 5.10. Omrežje je sestavljeno iz štirih podatkovnih izvorov, dveh usmerjevalnikov in srednje zmogljive povezave med njima.

Za izvore podatkov smo izbrali štiri tipične aplikacije, ki se uporabljajo v omrežjih IP in so navedene v tabeli 5.3. Te aplikacije imajo zelo tipične podatkovne pretoke, ki so značilni tudi za druge njim podobne aplikacije. Karakteristike prometa so navedene v tabeli 5.4, promet aplikacije FTP smo modelirali s prenosi datotek dolžine 100 kB.

Usmerjevalnika sta imela omejene čakalne vrste, hitrost povezave med njima pa je bila 2048 kbit/s z zakasnitvijo razširjanja signala 10 ms. Povezava je delovala v načinu polni dupleks. Promet je tekkel od izvorov preko prvega usmerjevalnika in se je končeval v drugem usmerjevalniku, v nasprotno stran so potovali le potrditveni paketi protokolov, ki jih uporabljajo. Opazovali smo zakasnitve prometa v smeri od vozlišča 1 proti vozlišču 2.



Slika 5.10 – Topologija simulacijskega omrežja

	<b>Govor</b>	<b>Video</b>	<b>WWW</b>
Protokolni sklad	RTP/UDP/IP	RTP/UDP/IP	TCP/IP
Minimalna velikost paketa	30 oktetov	640 oktetov	10 oktetov
Maksimalna velikost paketa	30 oktetov	640 oktetov	1400 oktetov
Porazdelitev velikosti paketov	Konstantna	Konstantna	Enakomerna
Porazdelitev dolžine sporočil	Konstantna	Konstantna	Eksponentna

Tabela 5.4 – Karakteristike prometnih izvorov

### 5.5.5.2 Simulacijski scenariji in parametri

S simulacijskimi scenariji smo poskušali zajeti čimveč možnih situacij. Za določitev posameznega simulacijskega scenarija smo izbrali tri simulacijske parametre: obremenjenost povezave, delež prometa, ki zahteva prenos v realnem času in način strežbe čakalnih vrst.

Za obremenjenost povezave smo si izbrali vrednosti 50 %, 90 % in 100 %, ki ustrezajo vrednostim  $\rho = 0.5$ , 0.9 in 1. Prva pomeni simulacijo nizko obremenjenega omrežja. Tu ne pričakujemo nobenih težav, saj je omrežnih virov dovolj za vse. Obremenjenost omrežja pri 90 % že utegne povzročati določene težave. Čeprav je omrežnih virov v povprečju dovolj za vse, pa se podatki pojavljajo v izbruhih, kar pomeni podaljšanje čakalnih vrst in s tem tudi višje povprečne zakasnitve paketov. Simulacija 100 % zasedene povezave predstavlja zasičeno omrežje, ki ravnokar omenjene težave še stopnjuje. Čakalne vrste se močno povečajo in zapolnijo. To pomeni še višje zakasnitve, več zavrženih paketov in močno poslabšanje razmer v omrežju.

S spreminjanjem deleža prometa aplikacij, ki zahtevajo prenos v realnem času, želimo zajeti čimveč možnih prometnih kombinacij. Ta delež se spreminja v območju od 10 % do 90 % in pomeni delež prometa v realnem času glede na celotni promet na povezavi in ne delež prometa v realnem času glede na zmogljivost povezave. Če za primer vzamemo 50 % obremenjenost povezave in 30 % delež prometa v realnem času, potem to pomeni, da promet v realnem času zaseda 15 % zmogljivosti povezave.

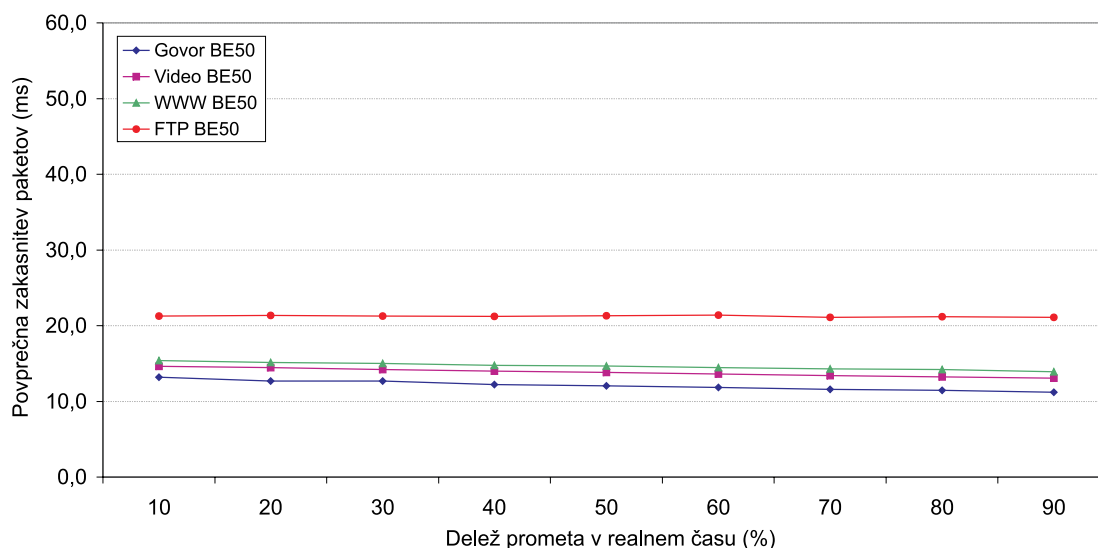
Simuliramo dva različna odpravnika. V prvem scenariju simuliramo osnovno stanje omrežij IP z FCFS odpravnikom, ki ga označimo z BE (Best Effort - po najboljših

zmožnostih). V drugem pa uporabimo strogo prednostno odpravljanje in ga označimo z DS (differentiated services). S tem v omrežje vpeljemo diferenciacijo storitev.

Scenarije smo predstavili grafično za vsako vrsto prometa posebej in jih označili glede na mehanizem strežbe čakalnih vrst in obremenjenost povezave. V vsakem grafu so skupaj predstavljeni vsi deleži realnočasovnega prometa. Na primer, scenarij z oznako BE90 označuje osnovno stanje omrežij IP z FCFS strežbo pri 90 % obremenjenosti povezave.

### 5.5.5.3 Simulacija osnovnega stanja omrežij IP

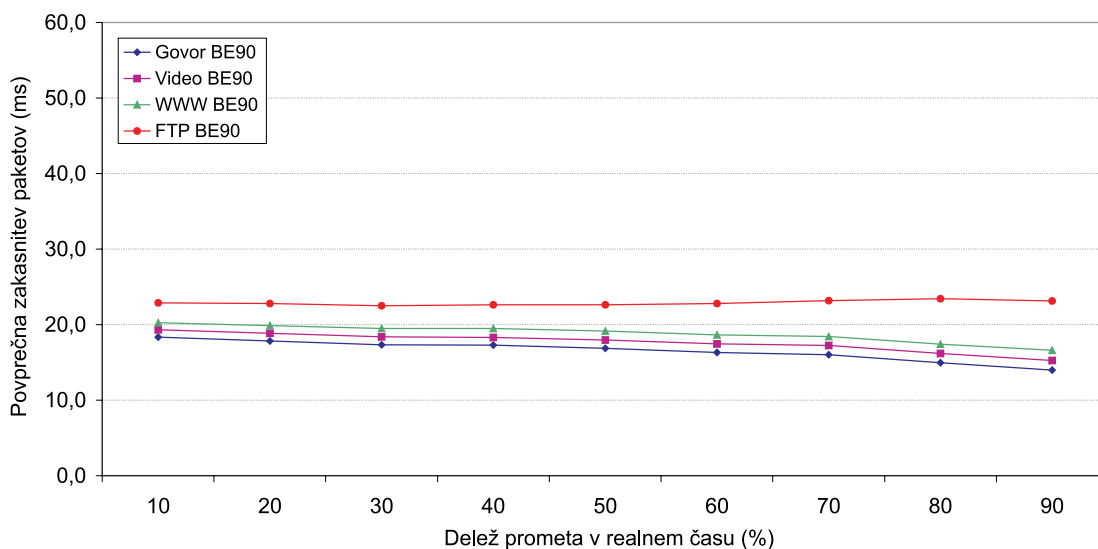
Scenariji BE50, BE90 in BE100 uporabljajo FCFS strežbo čakalnih vrst. Paket, ki se prvi uvrsti v čakalno vrsto, jo bo tudi prvi zapustil. To je zelo demokratičen način deljenja omrežnih virov, saj so vsi enakopravni. Težave pa nastanejo, ko določen delež prometa potrebuje več (ali manj) kot ostali, saj mu omrežje tega ne more zagotoviti. Rezultati simulacij teh scenarijev so prikazani na slikah 5.11 do 5.13.



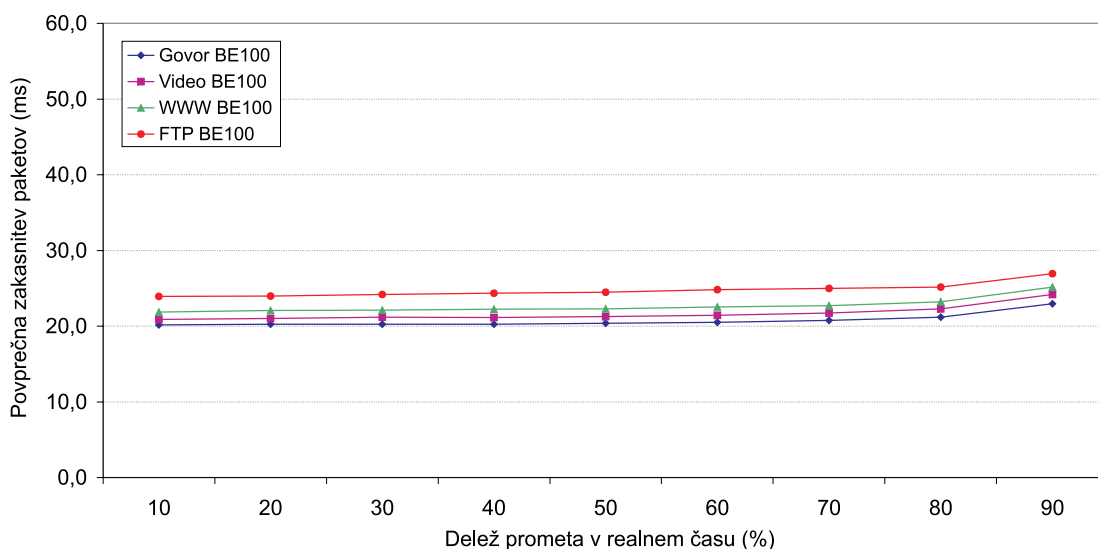
Slika 5.11 – Povprečne zakasnitve paketov za scenarij BE50

Rezultati simulacij prikazujejo povprečne vrednosti zakasnitev paketov pri 50 %, 90 % in 100 % obremenjeni povezavi, ki se med seboj skoraj ne razlikujejo. Zakasnitve so seveda najmanjše pri najnižji in največje pri najvišji obremenitvi povezave. To je posledica daljših čakalnih vrst. Zakasnitve se le malo razlikujejo tudi glede na aplikacijo. Največje so pri FTP, kar je posledica povprečno precej daljših paketov kot pri drugih aplikacijah. Iz istega vzroka imajo najmanjše zakasnitve paketi govora. Zakasnitve se tudi ne razlikujejo dosti v odvisnosti od deleža prometa v realnem času. To je zopet razumljivo, saj le ta ni obravnavan nič drugače kot preostali promet.

Seveda pa nas najbolj zanima na kakšne razmere v omrežju naleti najzahtevnejša aplikacija - govor. Tu je poleg povprečnih zakasnitev paketov govora pomembna tudi njihova porazdelitev. Kriteriji in meje za še zadovoljiv prenos govora so bili podrobno opisani in podani v [23]. Tu naj omenimo le najpomembnejše. Pri prenosu kodiranega govora je 5 % meja izgube paketov še za silo sprejemljiva. To pomeni, da je meja



Slika 5.12 – Povprečne zakasnitve paketov za scenarij BE90



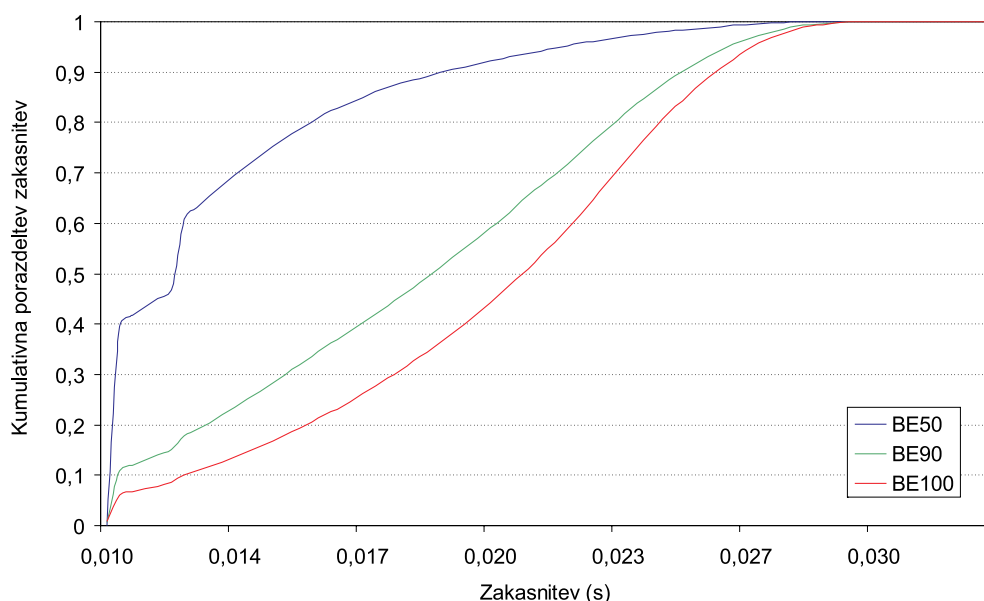
Slika 5.13 – Povprečne zakasnitve paketov za scenarij BE100

zakasnitve paketov govora omejena z zakasnitvijo, znotraj katere jih na cilj prispe vsaj 95 %. Ta meja pa je lahko precej višja kot povprečna zakasnitev teh paketov.

Tabela 5.5 kaže razlike med povprečno zakasnitvijo  $\bar{x}$  in 95 % zakasnitvijo, ki je v vseh primerih precej višja, v nekaterih primerih skoraj za dvakrat. Če tu upoštevamo še, da je zakasnitev povezave 10 ms, potem vidimo, da je spremenljivi del zakasnitve pri 95 % zakasnitvi tudi do več kot štirikrat večji od spremenljivega dela povprečne zakasnitve. Za mejo prispetja 99 % vseh paketov pa se ti odnosi še poslabšajo. Zaključimo lahko, da za prenos govora po BE scenariju povprečna zakasnitev paketov ni primeren kriterij. Bolj primerna mera je kumulativna porazdelitev zakasnitev, ki je za opisane scenarije prikazana na slikah 5.14 do 5.16. Ravno tako pa lahko zaključimo, da omrežja, ki uporabljajo FCFS strežbo čakalnih vrst, niso primerna za prenos podatkov zahtevnih aplikacij. Zakasnitve paketov so preveč spremenljive in preveč odvisne od

	10 % RT			50 % RT			90 % RT		
	$\bar{x}$	95 %	99 %	$\bar{x}$	95 %	99 %	$\bar{x}$	95 %	99 %
BE50	13.2	22.1	26.1	12.1	18.8	24.8	11.2	13.2	19.5
BE90	18.4	26.4	28.1	16.9	25.4	27.4	14.0	23.1	26.1
BE100	20.2	27.1	28.4	20.4	27.1	28.8	23.0	27.4	28.4

Tabela 5.5 – Razlika med povprečno zakasnitvijo paketov govora in zakasnitvijo znotraj katere pri BE scenarijih na cilj prispe vsaj 95 % ali 99 % govornih paketov [ms]



Slika 5.14 – Kumulativna porazdelitev zakasnitev govora za 10 % delež prometa v realnem času

stanja omrežja.

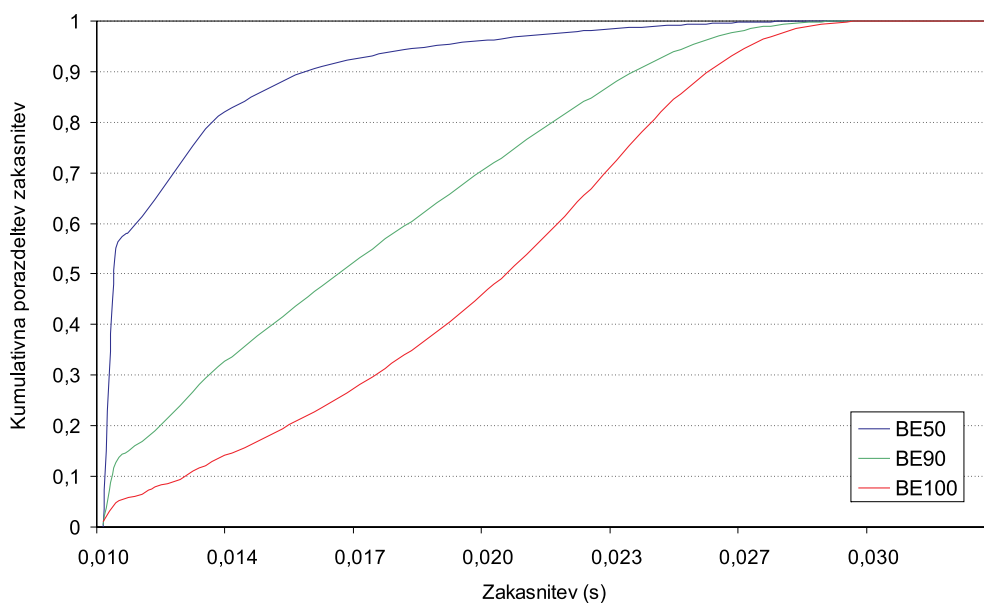
#### 5.5.5.4 Uvedba različnih stopenj kakovosti storitve

Razmere za zahtevnejše podatkovne pretoke lahko izboljšamo tako, da jih razvrstimo v storitvene razrede. Omrežje mora posameznim razredom nuditi različne prenosne pogoje in s tem različne stopnje kakovosti storitve. Simulacijski scenarij uvede štiri prometne razrede, ki jih strežemo po njihovi prednosti. Simulirali smo promet štirih tipičnih aplikacij, ki so našteje v tabeli 5.3 in opisane v razdelku 5.5.1, njihove prometne karakteristike pa so podane v tabeli 5.4.

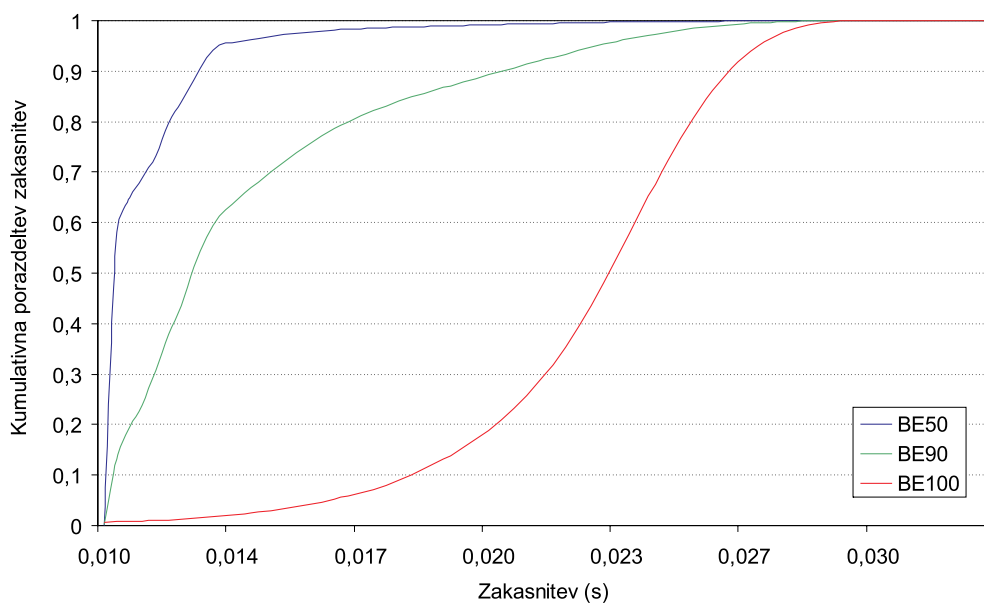
V omrežnih elementih posamezne pretoke razporedimo vsakega v svojo čakalno vrsto, ki jih strežemo strogo prednostno. Simulacije so potekale enako kot pri BE scenarijih in z enakimi simulacijskimi parametri, razlika je bila le v strežbi čakalnih vrst. Rezultati so prikazani na slikah 5.17 do 5.19.

Opazimo, da so povprečne zakasnitve paketov pri scenariju DS50 skoraj identične tistim v scenariju BE50. To je bilo tudi pričakovati, saj v napol praznem omrežju zagotavljanje različnih stopenj kakovosti storitve nima pravega učinka. Vse aplikacije





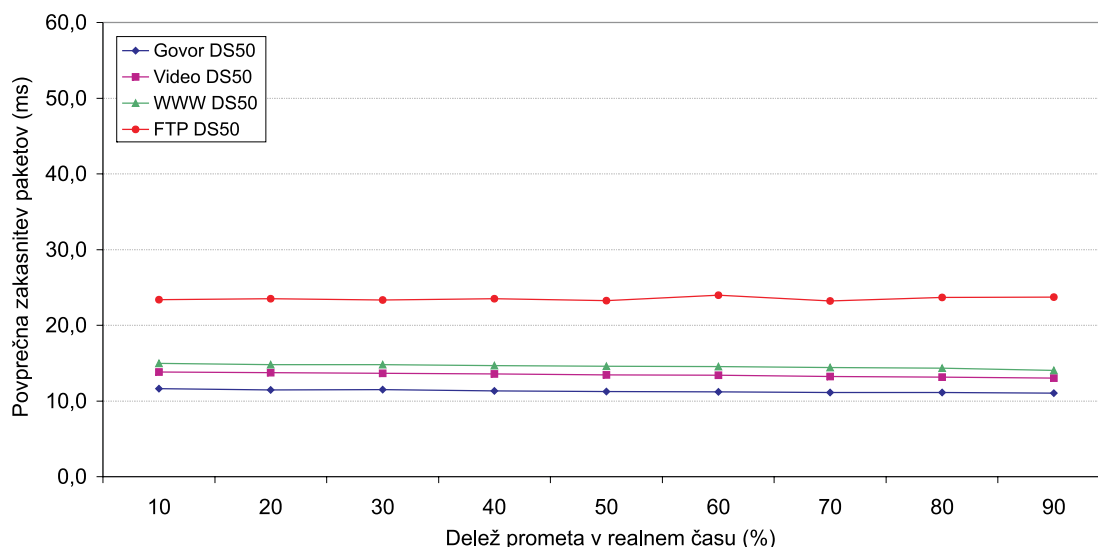
Slika 5.15 – Kumulativna porazdelitev zakasnitev govora za 50 % delež prometa v realnem času



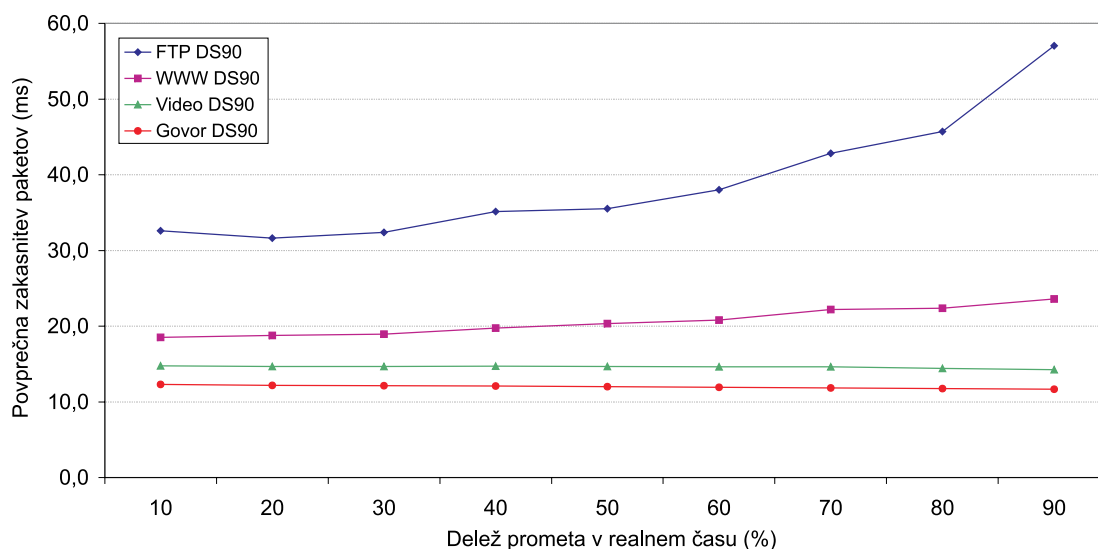
Slika 5.16 – Kumulativna porazdelitev zakasnitev govora za 90 % delež prometa v realnem času

imajo na voljo dovolj omrežnih virov in jim zanje ni potrebno tekrovati.

Precej drugačna pa je primerjava scenarijev DS90 in BE90 ter DS100 in BE100. Tu opazimo, da so se zakasnitve aplikacij razslojile glede na njihovo prednost. Zakasnitve paketov govora, videa in deloma tudi WWW padejo, močno pa se poveča zakasnitev FTP paketov, ki imajo najnižjo prednost. Pri scenariju DS90 je opazno znižanje zakasnitev govora in videa. WWW paketi svoje zakasnitve znižajo le pri nizkih deležih prometa v realnem času, medtem ko pri višjih deležih omenjenega prometa nji-



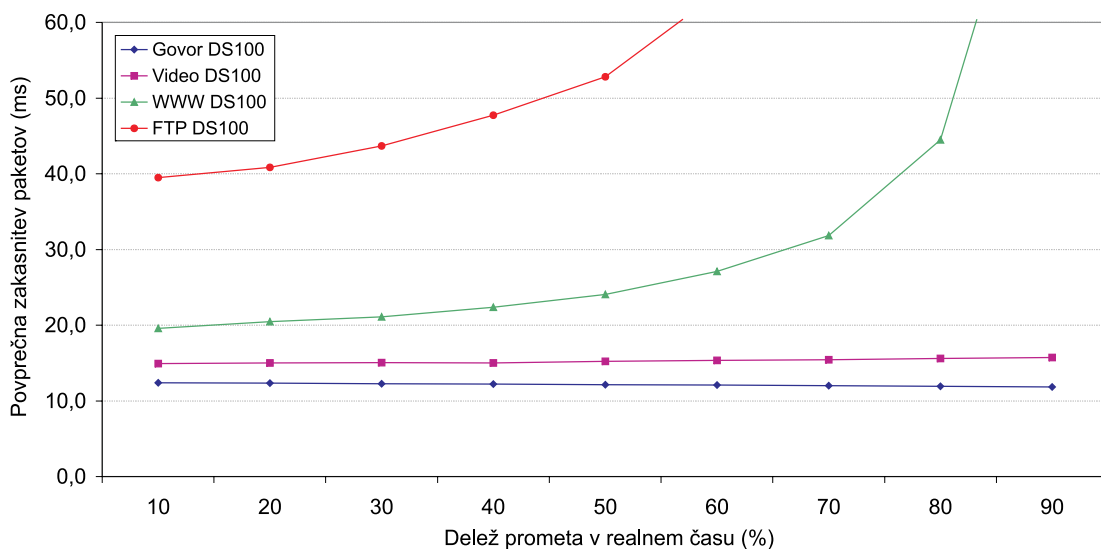
Slika 5.17 – Povprečne zakasnitve paketov za scenarij DS50



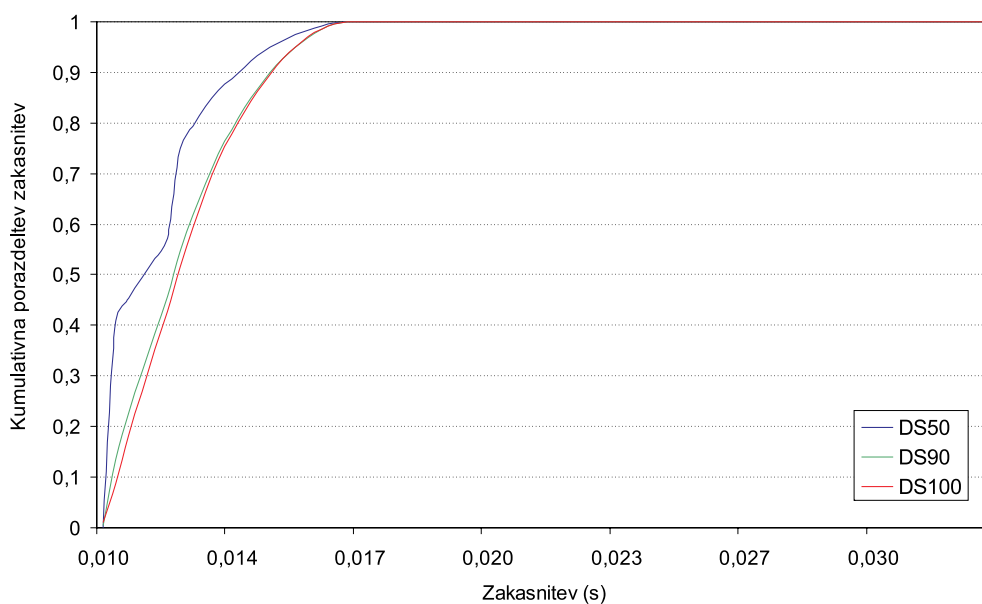
Slika 5.18 – Povprečne zakasnitve paketov za scenarij DS90

hova zakasnitev glede na BE90 nekoliko naraste. To je razumljivo, saj ima veliko paketov govora in videa prednost pred njimi. Scenarij DS100 (zasičenje omrežja) še poudari pomembnost prednosti. Govor in video skoraj ne utrpita dodatnih zakasnitev, za WWW pakete te nekoliko narastejo, zakasnitve FTP paketov pa se povzpnejo čez vse meje, kar smo glede na rezultate teoretične analize tudi pričakovali.

Če si zopet podrobneje ogledamo razmere za govor, vidimo, da so se absolutne povprečne zakasnitve močno zmanjšale. Ravno tako se je zmanjšala absolutna meja zakasnitve, v kateri na cilj prispe 95 % ali 99 % govornih paketov. Napram povprečni zakasnitvi se je nekoliko zmanjšalo tudi razmerje med spremenljivimi deli zakasnitev pri 95 % ali 99 % meji (tabela 5.6). Kumulativne porazdelitve zakasnitev pa so prikazane na slikah 5.20 do 5.22. Zaključimo lahko, da je DS shema precej bolj primerna za prenos podatkov zahtevnih aplikacij. Vendar pa se moramo zavedati, da se izboljšanje razmer



Slika 5.19 – Povprečne zakasnitve paketov za scenarij DS100

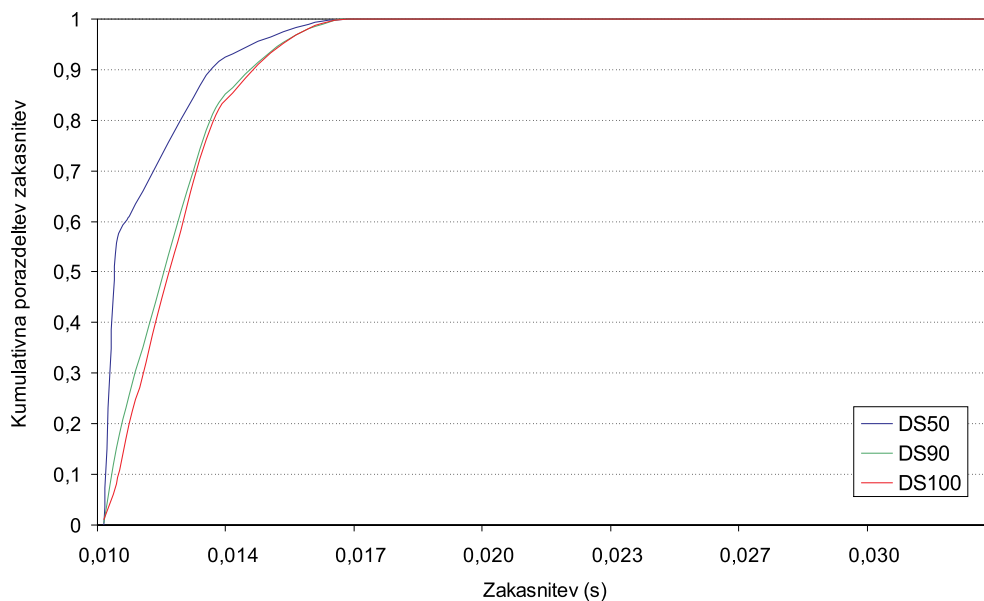


Slika 5.20 – Kumulativna porazdelitev zakasnitvev govora za 10 % delež prometa v realnem času

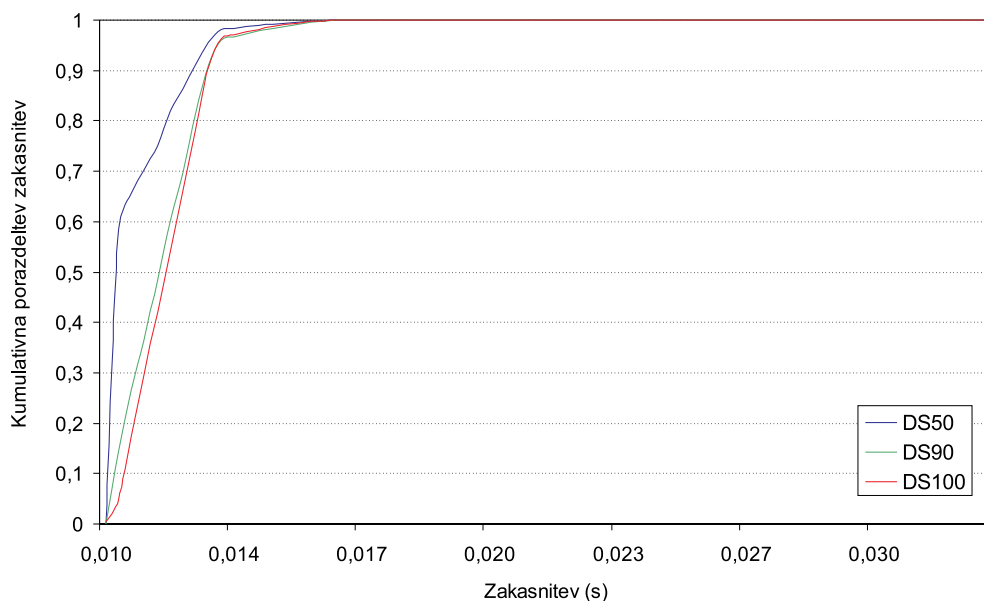
zgori na račun aplikacij, ki imajo nižjo prednost. Tem se razmere precej poslabšajo.

Na slikah 5.23 do 5.25 pa so prikazane razlike med povprečnimi zakasnitvami DS scenarijev napram istim BE scenarijem. Pokaže se, da pri 50 % obremenitvi omrežja razlike niso velike. Pridobijo vse aplikacije razen FTP. Pri 90 % obremenitvi omrežja se te razlike nekoliko povečajo, pri višjih deležih prometa v realnem času pa trpi tudi WWW. V zasičenju (100 % obremenitev omrežja) pa se te razlike še bolj poudarijo in se predvsem za FTP povečajo preko vseh razumnih mej.

Postavi se vprašanje, ali so pridobitve aplikacij z višjo prednostjo večje kot izgube aplikacij z manjšo prednostjo. Večkrat smo že omenili, da elastične aplikacije ne trpijo preveč zaradi spremenljivih razmer v omrežju. Tudi njihovo poslabšanje jih ne



Slika 5.21 – Kumulativna porazdelitev zakasnitev govora za 50 % delež prometa v realnem času



Slika 5.22 – Kumulativna porazdelitev zakasnitev govora za 90 % delež prometa v realnem času

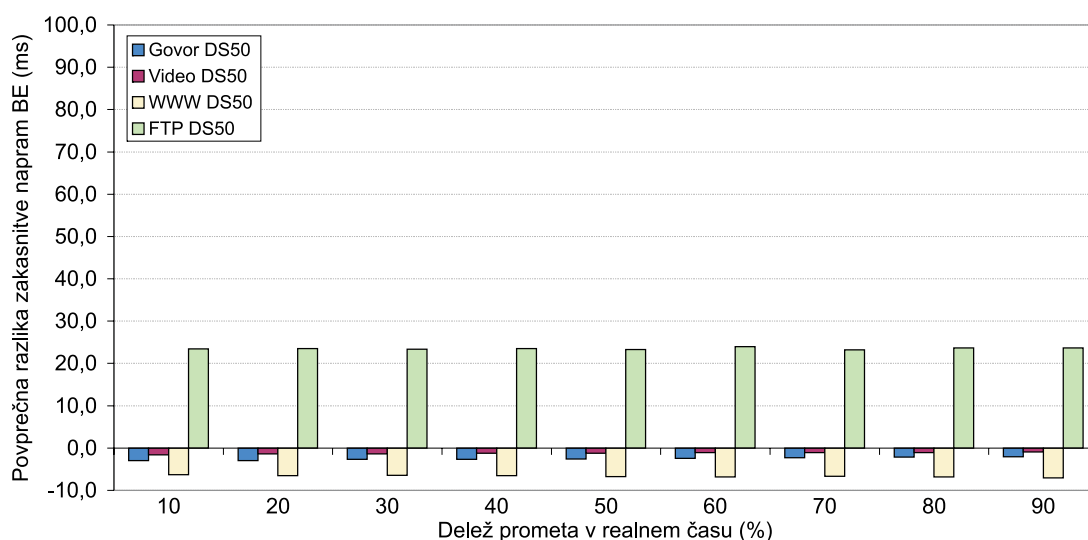
prizadene tako hudo kot toge aplikacije. Pogledjmo si to na primeru govora in FTP.

Pokazali smo že, da razmere za govor v BE shemi niso preveč primerne in da se z uvedbo DS močno popravijo. Tu smo gledali predvsem zakasnitve (povprečne vrednosti in njihove porazdelitve), ki so za nas najpomembnejši parameter kakovosti storitve pri prenosu govora. Pri FTP zakasnitve posameznih paketov niso najpomembnejše, bolj pomembna je propustnost in čas prenosa datoteke.

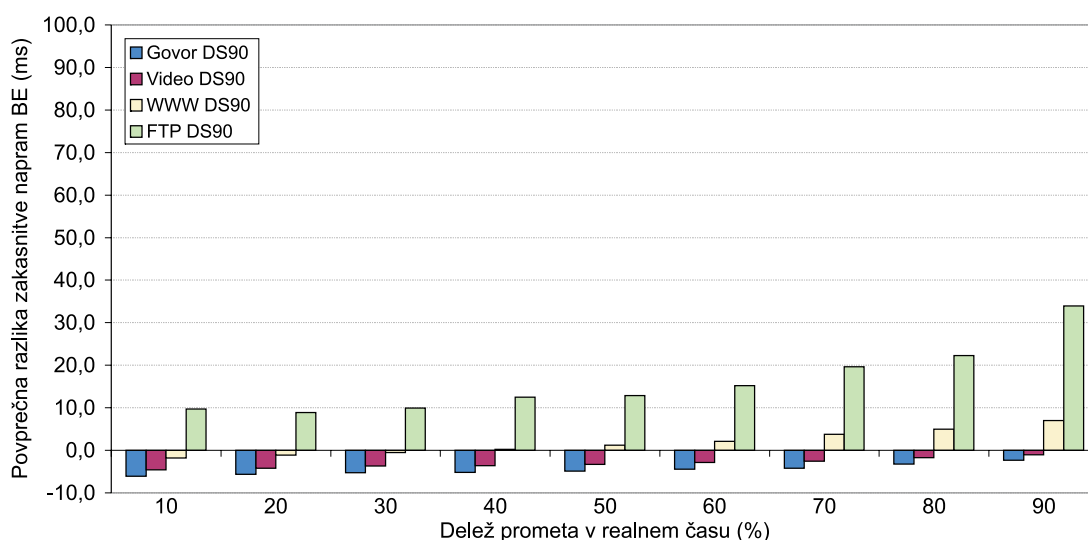
Na sliki 5.26 vidimo, da se propustnost omrežja za pakete FTP aplikacije skoraj ne

	10 % RT			50 % RT			90 % RT		
	$\bar{x}$	95 %	99 %	$\bar{x}$	95 %	99 %	$\bar{x}$	95 %	99 %
DS50	11.6	14.9	15.8	11.3	14.2	15.5	11.0	12.9	14.2
DS90	12.3	15.2	16.2	12.0	14.9	15.8	11.7	13.2	15.2
DS100	12.4	15.2	16.2	12.2	14.9	15.8	11.8	13.2	14.9

Tabela 5.6 – Razlika med povprečno zakasnitvijo paketov govora in zakasnitvijo znotraj katere pri DS scenarijih na cilj prispe vsaj 95 % ali 99 % govornih paketov [ms]

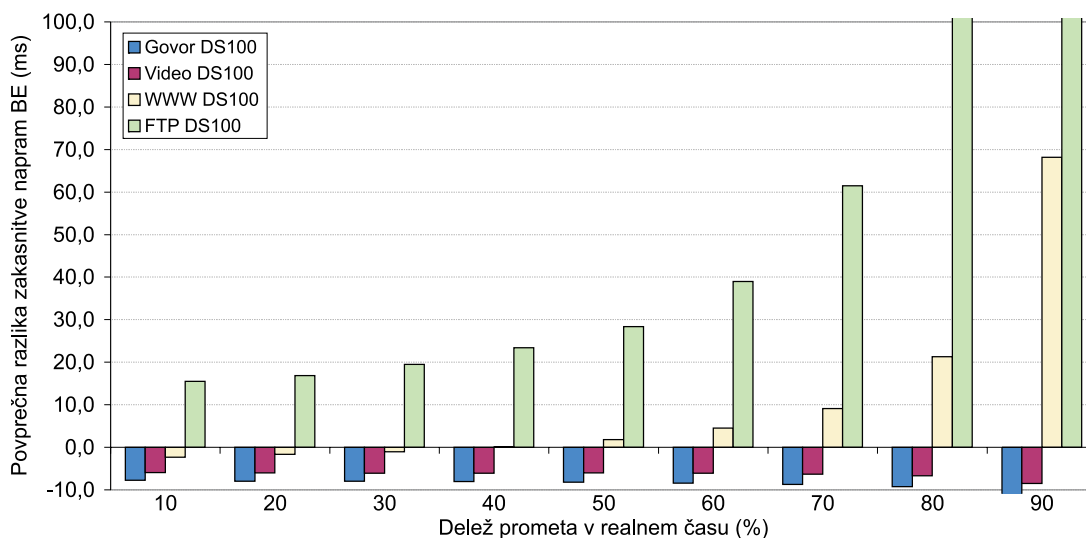


Slika 5.23 – Razlika med povprečno zakasnitvijo DS napram BE za 50 % obremenitev omrežja

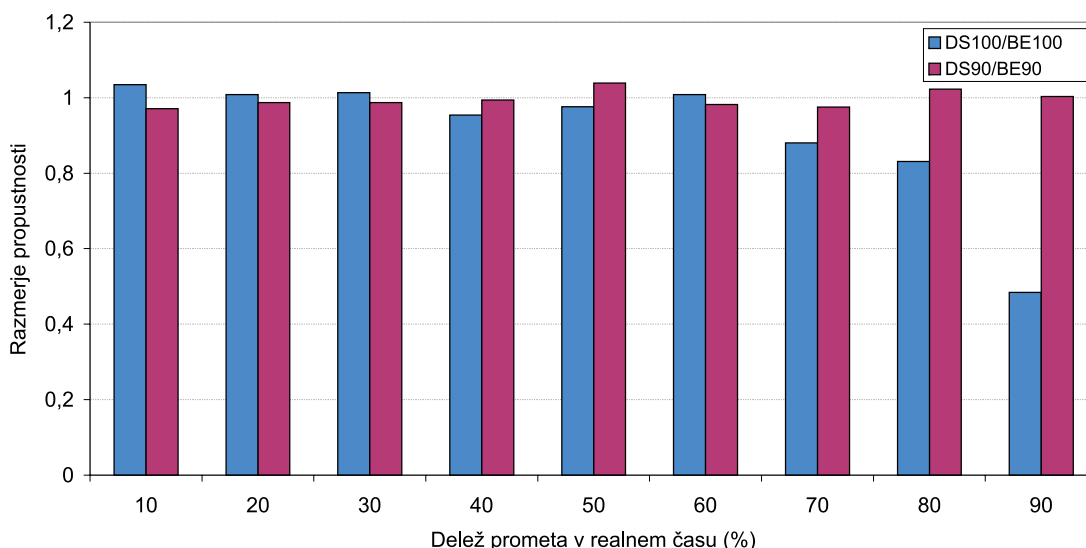


Slika 5.24 – Razlika med povprečno zakasnitvijo DS napram BE za 90 % obremenitev omrežja

spreminja in ostaja visoka. To še posebej velja za 90 % zasedeno omrežje, pri zasičenem omrežju pa propustnost pade šele pri zelo visokih deležih realnočasovnega prometa, ki pa je v realnih omrežjih malo verjeten. Slika 5.27 pa nam prikazuje čase prenosa 100



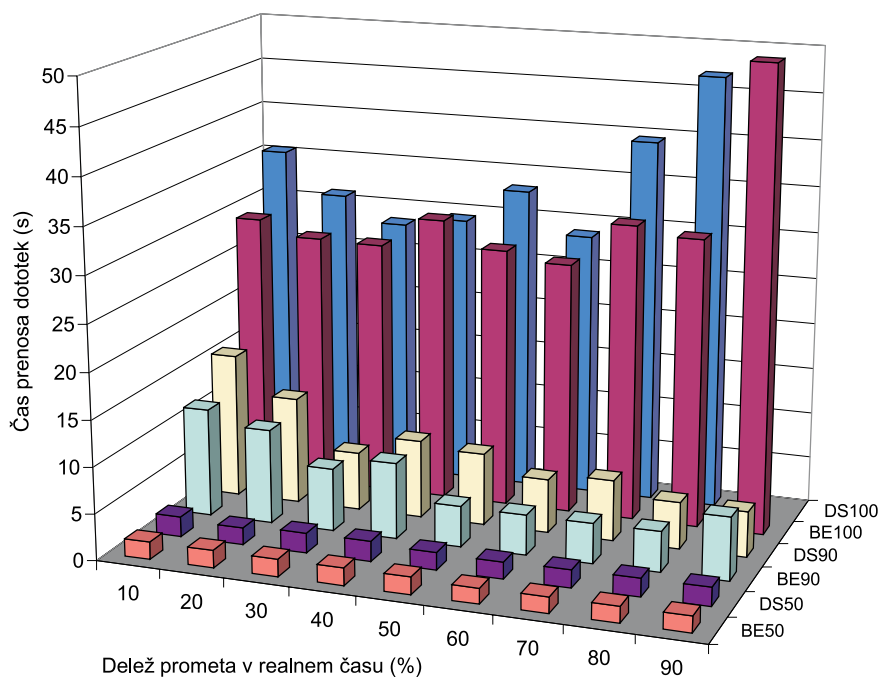
Slika 5.25 – Razlika med povprečno zakasnitvijo DS napram BE za 100 % obremenitev omrežja



Slika 5.26 – Propustnost visoko obremenjenega omrežja pri prenosu FTP paketov za scenarije BE90, BE100, DS90 in DS100

kB datoteke za vse simulirane scenarije. Opazimo, da se časi prenosa ene datoteke pri enaki obremenitvi omrežja skoraj ne razlikujejo. To pomeni, da diferenciacija in uvedba prednosti prenosa FTP datotek skoraj ne prizadene, pa čeprav smo jim podelili najnižjo prednost.

Na slikah 5.23 do 5.25 se zdi, da je poslabšanje zakasnitve za FTP veliko večje kot izboljšanje zakasnitev za govor. To se zdi še posebej očitno pri velikih deležih prometa v realnem času. Vendar pa je potrebno upoštevati, da pri 90 % prometa v realnem času FTP promet predstavlja le okoli 5 % odstotkov, govor pa kar okoli 45 % prometa na omrežju. Majhno izboljšanje zakasnitev za velik delež prometa je tako lahko pomembnejše od velikega poslabšanja za majhen delež prometa.



Slika 5.27 – Časi prenosov FTP datotek za vse scenarije BE in DS, pri različnih deležih prometa v realnem času

### 5.5.6 Komentar rezultatov

Iz rezultatov teoretične analize, še bolj pa iz rezultatov simulacij realnega sistema lahko zaključimo, da diferenciacija prometa na omrežju ugodno vpliva na prenosne parametre posameznih prednostnih razredov. Potrebna je previdna izbira prometnih pretokov, ki jih uvrščamo v posamezen prednostni razred. Obstaja namreč možnost, da se prenosni parametri prometov nižjih prednostnih razredov precej poslabšajo zaradi presežkov prometov višjih prednostnih razredov. V primeru našega predloga združevanja omrežij z DS in IS je ta možnost zelo majhna, saj je promet najvišjih prednostnih razredov omejen in nadzorovan ter ob neobvladanih pretokih po potrebi tudi preoblikovan.

Čeprav je stopnja kakovosti storitve, dobljena s strogo prednostno strežbo, relativna, nas to ne moti preveč, saj je storitev najvišjega prednostnega razreda vedno najboljša možna. Najvišji prednostni razred ima namreč na voljo celotno kapaciteto omrežja. Tu se po našem predlogu nahajajo zagotovljene storitve. Drugi prednostni razred ima še vedno zelo dobre razmere, kar ustreza storitvam nadzorovanega bremena. Ob predpostavki, da sta deleža zgornjih prometov na omrežju omejena, pa imajo tudi nižji prednostni razredi zadosti dobre razmere. Seveda dokler se držijo znotraj deleža kapacitete, ki jim po dogovoru pripada.

Iz rezultatov teoretične analize in simulacij zaključimo, da je strogo prednostna strežba primerna za izvedbo prve stopnje delovanja omrežja DIS po našem predlogu iz prejšnjega poglavja. Pogoje in rešitve za drugo stopnjo pa obravnavamo v naslednjem razdelku.

## 5.6 Analiza delovanja druge stopnje omrežja DIS

V prejšnjem razdelku smo pokazali, da lahko z uvedbo prednostnih razredov zagotovimo primerno diferenciacijo storitev v paketnih omrežjih. To smo pokazali tudi na primeru simulacije diferenciacije storitev na omrežju IP in zaključili, da je strogo prednostna strežba primerna za izvedbo prve stopnje omrežja DIS.

Za izvedbo druge stopnje omrežja DIS pa zahtevamo, da je odpravnik znotraj posameznega razreda zmožen razlikovati posamezne ali združene podatkovne pretoke ter jim zagotoviti pravičnost in nizko algoritemsko zakasnitev. Poleg tega pa mora biti tak odpravnik nezahteven: enostavno izvedljiv, nepotraten z viri in računsko nezahteven. Pri izbiri odpravnika moramo vedno narediti kompromis med pravičnostjo, algoritemsko zakasnitvijo in zahtevnostjo. Odpravniki, ki so pravični in imajo nizko algoritemsko zakasnitev, so navadno zahtevni, in obratno.

Kot zelo primeren kandidat za izvedbo druge stopnje omrežja DIS se izkaže odpravnik DRR. Ima računsko zahtevnost  $O(1)$ , ni potraten z viri, njegova pravičnost (glej tabelo 6.1) pa je primerljiva z zahtevnejšimi odpravniki kot sta SCFQ in WF<sup>2</sup>Q. Zaradi načina delovanja po principu časovnih okvirov pa ima nekaj slabšo algoritemsko zakasnitev kot nekateri primerljivi odpravniki (glej tabelo 6.2).

V tem razdelku s simulacijami pokažemo, da je odpravnik DRR primeren za zastavljeno nalogo. Najprej simuliramo samostojno delovanje odpravnika DRR in pokažemo, da je zmožen razlikovati posamezne pretoke in jim zagotoviti pravičnost, tudi ob prisotnosti neobvladanega pretoka. Rezultate primerjamo z FCFS odpravnikom. V nadaljevanju pa odpravnik DRR uvrstimo na drugo stopnjo omrežja DIS in s simulacijami pokažemo, da je odpravnik DRR primeren tudi za to nalogo.

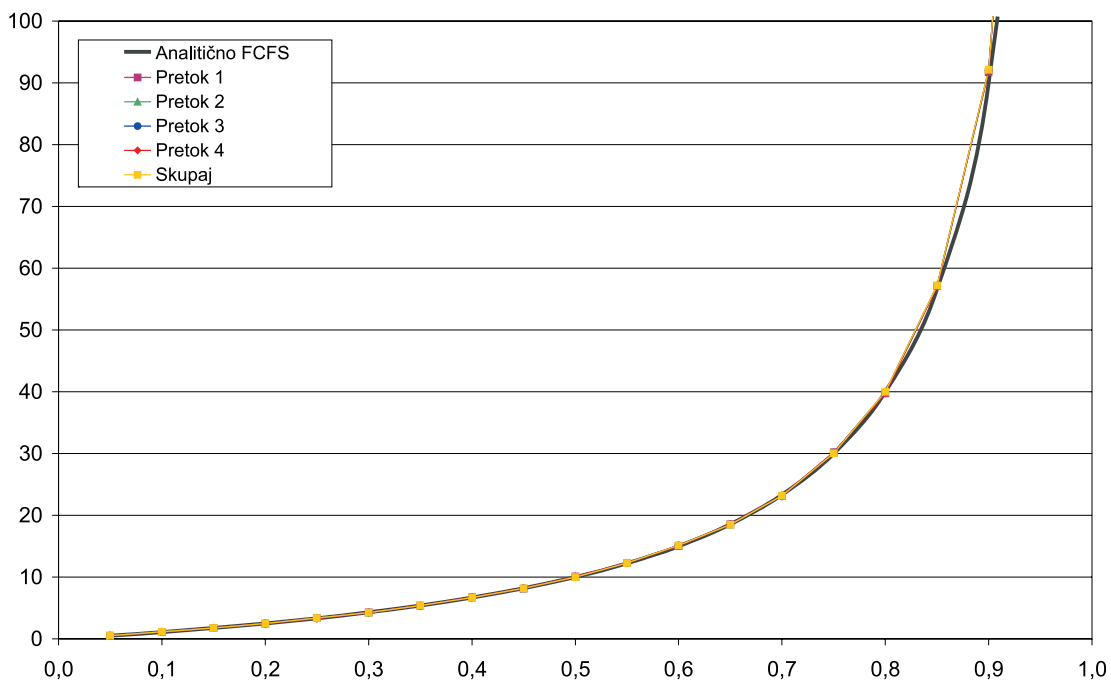
### 5.6.1 Simulacija samostojnega delovanja odpravnika DRR

Pri samostojnem delovanju simuliramo delovanje odpravnika DRR, kot je bilo opisano v razdelku 4.5.6. Vsakemu podatkovnemu pretoku  $i$  preko velikosti kvanta  $Q_i$  dodelimo določen del bitne hitrosti izhodne linije in ga uvrstimo v svojo čakalno vrsto. Če se vsi pretoki držijo dogovorjenih prenosnih parametrov (ne oddajajo več kot je bilo dogovorjeno), pričakujemo enako povprečno zakasnitev paketov kot pri FCFS odpravniku.

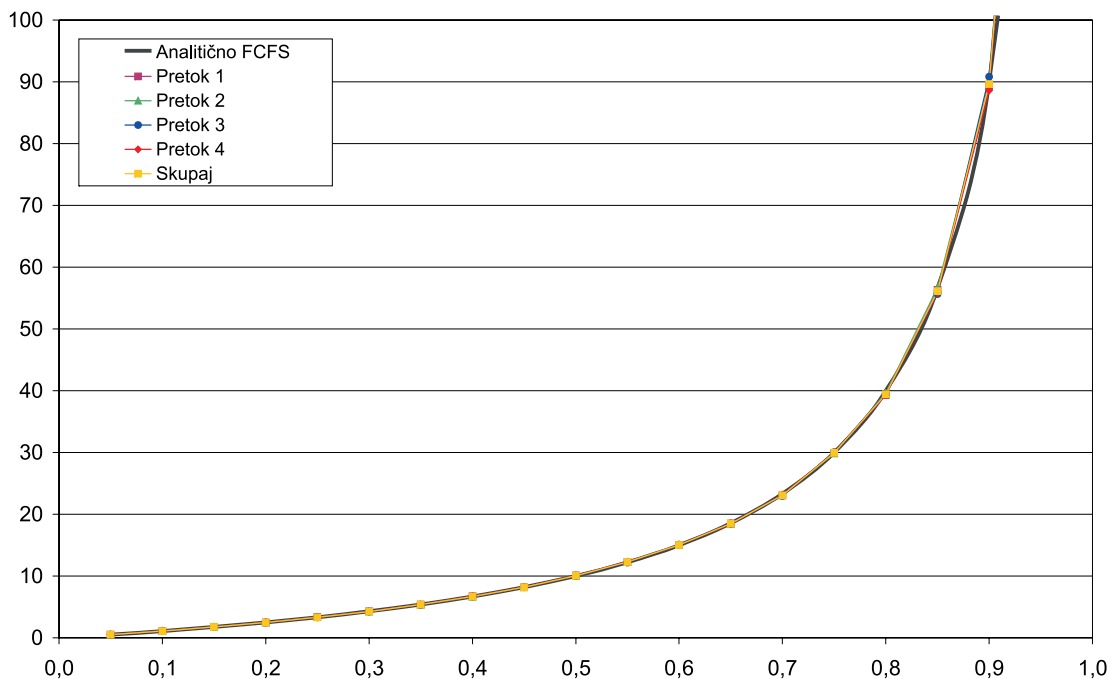
Simuliramo delovanje obeh odpravnikov pri eksponentni in enakomerni porazdelitvi dolžin paketov s štirimi aktivnimi podatkovnimi pretoki z bitnimi hitrostmi:  $r_1 = r_2 = r_3 = r_4$  in velikostmi kvantov:  $Q_1 = Q_2 = Q_3 = Q_4$ . Na slikah 5.28 do 5.31 so, v odvisnosti od izkoriščenosti sistema, za obe porazdelitvi prikazane povprečne zakasnitve paketov posameznih pretokov za odpravnik FCFS in DRR. Povprečna dolžina paketov pretokov za obe porazdelitvi je 1000 oktetov. Pri enakomerni porazdelitvi je minimalna velikost paketa 0, maksimalna pa 2000 oktetov. Pri eksponentni porazdelitvi je velikost kvanta 16000 oktetov<sup>8</sup>, pri enakomerni pa 2000. Bitna hitrost posameznega pretoka pa se spreminja tako, da se pri nespremenljivi bitni hitrosti izhodne linije izkoriščenost sistema  $\rho$  giblje med 0.05 in 1. Vidimo, da so povprečne zakasnitve paketov posameznih pretokov za odpravnik DRR praktično enake in so enake tudi z zakasnitvami odpravnik

<sup>8</sup>Takšna velikost kvanta zagotavlja pogoj  $Q \geq L_{max}$ . S predhodnimi simulacijami se je namreč izkazalo, da je bil najdaljši generiran paket med 15000 in 16000 oktetov.





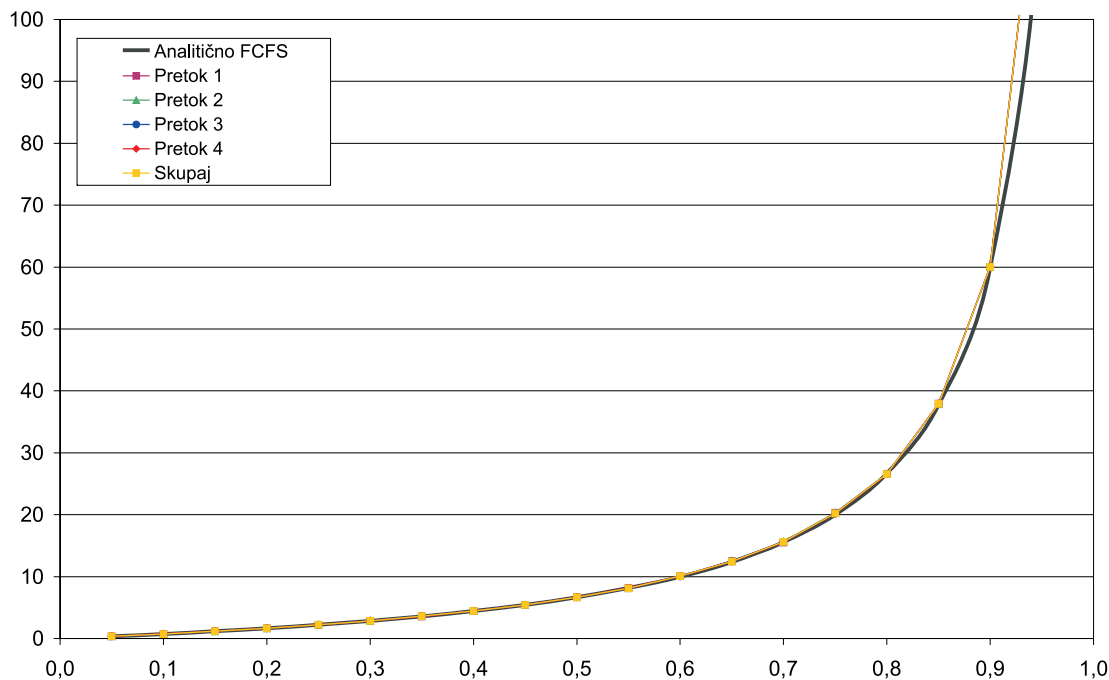
Slika 5.28 – Povprečne zakasnitve paketov v odpravniku FCFS, pri eksponentni porazdelitvi dolžin paketov in enakih bitnih hitrostih vseh pretokov



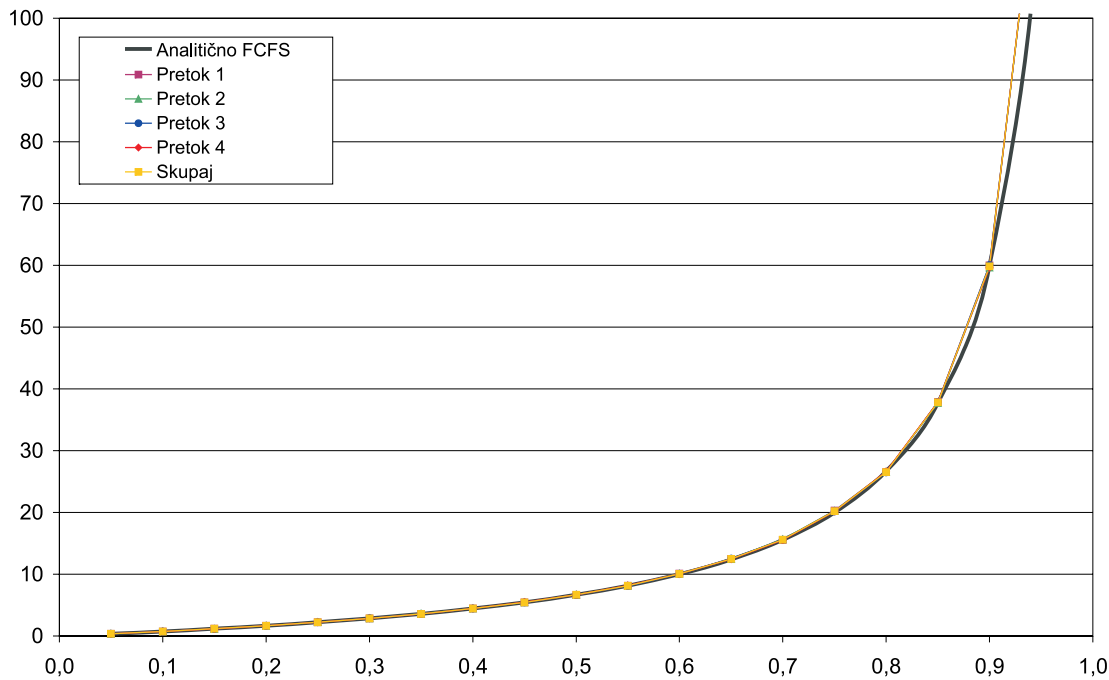
Slika 5.29 – Povprečne zakasnitve paketov v odpravniku DRR, pri eksponentni porazdelitvi dolžin paketov in enakih bitnih hitrostih vseh pretokov

FCFS. Rezultati za isti odpravnik, se glede na uporabljeno porazdelitev, praktično ne razlikujejo.

Poglejmo si še razmere v prisotnosti neobvladanega pretoka. Simuliramo delovanje obeh obravnavanih odpravnikov pri pretokih z bitnimi hitrostmi:  $r_1 = 5r$ ,

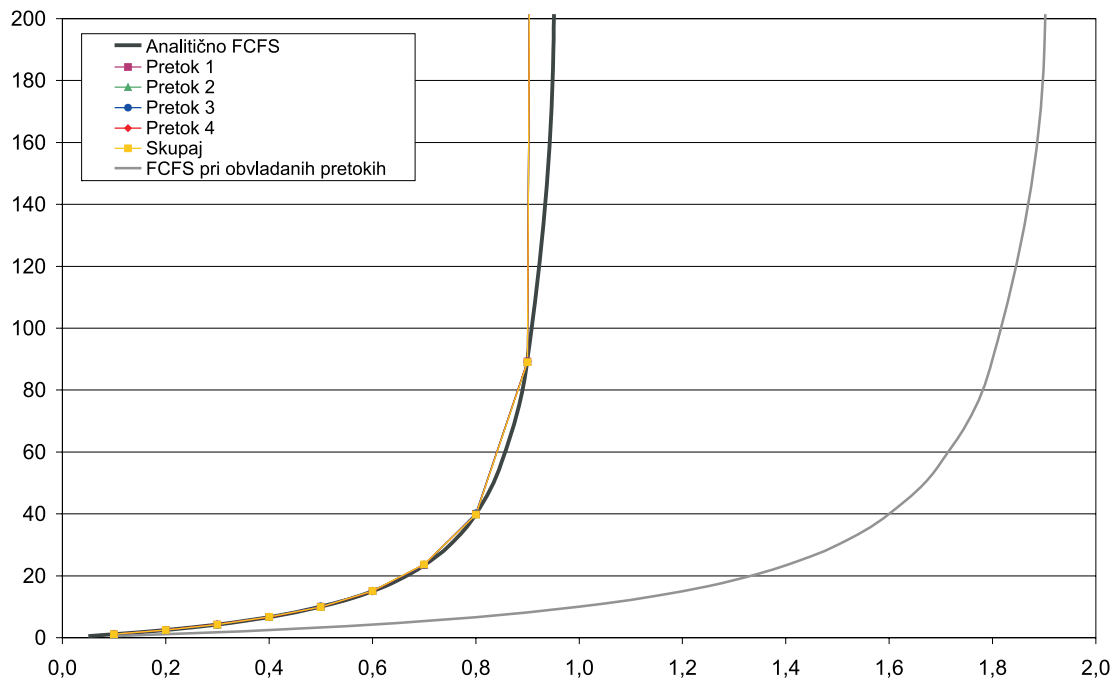


Slika 5.30 – Povprečne zakasnitve paketov v odpravniku FCFS, pri enakomerni porazdelitvi dolžin paketov in enakih bitnih hitrostih vseh pretokov

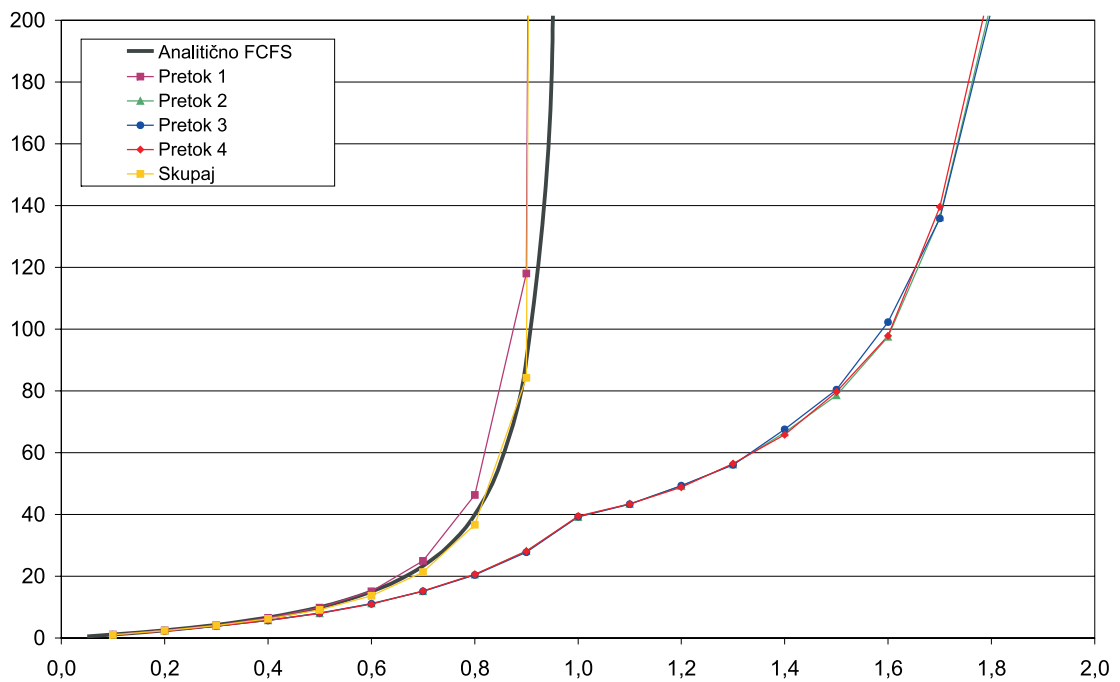


Slika 5.31 – Povprečne zakasnitve paketov v odpravniku DRR, pri enakomerni porazdelitvi dolžin paketov in enakih bitnih hitrostih vseh pretokov

$r_2 = r_3 = r_4 = r$  in velikostmi kvantov:  $Q_1 = Q_2 = Q_3 = Q_4$ . Opazimo, da pretok 1 pošilja petkrat toliko podatkov kot je njegov pravični delež (rezervacija) glede na velikost dodeljenega kvanta  $Q_1$  in ga zato označimo za neobvladanega. Na slika 5.32 do 5.34 so za eksponentno porazdelitev prikazane povprečne zakasnitve pake-

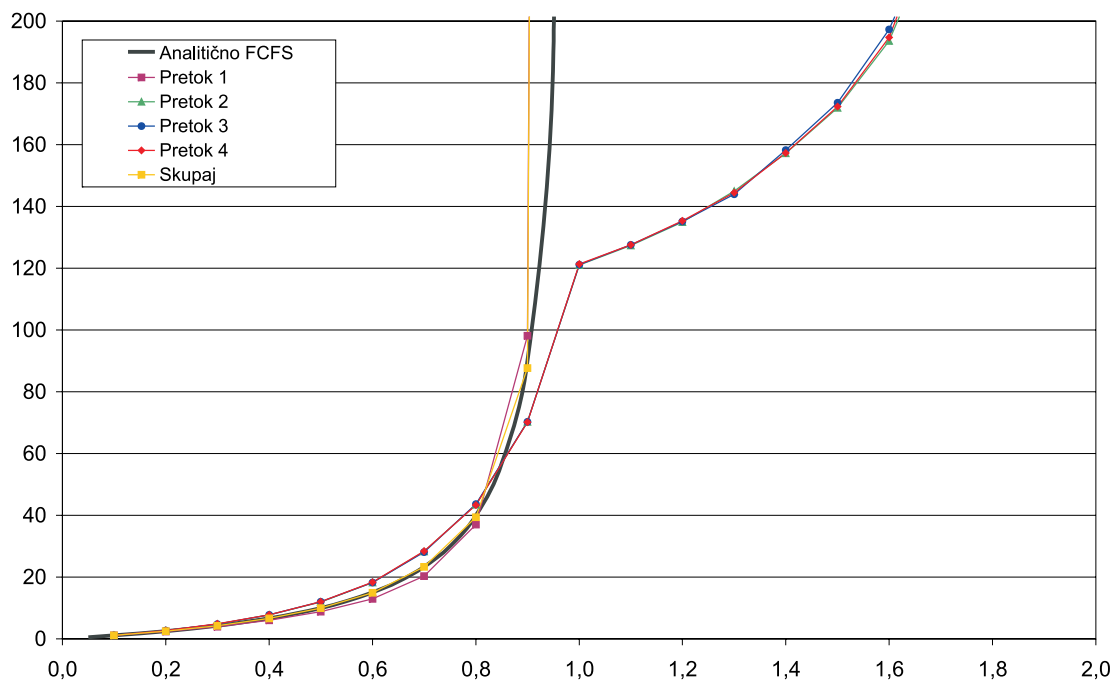


Slika 5.32 – Povprečne zakasnitve paketov v odpravniku FCFS pri eksponentni porazdelitvi dolžin paketov ob prisotnosti neobvladanega pretoka



Slika 5.33 – Povprečne zakasnitve paketov v odpravniku DRR z vrednostjo kvanta  $Q = 2000$  oktetov, pri eksponentni porazdelitvi dolžin paketov ob prisotnosti neobvladanega pretoka

tov posameznih pretokov za FCFS in odpravnik DRR v odvisnosti od izkoriščenosti sistema. Rezultati za enakomerno porazdelitev so izredno podobni rezultatom za eksponentno porazdelitev in jih zato ne navajamo posebej. Oblika krivulj je enaka, majhne razlike so le v vrednostih zakasnitev. Povprečna dolžina paketov je enaka kot pri prejšnji



Slika 5.34 – Povprečne zakasnitve paketov v odpravniku DRR z vrednostjo kvanta  $Q = 16000$  oktetov, pri eksponentni porazdelitvi dolžin paketov ob prisotnosti neobvladanega pretoka

simulaciji in znaša 1000 oktetov. Pri obeh porazdelitvah smo pri odpravniku DRR simulirali pri vrednostih kvantov 2000 in 16000 oktetov. Bitna hitrost posameznega pretoka se spreminja tako, da bi se v primeru, če bi bili vsi pretoki obvladani in bi delovali v okviru svojih rezervacij, izkoriščenost sistema  $\rho$  gibala med 0.05 in 1. Ker pa je pretok 1 neobvladan in skupaj z obvladanimi generira dvakrat toliko prometa, se vrednost  $\rho$  na grafih giblje med 0.05 in 2. Za vrednosti  $\rho > 1$  velja, da je v sistem pritekajočih paketov več kot jih je zmožen ta postreči.

Vidimo, da pri FCFS odpravniku pretok 1 dobi nepravilčen delež bitne hitrosti linije in s tem enakomerno slabša razmere tako sebi kot drugim pretokom. Povprečne zakasnitve so enake za vse pretoke - obvladane in neobvladane. Za koliko se razmere poslabšajo lahko vidimo na sliki 5.32, če primerjamo povprečne zakasnitve pretokov s sivo odebeljeno krivuljo, ki prikazuje zakasnitve, ki bi jih imeli pretoki v primeru, da bi bili vsi obvladani.

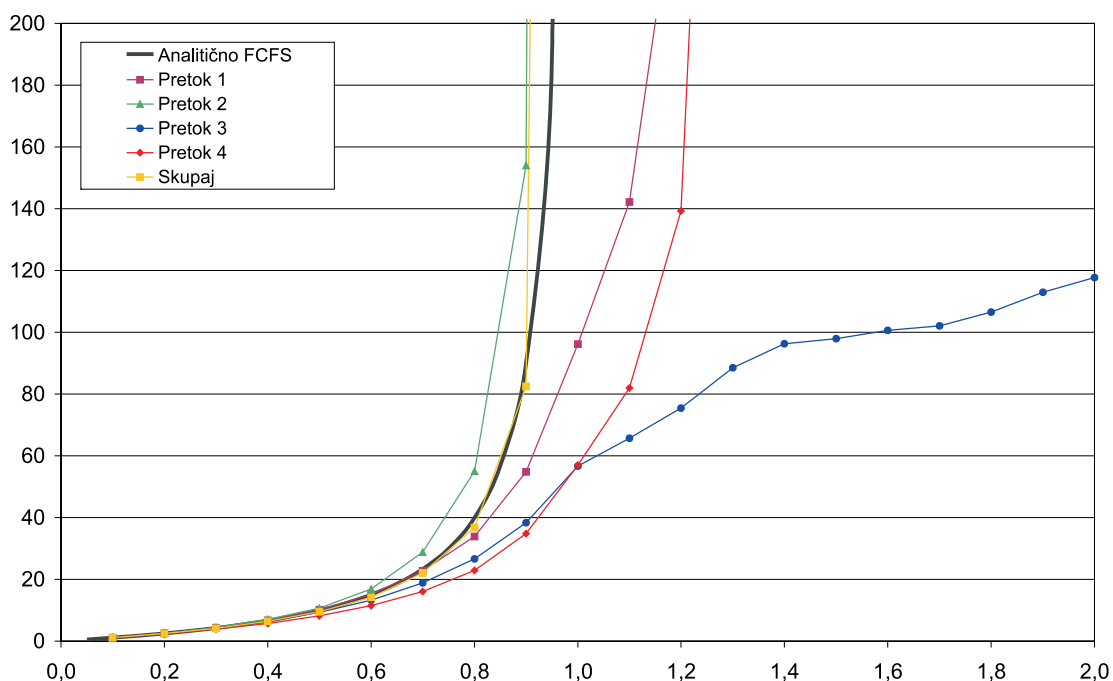
Pri odpravniku DRR te posledice ni, saj neobvladan pretok 1 dobi le tolikšen delež bitne hitrosti, kot je ima dodeljene glede na velikost kvanta  $Q_1$ , kar je prikazano na slikah 5.33 in 5.34. Simulirali smo razmere pri vrednostih kvantov 2000 in 16000 oktetov. Za oba primera opazimo, da so zakasnitve vseh pretokov (tudi neobvladanega) pri nižjih obremenitvah omrežja približno enake. Neobvladan pretok takrat namreč lahko uporabi neizkoriščen delež bitne hitrosti linije. To je ena od dobrih lastnosti odpravnika DRR, ki v primeru razpoložljivih kapacitet omogoča odpošiljanje tudi presežnih paketov (to so paketi, ki jih podatkovni pretok generira nad dogovorjeno vrednostjo).

Ko pa se obremenjenost omrežja bliža zasičenju pri vrednosti  $\rho = 1$ , zakasnitve neobvladanega pretoka 1 strmo narastejo, saj le ta za pošiljanje presežnih paketov nima več na voljo neizkoriščenega deleža bitne hitrosti linije. Za obvladane pretoke se razmere slabšajo počasneje in v skladu s pričakovanji. Obvladani pretoki dobijo

zadosten delež bitne hitrosti linije vse do vrednosti  $\rho = 2$ . Tudi ko je  $\rho > 1$  obvladanih pretokov to ne moti kaj dosti, saj imajo glede na svoj kvant vedno zagotovljen svoj delež bitne hitrosti linije.

Za zaključek simuliramo še delovanje odpravnika DRR pri pretokih z bitnimi hitrostmi:  $r_1 = r$ ,  $r_2 = 3r$ ,  $r_3 = r$  in  $r_4 = 3r$  in velikostmi kvantov:  $Q_1 = Q_2 = Q$ ,  $Q_3 = Q_4 = 3Q$ . Povprečna dolžina paketov pretokov je 1000 oktetov. Velikost kvantov  $Q_1$  in  $Q_2$  je 3000 oktetov, kvantov  $Q_3$  in  $Q_4$  pa 9000 oktetov. Glede na podane vrednosti bitnih hitrosti pretokov in njihovih pripadajočih kvantov lahko ugotovimo, da pretoka 1 in 4 delujeta v okviru rezervacij, pretok 2 je neobvladan, pretok 3 pa je glede na vrednost svojega kvanta podhranjen (pošilja manj od svoje rezervacije).

Na sliki 5.35 so prikazane povprečne zakasnitve paketov posameznih pretokov za eksponentno porazdelitev dolžin paketov. Vidimo, da se neobvladani pretok 2 obnaša kot je razloženo v primeru simulacije neobvladanih pretokov (sliki 5.33 in 5.34). Dokler so na voljo neizkoriščene rezervacije preostalih pretokov, so njegove zakasnitve podobne kot pri drugih pretokih, potem pa hitro narastejo. Pretoka 1 in 4, ki delujeta v okviru svojih rezervacij, ohranjata predvideno zakasnitev tudi malo nad mejno vrednostjo  $\rho = 1$ , saj je takrat na voljo še nekaj neizkoriščene rezervacije pretoka 3. Pretok 3 pa uživa nizke zakasnitve tudi še naprej, saj deluje pod svojimi rezervacijami vse do vrednosti  $\rho = 3$ . Rezultati za enakomerno porazdelitev so izredno podobni rezultatom za eksponentno porazdelitev in jih zato ne navajamo posebej. Oblika krivulj je enaka, majhne razlike so le v vrednostih zakasnitev.



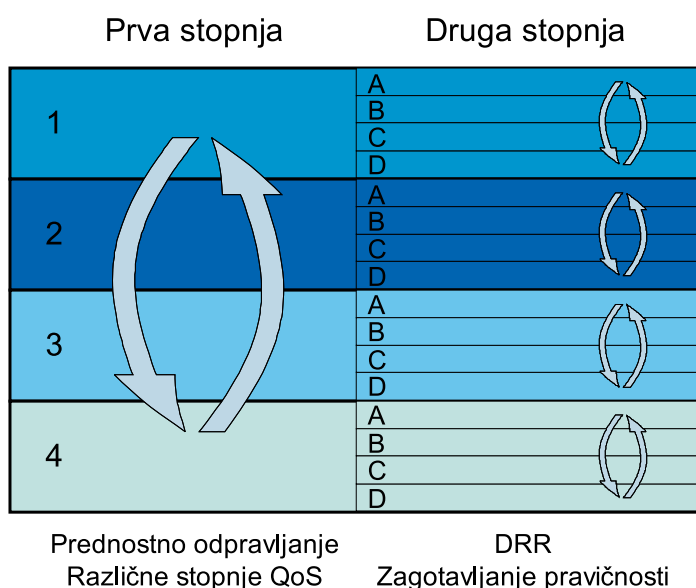
Slika 5.35 – Povprečne zakasnitve paketov v odpravniku DRR z vrednostjo kvantov  $Q_1 = Q_2 = 3000$  ter  $Q_3 = Q_4 = 9000$  oktetov, pri eksponentni porazdelitvi dolžin paketov

Zgornji rezultati kažejo, da je odpravnik DRR zmožen razlikovati med posameznimi pretoki in jim zagotoviti pravičnost v okviru njihovih rezervacij. Neobvladani pretoki ne morejo kvarno vplivati na obvladane, saj se zakasnitev poveča samo za prve, ki

tako škodijo le sebi. To potrjuje, da je odpravnik DRR primeren za drugo stopnjo odpravljanja v omrežjih DIS.

### 5.6.2 Simulacija dvostopenjskega delovanja odpravnika v omrežju DIS

Konfiguracija čakalnih vrst dvostopenjskega odpravnika v omrežju DIS je prikazana na sliki 5.36. Na prvi stopnji so štiri logične čakalne vrste, v katere se paketi vstavljajo in odpravljajo glede na pripadnost storitvenemu (prednostnemu) razredu. Vsaka izmed teh čakalnih vrst pa je še nadalje razdeljena v štiri čakalne vrste na drugi stopnji. Vanje se paketi vstavljajo glede na pripadnost pretoku in odpravljajo z drugostopenjskim odpravnikom.



Slika 5.36 – Konfiguracija čakalnih vrst odpravnika v omrežju DIS

V bistvu je to odpravnik s 16 čakalnimi vrstami označenimi od 1A do 4D, kjer cifra označuje storitveni (prednostni) razred, črka pa (združeni) pretok znotraj posameznega razreda. Vsak pretok ima svojo čakalno vrsto. Čakalne vrste se strežejo najprej glede na storitveni razred (prednosti od 1 do 4), kar zagotavlja različne stopnje kakovosti storitve, nato pa glede na podatkovni pretok. Odpravnik na drugi stopnji pa naj zagotovi pravičnost med pretoki istega storitvenega (prednostnega) razreda.

V naših simulacijah na prvi stopnji uporabljamo strogo prednostno odpravljanje, na drugi stopnji pa odpravljanje po principu DRR. V nadaljevanju s pomočjo simulacij primerjamo delovanje opisanega dvostopenjskega odpravnika z delovanjem enostopenjskega strogo prednostnega odpravljanja, ki je bilo analizirano v prejšnjih razdelkih.

#### 5.6.2.1 Rezultati simulacij dvostopenjskega odpravljanja in primerjava z enostopenjskim

V simulacijah uporabimo odpravnik s štirimi prednostnimi razredi, v katerih so po štiri čakalne vrste. Skupaj torej 16 čakalnih vrst. V simulacijah na prvi stopnji

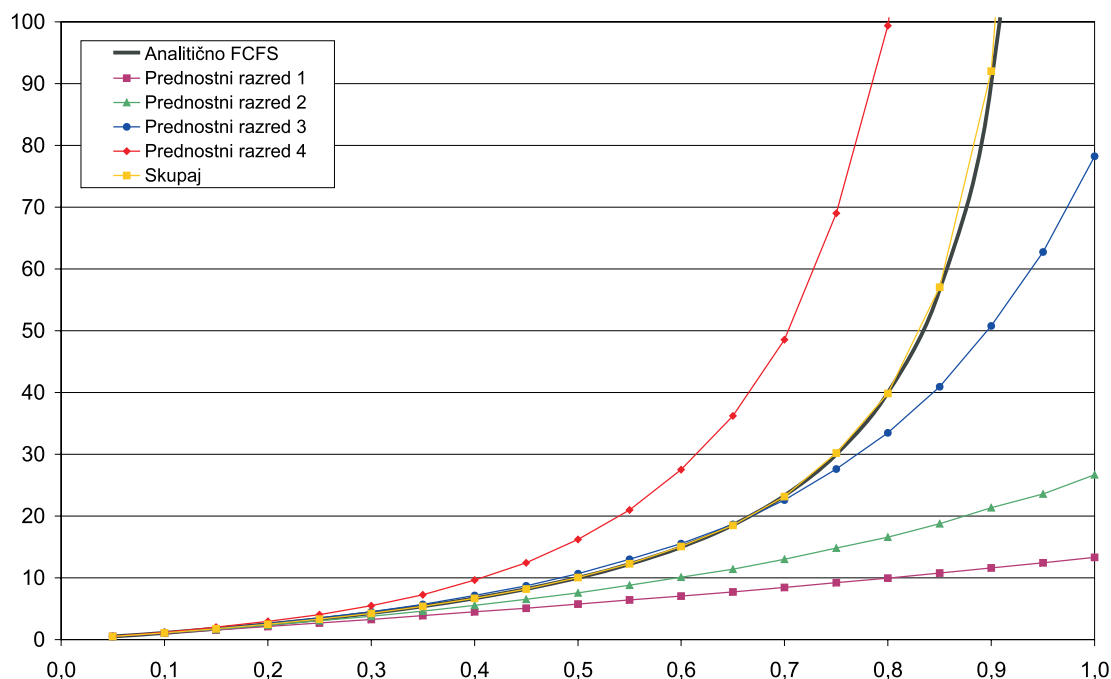
uporabljamo strogo prednostno odpravljanje, na drugi stopnji pa FCFS in odpravnik DRR. Z uporabo odpravnik FCFS na drugi stopnji celoten sistem v bistvu deluje kot enostopenjski odpravnik s strogimi prednostmi. Rezultate teh simulacij navajamo zaradi preverjanja pravilnosti delovanja dvostopenjskega odpravnik. V nadaljevanju predstavljamo le rezultate za eksponentno porazdelitev dolžin paketov, saj se je v razdelku 5.6.1 izkazalo, da med rezultati simulacij za eksponentno in enakomerno porazdelitev dolžin paketov ni bistvenih razlik. Na grafih so čakalne vrste znotraj posameznih razredov označene s ciframi od 1 do 4, pri tem pretoku A pripada čakalna vrsta 1, pretoku B pripada čakalna vrsta 2 in tako dalje.

Tudi pri dvostopenjskem delovanju z uporabo odpravnik DRR na drugi stopnji vsakemu podatkovnemu pretoku preko velikosti kvanta  $Q$  omogočimo rezervacijo določenega dela bitne hitrosti. Če se vsi pretoki držijo dogovorjenih prenosnih parametrov (ne oddajajo več kot je bilo dogovorjeno), predpostavimo da je povprečna zakasnitev paketov enaka kot pri enostopenjskem strogo prednostnem odpravljanju.

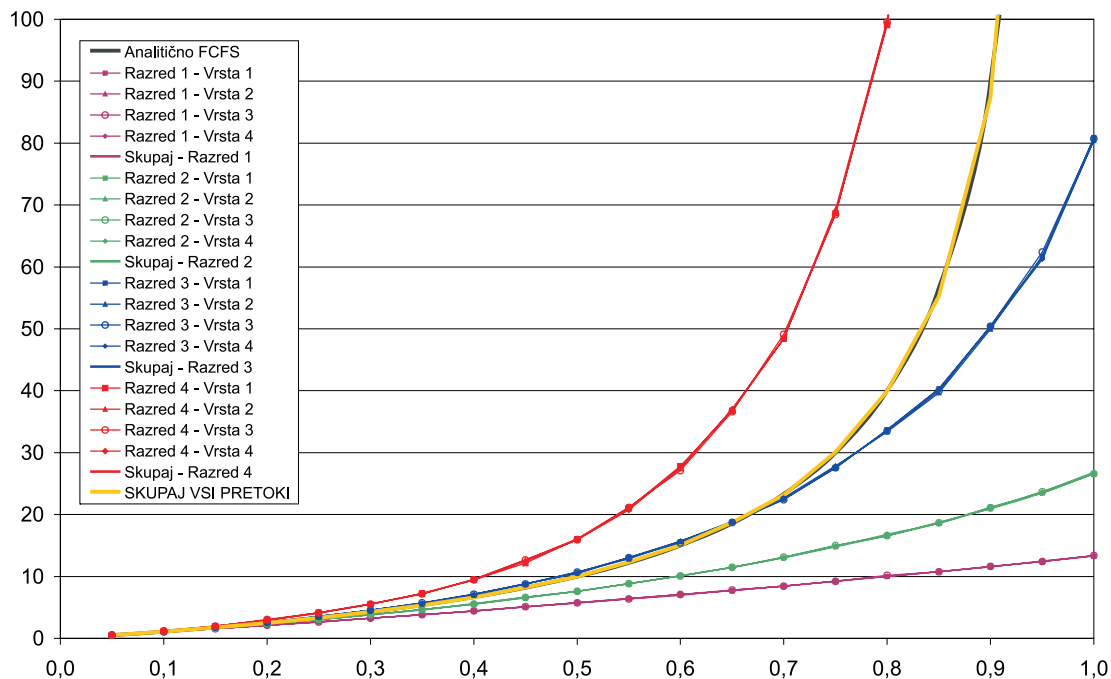
Simuliramo delovanje enostopenjskega in dvostopenjskega delovanja pri šestnajstih aktivnih pretokih z enakimi bitnimi hitrostmi:  $r_{1A} = r_{1B} = \dots = r_{4D}$  in ustreznimi velikostmi kvantov:  $Q_{1A} = Q_{1B} = \dots = Q_{4D}$ . Na slikah 5.37 do 5.40 so prikazane povprečne zakasnitve paketov posameznih pretokov za enostopenjsko in dvostopenjsko delovanja odpravnik glede na izkoriščenost sistema  $0.05 \leq \rho \leq 1$ . Porazdelitev dolžin paketov je eksponentna, povprečna dolžina paketov pretokov pa je 1000 oktetov. Vidimo, da so povprečne zakasnitve paketov posameznih pretokov znotraj enega prednostnega razreda pri dvostopenjskem delovanju praktično enake (krivulj posameznih pretokov istega prednostnega razreda med seboj ne ločimo) in so enake tudi povprečnim zakasnitvam paketov pri enostopenjskem delovanju. To dokazuje našo predpostavko o enakih zakasnitvah in tudi pravilnost delovanja dvostopenjskega odpravnik. Manjše odstopanje opazimo pri vrednosti kvanta  $Q_{i,j} = 3000$ , ko so povprečne zakasnitve paketov nekaj manjše. To je primer podkritične vrednosti kvanta  $Q_i < L_{max}$ , s čemer se podrobno ukvarja naslednje poglavje.

Poglejmo si še razmere v prisotnosti po enega neobvladanega pretoka v vsakem prednostnem razredu. Simuliramo delovanje dvostopenjskega odpravnik pri pretokih z bitnimi hitrostmi:  $r_{xA} = r_{xC} = r_{xD} = r$  in  $r_{xB} = 5r$  in enakimi velikostmi kvantov za vse pretoke:  $Q_{1A} = Q_{1B} = \dots = Q_{4D}$ . Vidimo, da pretoki  $xB$  pošiljajo petkrat toliko podatkov kot je njihov pravični delež glede na velikost dodeljenega kvanta. Na slikah 5.41 do 5.44 so prikazane povprečne zakasnitve paketov posameznih pretokov v odvisnosti od  $\rho$ , ki se giblje v mejah  $0.05 \leq \rho \leq 2$ . Porazdelitev dolžin paketov je eksponentna, povprečna dolžina paketov pretokov pa je 1000 oktetov. Poleg odpravnik FCFS in DRR smo na drugi stopnji tokrat uporabili še prednostno odpravljanje, kar je tudi prikazano na sliki 5.42.

Vidimo, da so pri odpravnik FCFS na drugi stopnji (slika 5.41) zakasnitve vseh pretokov enake, to pomeni, da pretoki B (glej razred x - vrsta 2 na sliki) dobijo nepravilni delež bitne hitrosti in s tem vplivajo na druge pretoke znotraj svojega prednostnega razreda. Na sliki 5.42 so prikazane tudi razmere, ko na drugi stopnji uporabimo prednostno odpravljanje. Neobvladan pretok v tem primeru vpliva samo na pretoke z manjšo prednostjo znotraj posameznega razreda in na vse pretoke v nižjih prednostnih razredih. Pri uporabi odpravnik DRR na drugi stopnji pa vidimo, da se razmere poslabšajo le za neobvladani pretok. Če primerjamo razmere za DRR pri vrednostih



Slika 5.37 – Povprečne zakasnitve paketov v enostopenjskem strogo prednostnem odpravniku s štirimi prednostnimi razredi

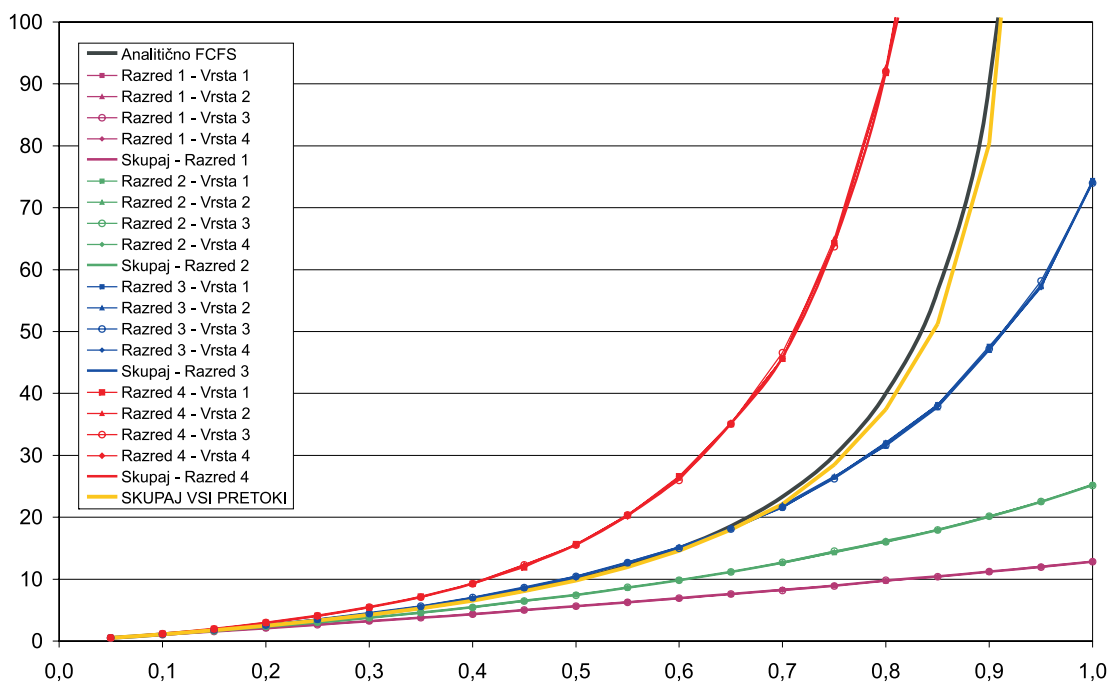


Slika 5.38 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom FCFS na drugi stopnji

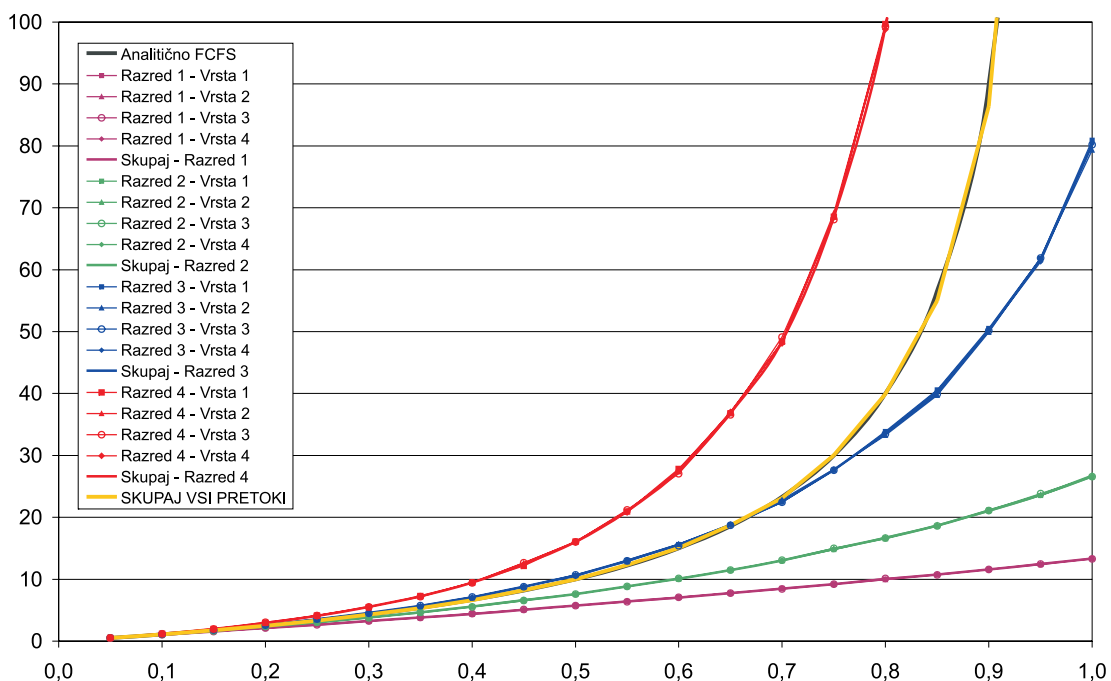
$Q_{i,j} = 3000$  in  $Q_{i,j} = 16000$  (glej sliki 5.43 in 5.44), opazimo, da se kar precej razlikujejo. Neobvladani pretoki pri nižjih vrednostih  $\rho$  izkoriščajo neuporabljene rezervacije drugih pretokov in zato je njihova zakasnitev primerljiva z obvladanimi pretoki.

Pri velikih vrednostih kvantov  $Q_{i,j}$  (glej DRR  $Q = 16000$ ) in pri nižjih vrednostih



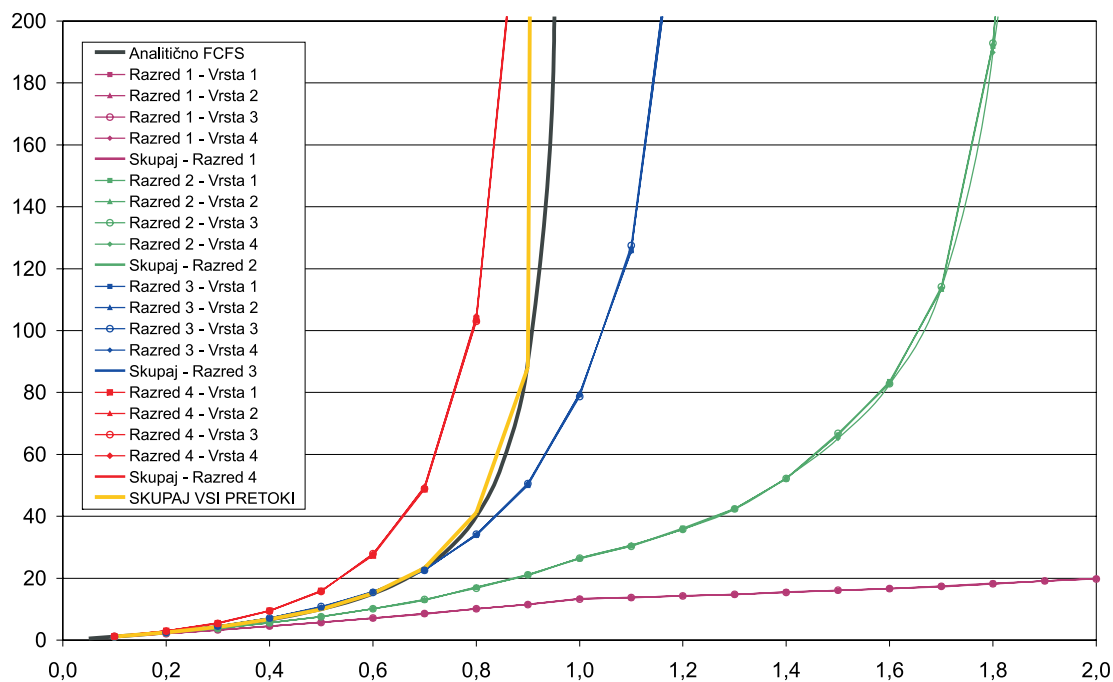


Slika 5.39 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 3000$ ) na drugi stopnji

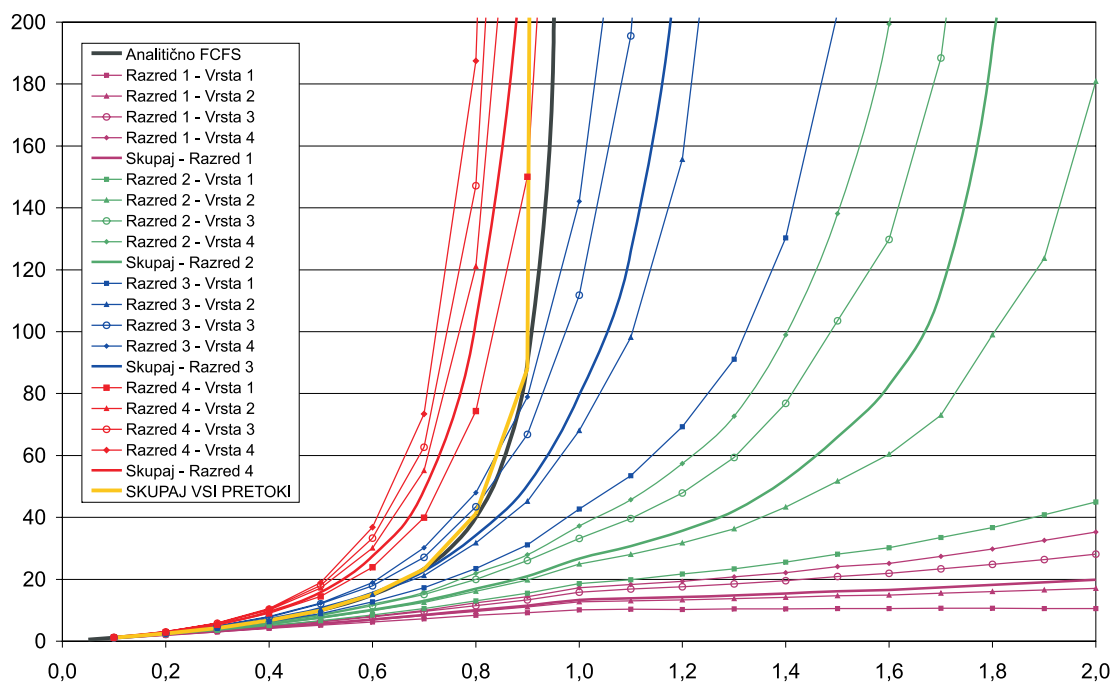


Slika 5.40 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 16000$ ) na drugi stopnji

$\rho$  je zakasnitev neobvladanih pretokov celo manjša od zakasnitve obvladanih pretokov. Neobvladani pretoki zaradi velike vrednosti kvanta v vsakem obhodu odpravijo precejšnje število svojih paketov. Obvladani pretoki takrat v vsakem obhodu v povprečju odpravijo petkrat manjše število paketov kot neobvladani. Ker v vsakem obhodu pre-

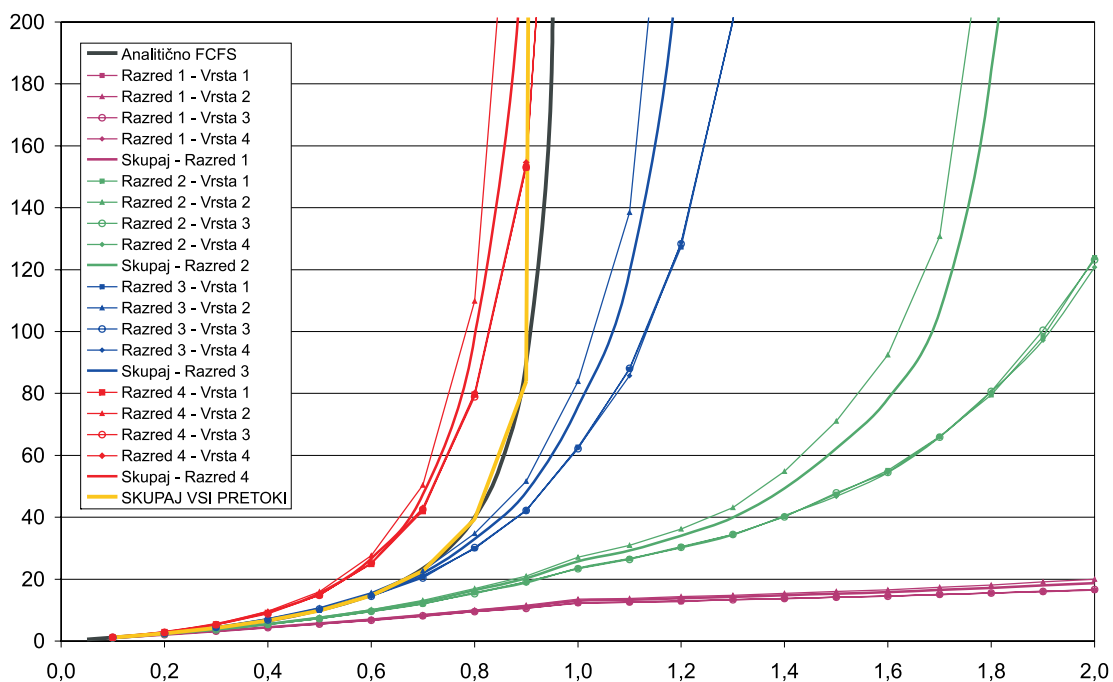


Slika 5.41 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom FCFS na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v vsakem razredu

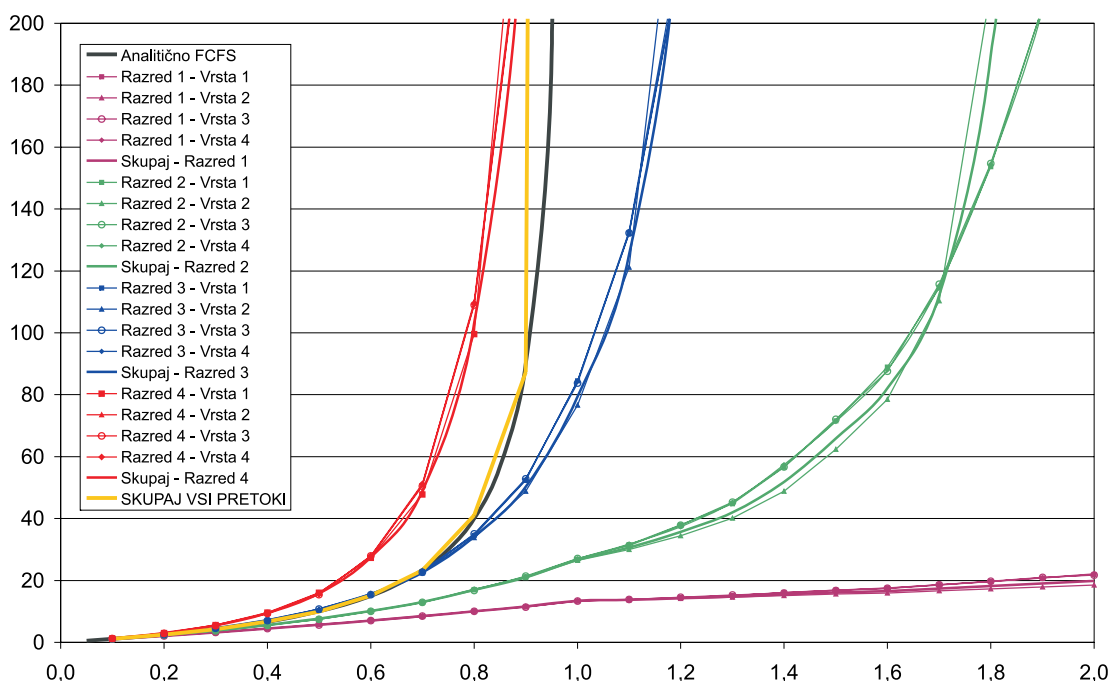


Slika 5.42 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na obeh stopnjah, pri neobvladanem pretoku B (vrsta 2) v vsakem razredu

toki med odpravljanjem paketov drugih pretokov čakajo na svojo priložnost, je njihova povprečna zakasnitev odvisna od količine prometa, ki se v vsakem obhodu odpravi iz drugih čakalnih vrst (drugih pretokov). Neobvladani pretoki na to priložnost v povprečju čakajo manj časa kot obvladani in zato je njihova povprečna zakasnitev



Slika 5.43 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 3000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v vsakem razredu



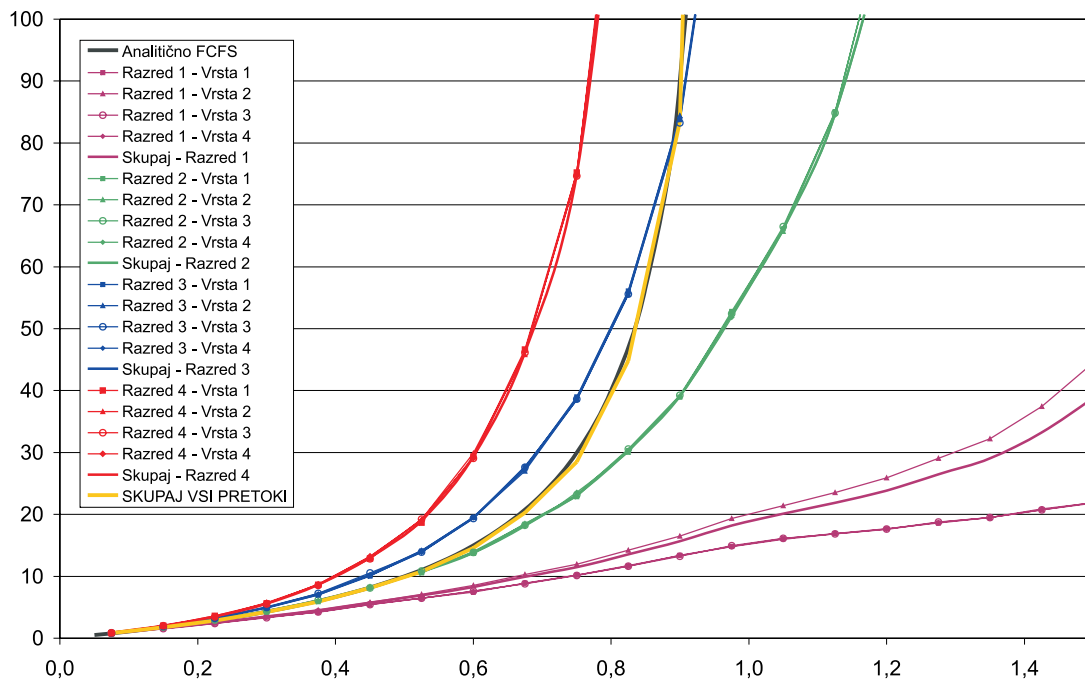
Slika 5.44 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 16000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v vsakem razredu

manjša. Pri višjih vrednostih  $\rho$  pa neobvladani pretoki nimajo na voljo neizkoriščenih rezervacij drugih pretokov. Njihove zakasnitve zato hitro narastejo in so v povprečju višje od zakasnitev obvladanih pretokov.

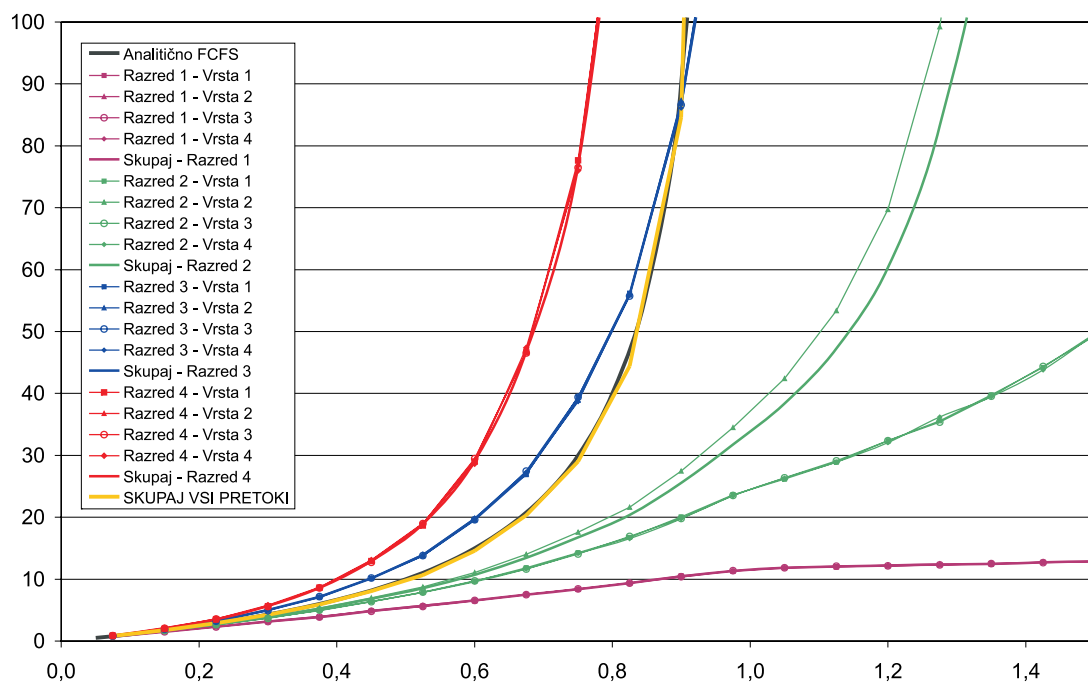
Pri majhnih vrednostih kvantov  $Q_{i,j}$  (glej DRR  $Q = 3000$ ) pa že sama velikost kvanta preprečuje velika nesorazmerja v številu odpravljenih paketov v vsakem obhodu. Zato je zakasnitev neobvladanih pretokov tudi pri nizkih vrednostih  $\rho$  večja od zakasnitve obvladanih pretokov. Simulacija pri  $Q = 3000$  pa pokaže tudi, da so povprečne zakasnitve vseh pretokov manjše kot pri vrednosti  $Q = 16000$ . Zmanjšanje velikosti kvanta ima torej vsaj dve dobri lastnosti: zmanjšanje povprečnih zakasnitev in izboljšanje pravičnosti med pretoki istega storitvenega razreda. S podkritičnim delovanjem odpravnik DRR pri vrednostih kvantov  $Q_i < L_{max}$  se podrobno ukvarja naslednje poglavje.

Simuliramo še delovanje odpravnik DRR pri enem neobvladanem pretoku v samo enem izmed prednostnih razredov. Imamo pretoke z bitnimi hitrostmi:  $r_{xA} = r_{xB} = r_{xC} = r_{xD} = r$  ter  $r_{yA} = r_{yC} = r_{yD} = r$  in  $r_{yB} = 9r$ , pri čemer  $x$  predstavlja razrede z obvladanimi pretoki in  $y$  razred z neobvladanim pretokom. Velikosti kvantov so za vse pretoke enake. Na slikah 5.45 do 5.50 so prikazane povprečne zakasnitve paketov posameznih pretokov za velikosti kvantov  $Q = 3000$  in  $Q = 16000$ . Porazdelitev dolžin paketov je eksponentna, povprečna dolžina paketov pretokov pa je 1000 oktetov.

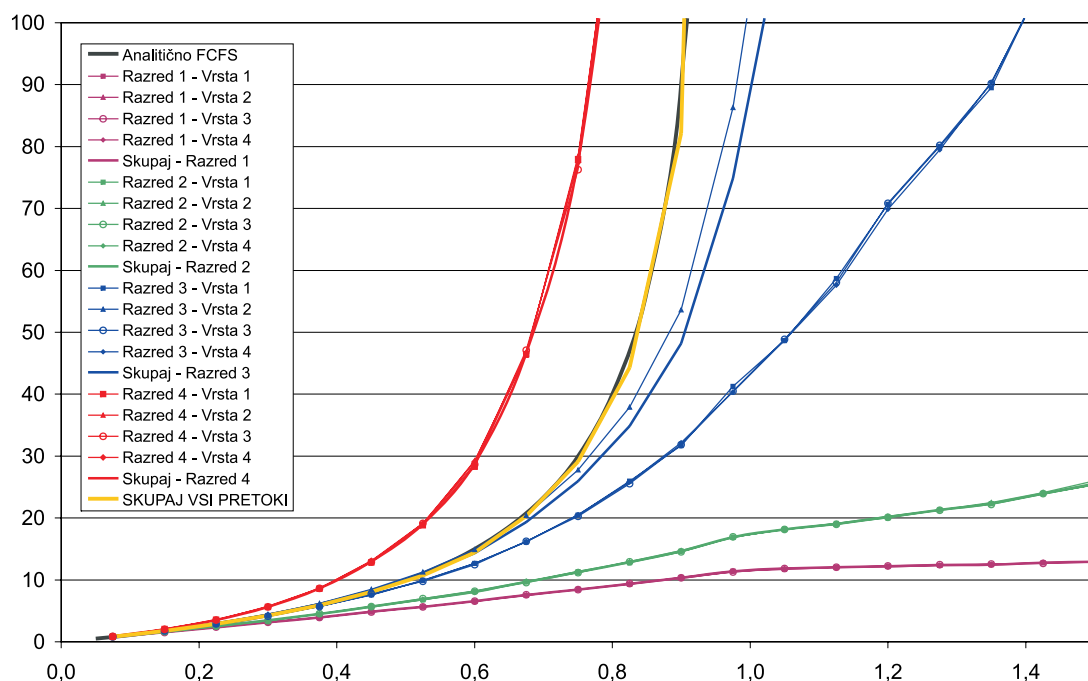
Ugotovimo, da so rezultati zelo podobni kot na slikah 5.43 in 5.44. Pri vrednosti kvanta  $Q = 3000$  ima neobvladani pretok slabše razmere kot obvladani, pri vrednosti kvanta  $Q = 16000$  pa velja enako kot pri prejšnjem primeru. Pri nizkih vrednostih  $\rho$  ima neobvladani pretok zaradi uporabe neizkoriščenih rezervacij drugih pretokov manjše povprečne zakasnitve, pri višjih vrednostih  $\rho$  pa slabše. To se še posebno lepo vidi na grafu za neobvladani pretok 2 v drugem prednostnem razredu pri vrednosti kvanta  $Q = 16000$ . Opazimo tudi, da neobvladani pretok v enem izmed prednostnih razredov poslabša razmere za vse pretoke v nižjih prednostnih razredih. To je posledica strogo



Slika 5.45 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 3000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 1

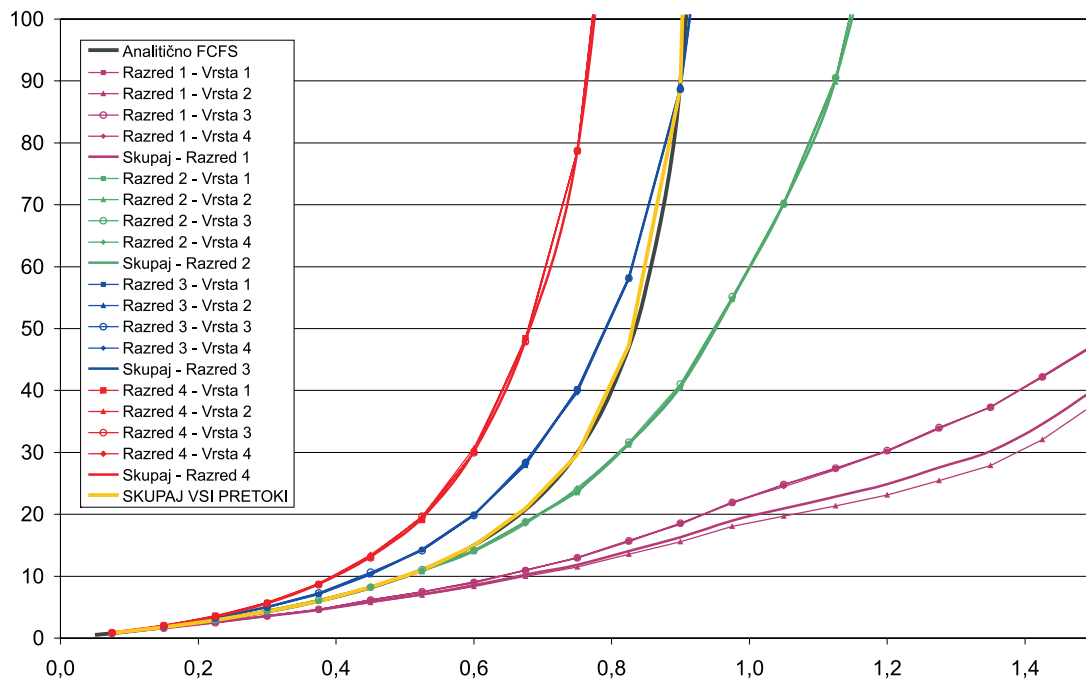


Slika 5.46 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 3000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 2

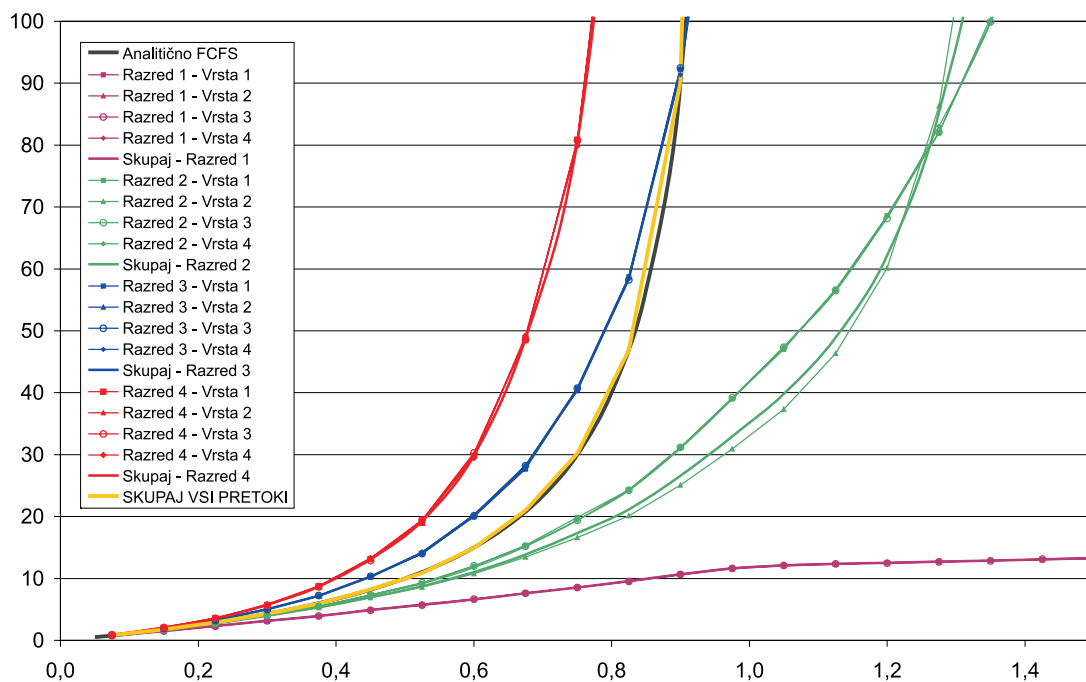


Slika 5.47 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 3000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 3

prednostnega odpravljanja na prvi stopnji. V omrežju DIS predpostavimo, da so v višjih prednostnih razredih pretoki, ki zahtevajo določeno jamstvo o stopnji kakovosti storitve. Promet teh pretokov je omejen po principu vedra z žetoni. Ti pretoki so zato

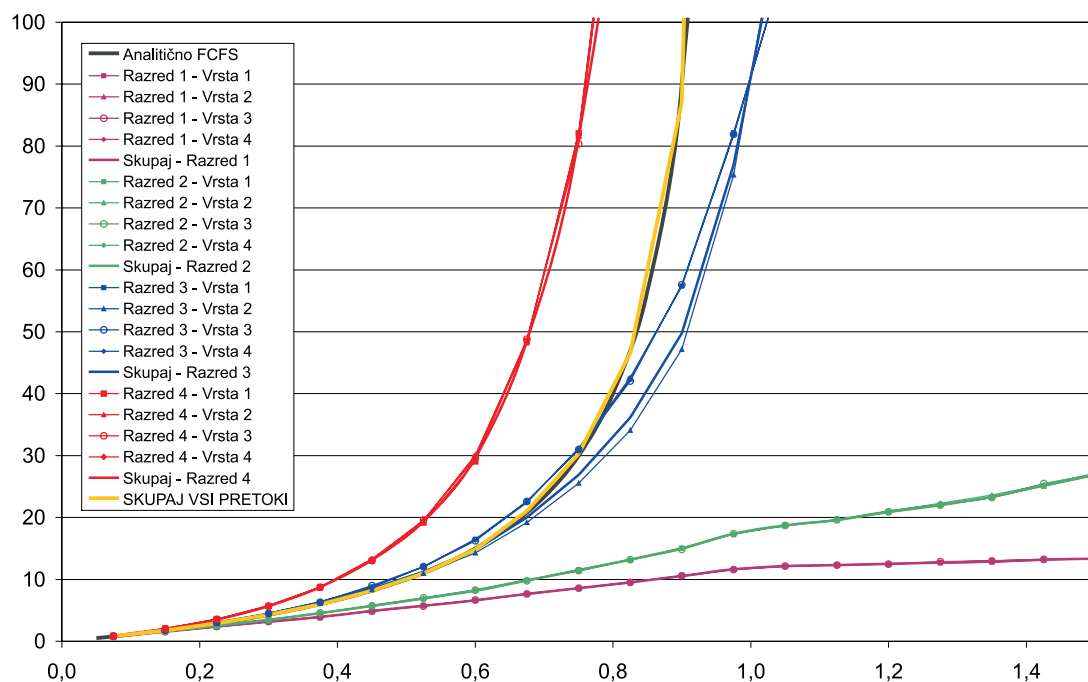


Slika 5.48 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 16000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 1

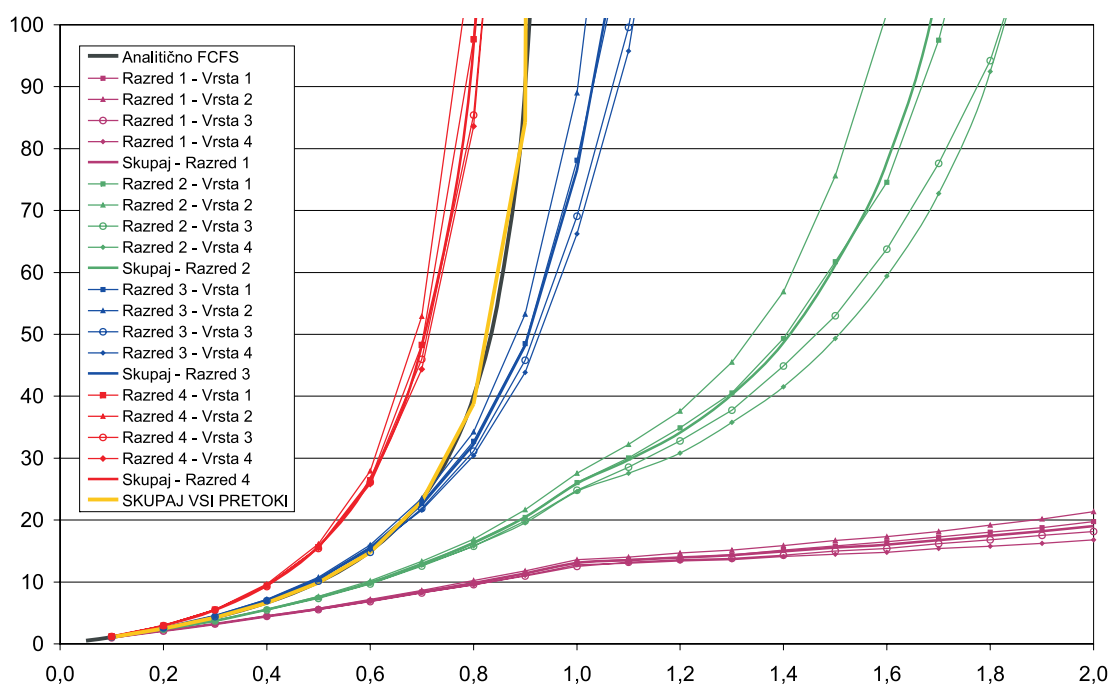


Slika 5.49 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 16000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 2

lahko "neobvladani" le kratkoročno in takrat je velikost odstopanja od rezervirane bitne hitrosti določena z velikostjo vedra, rezervirano bitno hitrostjo pretoka in bitno hitrostjo linije (glej dodatek A). V omrežju DIS torej ne more priti do razmer, v katerih bi bili



Slika 5.50 – Povprečne zakasnitve paketov v dvostopenjskem odpravniku s strogo prednostnim odpravnikom na prvi stopnji in odpravnikom DRR ( $Q = 16000$ ) na drugi stopnji, pri neobvladanem pretoku B (vrsta 2) v razredu 3



Slika 5.51 – Primerjava povprečnih zakasnitev paketov za dvostopenjski odpravnik pri različnih kombinacijah bitnih hitrosti in kvantov (rezervacij)

pretoki za višje prednostne razrede "neobvladani" daljše časovno obdobje. Pretoki v nižjih prednostnih razredih pa tako ali tako pričakujejo le relativne stopnje kakovosti storitve, ki jih prednostno odpravljanje tudi zagotavlja. Neobvladani pretoki tu lahko

nastopijo, vpliv pa imajo le na pretoke v nižjih prednostnih razredih.

Simuliramo še delovanje odpravnika DRR pri različnih kombinacijah bitnih hitrosti in kvantov. Rezervirane bitne hitrosti pretokov so:  $r_{xA} = r$ ,  $r_{xB} = 3r$ ,  $r_{xC} = r$  in  $r_{xD} = 3r$ . Velikostni kvantov pa so:  $Q_{xA} = Q_{xB} = Q$ ,  $Q_{xC} = Q_{xD} = 3Q$ . Na sliki 5.51 so prikazane povprečne zakasnitve paketov posameznih pretokov pri povprečni dolžini paketov pretokov 1000 oktetov. Velikost kvantov  $Q_{xA}$  in  $Q_{xB}$  je 3000 oktetov, kvantov  $Q_{xC}$  in  $Q_{xD}$  pa 9000 oktetov. Glede na podane vrednosti bitnih hitrosti pretokov in njihovih pripadajočih kvantov lahko ugotovimo, da pretoki  $xA$  in pretoki  $xD$  delujejo v okviru rezervacij, pretoki  $xB$  so neobvladani, pretoki  $xC$  pa so glede na vrednost svojega kvanta podhranjeni (pošiljajo manj od svojih rezervacij).

Na sliki 5.51 vidimo, da se neobvladani pretoki  $xB$  obnašajo kot je razloženo v primeru simulacije neobvladanih pretokov (glej komentar k slikam 5.43, 5.44 in 5.49). Dokler so na voljo neizkoriščene rezervacije preostalih pretokov v istem ali nižjih prednostnih razredih, so njihove zakasnitve podobne kot pri drugih pretokih v istem prednostnem razredu, potem pa hitro narastejo. Pretoki  $xA$  in  $xD$ , ki delujejo v okviru svojih rezervacij, ohranjajo predvideno zakasnitev tudi nad mejno vrednostjo  $\rho = 1$ , saj je takrat na voljo še nekaj neizkoriščene rezervacije pretokov  $xC$ . Najdlje pa uživajo nizke zakasnitve pretoki  $xC$ , saj delujejo pod svojimi rezervacijami.

### 5.6.3 Komentar simulacijskih rezultatov

Rezultati dvostopenjskega odpravljanje v omrežju DIS - strogo prednostno odpravljanje na prvi stopnji in DRR odpravljanje na drugi stopnji - zagotovi tako različne stopnje kakovosti storitve kot zaščito pretokov. Rezultate dvostopenjskega primerjamo z rezultati enostopenjskega strogo prednostnega odpravljanja. Če na drugo stopnjo postavimo FCFS odpravnik, je to namreč enakovredno enostopenjskemu strogo prednostnemu odpravljanju.

Na slikah 5.37 do 5.51 vidimo, da s strogo prednostnim odpravljanjem na prvi stopnji zagotovimo diferenciacijo med storitvenimi razredi. Podrobna analiza diferenciacije na osnovi prednostnih razredov je podana v prejšnjih razdelkih.

Več pozornosti je namenjene analizi zagotavljanja pravičnosti. Ugotovimo, da je znotraj posameznega storitvenega razreda pravičnost vedno zagotovljena, tudi ob prisotnosti neobvladanih pretokov. Manjše razlike se pojavijo pri različnih vrednostih kvantov. Izkazuje se, da je ugodnejša nižja vrednost kvanta, tako s stališča pravičnosti kot povprečnih zakasnitev. Glede na te ugotovitve smo se tudi odločili, da podrobneje raziščemo vpliv velikosti kvanta na delovanje odpravnika DRR, še posebej pri vrednostih, ki so manjše od maksimalne dolžine paketa ( $Q < L_{max}$ ). Ta analiza je podana v naslednjem poglavju.

Manjša slabost našega primera dvostopenjskega odpravljanja je le v poslabšanju razmer za nižje prednostne razrede, če je na višjih prisoten neobvladani pretok. Ker so v omrežju DIS pretoki višjih prednostnih razredov uravnavani z vedrom z žetonom, je to lahko problem le ob nepravilnem delovanju omrežja. Neobvladani pretoki v nižjih prednostnih razredih pa v glavnem škodujejo le sebi.

Zaključimo lahko, da je predlagano dvostopenjsko odpravljanje primerno za omrežje DIS in da z njim dosežemo želene in pričakovane rezultate.



## 6 Odpravnik SCDRR

Rezultati simulacij v prejšnjem poglavju nakazujejo, da z zmanjševanjem velikosti kvanta pri odpravniku DRR izboljšamo njegovo pravičnost in povprečno zakasnitev paketov. Ker sta za nas pravičnost in zakasnitev zelo pomembni meri, smo se odločili, da podrobno raziščemo vpliv zmanjšanja kvanta na lastnosti in delovanje odpravnika DRR.

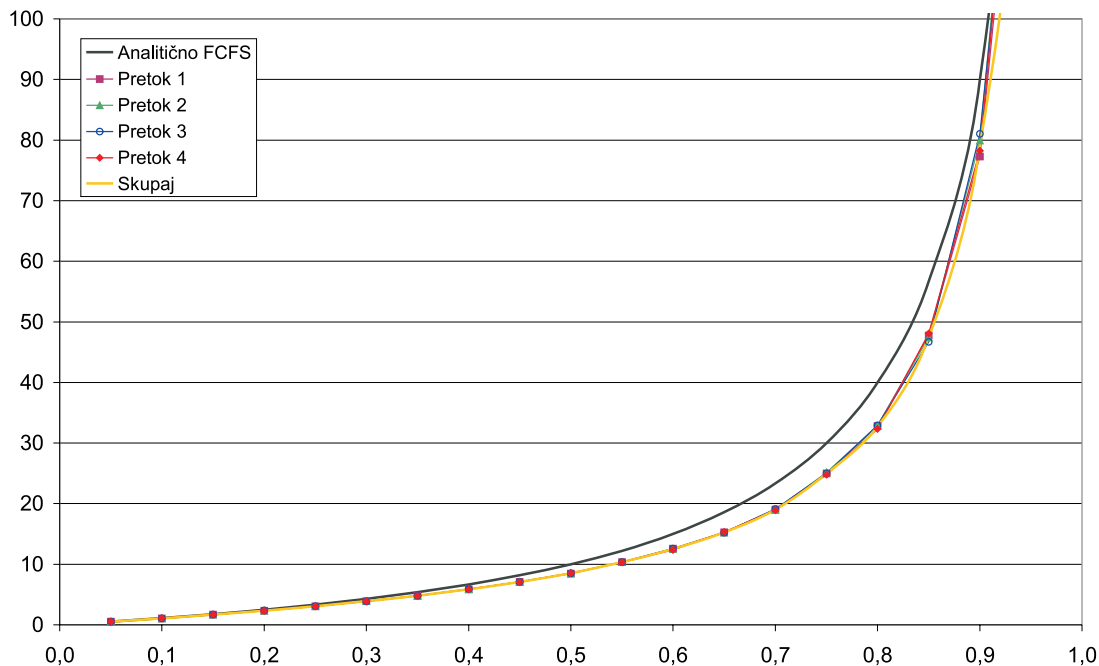
Slabost odpravnikov na osnovi časovnih okvirov, kamor spada tudi odpravnik DRR, je velika algoritemska zakasnitev. Ta je sestavni del zakasnitve v omrežnih napravah in s tem na celotni prenosni poti. Algoritemska zakasnitev lahko predstavlja pomemben delež teh zakasnitev. Za izboljšanje prenosnih parametrov, konkretno zakasnitve, želimo algoritemsko zakasnitev odpravnika DRR zmanjšati. Zanima nas možnost modifikacije delovanja odpravnika DRR, ki to omogoči s spremembo pogojev delovanja.

Algoritemska zakasnitev odpravnika DRR je odvisna od spremenljivk  $F$ ,  $Q_i$ ,  $R$  in  $L_{max}$  (glej definicije spremenljivk v razdelku 6.2 in tabelo 6.2). Na  $R$  in  $L_{max}$  navadno nimamo vpliva, saj sta podani z lastnostmi sistema in omrežja. Na velikost  $Q_i$  in posledično na  $F$  pa lahko preprosto vplivamo, saj sta parametra odpravnika. Lastnosti in delovanje odpravnika DRR so opisani v razdelku 4.5.6.

Vrednosti kvantov  $Q_i$  so pri DRR navzdol omejene z  $L_{max}$ , kar zagotavlja računsko zahtevnost  $O(1)$ . Z odpravo te omejitve lahko pričakujemo zmanjšanje algoritemske zakasnitve, s tem zakasnitev v čakalni vrsti in posledično tudi na celotni prenosni poti. V preskus te trditve smo opravili simulacije delovanja odpravnika DRR pri vrednostih  $Q_i$ , ki so manjše od  $L_{max}$ . Simulirali smo odpravnik DRR, ki streže štiri čakalne vrste z enakimi podatkovnimi pretoki, ki imajo eksponentno verjetnostno porazdelitev časov med prihodi paketov in eksponentno verjetnostno porazdelitev dolžin paketov. Povprečna dolžina paketov je 1000 oktetov, najdaljši paket, ki je prispel v času simulacije, pa je bil dolg 16000 oktetov.

Na sliki 6.1 so v odvisnosti od obremenjenosti omrežja prikazane povprečne zakasnitve paketov za vsako čakalno vrsto (pretok). Vrednost vseh kvantov je bila enaka in sicer  $Q_i = 1000(\forall i)$  ter je bila s tem enaka povprečni dolžini paketov. Za primerjavo so na sliki navedene tudi povprečne zakasnitve, ki jih imajo isti paketi, če namesto DRR uporabimo FCFS odpravnik. Opazimo, da je pri DRR za vrednosti  $Q_i$ , ki so manjše od  $L_{max}$ , zakasnitev opazno nižja kot pri FCFS. To nakazuje, da z odpravo omejitve najmanjše dovoljene vrednosti  $Q_i$  lahko znižamo algoritemsko zakasnitev odpravnika DRR. To nam potrjuje tudi slika 6.2, na kateri so v odvisnosti od obremenjenosti omrežja prikazane povprečne zakasnitve paketov za različne vrednosti  $Q_i$ .

Vrednost  $Q_i = 16000$  ustreza mejni vrednosti  $Q_i$ , ki je še ravno enaka  $L_{max}$  in zato predstavlja osnovno delovanje odpravnika DRR. Vidimo, da so v tem primeru povprečne zakasnitve paketov za DRR in FCFS odpravnik enake. Z nižanjem vrednosti



Slika 6.1 – Zakasnitve DRR pri podkritični vrednosti  $Q_i = 1000$

$Q_i$  pa povprečne zakasnitve padajo in so pri vrednosti  $Q_i = 1000$  že občutno nižje.

## 6.1 Analiza lastnosti in delovanja odpravnika SCDRR

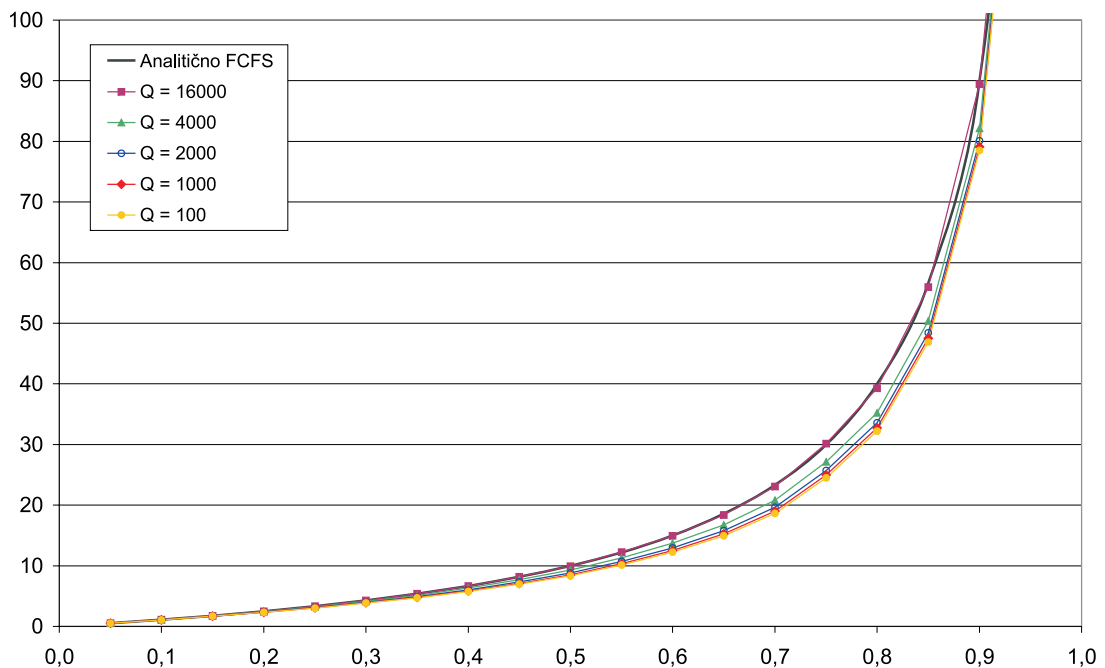
Zgornji rezultati potrjujejo našo domnevo o možnosti zmanjšanja algoritemska zakasnitve odpravnika DRR. Ker je zakasnitev zelo pomemben parameter kakovosti storitve nas to privede do potrebe po podrobnejši analizi delovanja odpravnika DRR pri podkritičnih vrednostih  $Q_i$ . Zato v tem poglavju predlagamo nov odpravnik, ki ga imenujemo *DRR s podkritičnim delovanjem* ali SCDRR (Sub-Critical DRR). Za podkritično delovanje štejeemo, ko so kvanti  $Q_i \leq L_{max} - 1$ .

Pričakujemo, da se s to modifikacijo odpravnika DRR zniža algoritemska zakasnitev in s tem tudi celotna zakasnitev. Pričakujemo, da se pravičnost takega odpravnika ne poslabša. Zaradi zmanjšanja  $Q_i$  pod kritično mejo pa lahko pričakujemo, da se poveča zahtevnost, saj sedaj ni več zagotovljeno, da se ob vsakem obhodu odpravil vsaj en paket.

V nadaljevanju najprej opišemo parametre in pogoje delovanja odpravnika SCDRR. Nato s podrobno teoretično analizo izpeljemo še njegovo pravičnost, algoritemsko zakasnitev in zahtevnost.

## 6.2 Parametri in robni pogoji delovanja SCDRR

Odpravnik SCDRR deluje enako kot odpravnik DRR (glej razdelek 4.5.6). Razlika je le v mejnih vrednostih kvantov  $Q_i$ . Podajmo spremenljivke, ki jih potrebujemo za opis in izpeljavo lastnosti in delovanja odpravnika SCDRR.

Slika 6.2 – Primerjava zakasnitev DRR pri različnih vrednostih  $Q_i$ 

- $R$  bitna hitrost linije na izhodu odpravnika,  
 $N$  število aktivnih čakalnih vrst,  
 $L_{max}$  dolžina najdaljšega možnega paketa.  
 $r_i$  bitna hitrost rezervirana za čakalno vrsto  $i$ ,  
 $w_i$  utež čakalne vrste  $i$ ,  
 $Q_i$  kvant dodeljen čakalni vrsti  $i$ ,  
 $DC_i$  števec primanjkljaja (Deficit Counter) čakalne vrste  $i$ , ki določa razliko med količino podatkov, ki bi jih čakalna vrsta  $i$  lahko odposlala in količino podatkov, ki jih je v resnici odposlala.  
 $F$  velikost okvira,  $F = \sum_i Q_i$

Ker si vse aktivne čakalne vrste delijo isto izhodno linijo, velja omejitev, da vsota rezerviranih bitnih hitrosti posameznih čakalnih vrst ne sme preseči bitne hitrosti izhodne linije:

$$\sum_i r_i \leq R \quad (6.1)$$

Naj bo  $r_{min}$  najmanjša izmed rezerviranih bitnih hitrosti za posamezno čakalno vrsto  $i$ :

$$r_{min} = \min_{\forall i} r_i$$

Ena izmed glavnih lastnosti vseh različic odpravnikov DRR je tudi njihova zmožnost zagotavljanja dogovorjene ravni storitve za vsako izmed čakalnih vrst. Zato je potrebno to lastnost zagotoviti tudi za našo različico algoritma. Vsaki čakalni vrsti dodelimo utež, ki je podana z:

$$w_i = \frac{r_i}{r_{min}}. \quad (6.2)$$

Opazimo da  $\forall i \in 1, 2, \dots, N$  velja naslednja pomembna zveza:  $w_i \geq 1$ .

Vsaki čakalni vrsti  $i$ , enako kot pri originalnem algoritmu DRR, dodelimo kvant  $Q_i \in \mathcal{N}$ . Definirajmo  $Q_{min}$  kot najmanjšega izmed vseh kvantov  $Q_i$ :

$$Q_{min} = \min_{\forall i} Q_i. \quad (6.3)$$

Potem lahko kvante posameznih čakalnih vrst izrazimo tudi kot:

$$Q_i = w_i Q_{min}. \quad (6.4)$$

Za razliko od originalnega algoritma DRR, kjer so kvanti  $Q_i$  vedno večji ali enaki največji možni dolžini paketa, pa pri algoritmu SCDRR velja, da so ti kvanti vedno manjši od največje možne dolžine paketov:

$$\forall i, 1 \leq Q_i \leq L_{max} - 1$$

Glede na to, da  $Q_i$  in s tem tudi  $Q_{min}$  ne moreta biti manjša od 1, moramo zagotoviti, da je  $L_{max} \geq 2$ . V nasprotnem primeru ( $L_{max} = 1$ ) bi imeli vse pakete enake, najmanjše možne dolžine. To pa bi pomenilo, da je odpravnik SCDRR postal WBRR odpravnik. Navedeni pogoji, na čelu s  $Q_i \leq L_{max} - 1$ , so v smislu določitve robnih vrednosti spremenljivk tudi glavna razlika med algoritmoma SCDRR in DRR.

Poudarimo naj, da na osnovi zgornjih definicij in robnih pogojev dokažemo, da zgornje meje mer za pravičnost (fairness) in zakasnitev algoritma DRR (latency), izpeljane v [93], niso pravilne. Pokažemo tudi, da naša različica algoritma (SCDRR) vodi do boljnih lastnosti v smislu pravičnosti in zakasnitve in sta zato naši meri za pravičnost in zakasnitev natančnejši kot meri podani v [93] in [42].

## 6.3 Pravičnost algoritma SCDRR

Splošna definicija pravičnosti odpravnikov je podana v razdelku 4.4.1. Vemo, da je najpravičnejši odpravnik GPS, glede na katerega se računa absolutna mera pravičnosti, ki jo označimo z AFM. Največkrat pa je lažje izračunati relativno mero pravičnosti, ki jo označimo z RFM. Za algoritem SCDRR zato izpeljemo obe meri in pokažemo, da sta obe strožji kot mere za primerljive algoritme, ki temeljijo na DRR. Izpeljava obeh mer pravičnosti odpravnika SCDRR je skupaj s podrobnimi rezultati podana v dodatku C. Tukaj navajamo le končne rezultate. Podani so tudi komentarji razlik med našimi natančnejšimi merami in merami podanimi v [93] in [42].

### 6.3.1 Relativna mera pravičnosti

V dodatku C smo s teoremom C.1 dokazali, da je RFM za odpravnik SCDRR omejena z izrazom:

$$RFM \leq Q_{min} + \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \quad (6.5)$$

**Komentar** Na desno stran neenačbe (6.5) lahko gledamo kot na funkcijo štirih spremenljivk:  $Q_{min}$ ,  $L_{max}$ ,  $w_i$  in  $w_j$ . Zadnji dve imata enak pomen in predstavljata uteži čakalnih vrst  $i$  in  $j$ . Mi jih bomo obravnavali kot dve neodvisni spremenljivki, ki imata isto definicijsko območje. Naš cilj je poiskati maksimum izraza

$$Q_{min} + \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \quad (6.6)$$

in s tem zgornjo mejo za RFM.

Ker iščemo samo lokalni maksimum, lahko naredimo analizo za vsako izmed spremenljiv posebej. Glede na lastnosti algoritma SCDRR imamo:

$$0 < Q_i \leq L_{max} - 1 \quad (\forall i)$$

$L_{max}$  je določen z lastnostmi sistema (omrežja) in ga zato lahko obravnavamo kot konstanto. Največja možna vrednost spremenljivke  $Q_{min}$  je  $L_{max} - 1$ . Da bi lahko poiskali maksimum glede na  $w_i$  in  $w_j$ , preuredimo izraz (6.6):

$$Q_{min} + \frac{L_{max} - 1}{w_i} + \frac{L_{max} - 1}{w_j} \quad (6.7)$$

Ker je glede na zahteve algoritma SCDRR  $L_{max} \geq 2$ , sta števca zgornjih ulomkov vedno večja ali enaka 1. Zato ima izraz (6.7) svoj maksimum, ko imenovalca  $w_i$  in  $w_j$  zavzameta svojo minimalno vrednost 1.

Zaključimo lahko, da pri podanem  $Q_{min}$  nastopi zgornja meja RFM, ko imata vsaj dve čakalni vrsti kvant enak  $Q_{min}$ . Veljavnost te trditve je zagotovljena z definicijo C.5, kjer nas zanima najslabši možni primer med dvema čakalnima vrstama.

Na splošno najslabši možni primer pa nastopi, ko  $Q_{min}$  zavzame svojo največjo možno vrednost  $L_{max} - 1$ . V tem posebnem primeru so vse uteži vseh čakalnih vrst enake 1. Ker velja  $1 \leq Q_i \leq L_{max} - 1$  in  $Q_i \geq Q_{min}$  ( $\forall i$ ), potem mora držati tudi  $Q_i = Q_{min}$  in  $w_i = 1$  ( $\forall i$ ). Zato najslabši možni primer nastopi, ko je  $Q_{min} = L_{max} - 1$  in  $w_i = 1$  ( $\forall i$ ). Takrat imamo  $RFM = 3L_{max} - 3$ .

Zgornja meja RFM za odpravnik DRR, izpeljana v [93], je podana z izrazom:

$$RFM_{DRR} \leq \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min} \quad (6.8)$$

ki je zelo podoben našemu izrazu (6.5) za SCDRR, izpeljanem v dodatku C. Potek izpeljave je v obeh primerih enak in tudi oba rezultata veljata tako za DRR kot za odpravnik SCDRR. Tekom izpeljave namreč nikjer ne upoštevamo omejitve velikosti kvanta  $Q$ . Ne navzdol, kot je primer pri DRR ( $Q_i \geq L_{max}$ ) in ne navzgor, kot je primer pri SCDRR ( $Q_{min} \leq L_{max} - 1$ ). Trdimo, da je naša zgornja meja za RFM (6.5) natančnejša, saj je izraz (6.8) izpeljan na osnovi napačnih predpostavk, kar je razloženo v dodatku C tekom naše izpeljave zgornje meje RFM.

Poudarimo naj, da je naša zgornja meja za RFM nižja kot meje podane v [93] ali drugih podobnih delih. To trditev lahko podkrepimo z naslednjimi dejstvi.

Prvo dejstvo je obstoj členov  $\frac{1}{w_i}$  in  $\frac{1}{w_j}$ , ki sta vedno pozitivna in se odštevata od zgornje meje za RFM ter jo na ta način znižujeta. V primeru, da je  $L_{max}$  velika

vrednost, je prispevek teh dveh členov relativno majhen, če pa je  $L_{max}$  manjša vrednost, je njun prispevek pomemben.

Drugo, bolj pomembno dejstvo pa leži v možnem razponu vrednosti  $Q_{min}$  pri obeh algoritmih. V našem algoritmu SCDRR velja, da je  $Q_i \leq L_{max} - 1$  ( $\forall i$ ) in posledično tudi  $Q_{min} \leq L_{max} - 1$ . V originalnem algoritmu DRR pa velja ( $\forall i$ ) in posledično  $Q_{min} \geq L_{max}$ . Posledica teh omejitev je, da je v najslabšem možnem primeru RFM za SCDRR ( $3L_{max} - 3$ ) še vedno boljša od najboljšega možnega primera za DRR ( $3L_{max} - 2$ ).

Zaključimo lahko, da je naša zgornja meja za RFM algoritma SCDRR vedno boljša od zgornje meje podane v [93] za algoritem DRR. Razlike med najslabšim možnim primerom za SCDRR in najboljšim možnim primerom za DRR je resda minimalna, vendar pa je razlika v normalnih pogojih delovanja velika.

### 6.3.2 Absolutna mera pravičnosti

Ker absolutna mera pravičnosti podaja obnašanje podanega odpravnika napram odpravniku GPS, je pomembno, da podamo tudi to. AFM naj bi predstavljala najprimernejšo mero pravičnosti, ki pa je težko izračunljiva. V ta namen bomo izrabili zvezo med RFM in AFM, ki je bila podana v [115].

**Lema 6.1** *Zveza med relativno in absolutno mero pravičnosti za katerikoli delovno intenzivni odpravnik je podana z neenačbo:*

$$AFM \leq \left(1 - \frac{Q_{min}}{F}\right) RFM \quad (6.9)$$

kjer  $F$  označuje velikost okvira po izrazu  $F = \sum_i Q_i$ .

Z uporabo leme 6.1, ki je dokazana v [115], lahko zapišemo, da za poljuben časovni interval in poljuben obhod odpravnika SCDRR velja: ■

$$AFM \leq \left(1 - \frac{Q_{min}}{F}\right) \left(\frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}\right) \quad (6.10)$$

Ker je ta meja tudi funkcija  $Q_{min}$ , je očitno, da je tudi AFM za odpravnik SCDRR manjša od AFM za odpravnik DRR. Praktično to pomeni, da je v smislu pravičnosti odpravnik SCDRR boljši približek GPS.

Glede na lemo 6.1 lahko zaključimo, da sta si pri majhnih vrednostih  $Q_{min}$  relativna in absolutna mera pravičnosti zelo blizu in se oddaljujeta, ko se  $Q_{min}$  veča. AFM je lahko majhna, četudi je ustrezna RFM velika.

Poglejmo si še člen  $\left(1 - \frac{Q_{min}}{F}\right)$  v neenačbi (6.10). Ta člen določa soodvisnost AFM in RFM in s tem tudi najboljši in najslabši možni primer za AFM. Analizo RFM, z njenim najboljšim in najslabšim možnim primerom, smo naredili že v razdelku C.2.1, zato se tu posvetimo le omenjenemu členu. Glede na to, da odštevamo vrednost  $\frac{Q_{min}}{F}$ , dosežemo minimum navedenega člena, ko je kvocient  $\frac{Q_{min}}{F}$  največji. Ker je  $F = \sum_i Q_i$ , lahko hitro pokažemo, da  $Q_{min}$  doseže svoj maksimum v primeru, ko so vsi kvanti  $Q_i$  enaki. Takrat je AFM najmanjši in relativno najdlje od RFM.

### 6.3.3 Primerjava pravičnosti nekaterih pomembnejših odpravnikov

V tabeli 6.1 podajamo primerjavo RFM med odpravnikom SCDRR in odpravniki SCFQ (WF<sup>2</sup>Q), PGPS (WFQ) in DRR. RFM za nekatere druge pomembnejše odpravnike pa so podane v tabeli 6.1.

Za odpravnik PGPS (WFQ) je v [98] dokazana naslednja neenačba:

$$RFM \leq \max_{i,j} \left( \max \left\{ C_j + \frac{L_{max}}{w_i} + \frac{L_j}{w_j}, C_i + \frac{L_{max}}{w_j} + \frac{L_i}{w_i} \right\} \right) \quad (6.11)$$

kjer  $L_i$  označuje najdaljši možni paket v čakalni vrsti  $i$ .  $C_i$  pa označuje presežek maksimalne normalizirane storitve, ki jo posamezna čakalna vrsta dobi od odpravnika PGPS, glede na storitev, ki bi jo dobila od odpravnika GPS.  $C_i$  je podana z:

$$C_i = \min \left\{ (N-1) \frac{L_{max}}{w_i}, \max_{1 \leq n \leq N} \frac{L_n}{w_n} \right\} \quad (6.12)$$

Iz neenačbe (6.11) vidimo, da je zgornja meja RFM za PGPS precej višja kot za SCDRR v (6.5). V izrazu (6.11) imamo namreč člen  $C_i$ , ki je večji kot primerljivi člen  $Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}$  v (6.5). Če podrobneje pogledamo definicijo  $C_i$  (6.12), vidimo, da vsebuje dva člena  $(N-1) \frac{L_{max}}{w_i}$  in  $\max_{1 \leq n \leq N} \frac{L_n}{w_n}$ . Oba sta večja od  $Q_{min}$ , zato je večji tudi njun minimum. Iz tega sledi, da je njun minimum večji tudi od izraza  $Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}$ . Zato je meja za RFM odpravnika SCDRR boljša kot pri PGPS odpravniku.

Odpravnik	Relativna mera pravičnosti
GPS	0
FIFO	$\infty$
Virtual Clock	$\infty$
SCFQ	$\frac{L_i}{w_i} + \frac{L_j}{w_j}$
PGPS (WFQ)	$\max_{i,j} \left( \max \left\{ C_j + \frac{L_{max}}{w_i} + \frac{L_j}{w_j}, C_i + \frac{L_{max}}{w_j} + \frac{L_i}{w_i} \right\} \right),$ $C_i = \min \left\{ (N-1) \frac{L_{max}}{w_i}, \max_{1 \leq n \leq N} \frac{L_n}{w_n} \right\}$
WF <sup>2</sup> Q	$\frac{L_i}{w_i} + \frac{L_j}{w_j}$
Round Robin	$\infty$
DRR [93]	$\frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min}$
Surplus Round Robin	$\frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min}$
Elastic Round Robin	$\frac{L_i}{w_i} + \frac{L_j}{w_j} + L_{max}$
SCDRR	$\frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}$

Tabela 6.1 – Primerjava meja za RFM za nekatere pomembnejše odpravnike

V [11] je Golestani dokazal obstoj naslednje meje za SCFQ odpravnik:

$$RFM \leq \frac{L_i}{w_i} + \frac{L_j}{w_j} \quad (6.13)$$

kjer  $L_i$  označuje najdaljši možni paket v čakalni vrsti  $i$ . Enako RFM ima tudi odpravnik  $WF^2Q$ .

Izraza (6.13) in (6.5) se razlikujeta le za člen  $Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}$ . Brez analize vrednosti tega člena zato ne moremo reči, kateri odpravnik je boljši. Če je vrednost tega člena manjša od nič, je boljši odpravnik SCDRR, če je pa večja od nič, je boljši odpravnik SCFQ.

Vemo, da velja  $1 \leq Q_{min} \leq L_{max} - 1$  in  $\frac{1}{w_i} \leq 1$ . Zatorej mora veljati tudi  $-2 \leq -\frac{1}{w_i} - \frac{1}{w_j} < 0$ . V tem primeru je lahko izraz  $Q_{min} - \frac{1}{w_i} - \frac{1}{w_j}$  negativen le, če je  $Q_{min} = 1$ , kar pa ni ravno pogost primer. Četudi velja, da je  $Q_{min} = 1$ , moramo zagotoviti, da so tudi vsi ostali kvanti  $Q_i$  blizu vrednosti  $Q_{min}$ . Odpravnik SCDRR je lahko boljši on SCFQ in  $WF^2Q$  odpravnikov v smislu meje za RFM, vendar le v specifičnih in redkih pogojih.

Shreedhar in Varghese sta v [93] izpeljala zgornjo mejo RFM za odpravnik DRR, ki je podana z izrazom (6.8). Kot smo že omenili, ta neenačba ni pravilna. Na tem mestu jo uporabljamo zgolj v namene primerjave.

Iz neenačbe (6.8) vidimo, da je meja RFM za DRR zelo podobna meji odpravnika SCDRR. Njuno primerjavo smo podali že v razdelku 6.3.1. Ne glede na to, da smo že dokazali, da je naša meja za RFM vedno boljša kot tista izračunana za DRR v [93], se nam zdi pomembno, da dodamo nekaj podrobnosti. V komentarju v razdelku 6.3.1 smo pokazali, da je pri SCDRR najslabši možni primer, ko sta vsaj dve uteži enaki 1 ( $w_i = w_j = 1$ ). Primerjajmo oba odpravnika še glede na možne vrednosti  $Q_{min}$ .

Analizirajmo meje za RFM obeh odpravnikov. Poglejmo si najslabši možni primer za SCDRR in najboljši možni primer za DRR. Dokazali smo že, da je najslabši možni primer za SCDRR, ko sta vsaj dve uteži ( $w_i = w_j = 1$ ) enaki 1. Takrat velja neenačba:  $RFM_{SCDRR} \leq 2L_{max} + Q_{min} - 2$ . Ker je pri SCDRR  $Q_{min}$  lahko največ  $L_{max} - 1$ , je v najslabšem možnem primeru:

$$RFM_{SCDRR} \leq 3L_{max} - 3. \quad (6.14)$$

Če enake kriterije uporabimo tudi za DRR ( $w_i = w_j = 1$ ), potem pri uporabi pravilnega izraza za pravičnost<sup>1</sup> (6.5) dobimo:  $RFM_{DRR} \geq 2L_{max} + Q_{min} - 2$ . Pri DRR je zaradi pričakovane računske zahtevnosti  $O(1)$  predpisano, da je lahko najmanjša vrednost za  $Q_{min}$  enaka  $L_{max}$ . Od tod dobimo:

$$RFM_{DRR} \geq 3L_{max} - 2. \quad (6.15)$$

Opazimo, da je najmanjša možna razdalja med mejami za RFM obeh odpravnikov enaka 1 v korist SCDRR. Majhna, vendar pomembna razlika. Pri odpravniku DRR je lahko  $Q_{min}$  le še večji, kar povečuje njegovo mejo za RFM. Nasprotno pa je pri odpravniku SCDRR  $Q_{min}$  lahko le še manjši, kar znižuje mejo za RFM. Razlika med mejami za RFM se med odpravnikoma tako le veča. V vsakem primeru pa ima odpravnik SCDRR nižjo RFM mejo kot odpravnik DRR. Enako ugotovimo tudi za odpravnik Surplus Round Robin in Elastic Round Robin. RFM za slednja je podana v tabeli 6.1.

Zaključimo lahko, da je odpravnik SCDRR pravičnejši od mnogih navedenih odpravnikov. Pomembno pa je predvsem, da je pravičnejši kot DRR.

<sup>1</sup>Meja RFM bi bila še večja, če bi uporabili izraz (6.14), izpeljan v [93].



## 6.4 Zakasnitev algoritma SCDRR

Splošna definicija algoritemske zakasnitve odpravnikov je podana v razdelku 4.4.2. Vemo, da je od vseh odpravnikov najprimernejši GPS odpravnik, ki ima algoritemsko zakasnitev enako 0. V tem razdelku izračunamo algoritemsko zakasnitev za naš algoritem SCDRR. Pokažemo, da je naša mera sicer manj stroga kot nekatere druge mere izračunane za osnovni odpravnik DRR, vendar bolj natančna. Pri izračunih drugih mer so namreč uporabljene napačne predpostavke [98] ali storjene matematične napake v izpeljavah [42].

Če želimo prenašati podatke v realnem času, mora odpravnik vsakemu podatkovnemu pretoku zagotoviti omejeno zakasnitev v omrežni napravi. Pri tem težimo tudi k čimbolj optimalni zasedenosti omrežnih virov. Ker je zakasnitev algoritma del celotne zakasnitve v omrežni napravi in s tem na celotni prenosni poti, mora biti ta zakasnitev ravno tako omejena.

Zgornjo mejo algoritemske zakasnitve za odpravnik SCDRR izpeljemo po zgledu izpeljav v delih Stiliadisa in Varma [98] ter Kanhereja in Sethuja [42]. Z uporabo dveh različnih pristopov pridemo do istega rezultata, kar potrjuje njegovo pravilnost.

V razdelku C.3 za zgornjo mejo algoritemske zakasnitve čakalne vrste  $i$  odpravnika SCDRR, ki jo označimo s  $\theta_i$ , izpeljemo naslednjo neenačbo:

$$\theta_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (6.16)$$

Stiliadis in Varma sta v [98] za zgornjo mejo algoritemske zakasnitve izpeljala neenačbo:

$$\theta_{i_{DRR}} \leq \frac{3F - 2Q_i}{R} \quad (6.17)$$

ki je zaradi napačnih predpostavk glede odnosov med  $DC_i$ ,  $Q_i$  in  $L_{max}$  ter napačnih robnih pogojev zelo nenatančna. Najprej v izrazu podobnem (C.71) predpostavita, da velja neenakost  $DC_i \leq Q_i$  ( $\forall_i$ ). Ta predpostavka je napačna, saj smo z lemo C.1 dokazali, da velja neenakost  $DC_i \leq L_{max} - 1$ . Potem pa na mestu, kjer ovrednotita storitev v časovnem intervalu  $(t_1, t_{k+1})$ , ne uporabita pravilne omejitve vrednosti  $DC_i$ , dokazane z lemo C.1. V našem izvajanju meje algoritemske zakasnitve smo pravilno upoštevali vse zgoraj naštetе relacije.

Izboljšano mejo algoritemske zakasnitve, podane z enačbo (6.17), sta v [42] podala Kanhere in Sethu, in jo podaja spodnja neenačba:

$$\theta_{i_{DRR}} \leq \frac{1}{R} \left( (F - Q_i) + (L_{max} - 1) \left( \frac{F}{Q_i} + N - 2 \right) \right) \quad (6.18)$$

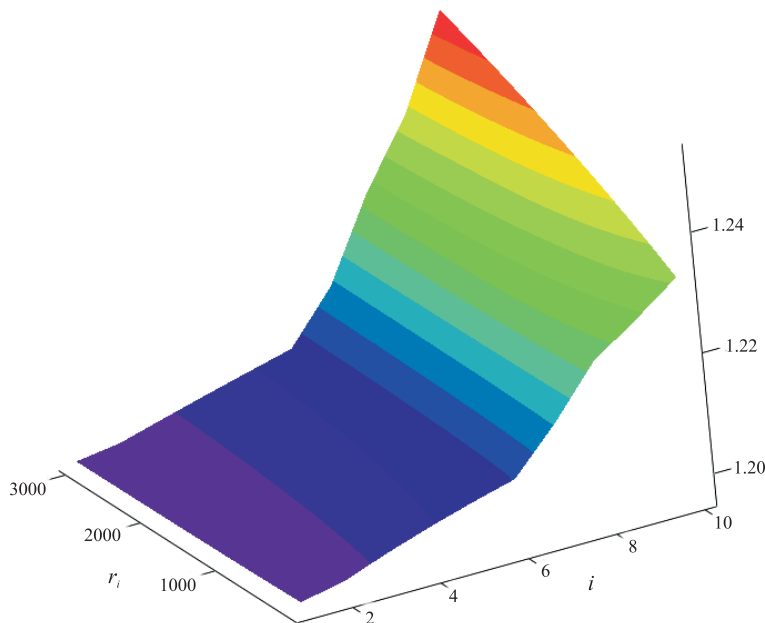
Tudi ta neenačba je nenatančna, saj je bila pri njeni izpeljavi storjena napaka pri upoštevanju neenakosti pri zamenjavi spremenljivk. Vrnimo se k izpeljavi zgornje meje algoritemske zakasnitve v razdelku C.3 in vrednosti izraza v oglatih oklepajih v neenačbi (C.60), ki jo še enkrat navajamo spodaj:

$$(\beta_i^k - \beta_i) - \frac{F - Q_i}{R} - \frac{(N - 1)(L_{max} - 1)}{R} + \frac{L_{max} - 1}{R} \left( 1 - \frac{F}{Q_i} \right) \quad (6.19)$$

Do te neenačbe sta naša izpeljava in izpeljava v dokazu Teorema 1 v [42] enaki. V tej točki izpeljave pa so avtorji v [42] linearno uporabili neenačbo  $r_i \leq \frac{Q_i}{F} R$ . Kot smo že pojasnili v dokazu teorema v C.2, je to matematično nepravilno, saj lahko izvedemo zamenjavo samo na pozitivnih členih razvitega izraza (6.19). Za pravilnost dokaza v [42] mora biti izraz (6.19) pozitiven za vse kombinacije nastopajočih spremenljivk. Ker to ne drži vedno, je njihov dokaz nepravilen. To trditev podkrepimo z nasprotnim primerom. Za izbrane vrednosti spremenljivk  $R = 4000$ ,  $N = 10$ ,  $Q_i = 200$ ,  $L_{max} = 1500$ ,  $k = 3$  ima izraz (6.19) vrednost -3.023. Za izračun časa  $\beta_i^k - \beta_i$  smo uporabili izraz

$$k \frac{F}{R} + \frac{(N-1)(L_{max}-1)}{R}$$

ki predstavlja najdaljše možno časovno obdobje, potrebno za zaključek  $k$  obhodov. Meja algoritemske zakasnitve, podana v [42], zato drži samo za nekatere kombinacije vrednosti  $k$  in  $Q_i$ . Napačna je za majhne vrednosti  $k$  in  $Q_i$  in je zaradi tega nepravilna.



Slika 6.3 – Relativna razlika zgornjih meja algoritemske zakasnitve izpeljane v tem delu (C.65) in podane v [42].  $Q = [10, 20, 40, 60, 80, 100, 200, 400, 600, 1000]$ ,  $L_{max} = 1500$ ,  $N = 10$

Če podrobneje pogledamo primerjavo obeh mej algoritemske zakasnitev, ki sta podani v [42] in z neenačbo (6.16), opazimo, da se z manjšanjem  $r_i$  in  $Q_i$  razlika med podanima mejama algoritemske zakasnitve manjša. Obratno velja za večanje  $r_i$  in  $Q_i$ . Oboje je vidno na sliki 6.3.

Tabela 6.2 prikazuje meje algoritemske zakasnitve različnih odpravnikov. Na prvi pogled ni možno natančno ugotoviti, kam se uvišča odpravnik SCDRR. Globlja analiza je zahtevna, saj so meje algoritemske zakasnitve posameznih odpravnikov funkcije več med seboj neprimerljivih spremenljivk. Za nas je najbolj zanimiva primerjava med mejami algoritemske zakasnitve odpravnikov SCDRR in DRR. Spomnimo se, da sta za DRR izpeljani meji v [98] in njena "izboljšana" verzija v [42]. Žal pa sta obe meji netočni in ju ne moremo neposredno primerjati z mejo izračunano za odpravnik SCDRR.

Odpravnik	Algoritemska zakasnitev
GPS	0
FIFO	$\infty$
PGPS (WFQ)	$\frac{L_{max}}{R} + \frac{L_{max}}{r_i}$
Virtual Clock Fair Queueing	$\frac{L_{max}}{R} + \frac{L_{max}}{r_i}$
SCFQ	$\frac{(N-1)L_{max}}{R} + \frac{L_{max}}{r_i}$
Frame Based Fair Queueing	$\frac{L_{max}}{R} + \frac{L_{max}}{r_i}$
DRR [42]	$\frac{1}{R} \left( (F - Q_i) + (L_{max} - 1) \left( \frac{F}{Q_i} + N - 2 \right) \right)$
Elastic Round Robin	$\frac{1}{R} \left( (F - Q_i) + (L_{max} - 1)(N - 1) \right)$
SCDRR	$Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{1}{r_i} - \frac{2}{R} + \frac{Q_i}{F r_i} N \right)$

Tabela 6.2 – Primerjava mej algoritemske zakasnitve za različne odpravnike

## 6.5 Analiza računske zahtevnosti SCDRR

Splošna definicija zahtevnosti odpravnikov je podana v razdelku 4.4.3. Tu podamo še računsko zahtevnost za algoritem SCDRR s stališča posameznega paketa.

Izračunamo, za kakšen faktor se pri SCDRR glede na osnovno različico DRR poveča število operacij potrebnih za odpravo posameznega paketa. Najprej navajamo splošne rezultate in za tem še rezultate za eksponentno in enakomerno verjetnostno porazdelitev dolžin paketov.

Najprej podajmo definicije in seznam spremenljivk. Nekatere definicije so v podobni obliki zapisane že v dodatku C. Tu jih zaradi preglednosti navajamo še enkrat.

### 6.5.1 Definicije in seznam spremenljivk

**Definicija 6.1** Čas pasivnosti čakalne vrste je vsak časovni interval, v katerem je čakalna vrsta ves čas prazna.

**Definicija 6.2** Čas aktivnosti čakalne vrste je časovni interval  $(t_0, t_1)$ , v katerem se v čakalni vrsti stalno nahaja vsaj en paket.

Čakalna vrsta preide iz pasivnega v aktivno stanje, ko ob času  $t_0$  prazna čakalna vrsta sprejme prvi paket. Čakalna vrsta preide iz aktivnega v pasivno stanje, ko jo ob času  $t_1$  v celoti zapusti zadnji čakajoči paket.

**Definicija 6.3** Obhod je interval, v katerem odpravnik eno za drugo obiše vse aktivne čakalne vrste in nad njimi izvede operacije, predpisane z vrsto strežbe čakalne vrste (odpravi paket(e), poveča DC, drugo).

**Definicija 6.4** Kvant  $Q$  določa število (kreditnih) bitov, do katerih je v vsakem obhodu upravičena posamezna aktivna čakalna vrsta.

Ker nadalje obravnavamo samo eno čakalno vrsto, smo zaradi preglednosti izpustili indeks  $i$ .

**Definicija 6.5** Števec primanjkljaja  $DC$  je spremenljivka stanja čakalne vrste. Določa največje število bitov, ki jih v tekočem obhodu lahko odpravimo iz čakalne vrste (postrežemo).

Vrednost  $DC$  vsake aktivne čakalne vrste se v vsakem obhodu najprej poveča za  $Q$  in nato zmanjša za vsoto bitov vseh postreženih paketov<sup>2</sup>. Vrednost  $DC$  se postavi na nič, ko se iz čakalne vrste odpravi zadnji paket in vrsta postane pasivna.

**Definicija 6.6** Operacija je definirana kot opravilo izvedeno v aktivni čakalni vrsti, ki ne povzroči povečanja  $DC$ . Samodejno povečanje  $DC$  za vrednost  $Q$  ob vsakem obhodu ne šteje kot operacija.

To pomeni, da za operacijo štejemo vsako postrežbo paketa, po kateri se vrednost  $DC$  zmanjša za dolžino paketa in vsak neuspešen obhod, pri katerem zaradi premajhne vrednosti  $DC$  ne moremo postreči nobenega paketa in zato vrednost  $DC$  ostane nespremenjena<sup>3</sup>. V enem obhodu zato izvedemo vsaj eno operacijo (postrežba enega ali nobenega paketa) ali več operacij (postrežba več paketov v istem obhodu). Definirajmo še ostale spremenljivke:

$f(l)$	gostota verjetnostne porazdelitev dolžin paketov,
$F(l)$	kumulativna verjetnostna porazdelitev dolžin paketov,
$\bar{L}$	povprečna dolžina paketov,
$l$	dolžina posameznega paketa,
$L_{min}$	najmanjša možna dolžina paketa,
$L_{max}$	največja možna dolžina paketa,
$M$	število vseh paketov, ki so prispeli v čakalno vrsto v času njene aktivnosti,
$m$	zaporedna številka paketa v čakalni vrsti v času njene aktivnosti $m = 1, 2, 3, \dots, M$ ,
$N_m$	največje možno število obhodov, potrebnih za oddajo $m$ -tega paketa,
$O_m$	povprečno število operacij, potrebnih za odpravo $m$ -tega paketa,
$n$	zaporedna številka obhoda, šteto od zadnje oddaje paketa iz te čakalne vrste $n = 0, 1, 2, \dots, N_m$ ,
$p_{m,n}$	verjetnost, da bo $m$ -ti paket postrežen v $n$ -tem obhodu,
$S_m$	verjetnost gotovega dogodka za $m$ -ti paket, ki mora biti enak 1 za vsak $m$ ,
$L_{m,n}$	povprečna dolžina $m$ -tega paketa, postreženega v $n$ -tem obhodu,
$DC$	trenutna vrednost števca primanjkljaja (Deficit Counter),
$DC_m$	povprečna vrednost $DC$ po oddaji $m$ -tega paketa,
$Q$	velikost kvanta (vrednost, ki se prišteje $DC$ ob vsakem obhodu).

<sup>2</sup>Ta vsota seveda ne more biti večja od same vrednosti  $DC$

<sup>3</sup>To se nanaša na vrednost  $DC$ , ki je že bila povečana za  $Q$

### 6.5.2 Robni pogoji

$DC_0 = 0$  - začetna vrednost  $DC$  je po definiciji delovanja DRR vedno enaka nič. To je vrednost  $DC$  ob času, ko vanjo prispe prvi paket in čakalna vrsta postane aktivna.

$L_{min} \leq Q \leq L_{max} - 1$  - vrednost kvanta  $Q$  naj bo večja ali enaka  $L_{min}$ , saj bi v nasprotnem primeru že za odpravo najkrajšega možnega paketa morali opraviti več kot eno operacijo (kar res ne bi imelo smisla), in manjša ali enaka  $L_{max} - 1$ , saj bi drugače imeli opraviti z osnovno različico DRR, kjer vedno velja  $Q \geq L_{max}$  in je računska zahtevnost  $O(1)$ .

### 6.5.3 Izračun števila potrebnih operacij

Analizo števila potrebnih operacij za odpravo posameznega paketa iz čakalne vrste začnemo ob času  $t_0$ , ki po definiciji iz 6.5.1 predstavlja čas prehoda čakalne vrste v aktivno stanje. Definiramo lahko začetni pogoj  $DC_0 = 0$ , ki predstavlja vrednost spremenljivke  $DC$  ob času  $t_0$ . Analizo začnemo pri  $m = 1$ , prvem paketu v čakalni vrsti. Najprej izračunamo največje možno število obhodov  $N_1$ , po katerih bo postrežen prvi paket.

$$N_1 = \left\lceil \frac{L_{max}}{Q} \right\rceil \quad (6.20)$$

Sedaj izračunajmo  $p_{1,1}$ , ki predstavlja verjetnost, da bo prvi paket ( $m = 1$ ) v čakalni vrsti odpravljen v prvem obhodu ( $n = 1$ ). To je enako verjetnosti, da je dolžina prvega paketa manjša ali enaka trenutni vrednosti  $DC$ , ki je v prvem obhodu enaka vrednosti kvanta  $Q$ . Namreč, ko je čakalna vrsta ob času  $t_0$  prešla v aktivno stanje, je bila vrednost  $DC = 0$ , ob vsakem obhodu se potem vrednost  $DC$  poveča za  $Q$ . Tako se je zgodilo tudi pri prvem obhodu, ko je bila vrednost  $DC$  povečana z nič na  $Q$ .

$$p_{1,1} = P(l \leq Q) = \int_0^Q f(l)dl = F(Q) \quad (6.21)$$

Na podoben način potem izračunamo tudi verjetnosti odprave prvega paketa v naslednjih obhodih ( $n > 1$ ), pri čemer ob vsakem obhodu vrednosti  $DC$  prištejemo vrednost kvanta  $Q$ . Tako dobimo:

$$p_{1,2} = P(Q < l \leq 2Q) = \int_Q^{2Q} f(l)dl = F(2Q) - F(Q) \quad (6.22)$$

⋮

$$\begin{aligned} p_{1,n} &= P((n-1)Q < l \leq nQ) \\ &= \int_{(n-1)Q}^{nQ} f(l)dl = F(nQ) - F((n-1)Q) \end{aligned} \quad (6.23)$$

⋮

$$\begin{aligned} p_{1,N_1} &= P((N_1-1)Q < l \leq N_1Q) \\ &= \int_{(N_1-1)Q}^{N_1Q} f(l)dl = F(N_1Q) - F((N_1-1)Q) \end{aligned} \quad (6.24)$$

Če upoštevamo enačbo (6.20) in dejstvo, da bo po  $N_1$  zaporednih obhodih vrednost  $DC$  najmanj  $N_1Q$  (seveda, če paket ni bil postrežen že prej), potem velja tudi:

$$DC \geq N_1Q \geq L_{max} \quad (6.25)$$

Preden nadaljujemo, na kratko osvežimo nekatere lastnosti verjetnostnih porazdelitev  $f(l)$  in  $F(l)$ . Če dolžine paketov ležijo med  $L_{min}$  in  $L_{max}$ , potem za vsako verjetnostno porazdelitev veljajo naslednji izrazi:

$$\begin{aligned} P(l < L_{min}) &= \int_0^{L_{min}} f(l)dl = F(L_{min}) = 0 \\ P(l > L_{max}) &= \int_{L_{max}}^{\infty} f(l)dl = F(\infty) - F(L_{max}) = 0 \\ P(L_{min} \leq l \leq L_{max}) &= \int_{L_{min}}^{L_{max}} f(l)dl = F(L_{max}) - F(L_{min}) = 1 \end{aligned} \quad (6.26)$$

Verjetnost, da bo paket krajši od  $L_{min}$  ali daljši od  $L_{max}$  je enaka nič, verjetnost, da se bo njegova dolžina nahajala med  $L_{min}$  in  $L_{max}$  pa je enaka ena. Zato lahko ob upoštevanju neenačbe (6.25) trdimo, da je odprava prvega paketa po  $N_1$  obhodih gotov dogodek  $S_1$ , saj takrat vedno velja  $l \leq DC$ . Verjetnost gotovega dogodka  $S_1$  lahko zapišemo kot:

$$S_1 = P(l \leq N_1Q) = P(l \leq L_{max}) = 1 \quad (6.27)$$

Gotov dogodek  $S_1$  pa lahko izračunamo tudi kot vsoto vseh možnih enostavnih dogodkov  $p_{1,n}$  po enačbi:

$$S_1 = \sum_{n=1}^{N_1} p_{1,n} = \int_0^Q f(l)dl + \int_Q^{2Q} f(l)dl + \dots + \int_{(N_1-1)Q}^{N_1Q} f(l)dl \quad (6.28)$$

Vsoto integralov s stikajočimi se mejami integracije (kot v zgornjem primeru) lahko zapišemo kot integral s skrajnimi mejami integracije  $(0, N_1Q)$ . Zato lahko enačbo (6.28) ob hkratnem upoštevanju lastnosti verjetnostne porazdelitve dolžin paketov zapišemo kot:

$$S_1 = \int_0^{N_1Q} f(l)dl = \int_{L_{min}}^{L_{max}} f(l)dl = 1 \quad (6.29)$$

iz česar vidimo, da nam vsota verjetnosti odprave prvega paketa  $p_{1,n}$  ( $n = 1, 2, \dots, N_1$ ) v posameznih obhodih  $n$  da ravno gotov dogodek  $S_1$ .

Sedaj že lahko izračunamo povprečno število operacij  $O_1$  potrebnih za oddajo prvega paketa:

$$O_1 = \sum_{n=1}^{N_1} np_{1,n} \quad (6.30)$$

Če želimo izračunati povprečno število operacij tudi za drugi paket v čakalni vrsti, moramo najprej izračunati povprečno vrednost  $DC$  po odpravi prvega paketa, ki jo označimo z  $DC_1$ . Za to pa potrebujemo vrednosti  $L_{1,n}$ , ki predstavljajo povprečno

dolžino prvega paketa, če je bil le ta postrežen v  $n$ -tem obhodu. Najprej izračunajmo  $L_{1,1}$  ali povprečno dolžino prvega paketa, če je bil le ta postrežen v prvem obhodu:

$$L_{1,1} = \frac{\int_0^Q l f(l) dl}{\int_0^Q f(l) dl} \quad (6.31)$$

Izračunati moramo povprečno dolžino paketa na intervalu  $(0, Q]$ , ki je podana z izrazom  $\int_0^Q l f(l) dl$  in predstavlja povprečno dolžino paketov, ki so lahko postreženi v prvem obhodu. Dobljeno vrednost moramo potem utežiti še z verjetnostjo, da bo dolžina prvega paketa padla v podane meje, ki ustreza izrazu  $\int_0^Q f(l) dl$ . Podobno potem izračunamo tudi preostale vrednosti  $L_{1,n}$ .

$$L_{1,2} = \frac{\int_Q^{2Q} l f(l) dl}{\int_Q^{2Q} f(l) dl} \quad (6.32)$$

⋮

$$L_{1,n} = \frac{\int_{(n-1)Q}^{nQ} l f(l) dl}{\int_{(n-1)Q}^{nQ} f(l) dl} \quad (6.33)$$

⋮

$$L_{1,N_1} = \frac{\int_{(N_1-1)Q}^{N_1Q} l f(l) dl}{\int_{(N_1-1)Q}^{N_1Q} f(l) dl} \quad (6.34)$$

Sedaj lahko izračunamo  $DC_1$  po izrazu (6.35), kjer spremenljivke  $dc_{1,n}$ , za  $n = 1, 2, \dots, N_1$ , predstavljajo preostanek trenutne vrednosti  $DC$ , če je bil prvi paket odpravljen v  $n$ -tem obhodu. Za izračun  $dc_n$  uporabimo izraz  $dc_n = nQ - L_{1,n}$ , ki predstavlja razliko med akumulirano vrednostjo  $DC$  v  $n$ -tem obhodu poskusa odprave prvega paketa in dolžino odpravljenega paketa, ki leži na intervalu  $(n-1)Q < l < nQ$ .

$$\begin{aligned} DC_1 &= p_{1,1}dc_{1,1} + p_{1,2}dc_{1,2} + \dots + p_{1,n}dc_{1,n} + \dots + p_{1,N_1}dc_{1,N_1} \\ &= p_{1,1}(Q - L_{1,1}) + p_{1,2}(2Q - L_{1,2}) + \dots + p_{1,n}(nQ - L_{1,n}) \\ &+ \dots + p_{1,N_1}(N_1Q - L_{1,N_1}) \\ &= \sum_{n=1}^{N_1} p_{1,n}(nQ - L_{1,n}) \end{aligned} \quad (6.35)$$

Izračun za drugi paket v čakalni vrsti poteka nekoliko drugače. Za razliko od prvega paketa, ki je ob prehodu čakalne vrste v aktivno stanje ob času  $t_0$  naletel na vrednost spremenljivke  $DC = DC_0 = 0$ , drugi paket ob času  $t_1$  naleti na vrednost spremenljivke  $DC = DC_1 \geq 0$ . To nekoliko spremeni izraze podane za  $m = 2$ .

Najprej izračunajmo največje možno število obhodov  $N_2$ , potrebnih za postrežbo drugega paketa:

$$N_2 = \left\lceil \frac{L_{max} - DC_1}{Q} \right\rceil \quad (6.36)$$

Ker imamo pri drugem paketu na voljo tudi ostanek  $DC_1$  od prvega paketa, velja da je  $N_2 \leq N_1$ . Sedaj lahko izračunamo verjetnosti  $p_{2,n}$  ( $n = 0, 1, 2, \dots, N_2$ ).

Zaradi  $DC_1 \geq 0$  obstaja verjetnost, da bo drugi paket postrežen v istem obhodu kot prvi; torej brez potrebe po dodajanju novega kvanta  $Q$ . To verjetnost definiramo kot  $p_{2,0}$  in znaša:

$$p_{2,0} = P(l \leq DC_1) = \int_0^{DC_1} f(l)dl = F(DC_1) \quad (6.37)$$

Na podoben način kot za prvi paket in ob upoštevanju vrednosti  $DC_1$  izračunamo tudi verjetnosti odprave drugega paketa v naslednjih obhodih ( $n > 0$ ), pri čemer ob vsakem obhodu  $DC$  prištejemo  $Q$ . Tako dobimo:

$$\begin{aligned} p_{2,1} &= P(DC_1 < l \leq DC_1 + Q) = \int_{DC_1}^{DC_1+Q} f(l)dl \\ &= F(DC_1 + Q) - F(DC_1) \end{aligned} \quad (6.38)$$

$$\begin{aligned} p_{2,2} &= P(DC_1 + Q < l \leq DC_1 + 2Q) = \int_{DC_1+Q}^{DC_1+2Q} f(l)dl \\ &= F(DC_1 + 2Q) - F(DC_1 + Q) \end{aligned} \quad (6.39)$$

⋮

$$\begin{aligned} p_{2,n} &= P(DC_1 + (n-1)Q < l \leq DC_1 + nQ) = \int_{DC_1+(n-1)Q}^{DC_1+nQ} f(l)dl \\ &= F(DC_1 + nQ) - F(DC_1 + (n-1)Q) \end{aligned} \quad (6.40)$$

⋮

$$\begin{aligned} p_{2,N_2} &= P(DC_1 + (N_2-1)Q < l \leq DC_1 + N_2Q) = \int_{DC_1+(N_2-1)Q}^{DC_1+N_2Q} f(l)dl \\ &= F(DC_1 + N_2Q) - F(DC_1 + (N_2-1)Q) \end{aligned} \quad (6.41)$$

Povprečno število operacij, potrebnih za odpravo drugega paketa, se ravno tako izračuna nekoliko drugače. Ker po definiciji operacije (glej 6.5.1) zanjo štejemo tudi odpravo paketa brez poprejšnjega povečanja  $DC$  za vrednost  $Q$ , moramo v izrazu (6.30) upoštevati še verjetnost, da bo drugi paket postrežen v istem obhodu kot prvi. Ta verjetnost je podana z izrazom  $p_{2,0}$ . V bistvu to pomeni, da je verjetnost, da bo drugi paket potreboval eno operacijo za odpravo, enaka vsoti verjetnosti  $p_{2,0} + p_{2,1}$ . Če za  $n$  vzamemo vrednosti  $n = 0, 1, 2, \dots, N_2$ , dobimo spodnji izraz za  $O_2$ .

$$O_2 = p_{2,0} + \sum_{n=1}^{N_2} np_{2,n} \quad (6.42)$$

Podobno kot pri  $p_{2,n}$  moramo tudi pri izračunu  $L_{2,n}$  in  $DC_2$  upoštevati vrednost  $DC_1$ . To nam da enačbe:



$$L_{2,0} = \frac{\int_0^{DC_1} l f(l) dl}{\int_0^{DC_1} f(l) dl} \quad (6.43)$$

$$L_{2,1} = \frac{\int_{DC_1}^{Q+DC_1} l f(l) dl}{\int_{DC_1}^{Q+DC_1} f(l) dl} \quad (6.44)$$

$$\vdots$$

$$L_{2,n} = \frac{\int_{(n-1)Q+DC_1}^{nQ+DC_1} l f(l) dl}{\int_{(n-1)Q+DC_1}^{nQ+DC_1} f(l) dl} \quad (6.45)$$

$$\vdots$$

$$L_{2,N_2} = \frac{\int_{(N_2-1)Q+DC_1}^{N_2Q+DC_1} l f(l) dl}{\int_{(N_2-1)Q+DC_1}^{N_2Q+DC_1} f(l) dl} \quad (6.46)$$

Za  $DC_2$  vrednosti  $dc_{2,n}$  izračunamo s pomočjo izraza  $dc_{2,n} = DC_1 + nQ - L_{2,n}$ , kar nam da:

$$\begin{aligned} DC_2 &= p_{2,0}dc_{2,0} + p_{2,1}dc_{2,1} + \dots + p_{2,n}dc_{2,n} + \dots + p_{2,N_2}dc_{2,N_2} \\ &= p_{2,0}(DC_1 - L_{2,0}) + p_{2,1}(DC_1 + Q - L_{2,1}) + \dots + \\ &+ p_{2,n}(DC_1 + nQ - L_{2,n}) + p_{2,N_2}(DC_1 + N_2Q - L_{2,N_2}) \\ &= \sum_{n=0}^{N_2} p_{2,n}(DC_1 + nQ - L_{2,n}) \end{aligned} \quad (6.47)$$

Za naslednje pakete v čakalni vrsti izračun poteka enako kot za drugega. Upoštevati moramo vrednost  $DC$  ob koncu odprave prejšnjega paketa. Za izračun vrednosti parametrov paketa  $m$  je tako potrebno upoštevati vrednost  $DC_{m-1}$ . Splošni izrazi so podani s spodnjimi enačbami, pri čemer moramo upoštevati, da vrednosti  $p_{1,0}$  in  $L_{1,0}$  glede na izpeljavo ne obstajata, za enostavnejšo izražavo splošnih rezultatov pa lahko privzamemo, da sta enaki nič, kar dobimo, če izračunamo vrednosti  $p_{m,n}$  in  $L_{m,n}$  pri  $m = 1$  in  $n = 0$ .

$$N_m = \left\lceil \frac{L_{max} - DC_{m-1}}{Q} \right\rceil \quad (6.48)$$

$$m = 1, 2, 3, \dots, M$$

$$n = 0, 1, 2, \dots, N_m$$

$$p_{m,0} = P(l \leq DC_{m-1}) = \int_0^{DC_{m-1}} f(l)dl = F(DC_{m-1}) \quad (6.49)$$

$$\begin{aligned} p_{m,n} &= P(DC_{m-1} + (n-1)Q < l \leq DC_{m-1} + nQ) \\ &= \int_{DC_{m-1} + (n-1)Q}^{DC_{m-1} + nQ} f(l)dl \\ &= F(DC_{m-1} + nQ) - F(DC_{m-1} + (n-1)Q) \end{aligned} \quad (6.50)$$

$$L_{m,0} = \frac{\int_0^{DC_{m-1}} lf(l)dl}{\int_0^{DC_{m-1}} f(l)dl} \quad (6.51)$$

$$L_{m,n} = \frac{\int_{(n-1)Q + DC_{m-1}}^{nQ + DC_{m-1}} lf(l)dl}{\int_{(n-1)Q + DC_{m-1}}^{nQ + DC_{m-1}} f(l)dl} \quad (6.52)$$

$$dc_{m,n} = DC_{m-1} + nQ - L_{m,n} \quad (6.53)$$

$$\begin{aligned} DC_m &= p_{m,0}dc_{m,0} + p_{m,1}dc_{m,1} + \dots + p_{m,n}dc_{m,n} + \dots + p_{m,N_m}dc_{n,N_m} \\ &= p_{m,0}(DC_{m-1} - L_{m,0}) + p_{m,1}(DC_{m-1} + Q - L_{m,1}) + \dots + \\ &+ p_{m,n}(DC_{m-1} + nQ - L_{m,n}) + \dots + \\ &+ p_{m,N_m}(DC_{m-1} + N_mQ - L_{m,N_m}) \\ &= \sum_{n=0}^{N_m} p_{m,n}(DC_{m-1} + nQ - L_{m,n}) \end{aligned} \quad (6.54)$$

$$O_m = p_{m,0} + \sum_{n=1}^{N_m} np_{m,n} \quad (6.55)$$

Na koncu lahko izračunamo še povprečno število operacij za odpravo posameznega paketa, če je bilo v času aktivnosti čakalne vrste število vseh postreženih paketov enako  $M$ .

$$\overline{O(M)} = \frac{1}{M} \sum_{m=1}^M O_m \quad (6.56)$$

Končna rezultata teoretične analize sta  $O_m$  in  $\overline{O(M)}$ .  $O_m$  predstavlja povprečno število operacij, ki so potrebne za odpravo  $m$ -tega paketa iz določene čakalne vrste tekom istega aktivnega obdobja. Predstavlja faktor povečanja števila operacij odpravnikoma SCDRR napram odpravniku DRR.  $\overline{O(M)}$  pa je povprečni faktor povečanja števila operacij, če je bilo iz čakalne vrste tekom istega aktivnega obdobja odpravljenih  $M$  paketov. Več o teh faktorjih in relacijah med odpravnikoma SCDRR in DRR pa v naslednjih razdelkih, kjer navajamo rezultate zahtevnosti odpravnikoma SCDRR pri eksponentni in enakomerni porazdelitvi dolžin paketov.

### 6.5.4 Zahtevnost pri eksponentni porazdelitvi dolžin paketov

Splošni primer uporabimo za izpeljavo rezultatov za posebne primere. Prvi je primer za eksponentno verjetnostno porazdelitev dolžin paketov, za katero lahko zapišemo naslednje izraze:

$$\begin{aligned} f(l) &= \lambda e^{-\lambda l} & L_{min} &= 0 & \bar{L} &= \frac{1}{\lambda} \\ F(l) &= 1 - e^{-\lambda l} & L_{max} &= \infty & N_m &= \infty \end{aligned}$$

Pri tem  $\lambda$  določa recipročno vrednost povprečne dolžine paketov v čakalni vrsti. Če upoštevamo še robne pogoje iz 6.5.2, opazimo, da se pri eksponentni porazdelitvi glede na vrednost kvanta  $Q$ , vedno nahajamo v podkritičnem režimu delovanja DRR. Ker je  $L_{max} = \infty$  namreč vedno velja, da je  $Q$  manjši od  $L_{max}$ .

Celotna izpeljava je podana v razdelku C.4 dodatka C. Tu pa podajamo le končne rezultate. Pri  $DC_0 = 0$ ,  $m = 1, 2, 3, \dots, M$  in  $n = 0, 1, 2, \dots, \infty$  velja:

$$DC_m = DC_{m-1} - \frac{1}{\lambda} + \frac{Qe^{\lambda Q}e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \quad (6.57)$$

$$p_{m,n} = \begin{cases} 1 - e^{-\lambda DC_{m-1}} & n = 0 \\ e^{-\lambda n Q} e^{-\lambda DC_{m-1}} (e^{\lambda Q} - 1) & n > 0 \end{cases} \quad (6.58)$$

$$L_{m,n} = \begin{cases} 0 & m = 1, n = 0 \\ \frac{1}{\lambda} + DC_{m-1} - \frac{DC_{m-1}}{1 - e^{-\lambda DC_{m-1}}} & m > 1, n = 0 \\ nQ + DC_{m-1} + \frac{1}{\lambda} - \frac{Qe^{\lambda Q}}{e^{\lambda Q} - 1} & m > 1, n > 0 \end{cases} \quad (6.59)$$

$$O_m = 1 + \frac{e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \quad (6.60)$$

Za izračun povprečnega števila operacij za  $m$ -ti paket ( $O_m$ ) potrebujemo vrednosti  $DC_m$ , ki pa se računajo iterativno po enačbi (6.57). To je zamudno in nepraktično. Iz primera C.4.4 in tabele C.3 vidimo, da se vrednost  $DC_m$  z višanjem indeksa  $m$  ustali pri neki stacionarni vrednosti. Podobno velja za  $O_m$  in  $\overline{O(m)}$ , kar kaže slika C.4.

Za primer, ko gre  $m \rightarrow \infty$ , lahko zapišemo poenostavitev izračuna  $DC_m$  z enačbo (C.108):

$$DC_m = -\frac{1}{\lambda} \ln \left[ \frac{1}{\lambda Q} (1 - e^{-\lambda Q}) \right] \quad (6.61)$$

Za majhne vrednosti  $m$  pa lahko vzamemo kar vrednost  $DC_1$  (C.109):

$$DC_m = DC_1 = \frac{Qe^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \quad (6.62)$$

in s tem pri izračunih  $DC_m$  pri  $m > 1$  zanemarimo verjetnosti  $p_{m,0}$ . Primerjava med iterativno izračunanimi vrednostmi  $\overline{O(m)}$  in vrednostmi, izračunanimi s pomočjo obeh

poenostavitev, je podana v grafu na sliki C.5. Na sliki C.6 pa še njihovo relativno odstopanje.

Iz podanih rezultatov je razvidno, da sta  $O_m$  in  $\overline{O(M)}$  največja pri  $m = 1$  in pri majhnih vrednostih  $M$ . Za majhne vrednosti  $M$  velja, da je omrežje malo obremenjeno. Takrat stanje čakalnih vrst stalno alternira med aktivnim in pasivnim. To pomeni, da je v večini primerov v posameznem stanju aktivnosti postreženih le nekaj paketov. Pri naraščajočem  $M$  pa so čakalne vrste vse bolj polne, časi aktivnosti so daljši in to je značilno za bolj obremenjeno omrežje. Takrat se tako  $O_m$  kot  $\overline{O(M)}$  približujeta neki asimptotični vrednosti. To je za nas ugoden rezultat, saj pri nizko obremenjenem omrežju število operacij za odpravo paketa ni toliko pomembno kot pri visoko obremenjenem omrežju.

V tabeli 6.3 lahko primerjamo rezultate za  $\overline{O(M)}$  pri različnih vrednostih kvantov za eksponentno porazdelitev dolžin paketov. Vidimo, da  $\overline{O(M)}$  ne narašča linearno s padanjem vrednosti kvanta. Pri povprečni dolžini paketov 1000 oktetov in enaki velikosti kvanta je faktor povečanja števila operacij pri SCDRR le 1,369. To je zelo ugoden rezultat, saj je pri DRR zahtevana vrednost kvanta kar 16000 oktetov. Pri 16-krat manjši velikosti kvanta je pri odpravniku SCDRR za odpravo enega paketa v povprečju potrebnih le 1,369-krat več operacij kot pri odpravniku DRR.

### 6.5.5 Zahtevnost pri enakomerni porazdelitvi dolžin paketov

Zapišimo izraze in vrednosti, ki ustrezajo enakomerni verjetnostni porazdelitvi dolžin paketov.

$$\begin{aligned} 0 < L_{min} \leq L_{max} & & \bar{L} &= \frac{1}{2}(L_{max} + L_{min}) \\ 0 < L_{max} \leq \infty & & L_{min} \leq Q < L_{max} & & N_m \leq \infty \end{aligned}$$

Zapišimo še izraze, ki veljajo za gostoto verjetnostne porazdelitve in kumulativno verjetnostno porazdelitev dolžin paketov.

$$\begin{aligned} f(l) &= \begin{cases} \frac{1}{L_{max} - L_{min}} & L_{min} \leq l \leq L_{max} \\ 0 & \text{drugje} \end{cases} \\ F(l) &= \begin{cases} 0 & l < L_{min} \\ \frac{l - L_{min}}{L_{max} - L_{min}} & L_{min} \leq l < L_{max} \\ 1 & l \geq L_{max} \end{cases} \end{aligned}$$

Ob upoštevanju robnih pogojev iz 6.5.2 zapišemo končne rezultate za enakomerno porazdelitev dolžin paketov. Celotna izpeljava je podana v razdelku C.5 dodatka C. Pri izpeljavi smo upoštevali zgornje definicije spremenljivk, začetni pogoj  $DC_0 = 0$  ter zalogo vrednosti indeksov  $m = 1, 2, 3, \dots, M$  in  $n = 0, 1, 2, \dots, N_m$ . Pri enakomerni porazdelitvi moramo upoštevati dva različna primera izračuna vrednosti spremenljivk glede na vrednost  $DC_{m-1}$  in sicer prvi primer nastopi, ko je  $DC_{m-1} < L_{min}$ , drugi pa ko je  $DC_{m-1} \geq L_{min}$ .

$$N_m = \left\lceil \frac{L_{max} - DC_{m-1}}{Q} \right\rceil \quad (6.63)$$

$$DC_m = \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 \right. \\ \left. + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right. \\ \left. + 2DC_{m-1}(L_{max} - L_{min} + Q(1 - N_m)) \right] \quad DC_{m-1} < L_{min} \quad (6.64)$$

$$DC_m = \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right. \\ \left. + 2DC_{m-1}(L_{max} - L_{min} - QN_m) \right] \quad DC_{m-1} \geq L_{min} \quad (6.65)$$

$$p_{m,n} = \begin{cases} 0 & n = 0, DC_{m-1} < L_{min} \\ \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}} & n = 0, DC_{m-1} \geq L_{min} \\ \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} & n = 1, DC_{m-1} < L_{min} \\ \frac{Q}{L_{max} - L_{min}} & n = 1, DC_{m-1} \geq L_{min} \\ \frac{Q}{L_{max} - L_{min}} & 1 < n < N_m \\ \frac{L_{max} - L_{min}}{L_{max} - (N_m - 1)Q - DC_{m-1}} & n = N_m \end{cases} \quad (6.66)$$

$$L_{m,n} = \begin{cases} 0 & n = 0, DC_{m-1} < L_{min} \\ \frac{DC_{m-1} + L_{min}}{2} & n = 0, DC_{m-1} \geq L_{min} \\ \frac{Q + DC_{m-1} + L_{min}}{2} & n = 1, DC_{m-1} < L_{min} \\ \frac{(2n-1)Q}{2} + DC_{m-1} & n = 1, DC_{m-1} \geq L_{min} \\ \frac{(2n-1)Q}{2} + DC_{m-1} & 1 < n < N_m \\ \frac{L_{max} + (N_m - 1)Q + DC_{m-1}}{2} & n = N_m \end{cases} \quad (6.67)$$

$$O_m = \frac{1}{L_{max} - L_{min}} \left[ N_m L_{max} - L_{min} + (1 - N_m) \left( DC_{m-1} + \frac{N_m Q}{2} \right) \right] \quad (6.68)$$

Tudi pri enakomerni porazdelitvi se vrednosti  $DC_m$ , ki so potrebne za izračun povprečnega števila operacij  $O_m$ , računajo iterativno po enačbah (6.64) in (6.65). Tudi tu v tabelah C.6 in C.9 opazimo, da se vrednost  $DC_m$  z višanjem indeksa  $m$  ustali pri neki stacionarni vrednosti. Podobno velja za  $O_m$  in  $\overline{O(m)}$ , kar kaže slika C.7.

Tudi za poenostavljeni izračun  $DC_m$ , ko gre  $m \rightarrow \infty$ , dobimo dva rezultata. Z  $DC1_m$  označimo primer pri pogoju  $DC_m < L_{min}$ , z  $DC2_m$  pa primer pri pogoju  $DC_m \geq L_{min}$ .

$$DC1_m = \frac{L_{max}^2 - L_{min}^2 + 2QL_{min} - 2N_mQL_{max} - N_mQ^2(1 - N_m)}{2Q(1 - N_m)} \quad (6.69)$$

$$DC2_m = \frac{L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m)}{2QN_m} \quad (6.70)$$

Pri izbiri primernejše izmed obeh poenostavitvev si pomagamo z grafom na sliki C.8. Podrobnejši opis najprimernejše izbire pa je opisan v razdelku C.5.6. Pomagamo si s presečiščem krivulj obeh poenostavitvev, ki ga glede na  $L_{min}$  izračunamo po izrazu:

$$L_{min,2} = QN_m \frac{+}{-} \sqrt{L_{max}^2 - 2N_mQL_{max} - N_mQ^2(1 - 2N_m)} \quad (6.71)$$

Mejno vrednost  $L_{min}$ , ki določa uporabo posamezne poenostavitve, lahko enostavno določimo s pomočjo korenov izraza (6.71). Glede na to, da je v večini omrežij vrednost  $L_{min}$  dokaj nizka (najmanjše dovoljene dolžine paketov so nekaj deset oktetov), je zato večinoma uporabljena kar poenostavitev  $DC2_m$  (glej sliko C.8).

Zanima nas še, kako poenostavljen izračun za  $DC_m$  vpliva na izračun števila operacij  $O_m$ , potrebnih za odpravo  $m$ -tega paketa. Ker v enačbi (C.136) za  $O_m$  nastopa člen  $DC_{m-1}$  moramo zopet upoštevati dva različna primera. Z  $Op1_m$  označimo poenostavljen izračun pri pogoju  $DC_m < L_{min}$ , pri čemer smo v enačbo  $O_m$  (C.136) vstavili poenostavljen izraz  $DC1_m$ . Pri  $Op2_m$  pa uporabimo izraz  $DC1_m$ , izračunan pri  $DC_m \geq L_{min}$ .

$$Op1_m = \frac{L_{min} + L_{max}}{2Q} \quad (6.72)$$

$$Op2_m = 1 - \frac{(1 - N_m)(L_{max}^2 - L_{min}^2 - N_mQ^2)}{2N_mQ(L_{max} - L_{min})} \quad (6.73)$$

Pri izbiri primernejše izmed obeh poenostavitev si pomagamo z grafom na sliki C.9. Podrobnejši opis najprimernejše izbire pa je opisan v razdelku C.5.7. Tudi tu si pomagamo s presečiščem krivulj obeh poenostavitev  $Op1_m$  in  $Op2_m$ , za katerega se izkaže, da leži pri isti vrednosti  $L_{min}$  kot presečišče poenostavitev  $DC_m$ .

Mejno vrednost  $L_{min}$ , ki določa uporabo posamezne poenostavitve, lahko torej enostavno določimo s pomočjo izraza (6.71). Glede na to, da je v večini omrežij vrednost  $L_{min}$  dokaj nizka (najmanjše dovoljene dolžine paketov so nekaj deset oktetov), je zato večinoma uporabljena kar poenostavitev  $Op2_m$ .

Iz podanih rezultatov je razvidno, da sta  $O_m$  in  $\overline{O(M)}$  največja pri  $m = 1$  in pri majhnih vrednostih  $M$ . Za majhne vrednosti  $M$  velja, da je omrežje malo obremenjeno. Takrat stanje čakalnih vrst stalno alternira med aktivnim in pasivnim. To pomeni, da je v večini primerov v posameznem stanju aktivnosti postreženih le nekaj paketov. Pri naraščajočem  $M$  pa so čakalne vrste vse bolj polne, časi aktivnosti so daljši in to je značilno za bolj obremenjeno omrežje. Takrat se tako  $O_m$  kot  $\overline{O(M)}$  približujeta neki asimptotični vrednosti. To je za nas ugoden rezultat, saj pri nizko obremenjenem omrežju število operacij za odpravo paketa ni toliko pomembno kot pri visoko obremenjenem omrežju.

V tabeli 6.4 lahko primerjamo rezultate za  $\overline{O(M)}$  pri različnih vrednostih kvantov za enakomerno porazdelitev dolžin paketov. Vidimo, da  $\overline{O(M)}$  ne narašča linearno s padanjem vrednosti kvanta. Pri povprečni dolžini paketov 1000 oktetov in enaki velikosti kvanta je faktor povečanja števila operacij pri SCDRR le 1,25. To je zelo ugoden rezultat, saj je pri DRR zahtevana vrednost kvanta kar 2000 oktetov. Pri dvakrat manjši velikosti kvanta je pri odpravniku SCDRR za odpravo enega paketa v povprečju potrebnih le 1,25-krat več operacij kot pri odpravniku DRR.

## 6.6 Simulacija samostojnega delovanja SCDRR

Teoretična analiza delovanja odpravnika SCDRR je narejena za primer samostojnega delovanja odpravnika. Na slikah 6.1 in 6.2 smo tudi s simulacijami pokazali, da ima odpravnik SCDRR nižje povprečne zakasnitve kot odpravnik DRR, na sliki 5.33 pa, da zagotavlja pravičnost med pretoki. Tudi tu je boljši kot odpravnik DRR. Te lastnosti so bile dokazane tudi z analitičnimi rezultati, ki kažejo, da sta tako algoritemska zakasnitev kot pravičnost za pretok  $i$  odvisni tudi od  $F$  in  $Q_i$ . Manjša kot sta  $F$  in  $Q_i$ , boljši sta meri za pravičnost in algoritemsko zakasnitev.

Pokažimo še, da se simulacijski rezultati za zahtevnost odpravnika SCDRR ujemajo z analitičnimi. Izvedemo simulacije za eksponentno in enakomerno porazdelitev dolžin paketov. Simuliramo delovanje odpravnika SCDRR pri različnih vrednostih kvantov. Pri eksponentni porazdelitvi s povprečno dolžino paketa 1000 oktetov vzamemo vrednosti kvantov  $Q = 100, 500, 1000, 2000, 4000, 16000$  oktetov; pri enakomerni porazdelitvi z  $L_{min} = 0$  in  $L_{max} = 2000$  pa vrednosti kvantov  $Q = 100, 500, 1000, 1500, 2000$ . Vsaka simulacija vsebuje milijon paketov. Analitične rezultate podamo z vrednostjo  $\overline{O(M)}$ , ki pomeni povprečno število operacij za odpravo posameznega paketa, če je v času aktivnosti čakalne vrste število vseh postreženih paketov enako  $M$ . Primerjava simulacijskih in analitičnih rezultatov je podana v spodnjih tabelah.

Q	Simulacije	Analitično $\overline{O(M)}$			
		$M = 10$	$M = 50$	$M = 100$	$M = 200$
100	10,060	10,093	10,057	10,053	10,051
500	2,215	2,240	2,219	2,216	2,215
1000	1,369	1,385	1,371	1,370	1,369
2000	1,067	1,073	1,069	1,068	1,068
4000	1,005	1,005	1,005	1,005	1,005
16000	1	1	1	1	1

Tabela 6.3 – Povprečno število operacij za odpravo enega paketa pri SCDRR in eksponentni porazdelitvi dolžin paketov

Q	Simulacije	Analitično $\overline{O(M)}$			
		$M = 10$	$M = 50$	$M = 100$	$M = 200$
100	10,026	10,073	10,034	10,030	10,027
500	2,125	2,163	2,132	2,129	2,127
1000	1,250	1,275	1,255	1,252	1,251
1500	1,001	1,025	1,005	1,002	1,001
2000	1	1	1	1	1

Tabela 6.4 – Povprečno število operacij za odpravo enega paketa pri SCDRR in enakomerni porazdelitvi dolžin paketov

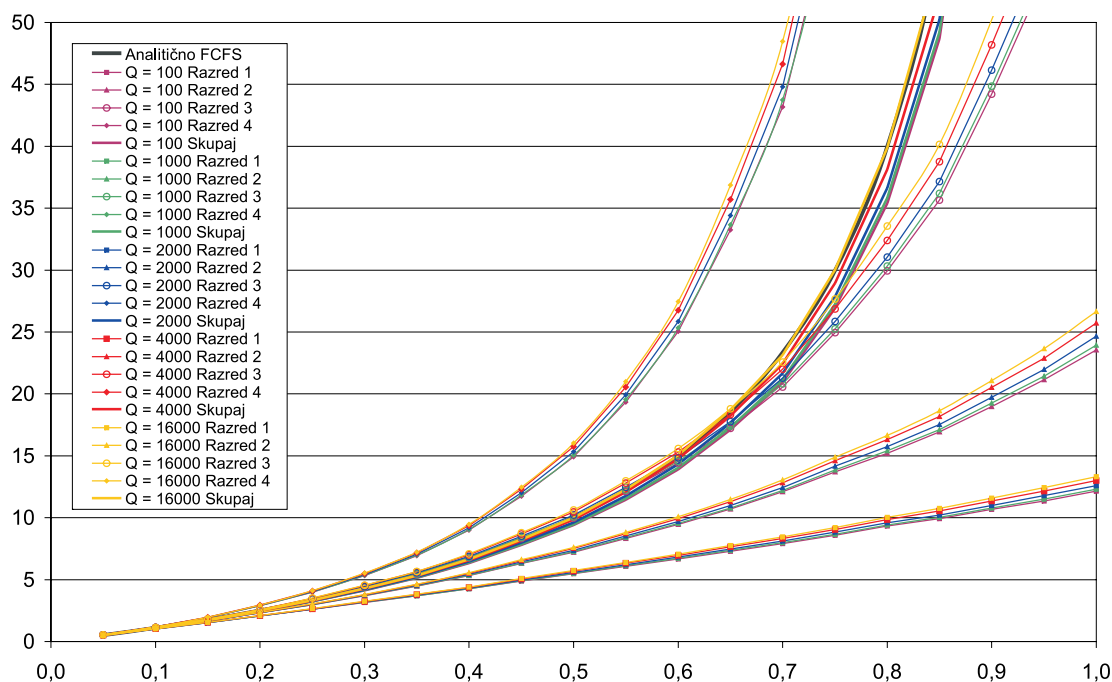
Simulacije izvedemo pri veliki obremenitvi omrežja  $\rho > 1$ . To nam zagotovi pogoje, v katerih so čakalne vrste odpravnika SCDRR večji del časa simulacije polne in zato je

povprečna vrednost  $\overline{M}$  preko več aktivnih obdobjih čakalne vrste velika. Zato moramo rezultate primerjati z  $\overline{O(M)}$  pri vrednosti  $M = 200$ . Vidimo, da so odstopanja minimalna in praktično zanemarljiva, kar potrjuje pravilnost analitičnih rezultatov.

## 6.7 Simulacija SCDRR na drugi stopnji omrežja DIS

Več pozornosti pa namenimo simulaciji delovanja odpravnika SCDRR v omrežju DIS. Na drugi stopnji predlaganega omrežja DIS namesto odpravnika DRR uporabimo odpravnik SCDRR. Nekatero rezultate smo predstavili že v prejšnjem poglavju (slike 5.39 do 5.51). Tukaj pa naredimo primerjavo rezultatov pri različnih vrednostih kvantov.

V simulacijah uporabimo dvostopenjski odpravnik DIS s štirimi prednostnimi razredi, v katerih so po štiri čakalne vrste - enakega kot v prejšnjem poglavju. Predstavljamo le rezultate za eksponentno porazdelitev dolžin paketov, saj se je v razdelku 5.6.1 izkazalo, da med rezultati simulacij za eksponentno in enakomerno porazdelitev dolžin paketov ni bistvenih razlik. Tudi pri dvostopenjskem delovanju z uporabo odpravnika SCDRR na drugi stopnji vsakemu podatkovnemu pretoku preko velikosti kvanta  $Q$  omogočimo rezervacijo določenega dela bitne hitrosti. Najdaljši paket, generiran tekom simulacij, je dolžine  $L_{max} = 16000$  oktetov. Vsi pretoki z dodeljenim kvantom, ki je manjši od  $L_{max}$ , delujejo v SCDRR režimu.

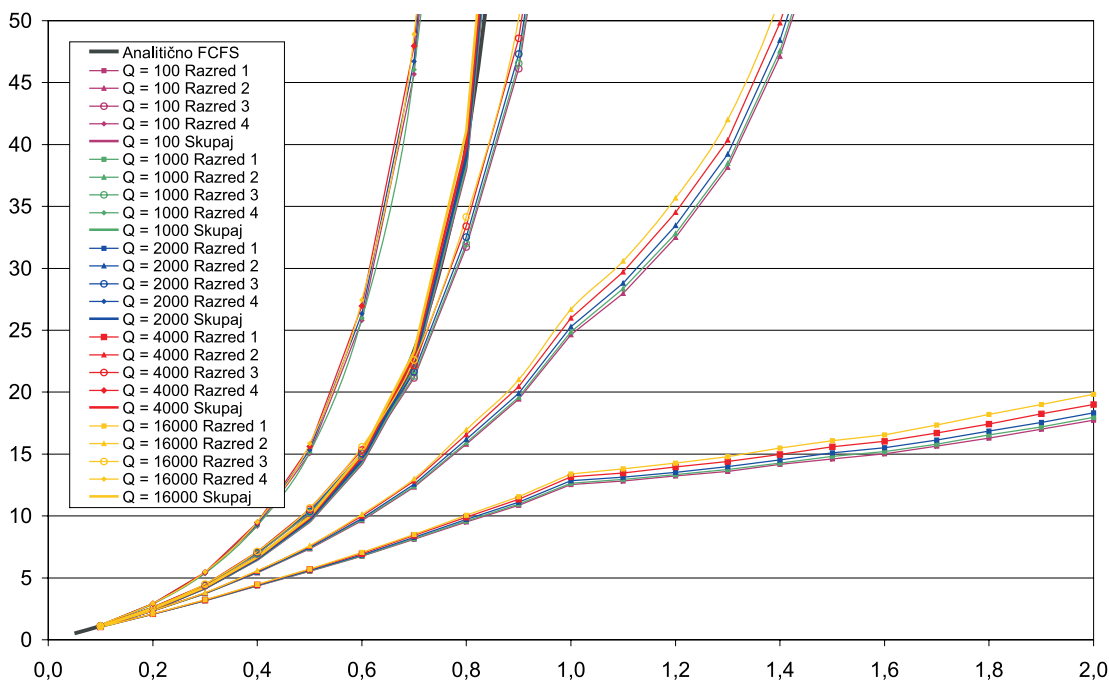


Slika 6.4 – Primerjava povprečnih zakasnitev paketov posameznih razredov za odpravnik SCDRR pri različnih vrednostih kvantov  $Q$

Simuliramo delovanje dvostopenjskega odpravnika pri šestnajstih aktivnih pretokih z enakimi bitnimi hitrostmi:  $r_{1A} = r_{1B} = \dots = r_{4D}$  in ustreznimi velikostmi kvantov:  $Q_{1A} = Q_{1B} = \dots = Q_{4D}$ . Povprečna dolžina paketov pretokov je 1000 oktetov. Vrednosti kvantov za posamezne simulacije so  $Q = 100, 1000, 2000, 4000, 16000$ . Pri



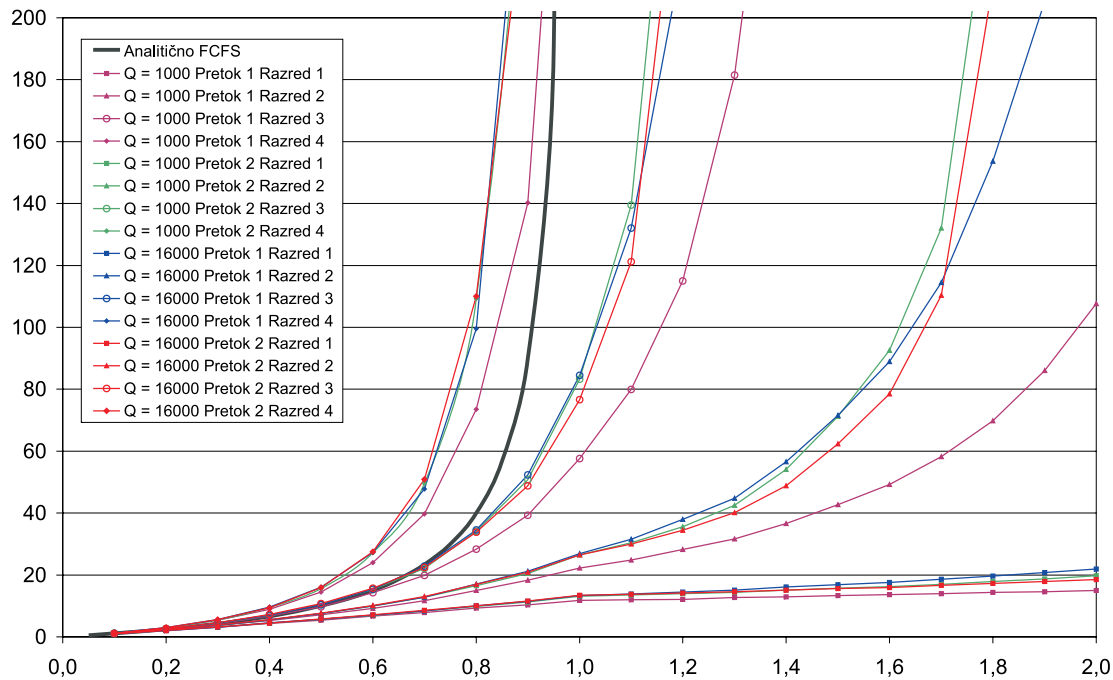
tem vrednost kvanta  $Q = 16000$  predstavlja delovanje v DRR režimu ( $Q \geq L_{max}$ ), preostale vrednosti kvantov pa delovanje v SCDRR režimu ( $Q < L_{max}$ ). Na sliki 6.4 so prikazane povprečne zakasnitve paketov posameznih razredov dvostopenjskega delovanja odpravnik SCDRR za različne vrednosti kvantov glede na izkoriščenost sistema  $0.05 \leq \rho \leq 1$ . Vidimo, da se z manjšanjem velikosti kvanta zmanjšajo tudi povprečne zakasnitve paketov v posameznih razredih. Če so vsi pretoki obvladani in delujejo v dogovorjenih mejah, potem enako velja tudi za posamezne pretoke. Če primerjamo delovanje odpravnik SCDRR pri vrednosti  $Q = 1000$  in odpravnik DRR pri vrednosti  $Q = 16000$ , opazimo, da je razlika v povprečni zakasnitvi precejšnja.



Slika 6.5 – Primerjava povprečnih zakasnitev paketov posameznih razredov pri različnih kvantih  $Q$  in neobvladanih pretokih  $x_B$

Poglejmo si še razmere v prisotnosti po enega neobvladanega pretoka v vsakem prednostnem razredu. Simuliramo delovanje dvostopenjskega odpravnik pri pretokih z bitnimi hitrostmi:  $r_{xA} = r_{xC} = r_{xD} = r$  in  $r_{xB} = 5r$  in enakimi velikostmi kvantov za vse pretoke:  $Q_{1A} = Q_{1B} = \dots = Q_{4D}$ . Pretoki  $x_B$  torej pošiljajo petkrat toliko podatkov kot je njihov pravični delež glede na velikost dodeljenega kvanta. Na sliki 6.5 so prikazane povprečne zakasnitve paketov posameznih pretokov v odvisnosti od  $\rho$ , ki se giblje v mejah  $0.05 \leq \rho \leq 2$ . Povprečna dolžina paketov v vseh pretokih je 1000 oktetov. Tudi v tem primeru se z manjšanjem velikosti kvanta zmanjšajo povprečne zakasnitve posameznih razredov.

Do sedaj smo opazovali povprečne zakasnitve vseh paketov znotraj enega prednostnega razreda. V primeru neobvladanih pretokov pa nas zanima tudi kako se le ti obnašajo v primerjavi z obvladanimi pretoki. Na sliki 6.6 so prikazane povprečne zakasnitve pretokov  $x_A$  in  $x_B$  (neobvladani) pri dveh različnih vrednostih kvantov:  $Q = 1000$  (SCDRR) in  $Q = 16000$  (DRR). V vsakem izmed razredov med seboj primerjamo zakasnitve pretokov  $x_A$  in  $x_B$  pri isti vrednosti kvanta. Pri  $Q = 1000$

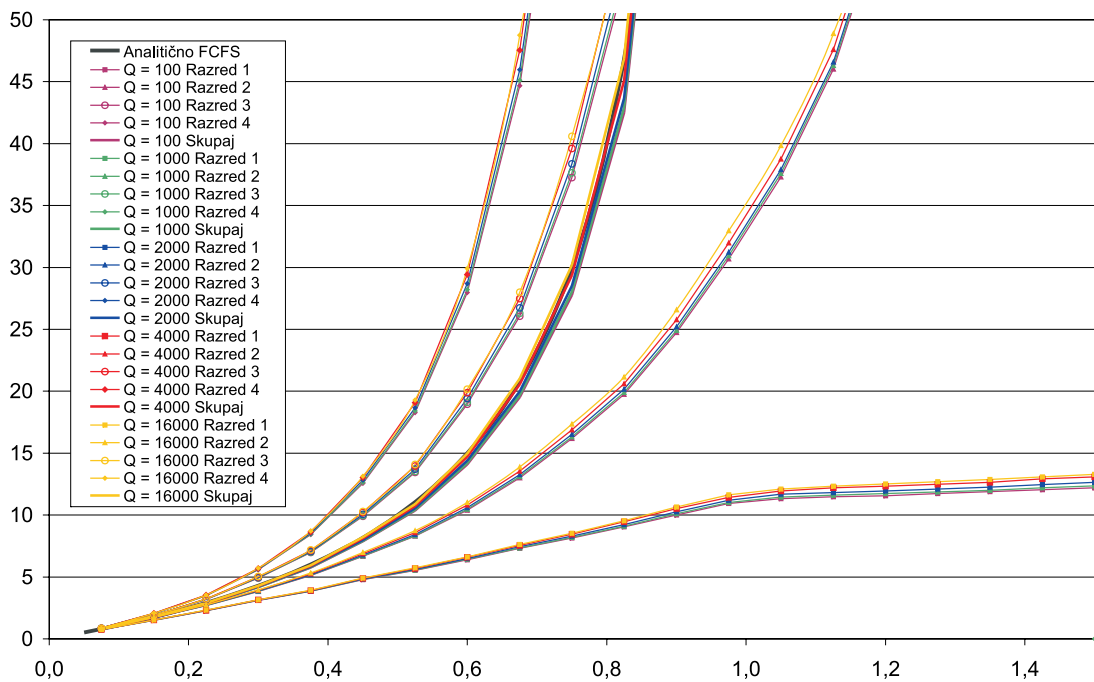


Slika 6.6 – Primerjava povprečnih zakasnitev paketov pretokov  $x_A$  in  $x_B$  pri vrednostih kvantov  $Q = 1000, 16000$ ; pretoki  $x_B$  so neobvladani

(SCDRR) primerjamo vijolične in zelene krivulje z istim markerjem, pri  $Q = 16000$  (DRR) pa modre in rdeče krivulje z istim markerjem.

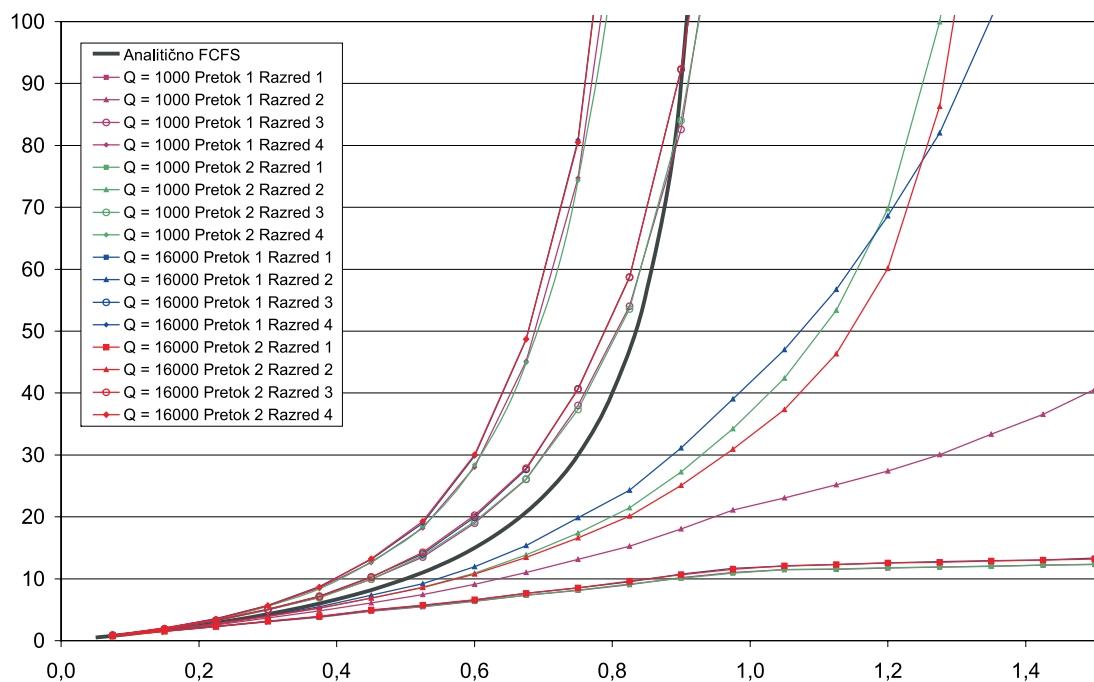
Pri  $Q = 16000$  in pri nižjih vrednostih  $\rho$  je zakasnitev neobvladanih pretokov celo manjša od zakasnitve obvladanih pretokov. Neobvladani pretoki zaradi velike vrednosti kvanta v vsakem obhodu odpravijo precejšnje število svojih paketov. Pri nizkih vrednostih  $\rho$  obvladani pretoki v vsakem obhodu v povprečju odpravijo petkrat manjše število paketov kot neobvladani. Ker v vsakem obhodu pretoki med odpravljanjem paketov drugih pretokov čakajo na svojo priložnost, je njihova povprečna zakasnitev odvisna od količine prometa, ki se v vsakem obhodu odpravi iz drugih čakalnih vrst (drugih pretokov). Neobvladani pretoki na to priložnost v povprečju čakajo manj časa kot obvladani in zato je njihova povprečna zakasnitev manjša. Pri višjih vrednostih  $\rho$  pa neobvladani pretoki nimajo na voljo neizkoriščenih rezervacij obvladanih pretokov. Njihove zakasnitve zato hitro narastejo in so v povprečju višje od zakasnitev obvladanih pretokov. To se lepo vidi iz modre (obvladani pretok) in rdeče (neobvladani pretok) krivulje s trikotnim markerjem (prednostni razred 2) na sliki 6.6. Pri  $Q = 1000$  pa že sama velikost kvanta preprečuje velika nesorazmerja v številu odpravljenih paketov v vsakem obhodu. Zato je zakasnitev neobvladanih pretokov tudi pri nizkih vrednostih  $\rho$  večja od zakasnitve obvladanih pretokov.

Simuliramo še delovanje odpravnika SCDRR pri enem neobvladanem pretoku v samo enem izmed prednostnih razredov. Imamo pretoke z bitnimi hitrostmi:  $r_{xA} = r_{xB} = r_{xC} = r_{xD} = r$  ter  $r_{yA} = r_{yC} = r_{yD} = r$  in  $r_{yB} = 9r$ , pri čemer  $x$  predstavlja razrede z obvladanimi pretoki in  $y$  razred z neobvladanim pretokom. Velikosti kvantov so za vse pretoke enake. Na sliki 6.7 so prikazane povprečne zakasnitve paketov posameznih pretokov za različne velikosti kvantov. Povprečna dolžina paketov pre-



Slika 6.7 – Primerjava povprečnih zakasnitev paketov posameznih razredov pri različnih kvantih  $Q$  in neobvladanem pretoku 2 v razredu 2

tokov je 1000 oktetov. Tudi v tem primeru se z manjšanjem velikosti kvanta zmanjšajo povprečne zakasnitve posameznih razredov.



Slika 6.8 – Primerjava povprečnih zakasnitev paketov pretokov  $x_A$  in  $x_B$  pri vrednostih kvantov  $Q = 1000, 16000$ ; pretok  $2B$  je neobvladan

Zanima nas tudi, kako se neobvladani pretoki obnašajo v primerjavi z obvladanimi pretoki. Na sliki 6.8 so prikazane povprečne zakasnitve pretokov  $x_A$  in  $x_B$  (neobvladani) pri dveh različnih vrednostih kvantov:  $Q = 1000$  (SCDRR) in  $Q = 16000$  (DRR). V vsakem izmed razredov med seboj primerjamo zakasnitve pretokov  $x_A$  in  $x_B$  pri isti vrednosti kvanta. Pri  $Q = 1000$  (SCDRR) primerjamo vijolične in zelene krivulje z istim markerjem, pri  $Q = 16000$  (DRR) pa modre in rdeče krivulje z istim markerjem.

Ugotovimo, da so rezultati zelo podobni kot na sliki 6.6. Pri  $Q = 1000$  ima neobvladani pretok slabše razmere kot obvladani, pri vrednosti kvanta  $Q = 16000$  pa velja enako kot pri prejšnjem primeru. Pri nizkih vrednostih  $\rho$  ima neobvladani pretok zaradi uporabe neizkoriščenih rezervacij drugih pretokov manjše povprečne zakasnitve, pri višjih vrednostih  $\rho$  pa slabše. To se še posebno lepo vidi iz modre (obvladani pretok) in rdeče (neobvladani pretok) krivulje s trikotnim markerjem (prednostni razred 2) na sliki 6.8.

## 6.8 Komentar simulacijskih rezultatov za SCDRR

Iz rezultatov simulacij je razvidno, da SCDRR prevzame večino prednosti odpravnika DRR. S podkritičnim delovanjem pri  $Q \leq L_{max} - 1$  pa izboljša algoritemsko zakasnitev odpravnika DRR in njegovo pravičnost. Prva trditev se pokaže v zmanjšanju povprečne zakasnitve paketov v odpravniku, druga pa v manjšem vplivu neobvladanih pretokov na obvladane pretoke znotraj istega prednostnega (storitvenega) razreda.

Nekaj slabše pa se SCDRR obnese pri meri za zahtevnost odpravnika. Z znižanjem vrednosti kvanta pod kritično vrednost  $Q = L_{max}$  zahtevnost ni več  $O(1)$ , ampak je v najslabšem primeru lahko tudi  $O(N)$ , pri čemer  $N$  predstavlja število aktivnih čakalnih vrst. To je precej neugoden rezultat, ki pa predstavlja le najslabši možni primer, ki ni zelo verjeten. S teoretično analizo smo dokazali in s simulacijami potrdili, da je faktor povprečnega povečanja števila operacij za odpravo enega paketa precej nižji od  $N$ . V primerih iz tabel 6.3 in 6.4 je pri povprečni dolžini paketov 1000 oktetov in vrednosti kvanta 1000 oktetov faktor povečanja 1,37 za eksponentno in 1,25 za enakomerno porazdelitev. Pri enakih podatkih se povprečna zakasnitev zmanjša za faktor 1,11.

Izboljšanja pravičnosti ne moremo podati v številkah. Je pa pomembno že dejstvo, da se pri SCDRR zmanjša škodljiv vpliv neobvladanih pretokov. Lahko pa ta vpliv predstavimo posredno preko zakasnitev. Pri DRR se izkaže, da imajo pri majhnih obremenitvah omrežja neobvladani pretoki celo manjše zakasnitve kot obvladani, pri SCDRR pa se to ne zgodi, obvladani pretoki so vedno na boljšem. Razmerja zakasnitev med neobvladanimi pretoki in obvladanimi pretoki se pri SCDRR gibljejo od 1,07 pri  $\rho = 0,3$  do 1,51 pri  $\rho = 0,9$  in 8,59 pri  $\rho = 1,43$  (glej sliko 6.8). Večjo težo imajo seveda faktorji pri večjih vrednostih  $\rho$ , saj so takrat razmere v omrežju bolj zaostrene kot pri manjših vrednostih  $\rho$ .

Menimo, da sta za zagotavljanje različnih stopenj kakovosti storitve algoritemska zakasnitev in pravičnost veliko pomembnejši lastnosti odpravnika kot njegova zahtevnost. Zato je uvedba odpravnika SCDRR smiselna in koristna.

## 7 Zaključek

V disertaciji smo predstavili možnosti zagotavljanja različnih stopenj kakovosti storitve v paketnih omrežjih s poudarkom na omrežjih IP. Po podanih definicijah kakovosti storitve, lastnostih paketnih omrežij in mehanizmov, ki nam omogočajo zagotavljanje različnih stopenj kakovosti storitve, smo podali predlog združevanja konceptov Diferenciranih in Integriranih storitev v omrežje DIS. S tem predlogom ohranjamo dobre lastnosti obeh združenih konceptov in hkrati izničimo njune slabosti.

Omrežje DIS deluje dvostopenjsko. Na prvi stopnji upošteva samo storitveni razred, ki mu pripadajo obravnavani paketi. Storitvene razrede obravnava prednostno; razredi z višjo prednostjo so deležni boljše, razredi z nižjo prednostjo pa slabše storitve omrežja. Na ta način zagotovimo različne stopnje kakovosti storitve. Na drugi stopnji omrežje DIS razločuje tudi med posameznimi pretoki znotraj storitvenih razredov. Zaželeno je namreč, da znotraj posameznega storitvenega razreda vsi pretoki prejemajo svoj pravični delež, ki je odvisen od njihovih rezervacij. Kot glavni parameter kakovosti storitve smo vzeli zakasnitev.

Rezultati teoretične analize in simulacij kažejo, da je predlagano strogo prednostno odpravljjanje primerno za prvo stopnjo omrežja DIS. Razlike v zakasnitvah med paketi v posameznih prednostnih razredih zagotovijo različne stopnje kakovosti storitve. Ob povečanem prometu, ki presega kapaciteto omrežja, razredi z nižjo prednostjo pri strogo prednostnem odpravljanju postanejo nestabilni. To pomeni, da njihove zakasnitve narastejo preko vseh meja. Izogniti se moramo primerom, ko je za to kriv povečan promet v razredih z višjo prednostjo. Zato lahko prometni pretoki višjih prednostnih razredov rezervirajo vire samo do nekega vnaprej določenega deleža. Rezervacija virov poteka preko rezervacijskega protokola, na primer RSVP. Simulacije z upoštevanjem višjih protokolnih plasti in prometnih karakteristik realnih aplikacij kažejo tudi, da lahko z njihovo pravilno razporeditvijo v prednostne razrede večina izmed njih precej pridobi, medtem ko preostale bistveno ne izgubijo. To smo pokazali na primeru velikega izboljšanja razmer za prenos govora in le malenkostnega poslabšanja razmer pri FTP prenosu datotek. Zakasnitve in njihova spremenljivost se za govor bistveno izboljšata. Pri FTP se zakasnitve sicer zelo povečajo, kar pa ne vpliva bistveno na delovanje aplikacije. Tu je bolj pomembna propustnost, ki se skorajda ne poslabša.

Tudi simulacije dvostopenjskega delovanja omrežja z odpravnikom DRR na drugi stopnji pokažejo, da odpravnik zagotovi pravičnost med pretoki znotraj posameznih prednostnih razredov. Tudi v prisotnosti neobvladanih pretokov se razmere za obvladane pretoke ne poslabšajo, kar smo tudi hoteli doseči. Rezultati simulacij kažejo tudi, da povprečne zakasnitve paketov padejo, če vrednosti kvantov zmanjšamo pod kritično vrednost pri  $Q = L_{max}$ . Zato predlagamo nov odpravnik SCDRR, ki izkoristi to dejstvo.

S teoretično analizo pridemo do analitičnih rezultatov za pravičnost, algoritemsko zakasnitev in zahtevnost odpravnik SCDRR. S simulacijami pokažemo tudi, da predlagani odpravnik SCDRR zagotovi pravičnost med pretoki znotraj posameznega prednostnega razreda tudi ob prisotnosti neobvladanih pretokov. Iz primerjave simulacijskih rezultatov pa vidimo tudi, da je SCDRR pravičnejši in ima manjšo algoritemsko zakasnitev kot odpravnik DRR, a večjo zahtevnost. Ker je po našem mnenju zmanjšanje zakasnitve in povečanje pravičnosti pomembnejše od povečanja zahtevnosti, smatramo da je uporaba odpravnik SCDRR upravičena in smotrna.

Čeprav zaključena celota, predstavlja disertacija osnovo za nadaljnje delo in raziskave. Na teoretičnem področju bi bila dobrodošla stohastična analiza zakasnitev odpravnik SCDRR. Poleg samih analitičnih rezultatov zakasnitve bi le te lahko primerjali z rezultati za algoritemsko zakasnitev. To bi nam omogočilo podati oceno natančnosti mere algoritemske zakasnitve. Teoretično analizo in simulacije bi kazalo razširiti z drugimi analitično obvladljivimi verjetnostnimi porazdelitvami dolžin paketov, simulacije pa tudi z realnim prometom. Razviti simulator pa bi lahko nadgradili s funkcionalnostjo višjih protokolnih skladov in zmožnostjo simulacij več med seboj povezanih omrežnih naprav.

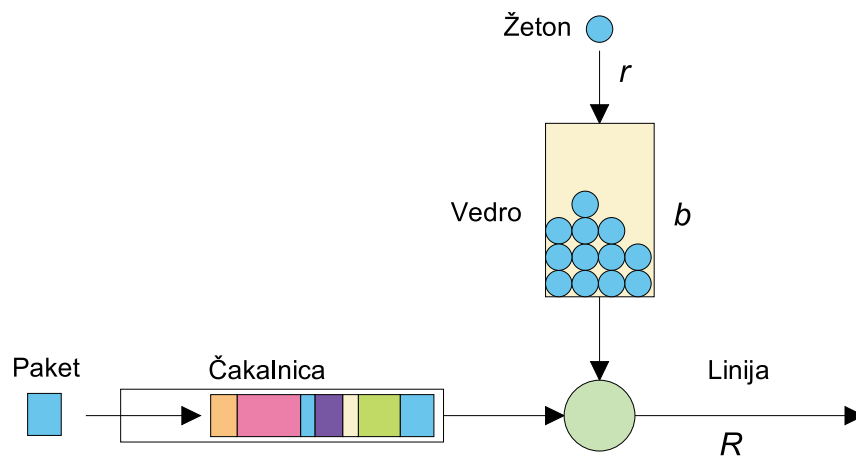
# Dodatek





## A Vedro z žetoni

Vedro z žetoni nam v omrežju služi za krmiljenje prometa (traffic policing) in njegovo oblikovanje (traffic shaper). Vedro z žetoni je definirano s spremenljivkama  $r$  in  $b$ . Spremenljivka  $r$  določa hitrost, s katero žetoni polnijo vedro, spremenljivka  $b$  pa globino vedra. Slika A.1 predstavlja model vedra z žetoni. Vedro se polni z žetoni s stalno hitrostjo  $r$  bitov/s. Če je vedro polno, se odvečni žetoni zavržejo. Prihajajoči paketi prispejo v čakalnico, od koder se posredujejo naprej le v primeru, da je v vedru dovolj žetonov. Paket dolžine  $L$  bitov se posreduje le, če je v vedru vsaj za  $L$  bitov žetonov. V nasprotnem primeru mora paket počakati v čakalnici, da se v vedru nabere dovolj žetonov. Ob posredovanju paketa dolžine  $L$  bitov se iz vedra odvzame za  $L$  bitov žetonov.



Slika A.1 – Model vedra z žetoni

Pri delovanju vedra z žetoni so pomembni tudi naslednji parametri:

- **Povprečna oddajna hitrost**  $\rho$  je dolgoročna oddajna hitrost, ki ne more biti večja od hitrosti, s katero žetoni polnijo vedro:  $\rho \leq r$ .
- **Vršna oddajna hitrost** je najvišja oddajna hitrost, ki jo smejo doseči paketi. Če ni določeno drugače, je to kar bitna hitrost linije  $R$ , lahko pa je tudi nižja od nje.
- **Velikost izbruha** (burst size) določa največjo količino podatkov, ki so lahko zaporedoma oddani. Maksimalna velikost izbruha je odvisna od  $r$ ,  $b$  in  $R$ . Predpostavimo, da je ob času  $t_0$  vedro polno. V njem se nahaja za  $b$  žetonov. Naj se v časovnem intervalu  $[t_0, t_0 + t]$  vedro popolnoma izprazni. V tem času je vanj prispelo še za  $rt$  žetonov. Če smo pakete oddajali s hitrostjo linije  $R$ , potem smo

v časovnem intervalu  $[t_0, t_0 + t]$  odposlali za  $Rt$  podatkov. Iz teh relacij lahko zapišemo enačbo  $b + rt = Rt$  in iz nje izrazimo čas  $t$ :

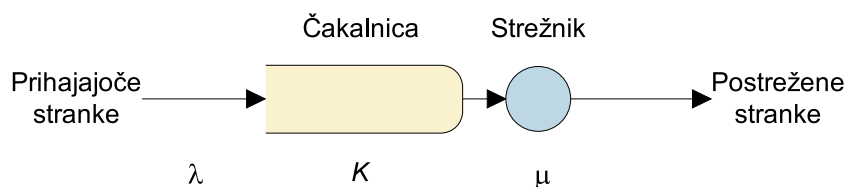
$$t = \frac{b}{R - r}$$

Do izpraznjenja vedra imamo torej na voljo za  $b + rt$  žetonov. Če v slednji izraz vstavimo zgoraj izračunani čas  $t$ , dobimo največjo velikost izbruha  $\sigma$ :

$$\sigma = b \left( 1 + \frac{r}{R - r} \right)$$

## B Osnove teorije čakalnih vrst

Teorija čakalnih vrst je obsežno in dokaj razvito področje. Na tem mestu podajamo le njene osnove, ki jih potrebujemo za teoretično analizo, na kateri temeljijo simulacije [48, 49].



Slika B.1 – Model čakalne vrste

Model in elementi čakalne vrste so prikazani na sliki B.1. Sestavljena je iz čakalnice in strežnika. Če obravnavamo paketna omrežja, čakalnica predstavlja pomnilnik, v katerem se nahajajo paketi, ki še niso prišli na vrsto za oddajo. Strežnik pa streže pakete, ko pridejo na vrsto za oddajo (o tem, kateri paket je na vrsti, odloča strežna disciplina). Parameter  $\lambda$  določa povprečno število prihodov strank v sistem,  $K$  je velikost čakalnice,  $\mu$  pa predstavlja povprečno število postreženih strank. Za paketna omrežja imamo naslednje enote:  $\lambda$  [paketi/s],  $K$  [paketi],  $\mu$  [paketi/s].

Preden nadaljujemo z obravnavo posameznih primerov čakalnih vrst, si pogledjmo, kako jih označujemo. Po Kendallu jih enolično določimo s kombinacijo črk A/B/c/K/p z naslednjimi pomeni:

- A** proces, ki opisuje čase med prihodi strank<sup>1</sup> v sistem,
- B** proces, ki opisuje čase strežbe strank v čakalnici,
- c** število strežnikov,
- K** velikost čakalnice,
- p** velikost populacije možnih strank.

Zadnja dva parametra oznake lahko izpustimo. S tem privzamemo njuno vrednost kot  $\infty$ . Razčistimo še pomen parametrov A in B, ki opisujeta proces prihodov ali strežbe strank. Opišemo ju lahko kot naključna procesa, definirana z naključnima spremenljivkama z verjetnostnima porazdelitvama  $A(t)$  in  $B(x)$

$$A(t) = P(\text{čas med prihodi} \leq t) \quad (\text{B.1})$$

$$B(x) = P(\text{čas strežbe} \leq x) \quad (\text{B.2})$$

---

<sup>1</sup>v primeru paketnih omrežij so stranke kar posamezni paketi

Parameter A torej določa porazdelitev časov med dvema zaporednima prihodoma strank v sistem (interarrival time), parameter B pa porazdelitev časov, potrebnih za strežbo posamezne stranke. Parametra navadno izbiramo iz naslednjega nabora oznak:

- D** čas med prihodi ali čas potreben za strežbo je determinističen ali konstanten,
- M** čas med prihodi ali čas potreben za strežbo ustreza Poissonovemu procesu (diskreten sistem) ali eksponentni porazdelitvi (zvezen sistem),
- G** splošna (katerakoli) porazdelitev,
- E** Erlangova porazdelitev.

Primer: čakalna vrsta z oznako M/D/1 ima eksponentno porazdelitev časov med prihodi in konstanten čas strežbe posamezne stranke. V sistemu je en strežnik, velikost čakalnice in populacije pa je neskončna.

## B.1 Poissonov proces

Najosnovnejši naključni proces, ki se uporablja v teoriji čakalnih vrst, je Poissonov proces. Čeprav ta največkrat ne odraža pravega stanja v omrežju, ga kljub temu uporabljamo zaradi njegove matematične obvladljivosti. Njegova dobra lastnost je, da je brez spomina, kar pomeni, da so zaporedni prihodi med seboj neodvisni in eksponentno porazdeljeni.

Poglejmo si nekatere rezultate izvajanj [49], ki nas pripeljejo do zelenih rezultatov. Verjetnost  $k$  prihodov v časovnem intervalu  $t$  je podana z izrazom

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad (\text{B.3})$$

pri čemer  $\lambda$  predstavlja povprečno število prihodov na sekundo. Povprečno število prihodov v času  $t$  je tako enako

$$E(K) = \sum_{k=0}^{\infty} k P_k(t) = \lambda t \quad (\text{B.4})$$

in pokažemo lahko, da je varianca  $\sigma_K^2 = \lambda t$ , kar je enako srednji vrednosti  $E(K)$ . Poleg števila prihodov v določenem času pa nas zanima tudi verjetnost, da se bo naslednji prihod zgodil v intervalu  $(0, t)$ , ki jo lahko zapišemo kot  $P(\tau \leq t)$  in je podana z izrazom

$$A(t) = P(\tau \leq t) = 1 - e^{-\lambda t} \quad (\text{B.5})$$

pri čemer  $\tau$  predstavlja čas, ki preteče med zadnjim in trenutnim prihodom.  $A(t)$  imenujemo tudi kumulativna porazdelitev verjetnosti. Z odvodom izraza  $A(t)$  dobimo gostoto verjetnostne porazdelitve časov med dvema zaporednima prihodoma  $a(t)$

$$a(t) = \lambda e^{-\lambda t} \quad (\text{B.6})$$

ki je dobro poznana eksponentna porazdelitev. Od tod lahko izračunamo še povprečen čas med prihodi  $E(\tau)$  in njihovo varianco  $E(\tau^2)$

$$E(\tau) = \bar{\tau} = \int_0^{\infty} ta(t)dt = \int_0^{\infty} t\lambda e^{-\lambda t}dt = \frac{1}{\lambda} \quad (\text{B.7})$$

$$E((\tau - \bar{\tau})^2) = \sigma_t^2 = \frac{1}{\lambda^2} \quad (\text{B.8})$$

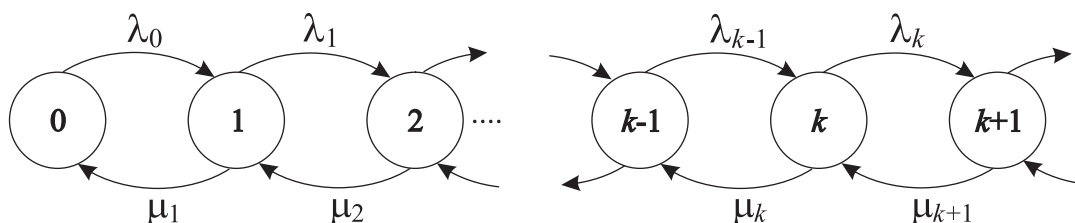
V paketnih omrežjih je čas strežbe paketa premosorazmeren njegovi dolžini. Če predpostavimo, da je porazdelitev dolžin paketov eksponentna, potem časi strežbe posameznega paketa  $x$  ustrezajo Poissonovemu procesu. Ob povprečnem številu posreženih paketov v časovni enoti  $\mu$  lahko zapišemo povprečen čas in varianco njihove strežbe z izrazoma

$$E(x) = \frac{1}{\mu} \quad \text{in} \quad E((x - E(x))^2) = \frac{1}{\mu^2}$$

EkspONENTNA porazdelitev ima še lepo lastnost, da je brez spomina ter zanimivo zvezo med srednjo vrednostjo  $E(\tau)$  in standardno deviacijo  $\sigma$ , ki sta enaki.

## B.2 Markovove verige ter proces rojstev in smrti

Za uspešno analizo čakalnih vrst si moramo na kratko ogledati orodja, s katerimi se je lotimo [49]. Čakalna vrsta se lahko nahaja v enem izmed stanj, ki so določena s trenutnim številom strank v čakalnici. Če predpostavimo, da lahko v čakalnico naenkrat vstopi samo ena stranka in da lahko sistem zapusti tudi samo ena stranka naenkrat, potem lahko prehode sistema iz enega stanja v drugega opišemo s *procesom rojstev in smrti*. To je poseben primer Markovovega procesa, kjer so prehodi dovoljeni samo med sosednjimi stanji. Na primer, iz stanja  $E_k$  lahko pridemo samo v stanja  $E_{k-1}$ ,  $E_k$  in  $E_{k+1}$ . Pri tem je s stanjem  $E_k$  določena populacija (število strank  $k$  v sistemu), prehod iz stanja  $E_k$  v stanje  $E_{k-1}$  pomeni *smrt* (stranka zapusti sistem), prehod iz stanja  $E_k$  v stanje  $E_{k+1}$  pa *rojstvo* (prihod stranke). Definiramo še hitrost rojstev  $\lambda_k$  in smrti  $\mu_k$ , ki povesta hitrost rojstev ali smrti, ko se sistem nahaja v stanju  $E_k$ . Povedano lahko ponazorimo s spodnjim diagramom prehodov med stanji sistema (slika B.2).

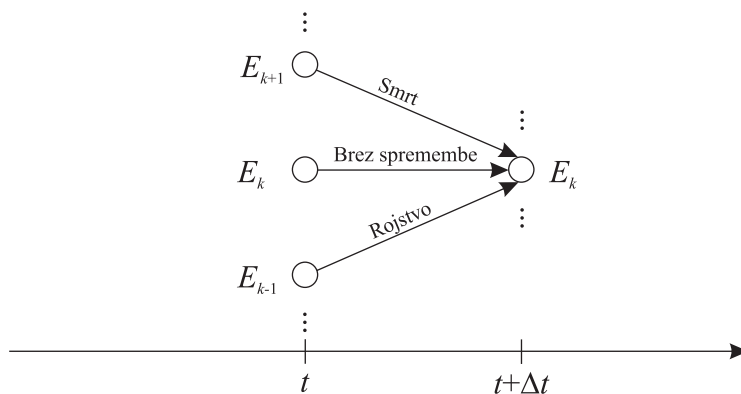


Slika B.2 – Diagram prehodov med stanji sistema, ki ustrezajo procesu rojstev in smrti

Za proces rojstev in smrti predpostavimo, da so rojstva in smrti med seboj neodvisna in da je verjetnost več hkratnih rojstev, smrti ali hkratnih rojstev in smrti zanemarljiva in reda  $o(\Delta t)$ . Tako lahko zapišemo naslednje verjetnosti

$$\begin{aligned}
P_{R_1}(\text{točno 1 rojstvo na } (t, t + \Delta t) \text{ pri populaciji } k) &= \lambda_k \Delta t + o(\Delta t) \\
P_{S_1}(\text{točno 1 smrt na } (t, t + \Delta t) \text{ pri populaciji } k) &= \mu_k \Delta t + o(\Delta t) \\
P_{R_2}(\text{točno 0 rojstev na } (t, t + \Delta t) \text{ pri populaciji } k) &= 1 - \lambda_k \Delta t + o(\Delta t) \\
P_{S_2}(\text{točno 0 smrti na } (t, t + \Delta t) \text{ pri populaciji } k) &= 1 - \mu_k \Delta t + o(\Delta t)
\end{aligned}$$

Pri analizi sistema se osredotočimo na dva zaporedna trenutka  $t$  in  $t + \Delta t$ . Če se sistem v trenutku  $t + \Delta t$  nahaja v stanju  $E_k$ , potem se je ob trenutku  $t$  lahko nahajal v enem izmed stanj  $E_{k-1}$ ,  $E_k$  ali  $E_{k+1}$ . Zgodil se je lahko eden izmed naslednjih medsebojno izključujočih se dogodkov. Možne prehode v stanje  $E_k$  lahko ponazorimo s sliko B.3.



Slika B.3 – Možne poti do stanja  $E_k$

1. Sistem se je ob času  $t$  nahajal v stanju  $E_k$ , stanje sistema v intervalu  $(t, t + \Delta t)$  se ni spremenilo.
2. Sistem se je ob času  $t$  nahajal v stanju  $E_{k-1}$ , v intervalu  $(t, t + \Delta t)$  se je zgodilo rojstvo.
3. Sistem se je ob času  $t$  nahajal v stanju  $E_{k+1}$ , v intervalu  $(t, t + \Delta t)$  je nastopila smrt.

Verjetnost  $P_k(t + \Delta t)$ , da se sistem ob času  $t + \Delta t$  nahaja v stanju  $E_k$ , lahko zapišemo kot vsoto verjetnosti zgornjih dogodkov, ki velja ob pogoju  $k \geq 1$

$$\begin{aligned}
P_k(t + \Delta t) &= P_k(t)(1 - \lambda_k \Delta t + o(\Delta t))(1 - \mu_k \Delta t + o(\Delta t)) + \\
&+ P_{k-1}(t)(\lambda_{k-1} \Delta t + o(\Delta t)) + \\
&+ P_{k+1}(t)(\mu_{k+1} \Delta t + o(\Delta t)) + \\
&+ o(\Delta t)
\end{aligned} \tag{B.9}$$

Za stanje  $E_0$  pa moramo zapisati svojo enačbo, saj tu smrt ni možna

$$\begin{aligned}
P_0(t + \Delta t) &= P_0(t)(1 - \lambda_0 \Delta t + o(\Delta t)) + \\
&+ P_1(t)(\mu_1 \Delta t + o(\Delta t)) + \\
&+ o(\Delta t)
\end{aligned} \tag{B.10}$$

Če sedaj enačbi preuredimo in delimo z  $\Delta t$ , dobimo naslednji par enačb

$$\begin{aligned} \frac{P_k(t + \Delta t) - P_k(t)}{\Delta t} &= -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \\ &+ \mu_{k+1}P_{k+1}(t) + \frac{o(\Delta t)}{\Delta t} \end{aligned} \quad (\text{B.11})$$

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda_0P_0(t) + \mu_1P_1(t) + \frac{o(\Delta t)}{\Delta t} \quad (\text{B.12})$$

ko gre  $\Delta t \rightarrow 0$  leva stran enačbe preide v odvod po času, izraz  $o(\Delta t)/\Delta t$  pa gre proti nič. Dobimo naslednji diferencialni enačbi

$$\begin{aligned} \frac{dP_k(t)}{dt} &= -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \\ &+ \mu_{k+1}P_{k+1}(t) \quad k \geq 1 \\ \frac{dP_0(t)}{dt} &= -\lambda_0P_0(t) + \mu_1P_1(t) \quad k = 0 \end{aligned} \quad (\text{B.13})$$

ki ju lahko rešimo s pomočjo  $z$ -transforma in Laplaceovega transforma [49]. Na koncu pridemo do dokaj zapletene in nič kaj ohrabrujoče rešitve

$$\begin{aligned} P_k(t) &= e^{-(\lambda+\mu)t} \left[ \rho^{(k-i)/2} I_{k-i}(at) + \rho^{(k-i-1)/2} I_{k+i+1}(at) + \right. \\ &+ \left. (1-\rho)\rho^k \sum_{j=k+i+3}^{\infty} \rho^{-j/2} I_j(at) \right] \end{aligned} \quad (\text{B.14})$$

pri čemer so  $\rho = \lambda/\mu$ ,  $a = 2\mu\rho^{1/2}$  in  $I_k(x)$  modificirane Besselove funkcije reda  $k$  prve vrste.

Rešitev časovne odvisnosti verjetnosti posameznih stanj čakalne vrste je dokaj zapletena. Zato se v nadaljevanju posvečamo obravnavi čakalnih vrst v ravnovesju. Če predpostavimo, da ob naraščajočem času  $t$  verjetnost  $P_k(t)$  teži k neki vrednosti  $p_k$ , to definiramo z izrazom

$$p_k \triangleq \lim_{t \rightarrow \infty} P_k(t) \quad (\text{B.15})$$

Potrebno je poudariti, da limitne verjetnosti  $p_k$  niso več funkcije časa kot to velja za  $P_k(t)$ . To seveda ne pomeni, da se stanje sistema ne bo spreminjalo s časom, le verjetnosti  $p_k$ , da ga bomo našli v stanju  $E_k$ , so konstantne in neodvisne od časa. Če torej sprejmemo obstoj limitnih ali ravnotežnih verjetnosti  $p_k$ , se verjetnost  $P_k(t)$  ne spreminja več, torej je njen odvod enak nič,  $dP_k(t)/dt = 0$  in enačbi (B.13) ob upoštevanju preideta v

$$\begin{aligned} 0 &= -(\lambda_k + \mu_k)p_k + \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1} \quad k \geq 1 \\ 0 &= -\lambda_0p_0 + \mu_1p_1 \quad k = 0 \end{aligned} \quad (\text{B.16})$$

Ob postavitvi vrednosti  $\lambda_k$ ,  $\mu_k$  in  $p_k$  na nič za  $k < 0$ , lahko enačbi (B.16) zapišemo z eno samo

$$0 = -(\lambda_k + \mu_k)p_k + \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1} \quad (\text{B.17})$$

Ob upoštevanju zakonov verjetnosti moramo zapisati še spodnji pogoj, saj je verjetnost, da se bo sistem nahajal v enem izmed stanj  $E_k$ , natančno ena.

$$\sum_{k=0}^{\infty} p_k = 1 \quad (\text{B.18})$$

Tako smo prišli do ravnotežne diferenčne enačbe za naš sistem. Če si še enkrat podrobno ogledamo sliko B.2, vidimo, da če okoli stanja  $E_k$  zarišemo obroč, enačba opisuje ravno vsoto verjetnosti vseh njegovih prihodov in odhodov, ki se v ravnotežnem stanju sistema izničijo in so enaki nič. Namesto obročev pa lahko med stanji povlečemo navpične črte in zapišemo naslednjo enačbo za mejo med stanji  $E_{k-1}$  in  $E_k$

$$\lambda_{k-1}p_{k-1} = \mu_k p_k \quad (\text{B.19})$$

Zgornji sistem enačb lahko rešimo po iterativni metodi, kjer najprej pri pogoju  $k = 0$  izrazimo  $p_1 = \frac{\lambda_0}{\mu_1}p_0$ , ta rezultat uporabimo za izračun  $p_2$  pri pogoju  $k = 1$ , pri čemer dobimo  $p_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}p_0$ . Ob nadaljevanju tega postopka kmalu ugotovimo splošno formulo za izračun  $p_k$  ali *ravnotežno rešitev verjetnosti stanj čakalne vrste*

$$p_k = \frac{\lambda_0\lambda_1 \dots \lambda_{k-1}}{\mu_1\mu_2 \dots \mu_k}p_0 = p_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \quad k = 0, 1, 2, \dots \quad (\text{B.20})$$

Ob upoštevanju pogoja (B.18) pa lahko zapišemo še enačbo za izračun verjetnosti  $p_0$  (prazne čakalne vrste)

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}} \quad (\text{B.21})$$

od tod pa lahko po enačbi (B.20) izračunamo posamezne verjetnosti stanj  $p_k$  čakalne vrste. Poudarimo naj, da je enačba (B.20) **osnovna enačba elementarne teorije čakalnih vrst**.

### B.3 Najpomembnejše lastnosti in parametri čakalnih vrst

Nekatere lastnosti čakalnih vrst so zelo splošne in neodvisne od obravnavanih procesov prihodov, strežbe in strežnih disciplin. Dovolj je že, da so čakalne vrste podvržene



splošnim statističnim zakonom in da njihovim spremenljivkam lahko določimo statistične vrednosti, kot sta na primer srednja vrednost in varianca [49], [50], [48].

Najpomembnejši je **Littlov teorem**, ki pravi: *Povprečno število strank  $N$  v čakalni vrsti je enako povprečnemu številu prihodov strank  $\lambda$  v ta sistem, pomnoženemu z njihovim povprečnim časom  $T$ , prebitem v tem sistemu.*

$$N = \lambda T \quad (\text{B.22})$$

Pri tem moramo podrobneje razdelati vrednosti  $N$  in  $T$ .  $N$  predstavlja povprečno število vseh strank v sistemu - tiste, ki so v čakalnici  $N_c$  in tiste, ki se trenutno strežejo  $N_s$ . Zato lahko zapišemo enačbo  $N = N_c + N_s$ . Analogno velja za povprečni čas  $T$ , ki ga stranke prebijejo v sistemu,  $T = W + \bar{x}$ , kjer je  $W$  povprečni čas prebit v čakalnici in  $\bar{x}$  povprečni čas strežbe. Littlov teorem tako lahko zapišemo tudi drugače, posebej za stranke v čakalnici in posebej za stranke v strežbi.

$$N = N_c + N_s = \lambda(W + \bar{x}) = \lambda W + \lambda \bar{x} = \lambda T \quad (\text{B.23})$$

Pomembni parametri čakalnih vrst so še  $\lambda$ , ki predstavlja povprečno število prihodov strank v sistem,  $\mu$  predstavlja povprečno število postreženih strank,  $\rho = \lambda/\mu$  pa izkoriščenost sistema. Veljati mora  $\rho < 1$ , sicer se v čakalnici nabere neskončno število strank in sistem postane nestabilen. V tem primeru velja  $\lambda > \mu$ , stranke v sistem prihajajo hitreje kot jih je ta zmožen streči.

### B.3.1 Čakalna vrsta M/M/1

Čakalna vrsta M/M/1 ima eksponentno porazdelitev časov med zaporednimi prihodi in časov strežbe posameznih strank. V sistemu je en strežnik, velikost čakalnice in populacije pa je neskončna. Dogajanje v čakalni vrsti lahko opišemo s procesom rojstev in smrti, ki ustrezajo naslednjim pogojem:

$$\begin{aligned} \lambda_k &= \lambda & k = 0, 1, 2, \dots \\ \mu_k &= \mu & k = 0, 1, 2, \dots \end{aligned}$$

kar pomeni, da so verjetnosti rojstev ali smrti (prehodi med stanji) za vsa stanja  $E_k$  enaki. To velja zaradi lastnosti eksponentne porazdelitve, ki je brez spomina, zato verjetnosti rojstev in smrti niso odvisne od prejšnjih rojstev in smrti. Enačbi (B.20) in (B.21) sedaj lahko zapišemo kot

$$p_k = p_0 \prod_{i=0}^{k-1} \frac{\lambda}{\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^k \quad k \geq 0 \quad (\text{B.24})$$

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k} \quad (\text{B.25})$$

Vsota konvergira pri pogoju  $0 \leq \rho = \lambda/\mu < 1$ , kar hkrati zagotovi, da je  $p_0 > 0$  (navadno imamo dodatno zahtevo, da se čakalna vrsta občasno izprazni). Končno dobimo enačbi za  $p_0$  in  $p_k$ :

$$p_0 = \frac{1}{1 + \frac{\lambda/\mu}{1 - \lambda/\mu}} = 1 - \frac{\lambda}{\mu} = 1 - \rho \quad (\text{B.26})$$

$$p_k = (1 - \rho)\rho^k \quad k \geq 0 \quad (\text{B.27})$$

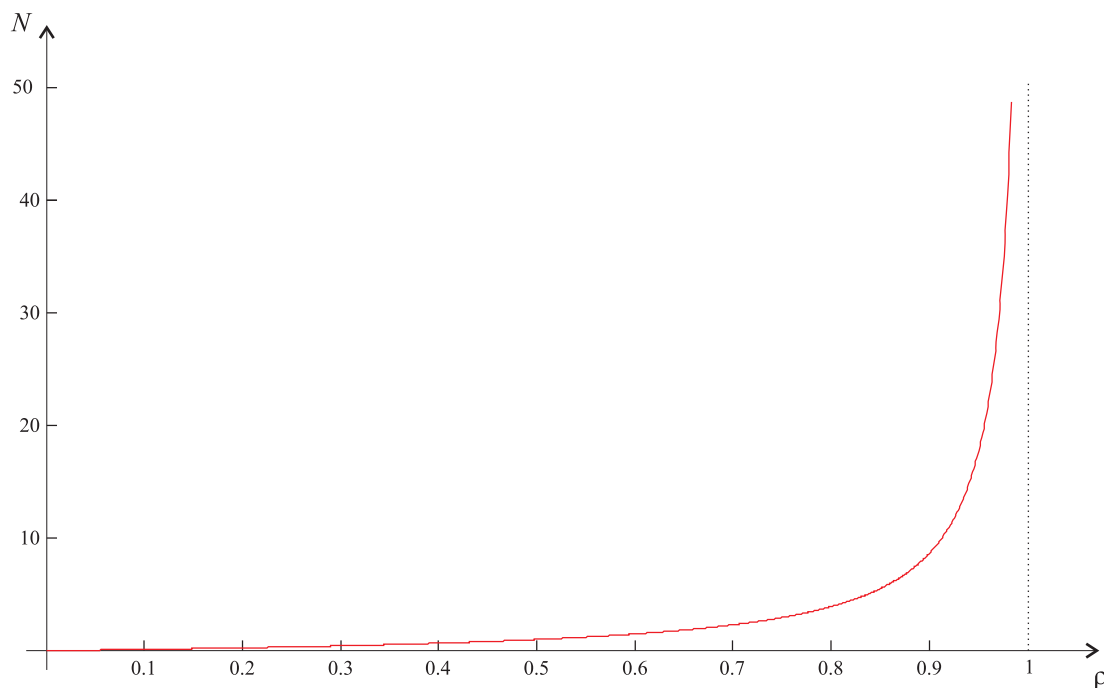
Z zgornjima enačbama smo dobili verjetnosti  $p_k$ , da se bo sistem nahajal v stanju  $E_k$ . Pri analizi čakalnih vrst pa nas zanimajo tudi parametri, ki neposredno zadevajo stranke v sistemu. Pomembni meri pri analizi čakalnih vrst sta povprečno število strank  $N$  v sistemu in povprečen čas  $T$ , ki ga te prebijejo v njem. Prvo mero preprosto izračunamo kot srednjo vrednost verjetnosti stanj sistema

$$N = \sum_{k=0}^{\infty} k p_k = (1 - \rho) \sum_{k=0}^{\infty} k \rho^k = \frac{\rho}{1 - \rho} \quad (\text{B.28})$$

Izrazimo še varianco stanj sistema

$$\sigma_N^2 = \sum_{k=0}^{\infty} (k - N)^2 p_k = \frac{\rho}{(1 - \rho)^2} \quad (\text{B.29})$$

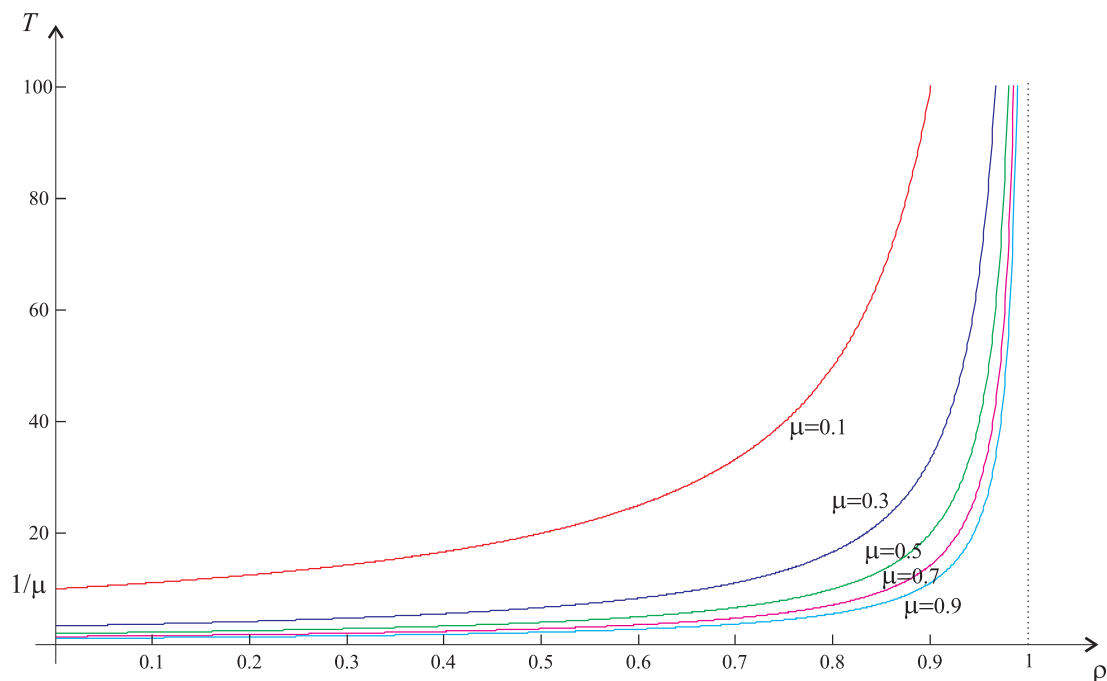
Na sliki B.4 je prikazan potek stanja sistema v odvisnosti od njegove izkoriščenosti  $\rho$ . Vidimo, da število strank v sistemu začne močno naraščati, ko se ta približuje polni izkoriščenosti. Obratovanje čakalne vrste M/M/1 pri polni izkoriščenosti  $\rho \rightarrow 1$  ni možno, saj se v njeni čakalnici nabere neskončno število strank  $N \rightarrow \infty$ .



Slika B.4 – Povprečno število strank v čakalni vrsti M/M/1

Povprečen čas  $T$ , ki ga stranke prebijejo v sistemu, pa dobimo preprosto iz Littlevega teorema in zanj veljajo enake ugotovitve kot za  $N$ . Ob približevanju polni izkoriščenosti čakalne vrste  $\rho \rightarrow 1$  velja  $T \rightarrow \infty$ .

$$T = \frac{N}{\lambda} = \left( \frac{\rho}{1-\rho} \right) \frac{1}{\lambda} = \frac{1/\mu}{1-\rho} \quad (\text{B.30})$$



Slika B.5 – Povprečen čas prebit v čakalni vrsti M/M/1

Zanimiva rezultata sta še verjetnost, da se bo v sistemu nahajalo  $k$  ali več strank, ki je podan z naslednjim izrazom:

$$P[\text{število strank} \geq k] = \sum_{i=k}^{\infty} p_i = \sum_{i=k}^{\infty} (1-\rho)\rho^i = \rho^k \quad (\text{B.31})$$

in pa verjetnostne porazdelitve časov, ki jih stranke prebijejo v sistemu ali čakalnici. Gostoti verjetnostne porazdelitve prebitih časov v sistemu in čakalnici sta podani z  $s(y)$  in  $w(y)$ , kumulativni verjetnostni porazdelitvi pa z  $S(y)$  in  $W(y)$ , pri čemer  $\delta(y)$  predstavlja enotni impulz.

$$s(y) = \mu(1-\rho)e^{-\mu(1-\rho)y} \quad (\text{B.32})$$

$$S(y) = 1 - e^{-\mu(1-\rho)y} \quad (\text{B.33})$$

$$w(y) = (1-\rho)\delta(y) + \lambda(1-\rho)e^{-\mu(1-\rho)y} \quad (\text{B.34})$$

$$W(y) = 1 - \rho e^{-\mu(1-\rho)y} \quad (\text{B.35})$$

### B.3.2 Čakalna vrsta M/M/1/K

Bolj realne razmere dobimo pri omejenem prostoru v čakalnici. Če predpostavimo, da je v čakalnici prostora za  $K$  strank (vključno s tisto, ki se streže), potem so vse stranke, ki pridejo do sistema, ko je čakalnica polna, zavrnjene in niso nikoli postrežene. V primeru paketnih omrežij to pomeni, da je v pomnilniku, ki hrani čakaajoče pakete, prostora le za določeno število paketov. Če prihajajoči paket naleti na polno zaseden pomnilnik, ga mora naprava zavrniti (ga zavrže). Če preskočimo izpeljavo, ki je podana v [49], in navedemo le rezultate, dobimo naslednje izraze za verjetnost, da se sistem nahaja v stanju  $k$ :

$$p_k = \begin{cases} \frac{1-\rho}{1-\rho^{K+1}}\rho^k & 0 \leq k \leq K \\ 0 & \text{drugače} \end{cases} \quad (\text{B.36})$$

Za povprečno število strank v sistemu:

$$N = \sum_{k=0}^K k p_k = \sum_{k=0}^K k \frac{1-\rho}{1-\rho^{K+1}} \rho^k \quad (\text{B.37})$$

Za povprečni čas prebit v sistemu, pri čemer parameter  $\Lambda$  predstavlja povprečno število strank, ki so bile sprejete v sistem in  $p_K$  verjetnost, da je bila čakalnica polna.

$$T = \frac{N}{\Lambda} = \frac{N}{\lambda(1-p_K)} \quad (\text{B.38})$$

### B.3.3 Čakalna vrsta M/G/1

Čakalna vrsta M/G/1 obravnava primere, pri katerih strežni časi niso porazdeljeni eksponentno (M), ampak splošno (G). Verjetnostna porazdelitev časov strežbe je podana z izrazom:

$$B(x) = P(\text{čas strežbe} \leq x)$$

V paketnih omrežjih je to bolj realen primer, saj ne moremo pričakovati, da bodo dolžine paketov, ki so premosorazmerne s časom njihove strežbe, eksponentno porazdeljene. V globljo analizo čakalne vrste M/G/1 se ne bomo spuščali in na tem mestu navajamo le najpomembnejše rezultate, povzete po [49].

Za razliko od obravnave čakalne vrste M/M/1 se tu ne posvečamo analizi verjetnosti posameznih stanj sistema, ampak naredimo analizo s stališča stranke, ki se trenutno streže. Za to moramo definirati naslednje spremenljivke:

$C_n$	$n$ -ta stranka, ki prispe v sistem,
$q_n$	število strank v sistemu po odhodu stranke $C_n$ in njena srednja vrednost $\bar{q}$ ,
$v_n$	število strank, ki prispe v sistem med strežbo stranke $C_n$ ,
$x_n$	čas strežbe stranke $C_n$ in njena srednja vrednost $\bar{x}$ .

Po analizi, podani v [49], dobimo naslednji rezultat, ki je dobro poznan pod imenom *Pollaczek-Kninchin-ova* formula.

$$\bar{q} = \rho + \frac{\lambda^2 \overline{x^2}}{2(1 - \rho)} \quad (\text{B.39})$$

pri čemer je  $\rho = \lambda \bar{x}$ . Če pa upoštevamo še zvezo  $\overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2$ , potem lahko enačbo (B.42) zapišemo tudi kot funkcijo variacijskega koeficienta  $C_b$ , ki se izračuna po formuli  $C_b = \sigma_b / \bar{x}$ , pri čemer indeks  $b$  označuje gostoto verjetnostne porazdelitev  $b(x)$ .

$$\bar{q} = \rho + \rho^2 \frac{1 + C_b^2}{2(1 - \rho)} \quad (\text{B.40})$$

Iz zgornjih enačb lahko izpeljemo še povprečno število strank v sistemu  $N$ , povprečen čas  $T$ , prebit v sistemu, in povprečen čas  $W$ , prebit v čakalnici. Povprečno število strank v sistemu  $N$  je kar enako  $\bar{q}$ .  $T$  lahko izračunamo s pomočjo Littleve formule  $N = \lambda T$ , iz katere sledi:

$$T = \frac{N}{\lambda} = \bar{x} + \frac{\rho \bar{x}(1 + C_b^2)}{2(1 - \rho)} \quad (\text{B.41})$$

Povprečen čas  $W$ , prebit v čakalnici, pa je ravno povprečen čas prebit v sistemu manj povprečen čas strežbe ali  $W = T - \bar{x}$ . To nam da izraz

$$W = \frac{\rho \bar{x}(1 + C_b^2)}{2(1 - \rho)} \quad (\text{B.42})$$

Če zgornje enačbe prenesemo na čakalno vrsto M/M/1 in upoštevamo, da je  $C_b$  za eksponentno porazdelitev ravno 1, potem dobimo izraze, ki smo jih podali v razdelku B.3.1.



## C Analiza lastnosti odpravnika SCDRR

Delovanje odpravnika SCDRR je enako kot delovanje odpravnika DRR, razlika je le v mejnih vrednostih kvantov  $Q_i$ . Pri DRR je vrednost kvanta omejena navzdol ( $Q_i \geq L_{max}$ ), pri SCDRR pa navzgor ( $Q_{min} \leq L_{max} - 1$ ). Delovanje odpravnika DRR je predstavljeno v razdelku 4.5.6.

### C.1 Parametri SCDRR

Podajmo spremenljivke, ki jih potrebujemo za opis in izpeljavo lastnosti in delovanja odpravnika SCDRR:

$R$	bitna hitrost linije na izhodu odpravnika,
$N$	število aktivnih čakalnih vrst,
$L_{max}$	dolžina najdaljšega možnega paketa.
$r_i$	bitna hitrost rezervirana za čakalno vrsto $i$ ,
$w_i$	utež čakalne vrste $i$ ,
$Q_i$	kvant dodeljen čakalni vrsti $i$ ,
$DC_i$	števec primanjkljaja (Deficit Counter), ki določa razliko med količino podatkov, ki bi jih čakalna vrsta $i$ lahko odposlala in količino podatkov, ki jih je v resnici odposlala.
$F$	velikost okvira, $F = \sum_i Q_i$

### C.2 Pravičnost algoritma SCDRR

Splošna definicija pravičnosti odpravnikov je bila podana v razdelku 4.4.1. Za naš SCDRR algoritem izračunamo obe meri pravičnosti (RFM in AFM) ter pokažemo, da sta ti meri za SCDRR strožji kot mere za primerljive algoritme, ki temeljijo na DRR.

Poleg zgornjih definicij in spremenljivk moramo pri izpeljavi RFM in AFM za SCDRR algoritem podati še spodnje spremenljivke in definicije. Pri tem upoštevamo, da ima lahko vsak podatkovni pretok svojo čakalno vrsto.

$SENT_i(t_1, t_2)$	količina podatkov, ki se v časovnem intervalu $(t_1, t_2)$ odpošljejo iz čakalne vrste $i$
$SENT_i(m)$	količina podatkov, ki se odpošljejo iz čakalne vrste $i$ v obhodu $m$
$SENT_i(p, m)$	količina podatkov, ki se odpošljejo iz čakalne vrste $i$ v obhodih $p$ do vključno $m$

Glede na [115] podajmo sledeče definicije:

**Definicija C.1** Mejo absolutne pravičnosti za pretok  $i$  v časovnem intervalu  $(t_1, t_2)$  označimo z:

$$AF_i(t_1, t_2) = \left| \frac{SENT_i(t_1, t_2)}{w_i} - \frac{SENT_{gps_i}(t_1, t_2)}{w_{gps_i}} \right| \quad (C.1)$$

kjer s  $SENT_{gps_i}$  označimo količino storitve, ki bi jo pretok  $i$  prejel od odpravnik GPS, in z  $w_{gps_i}$  utež pretoka  $i$  v GPS odpravniku.

**Definicija C.2** Absolutna pravičnost (AF) algoritma v časovnem intervalu  $(t_1, t_2)$  in absolutna mera pravičnosti (AFM) algoritma sta podani z:

$$AF(t_1, t_2) = \max_{\forall i} AF_i(t_1, t_2) \quad (C.2)$$

$$AFM = \max_{\forall (t_1, t_2)} AF(t_1, t_2) \quad (C.3)$$

Za razliko od absolutne mere pravičnosti je relativna mera pravičnosti smiselno definirana le kot primerjava storitve dveh aktivnih čakalnih vrst (pretokov) v nekem časovnem intervalu. Tako za AFM kot za RFM velja, da so vse obravnavane čakalne vrste aktivne ves čas podanega časovnega intervala.

**Definicija C.3** Relativno pravičnost za par čakalnih vrst  $(i, j)$  v časovnem intervalu  $(t_1, t_2)$  označimo z  $RF_{(i,j)}(t_1, t_2)$  in je podana z izrazom:

$$RF_{(i,j)}(t_1, t_2) = \left| \frac{SENT_i(t_1, t_2)}{w_i} - \frac{SENT_j(t_1, t_2)}{w_j} \right| \quad (C.4)$$

**Definicija C.4** Relativno pravičnost za čakalno vrsto  $i$  v časovnem intervalu  $(t_1, t_2)$  označimo z  $RF_i(t_1, t_2)$ :

$$RF_i(t_1, t_2) = \max_{\forall j} RF_{(i,j)}(t_1, t_2) \quad (C.5)$$

**Definicija C.5** Relativna pravičnost (RF) algoritma v časovnem intervalu  $(t_1, t_2)$  in relativna mera pravičnosti (RFM) algoritma sta podani z izrazoma:

$$RF(t_1, t_2) = \max_{\forall i} RF_i(t_1, t_2) \quad (C.6)$$

$$RFM = \max_{\forall (t_1, t_2)} RF(t_1, t_2) \quad (C.7)$$

**Definicija C.6** Odpravnik je **pravičen**, če sta tako AFM kot RFM omejena z neko konstanto majhne vrednosti.

### C.2.1 Relativna mera pravičnosti

Nadaljujmo s podrobno analizo relativne pravičnosti SCRR algoritma.

**Lema C.1** Za vsa cela števila  $i$  drži naslednja trditev:

$$0 \leq DC_i \leq L_{max} - 1 \quad (C.8)$$

kjer z  $L_{max}$  označimo najdaljši paket, ki lahko prispe v sistem.



**Komentar** Na tem mestu želimo poudariti nekatere lastnosti parametra  $DC_i$ . Če na  $DC_i$  gledamo kot na funkcijo, lahko rečemo, da le ta preslika nenegativna cela števila v nenegativna cela števila. Ne moremo reči, ali funkcija  $DC_i$  s časom narašča ali pada, lahko pa z gotovostjo trdimo, da preslika  $\mathcal{N}$  v  $\mathcal{N}$ , torej  $DC_i: \mathcal{N} \rightarrow \mathcal{N}$ . Ravno tako je potrebno poudariti, da je zato  $DC_i$  diskretna spremenljivka. To pomeni, da lahko  $DC_i$ , na primer, zavzame ali vrednost  $L_{max} - 1$  ali vrednost  $L_{max}$ , ne more pa zavzeti vmesnih vrednosti. Z lemo C.1 torej trdimo, da lahko pri SCDRR algoritmu (in pri vseh algoritmih izpeljanih iz DRR)  $DC_i$  zavzame vrednosti, ki niso večje od  $L_{max} - 1$ . Lema 1 v [93] te lastnosti spremenljivke  $DC_i$  ne razjasni - trdijo, da velja sledeča:  $0 \leq DC_i < L_{max}$ . Slednja trditev bi bila pravilna le, če bi bila  $DC_i$  zvezna funkcija (spremenljivka). Tako v našem SCDRR algoritmu kot v originalnem DRR algoritmu pa je  $DC_i$  diskretna funkcija, zato je izraz (C.8) primernejši od istega izraza v [93].

**Dokaz - Lema C.1** Ob zagonu SCDRR algoritma se  $DC_i$  postavi na 0 za vsak  $i$ . V prvem obhodu lahko vsaka aktivna čakalna vrsta oddaja pakete toliko časa, dokler je vsota njihovih dolžin manjša ali enaka kvantu, ki je dodeljen tej čakalni vrsti:  $SENT_i(1) \leq Q_i$ . Vrednost  $DC_i$  po prvem obhodu pa je:  $Q_i - SENT_i(1) = DC_i \geq 0$ .

Opazujmo sedaj  $m$ -ti obhod. Opazimo, da se vrednost  $DC_i$  spremeni le, če se iz čakalne vrste  $i$  odpravi kakšen paket. V primeru, da se iz čakalne vrste  $i$  odpravijo vsi paketi in se čakalna vrsta sprazni, se vrednost  $DC_i$  nastavi na 0. V nasprotnem primeru je vrednost  $DC_i$  ob koncu obhoda  $m$  naslednja:

$$DC_i(m|_{end}) = DC_i(m|_{beginning}) + Q_i - SENT_i(m) \quad (C.9)$$

Na začetku obhoda  $m + 1$  je količina storitve, ki je na voljo čakalni vrsti  $i$ , enaka vsoti  $Q_i$  in vrednosti  $DC_i$  ob koncu obhoda  $m$ :

$$DC_i((m + 1)|_{beginning}) = DC_i(m|_{end}) + Q_i \quad (C.10)$$

Ob koncu obhoda  $m$  v čakalni vrsti  $i$  imamo dve možni situaciji:

- Če se v čakalni vrsti  $i$  še vedno nahajajo paketi, potem mora biti prvi paket v čakalni vrsti  $i$  večji od trenutne vrednosti  $DC_i$ . Ker vemo, da je dolžina paketov omejena na  $L_{max}$ , sledi, da je  $DC_i$  ob koncu vsakega obhoda strogo manjši od  $L_{max}$ . Predpostavimo nasprotno. Naj bo  $DC_i$  večji ali enak  $L_{max}$ . V tem primeru bo prvi paket v čakalni vrsti, ker ne more biti večji od  $L_{max}$ , odposlan. Ko se v tekočem obhodu iz čakalne vrste  $i$  odpošlje eden ali več paketov,  $DC_i$  neobhodno postane manjši od  $L_{max}$ . Vsakokrat kadar je  $DC_i$  večji ali enak  $L_{max}$ , se iz čakalne vrste  $i$  odpošlje eden ali več paketov in na koncu vsakega obhoda je vrednost  $DC_i$  manjša od  $L_{max}$ . Neenačba (C.8) torej drži.
- Če v čakalni vrsti  $i$  ni več paketov za prenos, se vrednost  $DC_i$  nastavi na 0. Ker je  $DC_i = 0$  velja:  $DC_i \leq L_{max} - 1$  in neenačba (C.8) drži.

■

Glede na SCDRR algoritem se na začetku vsakega obhoda  $DC_i$  nastavi na:

$$DC_i((k + 1)|_{beginning}) = DC_i(k|_{end}) + Q_i$$

kjer  $k$  označuje katerikoli obhod algoritma. Dokazali smo, da ob koncu kateregakoli obhoda  $DC_i$  ne more biti večja ali enaka  $L_{max}$ , lahko je kvečjemu enaka  $L_{max} - 1$ . Ker je pri SCDRR algoritmu tudi vrednost  $Q_i$  manjša od  $L_{max}$ , lahko je kvečjemu enaka  $L_{max} - 1$ , dobimo naslednjo neenačbo:

$$DC_i \leq (L_{max} - 1) + (L_{max} - 1) = 2L_{max} - 2 < 2L_{max} \quad (\text{C.11})$$

Zgornja neenačba velja tudi za maksimalne vrednosti  $DC_i$  in  $Q_i$ , za vse čase in za vse indekse  $i$ .

To pomeni, da v istem obhodu ne moremo poslati dva zaporedna paketa maksimalne dolžine iz iste čakalne vrste. To je prednost SCDRR algoritma, saj originalni DRR algoritem te omejitve nima. To je očitna slabost DRR-a, ki lahko vodi do velike zakasnitve algoritma in zelo slabih razmer glede pravičnosti. Z omejitvijo velikosti kvantov  $Q_i$  ( $\forall i$ ), ki ne smejo preseči vrednosti  $L_{max} - 1$ , zagotovimo, da se lahko v istem obhodu pošlje največ en paket maksimalne dolžine in s tem izboljšamo zakasnitev algoritma in njegovo pravičnost.

Odpravnik krožno streže vse aktivne čakalne vrste. V nadaljevanju določimo količino podatkov, ki jih lahko odpošlje čakalna vrsta  $i$  v kateremkoli časovnem intervalu.

**Lema C.2** *Preučimo delovanje SCDRR algoritma za vsak časovni interval  $(t_1, t_2)$ , v katerem je čakalna vrsta  $i$  nenehno aktivna. Če je  $m$  število obhodov odpravnika, ki jih je v danem časovnem intervalu prejela čakalna vrsta  $i$ , potem velja naslednja neenačba:*

$$mQ_i - L_{max} + 1 \leq SENT_i(1, m) \leq mQ_i + L_{max} - 1 \quad (\text{C.12})$$

**Komentar** Na tem mestu bi radi poudarili, da v nobenem od nam poznanih del meje podane neenačbe za izračun odposlane količine podatkov niso tako ozke, kot v naši neenačbi (C.12). V nobenem izmed teh del namreč ni bilo upoštevano, da je  $DC_i$  funkcija diskretne spremenljivke, ki lahko zavzame le vrednosti, ki niso večje od  $L_{max} - 1$ . Povsod je bilo upoštevano, da je  $DC_i < L_{max}$ . Mi pa trdimo, da je točnejša vrednost  $DC_i \leq L_{max} - 1$ , saj  $DC_i$  zaradi svoje diskretnosti ne more zavzeti vrednosti, kot je na primer  $L_{max} - 0,001$ . Ta lastnost  $DC_i$  je obširneje razložena v komentarju na strani 143.

Upoštevaajoč dejstvo, da je ob koncu vsakega obhoda vrednost  $DC_i$  lahko le v mejah  $0 \leq DC_i \leq L_{max} - 1$ , dokažemo, da je naša meja za mero pravičnosti bolj pravilna in strožja od meje, izpeljane v [93], ki temelji na napačni predpostavki  $DC_i < L_{max}$ .

**Dokaz - Lema C.2** Z obhodom označujemo vsako priložnost strežbe čakalne vrste  $i$  v časovnem intervalu  $(t_1, t_2)$ . Obravnavani obhodi tečejo od indeksa 1 do indeksa  $m$ . Zaradi prikladnosti naj čas  $t_1$  označuje konec domnevnega obhoda 0. Z  $DC_i(k)$  pa označimo vrednost  $DC_i$  ob koncu obhoda  $k$ . Glede na lastnosti predlaganega SCDRR algoritma velja:

$$SENT_i(k) + DC_i(k) = Q_i + DC_i(k - 1)$$

Utemeljitev za zgornji izraz izhaja iz samega algoritma, za katerega velja:

$$DC_i(k) = DC_i(k - 1) + Q_i - (\text{kolicina podatkov odposlana v obhodu}) k$$

Vidimo, da je to samo drugačen zapis prejšnjega izraza. Pri zadnjem izrazu predpostavljamo, da je čakalna vrsta aktivna znotraj celotnega časovnega intervala  $(t_1, t_2)$ . V obhodu  $k$  je tako celotna količina storitve, ki pripada čakalni vrsti  $i$ , enaka  $Q_i + DC_i(k-1)$ . V obhodu  $k$  se iz čakalne vrste  $i$  odpošlje  $SENT_i(k)$  podatkov. Ker se čakalna vrsta  $i$  v časovnem intervalu  $(t_1, t_2)$  nikoli ne izprazni, se preostanek shrani v  $DC_i(k)$ . Glede na povedano lahko zapišemo:

$$SENT_i(k) = Q_i + DC_i(k-1) - DC_i(k) \quad (C.13)$$

S seštevanjem zadnjega izraza preko vseh  $m$  obhodov dobimo:

$$\begin{aligned} \sum_{k=1}^m SENT_i(k) &= \sum_{k=1}^m Q_i + \sum_{k=1}^m [DC_i(k-1) - DC_i(k)] \\ &= mQ_i + [(DC_i(0) - DC_i(1)) + (DC_i(1) - DC_i(2)) + \dots] \end{aligned}$$

Če pobliže pogledamo zadnjo vsoto na desni strani zgornje enačbe, vidimo, da se vsi vmesni členi vsote izničijo. Ostaneta le prvi in zadnji člen.

$$\sum_{k=1}^m SENT_i(k) = mQ_i + DC_i(0) - DC_i(m) \quad (C.14)$$

Ko  $DC_i(0)$  zavzame svojo največjo možno vrednost  $L_{max} - 1$  in  $DC_i(m)$  zavzame svojo najmanjšo možno vrednost 0, je razlika  $DC_i(0) - DC_i(m)$  maksimalna. Z uporabo rezultata leme C.1 dobimo:

$$SENT_i(1, m) \leq mQ_i + L_{max} - 1 \quad (C.15)$$

Ko pa  $DC_i(0) = 0$  zavzame svojo najmanjšo možno vrednost 0 in  $DC_i(m) = L_{max} - 1$  zavzame svojo največjo možno vrednost  $L_{max} - 1$ , je razlika  $DC_i(0) - DC_i(m)$  minimalna, tako imamo:

$$SENT_i(1, m) \geq mQ_i - L_{max} + 1 \quad (C.16)$$

Z združitvijo zadnjih dveh neenačb dobimo neenačbo (C.12). ■

S pomočjo definicij C.3, C.4 in C.5 lahko sedaj izpeljemo relativno mero pravičnosti za predlagani SCDRR algoritem.

**Lema C.3** *Osnovna lastnost SCDRR algoritma je, da v poljubnem časovnem intervalu, v katerem sta čakalni vrsti  $i$  in  $j$  stalno aktivni, velja: med poljubnima dvema priložnostma za strežbo čakalne vrste  $i$  mora imeti čakalna vrsta  $j$  vsaj eno priložnost strežbe.*

**Dokaz - Lema C.3** Dokaz izvedemo neposredno iz načina strežbe odpravnikar s krožnim dodeljevanjem. Predpostavimo, da je čakalna vrsta  $i$  že imela eno priložnost strežbe. Ob koncu te priložnosti se je indeks  $i$  prestavil na konec seznama aktivnih čakalnih vrst. Ker je čakalna vrsta  $j$  v tem času ravno tako aktivna, mora biti njen indeks na istem seznamu kot indeks  $i$ . In ker je slednji sedaj zadnji na seznamu, mora biti indeks  $j$  na tem seznamu pred njim. To pomeni, da bo čakalna vrsta  $j$  dobila svojo priložnost strežbe pred čakalno vrsto  $i$ . ■

**Teorem C.1** Za poljuben časovni interval in poljubni obhod odpravnikar SCDRR velja:

$$RFM = \left| \frac{SENT_i(m)}{w_i} - \frac{SENT_j(m')}{w_j} \right| \leq \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} + Q_{min} - \frac{1}{w_i} - \frac{1}{w_j} \quad (C.17)$$

kjer z  $m$  označimo število obhodov, ki jih je v časovnem intervalu  $(t_1, t_2)$  prejela čakalna vrsta  $i$ , in z  $m'$  število obhodov, ki jih je v istem časovnem intervalu prejela čakalna vrsta  $j$ .

**Dokaz - Teorem C.1** Vzemimo poljuben časovni interval  $(t_1, t_2)$  v poljubnem obhodu odpravnikar SCDRR in poljubni dve čakalni vrsti  $i$  in  $j$ , ki sta aktivni celoten čas podanega intervala. Glede na lemo C.3 lahko zapišemo:  $|m - m'| \leq 1$ . In glede na lemo C.2 imamo:

$$\frac{SENT_i(1, m)}{w_i} \leq \frac{mQ_i}{w_i} + \frac{L_{max}}{w_i} - \frac{1}{w_i} \quad (C.18)$$

$$\frac{SENT_j(1, m')}{w_j} \geq \frac{m'Q_j}{w_j} - \frac{L_{max}}{w_j} + \frac{1}{w_j} \quad (C.19)$$

Uporabili smo le enostranske neenačbe iz leme reflexmasent: desno stran za čakalno vrsto  $i$  in levo za čakalno vrsto  $j$ . V enačbah (C.18) in (C.19) opazimo, da sta pomembna člena  $\frac{Q_i}{w_i}$  in  $\frac{Q_j}{w_j}$ . Ker velja, da je  $\frac{Q_i}{w_i} = Q_{min}$ , imamo:

$$\frac{SENT_i(1, m)}{w_i} \leq mQ_{min} + \frac{L_{max}}{w_i} - \frac{1}{w_i} \quad (C.20)$$

$$\frac{SENT_j(1, m')}{w_j} \geq m'Q_{min} - \frac{L_{max}}{w_j} + \frac{1}{w_j} \quad (C.21)$$

Iz tega sledi:

$$\begin{aligned} & \left| \frac{SENT_i(1, m)}{w_i} - \frac{SENT_j(1, m')}{w_j} \right| \\ & \leq \left| mQ_{min} + \frac{L_{max}}{w_i} - \frac{1}{w_i} - m'Q_{min} + \frac{L_{max}}{w_j} - \frac{1}{w_j} \right| \\ & \leq |Q_{min}(m - m')| + \left| \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \right| \end{aligned} \quad (C.22)$$

$$= Q_{min}|m - m'| + \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \quad (C.23)$$

V neenačbi (C.23) smo pri zadnjih dveh členih ohranili minus. To lahko opravičimo, če dokažemo, da je izraz  $\frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j}$  vedno večji ali enak nič. Izraz preuredimo v:

$$\frac{L_{max} - 1}{w_i} + \frac{L_{max} - 1}{w_j}$$

in ker velja, da je  $L_{max} \geq 2$  in  $w_i, w_j \geq 1$ , je naša trditev dokazana. Sedaj dobimo:

$$\left| \frac{SENT_i(1, m)}{w_i} - \frac{SENT_j(1, m')}{w_j} \right| \leq Q_{min} |m - m'| + \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \quad (C.24)$$

Z uporabo leme C.3, ki pravi, da velja  $|m - m'| \leq 1$ , in neenačbe (C.24) lahko zaključimo dokaz teorema:

$$\left| \frac{SENT_i(1, m)}{w_i} - \frac{SENT_j(1, m')}{w_j} \right| \leq Q_{min} + \frac{L_{max}}{w_i} + \frac{L_{max}}{w_j} - \frac{1}{w_i} - \frac{1}{w_j} \quad (C.25)$$

■

### C.3 Zakasnitev algoritma SCDRR

Zgornjo mejo algoritemske zakasnitve za odpravnik SCDRR izpeljemo po zgledu izpeljav v delih Stiliadisa in Varne [98] ter Shreedharja in Vargheseja [93]. Z uporabo dveh različnih pristopov k izpeljavi pokažemo, da se ta meja ne spreminja ne glede na uporabljen pristop.

Najprej podajmo spremenljivke, definicije in pogoje za SCDRR. Druge uporabljene spremenljivke so enake kot v razdelku C.2, v katerem smo podali pravičnost odpravnika SCDRR.

**Definicija C.7** *Aktivno obdobje čakalne vrste (active period) je najdaljši časovni interval, v katerem se v čakalni vrsti nahaja ali streže vsaj en paket.*

**Definicija C.8** *Delavno obdobje (busy period) čakalne vrste je najdaljši časovni interval, v katerem bi bila čakalna vrsta aktivna, če bi se stregla natanko z njeno rezervirano bitno hitrostjo.*

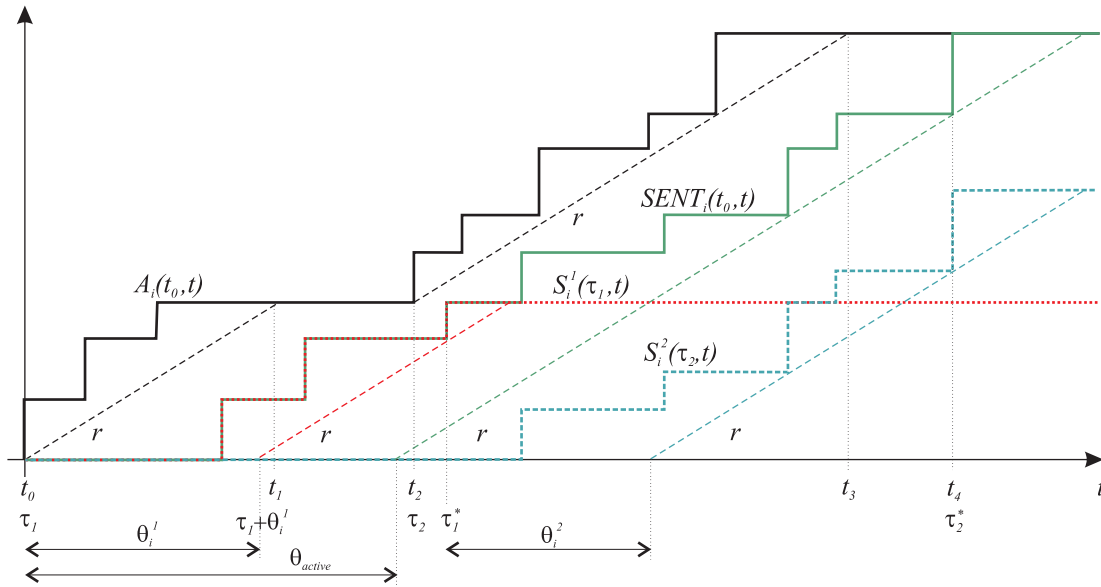
Aktivno obdobje čakalne vrste se razlikuje in je navadno daljše od njenega delavnega obdobja. Eno aktivno obdobje namreč lahko vsebuje več delavnih obdobj. Aktivno obdobje odseva resnično delovanje odpravnika, saj je storitev posamezne čakalne vrste odvisna od števila aktivnih čakalnih vrst. Delavno obdobje čakalne vrste pa je matematični konstrukt, ki nam pove, koliko časa bi bila čakalna vrsta aktivna ob strežbi natanko z njeno rezervirano bitno hitrostjo.

Definirajmo naslednje parametre in spremenljivke:

$\alpha_i$	začetek delavnega obdobja čakalne vrste $i$
$t > \alpha_i$	časovni trenutek za katerega velja, da je v časovnem intervalu $(\alpha_i, t)$ čakalna vrsta $i$ stalno delavna
$A_i(\alpha_i, t)$	prihodi v čakalno vrsto $i$ v časovnem intervalu $(\alpha_i, t)$ [bit]
$SENT_i(\alpha_i, t)$	storitev čakalne vrste $i$ v časovnem intervalu $(\alpha_i, t)$ [bit]
$S_i(\alpha_i, t)$	število bitov, ki so v čakalno vrsto $i$ prispeli po času $\alpha_i$ in so iz nje odpravljeni v časovnem intervalu $(\alpha_i, t)$
$\tau_j$	začetek $j$ -tega delavnega obdobja čakalne vrste
$\tau_j^*$	časovni trenutek, v katerem je odpravljen zadnji bit, ki je prispel v času $j$ -tega delavnega obdobja
$S_i^j(\tau_j, t)$	število bitov, ki so v čakalno vrsto prispeli v $j$ -tem delavnem obdobju v časovnem intervalu $(\tau_j, t)$

Podajmo še alternativno in bolj intuitivno definicijo delavnega obdobja čakalne vrste  $i$ . Delavno obdobje čakalne vrste  $i$  je najdaljši časovni interval  $(\alpha_i, t)$  za katerega velja, da nakopičeni prihodi v čakalno vrsto  $i$  v podanem intervalu nikoli ne padejo pod storitev, ki bi jo čakalna vrsta prejela ob strežbi natanko z njeno rezervirano bitno hitrostjo  $r_i$ :  $A_i(\alpha_i, t) \geq r_i(t - \alpha_i)$ .

Omeniti je potrebno tudi, da veličini  $S_i(\alpha_i, t)$  in  $SENT_i(\alpha_i, t)$  nista enaki. Tekom določenega časovnega intervala lahko odpravnik še vedno streže pakete, ki so prispeli v prejšnjem delavnem obdobju. Zato  $S_i(\alpha_i, t)$  in  $SENT_i(\alpha_i, t)$  nista enaki, pri čemer mora veljati:  $S_i(\alpha_i, t) \leq SENT_i(\alpha_i, t)$ . Relacije med zgornjimi parametri in spremenljivkami so dobro razvidne iz slike C.1



Slika C.1 – Aktivna in delavna območja čakalne vrste  $i$   $\mathcal{LR}$  odpravnik

**Definicija C.9** Množica  $T$  je unija vseh časovnih trenutkov, v katerih odpravnik konča strežbo ene čakalne vrste in začne s strežbo naslednje. Množico vseh časovnih trenutkov, ob katerih začne odpravnik streči čakalno vrsto  $i$ , označimo s  $T_i$ . Velja:

$$T = \bigcup_{i=1}^N T_i$$

**Definicija C.10** Odpravnik  $S$  pripada razredu  $\mathcal{LR}$  odpravnikov, če in samo če za vse časovne trenutke  $t$  po času  $\alpha_i$  v  $j$ -tem delavnem obdobju velja, da:

$$S_i^j(\alpha_i, t) \geq \max\{0, r_i(t - \alpha_i - \theta_i)\} \quad (\text{C.26})$$

kjer je  $\theta_i$  najmanjša nenegativna konstanta, ki zadovolji zgornjo neenačbo in predstavlja zgornjo mejo algoritemske zakasnitve odpravnik.

**Lema C.4** Naj bo  $(\alpha_i, \tau)$  delavno obdobje čakalne vrste  $i$ . Vsi paketi, ki so prispeli v čakalno vrsto  $i$  v času tega delavnega obdobja, bodo odpravljeni do časa  $\tau + \theta_i$ .

**Dokaz - Lema C.4** Upoštevajoč definicijo parametrov in spremenljivk na strani 147 naj bo  $t$  časovni trenutek, v katerem se odpravi zadnji bit zadnjega paketa  $j$ -tega delavnega obdobja čakalne vrste  $i$ . Takrat velja:

$$A_i(\alpha_i, \tau) = S_i(\alpha_i, t) \quad (\text{C.27})$$

Glede na definicijo delavnega obdobja lahko zapišemo:

$$A_i(\alpha_i, \tau) = r_i(\tau - \alpha_i) \quad (\text{C.28})$$

Iz definicije  $\mathcal{LR}$  odpravnikov in iz enačb (C.27) in (C.28) sledi:

$$r_i(t - \alpha_i - \theta_i) - r_i(\tau - \alpha_i) \leq 0 \quad (\text{C.29})$$

ali ekvivalentno

$$t \leq \tau + \theta_i \quad (\text{C.30})$$

kar dokazuje, da je  $\theta_i$  zgornja meja algoritemske zakasnitve odpravnika. ■

Zgornja definicija algoritemske zakasnitve temelji na delavnem obdobju čakalne vrste. V praksi pa je lažje analizirati odpravnike glede na aktivna obdobja. Naslednja lema povezuje obe obdobji in nam tako pomaga pri lažji analizi algoritemske zakasnitve. Omogoči nam oceno algoritemske zakasnitve  $\mathcal{LR}$  odpravnika zgolj na analizi aktivnih obdobji čakalne vrste.

**Lema C.5** Naj  $(\alpha_i, t)$  označuje časovni interval, v katerem je čakalna vrsta  $i$  ves čas delavna. Če je storitev čakalne vrste za pakete, ki so prispeli v časovnem intervalu  $(\alpha_i, t)$ , omejena za vsak časovni trenutek  $s$ ,  $\alpha_i < s \leq t$  po pogoju:

$$S_i(\alpha_i, s) \geq \max\{0, r_i(s - \alpha_i - \theta_i)\} \quad (\text{C.31})$$

potem je tak odpravnik  $\mathcal{LR}$  odpravnik z algoritemsko zakasnitvijo, ki je manjša ali enaka  $\theta_i$ . ■

Zgornja lema je dokazana v [98] in določa, da če odpravnik lahko omeji storitev čakalne vrste v njenem delavnem obdobju, potem je to  $\mathcal{LR}$  odpravnik z algoritemsko zakasnitvijo, ki ne presega tiste, ki jo izračunamo iz delavnega obdobja. Poudariti pa je potrebno, da nasprotno ne velja.

V zgornji lemi je skrita stroga definicija algoritemske zakasnitve  $\theta_i$ . Zaradi njene pomembnosti jo podajmo posebej.

**Definicija C.11** Algoritemska zakasnitev za čakalno vrsto  $i$  je najmanjša nenegativna konstanta  $\theta_i$ , ki za vsa delavna obdobja zadovoljuje neenačbo:

$$S_i(\alpha_i, t) \geq \max\{0, r_i(t - \alpha_i - \theta_i)\} \quad (\text{C.32})$$

Glede na zgornjo definicijo algoritemska zakasnitev ni le čas, ki je potreben, da se iz čakalne vrste  $i$  odpošlje prvi prispeli paket, ampak je algoritemska zakasnitev  $\mathcal{LR}$  odpravnika čas, ki je potreben, da storitev čakalne vrste  $i$  postane enaka njeni rezervirani bitni hitrosti.

Definirajmo naslednji spremenljivki:

$\beta_i$  časovni trenutek, v katerem čakalna vrsta  $i$  postane aktivna  
 $t > \beta_i$  časovni trenutek, za katerega velja, da je v časovnem intervalu  $(\beta_i, t)$  čakalna vrsta  $i$  stalno delavna

**Definicija C.12**  $\theta'_i$  je najmanjša nenegativna konstanta, ki zadovoljuje izraz:

$$SENT_i(\beta_i, t) \geq \max\{0, r_i(t - \beta_i - \theta'_i)\} \quad (\text{C.33})$$

**Posledica** Čeprav se lahko zgodi, da časovni interval  $(\beta_i, t)$  ni neprekinjeno delavno obdobje čakalne vrste  $i$ , drži sledeče:

$$\theta_i \leq \theta'_i \quad (\text{C.34})$$

npr., algoritemska zakasnitev  $\theta_i$ , kot je bila definirana, je za vsako čakalno vrsto  $i$  omejena z  $\theta'_i$ .

Zgornja relacija med dodatkom in lemo C.4 je dokazana v [98].

Slika C.1 grafično predstavi algoritemsko zakasnitev, delavno in aktivno obdobje čakalne vrste ter glavne parametre  $\mathcal{LR}$  odpravnikov. Označeni sta meji algoritemske zakasnitve  $\theta_i^1$  in  $\theta_i^2$  za prvo in drugo delavno obdobje. Kot algoritemska zakasnitev čakalne vrste  $i$  se vzame manjša izmed obeh. Slika ravno tako potrjuje trditve leme C.4 in dodatka - algoritemska zakasnitev delavnega območja je omejena z algoritemsko zakasnitvijo aktivnega obdobja. Vidimo, da velja  $\theta_i^1 \leq \theta_i^{\prime 1}$ , kot je navedeno v izrazu (C.34). Naj na kratko obrazložimo še druge spremenljivke na sliki C.1:

$(t_0, t_1)$  prvo delavno obdobje čakalne vrste  $i$   
 $(t_2, t_3)$  drugo delavno obdobje čakalne vrste  $i$   
 $(t_0, t_4)$  aktivno obdobje čakalne vrste  $i$

Naj omenimo, da v časovnem intervalu  $(\tau_j, t)$  velja:

$$S_i^j(\tau_j, t) \geq \max\{0, r_i(t - \tau_j - \theta_i^j)\}$$

Nadaljujmo z izračunom zgornje meje algoritemske zakasnitve, ki temelji na aktivnem obdobju čakalne vrste  $i$ .

Algoritemska zakasnitev za čakalno vrsto  $i$   $\mathcal{LR}$  odpravnika je navzgor omejena. Ob ustrezno glajenem prometu to velja tudi za zakasnitev v omrežnih napravah, posledično za zakasnitev na celotni prenosni poti in za potrebno velikost medpomnilnikov. Za odpravnik DRR je bil teorem dokazan s strani Stiliadisa in Varme v [98]. Za odpravnik SCDRR pa je teorem in njegov dokaz naveden spodaj.

**Teorem C.2** Odpravnik SCDRR pripada razredu  $\mathcal{LR}$  odpravnikov z zgornjo mejo algoritemske zakasnitve  $\theta_i$  za čakalno vrsto  $i$ , ki je podana z izrazom:

$$\theta_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{Fr_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (\text{C.35})$$



**Dokaz - Teorem C.2** Ker lahko zgornjo mejo algoritemske zakasnitve  $\mathcal{LR}$  odpravnikov izračunamo na osnovi aktivnega obdobja čakalne vrste, glej (C.34), teorem dokažemo s tem, da pokažemo, da velja:

$$\theta'_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{Fr_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (\text{C.36})$$

V [45] je izpeljana meja algoritemske zakasnitve odpravnika Elastic Round Robin. Dokazano je, da če je algoritemska zakasnitev  $\theta'_i$  v časovnem intervalu  $(\beta_i, t)$  enaka svoji zgornji meji, potem mora veljati:

1.  $\beta_i \in T$
2.  $t \in T_i$

Ni težko preveriti, da zgornja pogoja veljata za vse odpravnike s krožnim dodeljevanjem, vključno z našim odpravnikom SCDRR. Z  $\beta_i^k$  označimo časovni trenutek, ki predstavlja začetek  $k$ -tega obhoda odpravnika za čakalno vrsto  $i$ . Opazimo, da velja sledeče:  $\beta_i^k \in T_i$ .

Iz zapisanega in upoštevajoč metodologijo, uporabljeno v [42], moramo za izračun zgornje meje algoritemske zakasnitve odpravnika SCDRR upoštevati spodnje časovne intervale za vse obhode  $k$ , ki jih je deležna čakalna vrsta  $i$ :

- $(\beta_i, \beta_i^1)$  časovni interval od trenutka, ko je čakalna vrsta  $i$  postala aktivna, do trenutka njenega prvega obhoda
- $(\beta_i^1, \beta_i^2)$  časovni interval med prvim in drugim obhodom za čakalno vrsto  $i$
- $(\beta_i^2, \beta_i^3)$  časovni interval med drugim in tretjim obhodom za čakalno vrsto  $i$
- ...

Upoštevati je potrebno, da ni nujno, da časovni trenutek  $\beta_i$  sovpada z začetkom ali koncem posameznega obhoda v čakalni vrsti  $i$ .

Za lažjo izpeljavo dokaza podajmo spremenljivke:

- $k_0$  označuje zaporedno številko obhoda, ki je bil v teku v časovnem trenutku  $\beta_i$  ali obhoda, ki se konča ravno v trenutku  $\beta_i$
- $t_h$  označuje časovni trenutek konca obhoda  $(k_0 + h - 1)$  in hkrati začetka obhoda  $(k_0 + h)$
- $SENT_i(m)$  označuje količino storitve čakalne vrste  $i$  v obhodu  $m$
- $DC_i(m)$  označuje vrednost v števcu primanjkljaja čakalne vrste  $i$  ob koncu obhoda  $m$

V lemi C.1 smo izpeljali meje števca primanjkljaja, ki so podane z izrazom:  $0 \leq DC_i \leq L_{max} - 1$ . Iz samega delovanja SCDRR algoritma lahko zapišemo naslednjo enačbo:

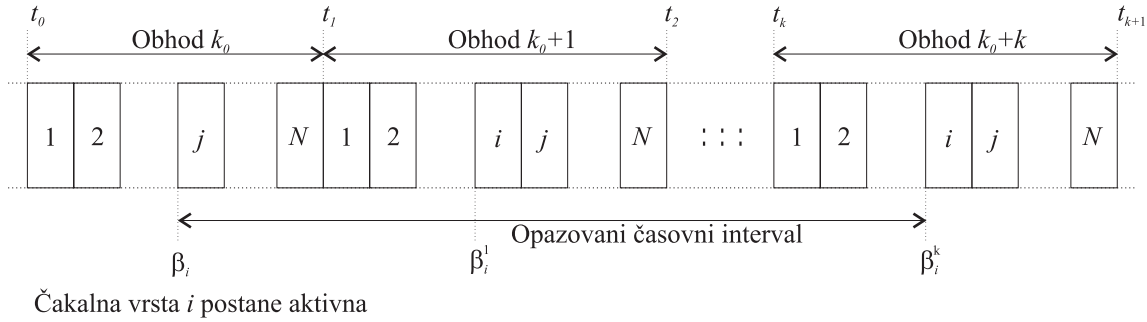
$$SENT_i(m) = Q_i + DC_i(m - 1) - DC_i(m) \quad (\text{C.37})$$

Predpostavimo, da časovni trenutek  $\beta_i$ , ko čakalna vrsta  $i$  postane aktivna, sovpada s časovnim trenutkom, ko neka druga čakalna vrsta začne z obhodom  $k_0$ . Takrat lahko množico vseh aktivnih čakalnih vrst razdelimo v dve podmnožici,  $G_\alpha$  in  $G_\beta$ , ki sta definirani kot:

- $G_\alpha$  množica vseh aktivnih čakalnih vrst, ki so postrežene v časovnem intervalu  $(\beta_i, t_1)$  - po tem, ko čakalna vrsta  $i$  postane aktivna
- $G_\beta$  množica vseh aktivnih čakalnih vrst, ki so postrežene v časovnem intervalu  $(t_0, \beta_i)$  - preden je čakalna vrsta  $i$  postala aktivna

Opazimo, da čakalna vrsta  $i$  ni v nobeni izmed množic, saj prvič pride na vrsto za strežbo šele v obhodu  $k_0 + 1$ .

Za izpeljavo zgornje meje algoritemske zakasnitve moramo analizirati časovni interval  $(\beta_i, \beta_i^k)$  in mejo zakasnitve izraziti kot funkcijo spremenljivk SCDRR algoritma.



Slika C.2 – Podroben pogled na obravnavani časovni interval  $(\beta_i, \beta_i^k)$

Za lažji izračun interval  $(\beta_i, \beta_i^k)$  razdelimo na tri intervale:  $(\beta_i, t_1)$ ,  $(t_1, t_k)$ ,  $(t_k, \beta_i^k)$ . Spomnimo se, da časovni trenutek  $t_1$  označuje konec obhoda  $k_0$  in začetek obhoda  $k_1$ . Nadaljnja analiza je razdeljena v tri dele, od katerih vsak obravnava enega izmed treh podanih časovnih intervalov.

- $(\beta_i, t_1)$  : Na začetku tega intervala čakalna vrsta  $i$  postane aktivna, vendar pa je zaradi delovanja SCDRR algoritma prvič postrežena v obhodu  $k_0 + 1$ . V podanem časovnem intervalu so postrežene vse čakalne vrste iz podmnožice  $G_\alpha$ . Trajanje podanega časovnega intervala izračunamo iz časov, potrebnih za postrežbo vseh čakalnih vrst iz podmnožice  $G_\alpha$ . Z uporabo enačbe (C.37) in njeno vsoto preko vseh čakalnih vrst iz podmnožice  $G_\alpha$  dobimo:

$$\begin{aligned} t_1 - \beta_i &= \frac{1}{R} \sum_{j \in G_\alpha} SENT_j(k_0) = \\ &= \frac{1}{R} \sum_{j \in G_\alpha} [Q_j + DC_j(k_0 - 1) - DC_j(k_0)] \end{aligned} \quad (C.38)$$

- $(t_1, t_k)$  :  $t_1$  označuje začetek obhoda  $k_0 + 1$ ,  $t_k$  pa konec obhoda  $k_0 + k - 1$  (in začetek obhoda  $k_0 + k$ ). V tem časovnem intervalu imamo  $k - 1$  celih obhodov, zato je vsaka aktivna čakalna vrsta postrežena  $k - 1$ -krat. Razdelimo interval  $(t_1, t_k)$  v  $k - 1$  podintervalov, po enega za vsakega izmed obhodov. Zapišemo lahko:

$$(t_1, t_k) = \bigcup_{h=1}^{k-1} (t_h, t_{h+1}) \quad (C.39)$$

V časovnih intervalih  $(t_h, t_{h+1})$ , ki predstavljajo obhode  $k_0 + h$ , vsaka čakalna vrsta dobi po eno priložnost strežbe. S seštevanjem preko vseh aktivnih čakalnih vrst dobimo:

$$t_{h+1} - t_h = \frac{1}{R} \sum_{j=1}^N SENT_j(k_0 + h) \quad (C.40)$$

$$= \frac{1}{R} \sum_{j=1}^N Q_j + \frac{1}{R} \sum_{j=1}^N [DC_j(k_0 + h - 1) - DC_j(k_0 + h)] \quad (C.41)$$

Glede na izraz (C.39) pa lahko zapišemo naslednjo enačbo:

$$t_k - t_1 = \sum_{h=1}^{k-1} (t_{h+1} - t_h) \quad (C.42)$$

$$= \frac{1}{R} (k-1)F + \frac{1}{R} \sum_{h=1}^{k-1} \sum_{j=1}^N [DC_j(k_0 + h - 1) - DC_j(k_0 + h)] \quad (C.43)$$

kjer  $F$  predstavlja velikost okvira algoritma:  $F = \sum_i Q_i$ . Če razvijemo dvojno vsoto iz izraza (C.43), dobimo:

$$\begin{aligned} & \sum_{h=1}^{k-1} \sum_{j=1}^N [DC_j(k_0 + h - 1) - DC_j(k_0 + h)] = \\ & = DC_1(k_0) - DC_1(k_0 + 1) + DC_2(k_0) - DC_2(k_0 + 1) + \dots + DC_N(k_0) - DC_N(k_0 + 1) \\ & + DC_1(k_0 + 1) - DC_1(k_0 + 2) + DC_2(k_0 + 1) - DC_2(k_0 + 2) + \dots + DC_N(k_0 + 1) \\ & - DC_N(k_0 + 2) + DC_1(k_0 + 2) - DC_1(k_0 + 3) + DC_2(k_0 + 2) - DC_2(k_0 + 3) \\ & + \dots + DC_N(k_0 + 2) - DC_N(k_0 + 3) + \dots + DC_1(k_0 + k - 2) - DC_1(k_0 + k - 1) \\ & + DC_2(k_0 + k - 2) - DC_2(k_0 + k - 1) + \dots + DC_N(k_0 + k - 2) - DC_N(k_0 + k - 1) \end{aligned}$$

Večina sumandov v zgornjem izrazu se med seboj izniči in ostane nam samo vsota  $\sum_{j=1}^N [DC_j(k_0) - DC_j(k_0 + k - 1)]$ . Enačbo (C.42) tako lahko zapišemo kot:

$$t_k - t_1 = \frac{k-1}{R} F + \frac{1}{R} \sum_{j=1}^N [DC_j(k_0) - DC_j(k_0 + k - 1)] \quad (C.44)$$

- $(t_k, \beta_i^k)$  : V zadnjem časovnem intervalu so postrežene le čakalne vrste iz podmnožice  $G_\beta$ . Dobimo sledeči izraz:

$$\begin{aligned} \beta_i^k - t_k & = \frac{1}{R} \sum_{j \in G_\beta} SENT_j(k_0 + k) \\ & = \frac{1}{R} \sum_{j \in G_\beta} [Q_j + DC_j(k_0 + k - 1) - DC_j(k_0 + k)] \quad (C.45) \end{aligned}$$

Z združitvijo enačb (C.38), (C.44) in (C.45) dobimo:

$$\beta_i^k - \beta_i = (\beta_i^k - t_k) + (t_k - t_1) + (t_1 - \beta_1) =$$

$$\begin{aligned}
&= \frac{1}{R} \sum_{j \in G_\beta} Q_j + \frac{1}{R} \sum_{j \in G_\beta} [DC_j(k_0 + k - 1) - DC_j(k_0 + k)] + \\
&+ \frac{k-1}{R} F + \frac{1}{R} \sum_{j=1}^N [DC_j(k_0) - DC_j(k_0 + k - 1)] \\
&+ \frac{1}{R} \sum_{j \in G_\alpha} Q_j + \frac{1}{R} \sum_{j \in G_\alpha} [DC_j(k_0 - 1) - DC_j(k_0)] \quad (C.46)
\end{aligned}$$

Ker čakalna vrsta postane aktivna tekom obhoda  $k_0$ , velja:

$$DC_i(k_0) = 0 \quad (C.47)$$

Vsoto kvantov, ki pripadajo čakalnim vrstam iz podmnožic  $G_\alpha$  in  $G_\beta$ , pa lahko enostavneje zapišemo kot:

$$\sum_{j \in G_\beta} Q_j + \sum_{j \in G_\alpha} Q_j = \sum_{\forall j} Q_j - Q_i = F - Q_i \quad (C.48)$$

Zgornji izraz sledi neposredno iz definicije podmnožic  $G_\beta$  in  $G_\alpha$  ter iz dejstva da čakalna vrsta  $i$  ni vključena v nobeno izmed njih.

Najprej poiščimo zgornjo mejo izraza:

$$\begin{aligned}
\sum_{j \in G_\beta} [DC_j(k_0 + k - 1) - DC_j(k_0 + k)] + \sum_{j=1}^N [DC_j(k_0) - DC_j(k_0 + k - 1)] + \\
+ \sum_{j \in G_\alpha} [DC_j(k_0 - 1) - DC_j(k_0)] \quad (C.49)
\end{aligned}$$

Ker lahko množico vseh aktivnih čakalnih vrst razdelimo na podmnožici  $G_\alpha$  in  $G_\beta$  ter čakalno vrsto  $i$ , lahko izraz (C.49) preoblikujemo v:

$$\begin{aligned}
\sum_{j \in G_\beta} [DC_j(k_0 + k - 1) - DC_j(k_0 + k)] + \sum_{j \in G_\beta} [DC_j(k_0) - DC_j(k_0 + k - 1)] + \\
+ \sum_{j \in G_\alpha} [DC_j(k_0) - DC_j(k_0 + k - 1)] + \sum_{j \in G_\alpha} [DC_j(k_0 - 1) - DC_j(k_0)] + \\
+ DC_i(k_0) - DC_i(k_0 + k - 1) \quad (C.50)
\end{aligned}$$

Upoštevajoč (C.47) in s krajšanjem nekaterih sumandov dobimo:

$$\begin{aligned}
\sum_{j \in G_\beta} [DC_j(k_0) - DC_j(k_0 + k)] + \sum_{j \in G_\alpha} [DC_j(k_0 - 1) - DC_j(k_0 + k - 1)] \\
- DC_i(k_0 + k - 1) \quad (C.51)
\end{aligned}$$

Število čakalnih vrst v podmnožicah  $G_\alpha$  in  $G_\beta$  je enako  $N - 1$ . Glede na lemo C.1 v razdelku C.2.1 je lahko maksimalna vrednost posameznega števca deficita enaka  $L_{max} - 1$ . Glede na povedano drži sledeče:

$$\frac{1}{R} \sum_{j \in G_\beta} [DC_j(k_0 + k - 1) - DC_j(k_0 + k)]$$

$$\begin{aligned}
& + \frac{1}{R} \sum_{j=1}^N [DC_j(k_0) - DC_j(k_0 + k - 1)] \\
& + \frac{1}{R} \sum_{j \in G_\alpha} [DC_j(k_0 - 1) - DC_j(k_0)] \\
& \leq \frac{(N-1)(L_{max} - 1)}{R} - \frac{1}{R} DC_i(k_0 + k - 1)
\end{aligned} \tag{C.52}$$

Z uporabo enačb (C.46), (C.47), (C.48) in (C.52) dobimo:

$$\beta_i^k - \beta_i \leq \frac{k-1}{R} F + \frac{F - Q_i}{R} + \frac{(N-1)(L_{max} - 1)}{R} - \frac{1}{R} DC_i(k_0 + k - 1) \tag{C.53}$$

Iz zgornje neenačbe lahko izrazimo  $(k-1)$ :

$$(k-1) \geq \frac{R}{F} (\beta_i^k - \beta_i) - \frac{F - Q_i}{F} - \frac{1}{F} (N-1)(L_{max} - 1) + \frac{1}{F} DC_i(k_0 + k - 1) \tag{C.54}$$

Ne pozabimo, da čakalna vrsta  $i$  dobi prvo priložnost strežbe v obhodu  $(k_0 + 1)$  in da ima v časovnem intervalu  $(\beta_i, \beta_i^k)$   $k - 1$  priložnosti strežbe. Spomnimo se tudi izraza (C.37), kjer z  $DC_i(m)$  označujemo vrednost spremenljivke  $DC_i$  ob koncu  $m$ -tega obhoda.

S seštevanjem izraza (C.37) preko vseh  $k - 1$  obhodov v časovnem intervalu  $(\beta_i, \beta_i^k)$  dobimo:

$$SENT_i(\beta_i, \beta_i^k) = \sum_{m=1}^{k-1} SENT_i(m) \tag{C.55}$$

$$SENT_i(\beta_i, \beta_i^k) = (k-1)Q_i + \sum_{m=1}^{k-1} [DC_i(k_0 + m - 1) - DC_i(k_0 + m)] \tag{C.56}$$

Zadnja vsota izraza (C.55) se skrči na razliko njenega prvega in zadnjega sumanda. Dobimo enačbo:

$$SENT_i(\beta_i, \beta_i^k) = (k-1)Q_i + [DC_i(k_0) - DC_i(k_0 + k - 1)]. \tag{C.57}$$

Z uporabo pogoja (C.47) imamo:

$$SENT_i(\beta_i, \beta_i^k) = (k-1)Q_i - DC_i(k-1 + k_0). \tag{C.58}$$

Z vstavitvijo že izračunane meje maksimalne vrednosti za  $(k-1)$  v (C.54) dobimo spodnjo neenačbo:

$$\begin{aligned}
SENT_i(\beta_i, \beta_i^k) & \geq \frac{Q_i R}{F} (\beta_i^k - \beta_i) - \frac{Q_i (F - Q_i)}{F} - \frac{Q_i}{F} (N-1)(L_{max} - 1) \\
& + \frac{Q_i}{F} DC_i(k_0 + k - 1) - DC_i(k_0 + k - 1)
\end{aligned} \tag{C.59}$$

Ker velja  $DC_i(m) \leq L_{max} - 1$ , lahko zapišemo:

$$\begin{aligned}
SENT_i(\beta_i, \beta_i^k) & \geq \frac{Q_i R}{F} \left[ (\beta_i^k - \beta_i) - \frac{F - Q_i}{R} - \frac{(N-1)(L_{max} - 1)}{R} \right] \\
& + \frac{L_{max} - 1}{R} \left( 1 - \frac{F}{Q_i} \right)
\end{aligned} \tag{C.60}$$

Zgradba neenačbe (C.60) je podobna zgradbi neenačbe v definiciji algoritemske zakasnitve za aktivno obdobje v (C.33). Z uporabo spremenljivk, uporabljenih v tem dokazu, lahko neenačbo (C.33) zapišemo kot:

$$SENT_i(\beta_i, \beta_i^k) \geq \max\{0, r_i(\beta_i^k - \beta_i - \theta'_i)\}$$

Ker velja neenačba  $r_i \leq \frac{Q_i}{F}R$ , lahko faktor  $\frac{Q_i}{F}R$  v izrazu (C.60) nadomestimo z  $r_i$  le v primeru, da je celoten izraz v oglatih oklepajih večji ali enak nič. Ker ne poznamo točne vrednosti izraza v oglatih oklepajih, moramo za pravilno nadaljevanje razviti desno stran neenačbe (C.60).

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq \frac{Q_i R}{F}(\beta_i^k - \beta_i) - Q_i + \frac{Q_i R}{F} \frac{Q_i}{R} - \frac{Q_i N(L_{max} - 1)}{F} \\ &+ \frac{Q_i(L_{max} - 1)}{F} + \frac{Q_i(L_{max} - 1)}{F} - (L_{max} - 1) \end{aligned} \quad (C.61)$$

Zamenjavo  $\frac{Q_i}{F}R$  z  $r_i$  lahko izvedemo le pri členih, ki so večji ali enaki nič. V tem primeru znižujemo vrednost desne strani neenačbe, ki tako še vedno drži. Če pa zamenjavo izvedemo pri negativnih členih, pa s tem povečujemo vrednost desne strani neenačbi, ki s tem ne velja več. Z uporabo omenjene metodologije dobimo neenačbo:

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq r_i(\beta_i^k - \beta_i) - Q_i + Q_i \frac{r_i}{R} - \frac{Q_i N}{F}(L_{max} - 1) \\ &+ \frac{r_i}{R}(L_{max} - 1) + (L_{max} - 1) \left( \frac{r_i}{R} - 1 \right) \end{aligned} \quad (C.62)$$

Neenačbo preoblikujemo v:

$$SENT_i(\beta_i, \beta_i^k) \geq r_i \left[ (\beta_i^k - \beta_i) - Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) - (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \right] \quad (C.63)$$

V zgornjih izvajanjih smo uporabili dejstvo, da meja algoritemske zakasnitve doseže svojo največjo vrednost, ko je  $DC_i(k-1+k_0)$  enak  $L_{max}-1$ . V tem primeru je namreč časovni interval  $(\beta_i, \beta_i^k)$ , za določen  $k$ , najdaljši.

Pri primerjavi neenačb (C.33) in (C.63) opazimo, da mora za zadovoljitev pogoja v Definiciji C.12 veljati:

$$\theta'_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (C.64)$$

kjer je  $\theta'_i$  najmanjša nenegativna konstanta, ki zadovolji neenačbi (C.33). In končno preko neenačbe (C.34) dobimo:

$$\theta_i \leq \theta'_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (C.65)$$

S tem smo dokazali trditve teorema. ■

Z drugačnim pristopom k izpeljavi zgornje meje algoritemske zakasnitve bi morali dobiti enak rezultat. V nadaljevanju zato izpeljemo zgornjo mejo algoritemske zakasnitve za SCDRR še z drugačnim pristopom in dokažemo, da je rezultat enak tistemu, ki smo ga dobili v teoremu C.2.

**Teorem C.3** Naj bo  $\beta_i$  začetek aktivnega obdobja čakalne vrste  $i$  za odpravnik SCDRR. Za vsak časovni trenutek  $t$  tekom aktivnega obdobja  $(\beta_i, t)$  drži naslednja neenačba:

$$\theta'_i \leq Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) + (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \quad (\text{C.66})$$

kjer  $\theta'_i$  označuje algoritemsko zakasnitev, izračunano za obravnavano aktivno obdobje.

**Dokaz - Teorem C.3** Vse uporabljene označbe so enake kot v teoremu C.2, sam dokaz pa večidel sledi postopku dokaza istega teorema. Za lažje sledenje dokazu ponovimo le najpomembnejša dejstva. Opazujemo aktivno obdobje čakalne vrste  $i$   $(\beta_i, t)$ . Recimo, da je v tem časovnem intervalu čakalna vrsta  $i$  deležna  $k$  obhodov. S  $t_h$  označimo časovne trenutke, ob katerih se začne  $h$ -ti obhod. Zanima nas torej časovni interval  $(\beta_i, \beta_i^k)$ . Zaradi lažjega izračuna za začetek vzemimo časovni interval  $(t_1, t_{k+1})$ , interval v katerem odpravnik opravi  $k$  celotnih obhodov (glej sliko C.2). V tem času čakalna vrsta  $i$  opravi storitev, ki je enaka:

$$SENT_i(t_1, t_{k+1}) = kQ_i - DC_i(k_0 + k) \quad (\text{C.67})$$

o čemer se lahko prepričamo, če seštejemo izraz (C.37) preko vseh  $k$  obhodov, in hkrati upoštevamo še pogoj (C.47). Argument  $k_0$  v enačbi nastopa, ker smo predpostavili, da je v času  $\beta_i^k$ , ko je čakalna vrsta  $i$  postala aktivna, v odpravniku tekel obhod  $k_0$ .

Za izračun vrednosti  $SENT_i(\beta_i, \beta_i^k)$  delno sledimo dokazu v teoremu C.2. Začnimo pa z izračunom časa, ki poteče tekom obhoda  $h$ . Za vsak interval  $(t_h, t_{h+1})$  lahko zapišemo:

$$t_{h+1} - t_h = \frac{1}{R} \sum_{j=1}^N SENT_j(k_0 + h) \quad (\text{C.68})$$

$$= \frac{1}{R} \sum_{j=1}^N Q_j + \frac{1}{R} \sum_{j=1}^N [DC_j(k_0 + h - 1) - DC_j(k_0 + h)] \quad (\text{C.69})$$

S seštevanjem zgornjega izraza preko vseh  $k$  obhodov dobimo:

$$t_{k+1} - t_1 = \frac{kF}{R} + \frac{1}{R} \sum_{j=1}^N DC_j(k_0) - \frac{1}{R} \sum_{j=1}^N DC_j(k_0 + k) \quad (\text{C.70})$$

kjer  $F$  predstavlja velikost okvira odpravnika:  $F = \sum_i Q_i$ . Z uporabo leme C.1 v razdelku C.2.1 lahko izrazimo najslabši možni primer; dobimo naslednjo neenačbo:

$$t_{k+1} - t_1 \leq k \frac{F}{R} + \frac{1}{R} (N - 1) (L_{max} - 1) \quad (\text{C.71})$$

ali ekvivalentno,

$$k \geq \frac{(t_{k+1} - t_1)R}{F} - (N - 1) \frac{L_{max} - 1}{F} \quad (\text{C.72})$$

Z upoštevanjem neenačbe (C.72) in enačbe (C.67) dobimo:

$$\begin{aligned} SENT_i(t_1, t_{k+1}) &\geq Q_i \left( \frac{(t_{k+1} - t_1)R}{F} - (N-1) \frac{L_{max} - 1}{F} \right) - DC_i(k_0 + k) \\ &\geq Q_i \frac{R}{F} (t_{k+1} - t_1) - \frac{Q_i}{F} (N-1)(L_{max} - 1) - DC_i(k_0 + k) \end{aligned} \quad (C.73)$$

V najslabšem možnem primeru čakalna vrsta  $i$  v vsakem izmed  $k$  obhodov pride na vrsto zadnja. V tem primeru časovna trenutka  $\beta_i$  and  $t_1$  sovpadata. Za izračun storitve čakalne vrste  $i$  v časovnem intervalu  $(\beta_i, \beta_i^k)$  moramo v zgornjih izrazih časovni trenutek  $t_{k+1}$  (zaključek obhoda  $k$ ) nadomestiti s časovnim trenutkom  $\beta_i^k$ , ko v  $k$ -tem obhodu čakalna vrsta  $i$  dobi priložnost strežbe. Razlika v dolžini časovnih intervalov  $(t_1, t_{k+1})$  in  $(\beta_i, \beta_i^k)$ , je le v času, ki ga v  $k$ -tem obhodu čakalna vrsta  $i$  porabi za strežbo. Velja, da je  $t_{k+1} \geq \beta_i^k$ , zato moramo od že izračunanega izraza (C.73) odšteti storitev čakalne vrste  $i$  v časovnem intervalu  $(\beta_i^k, t_{k+1})$ . Z upoštevanjem vsega naštetega dobimo sledečo neenačbo:

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq Q_i \frac{R}{F} (t_{k+1} - t_1) - \frac{Q_i}{F} (N-1)(L_{max} - 1) - DC_i(k_0 + k) \\ &\quad - Q_i - DC_i(k_0 + k - 1) + DC_i(k_0 + k) \\ &\geq Q_i \frac{R}{F} (\beta_i^k + \tau - \beta_i) - \frac{Q_i}{F} (N-1)(L_{max} - 1) - DC_i(k_0 + k) \\ &\quad - Q_i - DC_i(k_0 + k - 1) + DC_i(k_0 + k) \end{aligned} \quad (C.74)$$

kjer je  $\tau$  časovna razlika med  $t_{k+1}$  in  $\beta_i^k$ . Upoštevaajoč enačbo (C.37) lahko zapišemo sledeče:

$$\tau = \frac{SENT_i(k_0 + k)}{R} = Q_i + DC_i(k_0 + k - 1) - DC_i(k_0 + k) \quad (C.75)$$

Omenimo naj, da enačba (C.71) že upošteva pogoj:  $DC_i(k_0 + k) = 0$ , ki je bil identificiran kot najslabši možni primer. Z uporabo enačbe (C.75) v neenačbi (C.74) dobimo:

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq Q_i \frac{R}{F} (\beta_i^k - \beta_i) + \frac{Q_i}{F} Q_i + \frac{Q_i}{F} DC_i(k_0 + k - 1) \\ &\quad - \frac{Q_i}{F} N(L_{max} - 1) + \frac{Q_i}{F} (L_{max} - 1) - DC_i(k_0 + k - 1) - Q_i \end{aligned} \quad (C.76)$$

Sedaj z uporabo enačbe (C.8) in zveze  $\frac{Q_i}{F} \geq \frac{r_i}{R}$  dobimo:

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq Q_i \frac{R}{F} (\beta_i^k - \beta_i) + \frac{Q_i}{F} Q_i - (L_{max} - 1) \left(1 - \frac{Q_i}{F}\right) \\ &\quad - \frac{Q_i}{F} N(L_{max} - 1) + \frac{Q_i}{F} (L_{max} - 1) - Q_i \\ &\geq r_i (\beta_i^k - \beta_i) - Q_i \left(1 - \frac{r_i}{R}\right) - (L_{max} - 1) + \frac{r_i}{R} (L_{max} - 1) \\ &\quad - \frac{Q_i}{F} N(L_{max} - 1) + \frac{r_i}{R} (L_{max} - 1) \end{aligned} \quad (C.77)$$



Neposredno iz prejšnje neenačbe lahko zapišemo:

$$\begin{aligned} SENT_i(\beta_i, \beta_i^k) &\geq r_i \left[ (\beta_i^k - \beta_i) - \frac{Q_i}{r_i} \left( 1 - \frac{r_i}{R} \right) - (L_{max} - 1) \left( \frac{Q_i}{F} N + 1 - 2 \frac{r_i}{R} \right) \right] \\ &\geq r_i \left[ (\beta_i^k - \beta_i) - Q_i \left( \frac{1}{r_i} - \frac{1}{R} \right) - (L_{max} - 1) \left( \frac{Q_i}{F r_i} N + \frac{1}{r_i} - \frac{2}{R} \right) \right] \end{aligned} \quad (\text{C.78})$$

S tem smo dokazali teorem glede na definicije algoritemske zakasnitve  $\mathcal{LR}$  odpravnikov, podane v (C.33) in (C.34). ■

**Komentar** Zgornji teorem temelji na ideji izpeljave meje algoritemske zakasnitve za odpravnik DRR s strani Stiliadisa in Varme v [98]. Potrebno pa je opozoriti na napake, ki so prisotne v njuni izpeljavi. Najprej sta v izrazu podobnemu (C.71) predpostavila, da velja neenakost  $DC_i \leq Q_i$  ( $\forall_i$ ). Ta predpostavka je napačna, saj smo z lemo C.1 dokazali, da velja neenakost  $DC_i \leq L_{max} - 1$ . Potem pa na mestu, kjer sta ovrednotila storitev v časovnem intervalu  $(t_1, t_{k+1})$ , nista uporabila pravilne omejitve vrednosti  $DC_i$ , dokazane z lemo C.1. V našem izvajanju meje algoritemske zakasnitve pa smo upoštevali vse zgoraj naštetе omejitve.

## C.4 Zahtevnost SCDRR pri eksponentni porazdelitvi dolžin paketov

Z analizo splošnega primera v razdelku 6.5 smo prišli do izrazov, ki jih uporabimo pri analizi posebnih primerov. Prvi tak primer je analiza pri eksponentni verjetnostni porazdelitvi dolžin paketov.

### C.4.1 Definicije

Pri nadaljnji analizi uporabljamo spremenljivke in definicije iz 6.5.1, pri čemer lahko za nekatere izmed njih zapišemo izraze, ki ustrezajo eksponentni verjetnostni porazdelitvi dolžin paketov.

$$\begin{array}{lll} f(l) = \lambda e^{-\lambda l} & L_{min} = 0 & \bar{L} = \frac{1}{\lambda} \\ F(l) = 1 - e^{-\lambda l} & L_{max} = \infty & N_m = \infty \end{array}$$

Če upoštevamo še robne pogoje iz 6.5.2, opazimo, da se pri eksponentni porazdelitvi, glede na vrednost kvanta  $Q$ , vedno nahajamo v podkritičnem režimu delovanja DRR. Ker je  $L_{max} = \infty$  namreč vedno velja, da je  $Q$  manjši od  $L_{max}$ .

Pri analizi računske zahtevnosti pri eksponentni verjetnostni porazdelitvi dolžin paketov spremenljivka  $\lambda$  določa recipročno vrednost povprečne dolžine paketov v čakalni vrsti.

### C.4.2 Izračun števila potrebnih operacij

Za izračun uporabimo kar enačbe od (6.48) do (6.54), pri tem pa upoštevamo zgornje lastnosti in enačbe za eksponentno verjetnostno porazdelitev dolžin paketov. Ker je največja možna dolžina paketa  $L_{max}$  neskončna in vrednost kvanta  $Q$  končna vrednost, so tudi vse vrednosti  $N_m$  neskončne.

$$N = N_m = \infty \quad (\forall m \geq 0) \quad (\text{C.79})$$

Analizo zopet pričnemo pri prvem paketu v čakalni vrsti, torej pri  $m = 1$ . Veljajo enaki robni pogoji kot pri splošni analizi. Začetna vrednost  $DC$  ob času  $t_0$  je enaka nič  $DC = DC_0 = 0$ , vrednost kvanta  $Q$  pa se nahaja med najmanjšo in največjo možno dolžino paketa  $0 \leq Q < \infty$ . Najprej izrazimo  $p_{1,n}$  pri  $n = 1, 2, \dots, \infty$ .

$$\begin{aligned} p_{1,n} &= P((n-1)Q < l \leq nQ) = F(nQ) - F((n-1)Q) \\ &= 1 - e^{-\lambda nQ} - 1 + e^{-\lambda(n-1)Q} = e^{-\lambda(n-1)Q} - e^{-\lambda nQ} \\ &= e^{-\lambda nQ} (e^{\lambda Q} - 1) \end{aligned} \quad (\text{C.80})$$

Preden nadaljujemo, preverimo ali je verjetnost gotovega dogodka  $S_1$  za podane vrednosti  $p_{1,n}$  res enaka 1. Če uporabimo izraz (6.28) pri  $N_m = \infty$ , lahko zapišemo:

$$S_1 = \sum_{n=1}^{\infty} p_{1,n} = \sum_{n=1}^{\infty} e^{-\lambda nQ} (e^{\lambda Q} - 1) = (e^{\lambda Q} - 1) \sum_{n=1}^{\infty} e^{-\lambda nQ} \quad (\text{C.81})$$

Iz teorije neskončnih geometričnih vrst poznamo naslednji zvezi:

$$\sum_{n=0}^{\infty} x^n = \begin{cases} \frac{1}{1-x} & |x| < 1 \\ \text{divergira} & |x| \geq 1 \end{cases} \quad (\text{C.82})$$

$$\sum_{n=1}^{\infty} x^n = \begin{cases} \frac{x}{1-x} & |x| < 1 \\ \text{divergira} & |x| \geq 1 \end{cases} \quad (\text{C.83})$$

Preverimo ali neskončna vrsta v našem primeru konvergira. Spremenljivki  $x$  iz zgornjih izrazov v enačbi (C.81) ustreza izraz  $e^{-\lambda Q}$ , zato za konvergenco mora veljati pogoj  $|e^{-\lambda Q}| < 1$ . Ker ta pogoj lahko zapišemo tudi kot  $-\lambda Q < \ln 1 = 0$  in ker sta tako  $\lambda$  kot  $Q$  pozitivni realni števili, je zahtevani pogoj vedno izpolnjen in vrsta konvergira za vse njune kombinacije.  $S_1$  je torej:

$$S_1 = \sum_{n=1}^{\infty} p_{1,n} = (e^{\lambda Q} - 1) \frac{e^{-\lambda Q}}{1 - e^{-\lambda Q}} = \frac{1 - e^{-\lambda Q}}{1 - e^{-\lambda Q}} = 1 \quad (\text{C.84})$$

Sedaj lahko izračunamo povprečno število operacij  $O_1$  potrebnih za oddajo prvega paketa:

$$O_1 = \sum_{n=1}^{\infty} n p_{1,n} \quad (\text{C.85})$$

Zopet se poslužimo teorije neskončnih vrst, ki nam da rezultat:

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(x-1)^2} \quad (\text{C.86})$$

Za preverjanje konvergence zgornje vrste uporabimo *D'Alembertov kriterij*, ki velja za vrste s pozitivnimi členi, čemur naša vrsta zadošča. Kriterij se glasi: "Če imamo neskončno vrsto s členi  $a_1 + a_2 + \dots + a_n + \dots$ , velja, da vrsta konvergira, če so od nekega mesta naprej vsa razmerja  $\frac{a_{n+1}}{a_n} < 1$  ali zapisano drugače  $\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} < 1$ ". Za naš primer velja:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} &= \lim_{n \rightarrow \infty} \frac{(n+1)e^{-\lambda(n+1)Q}}{ne^{-\lambda nQ}} = \lim_{n \rightarrow \infty} \frac{(n+1)e^{-\lambda Q}}{n} \\ &= \lim_{n \rightarrow \infty} \left( e^{-\lambda Q} + \frac{e^{-\lambda Q}}{n} \right) = \lim_{n \rightarrow \infty} e^{-\lambda Q} + \lim_{n \rightarrow \infty} \left( \frac{e^{-\lambda Q}}{n} \right) \\ &= e^{-\lambda Q} < 1 \end{aligned} \quad (\text{C.87})$$

Ker tudi ta vrsta zadošča konvergenčnim pogojem, izračunajmo njeno vrednost, ki nam ob upoštevanju izraza (C.86) da:

$$\begin{aligned} O_1 &= \sum_{n=1}^{\infty} np_{1,n} = \sum_{n=1}^{\infty} ne^{-\lambda nQ} (e^{\lambda Q} - 1) = (e^{\lambda Q} - 1) \sum_{n=1}^{\infty} ne^{-\lambda nQ} \\ &= (e^{\lambda Q} - 1) \frac{e^{-\lambda Q}}{(e^{-\lambda Q} - 1)^2} = (e^{\lambda Q} - 1) \frac{e^{-\lambda Q}}{e^{-2\lambda Q} (e^{\lambda Q} - 1)^2} \\ &= \frac{e^{\lambda Q}}{e^{\lambda Q} - 1} \end{aligned} \quad (\text{C.88})$$

Za izračun števila operacij  $O_2$ , potrebnih za odpravo drugega paketa, zopet potrebujemo povprečno vrednost  $DC$ , ki ostane po odpravi prvega paketa in jo označimo z  $DC_1$ . Za njen izračun pa moramo najprej izračunati vrednosti  $L_{1,n}$  pri  $n = 1, 2, \dots, \infty$ . V enačbo (6.52) za splošni primer vnesemo izraze, ki veljajo za eksponentno verjetnostno porazdelitev dolžin paketov ter upoštevamo še  $m = 1$  in  $n = 1, 2, \dots, \infty$ . Sedaj lahko izračunamo nedoločene integrale dobljenih izrazov in vanje vnesemo podane meje.

$$\begin{aligned} L_{1,n} &= \frac{\int_{(n-1)Q}^{nQ} lf(l)dl}{\int_{(n-1)Q}^{nQ} f(l)dl} = \frac{\int_{(n-1)Q}^{nQ} l\lambda e^{-\lambda l} dl}{\int_{(n-1)Q}^{nQ} \lambda e^{-\lambda l} dl} = \frac{-\lambda e^{-\lambda l} \left( \frac{\lambda l + 1}{\lambda^2} \right) \Big|_{(n-1)Q}^{nQ}}{\frac{-\lambda e^{-\lambda l}}{\lambda} \Big|_{(n-1)Q}^{nQ}} \\ &= \frac{e^{-\lambda(n-1)Q} \left( \frac{\lambda(n-1)Q + 1}{\lambda} \right) - e^{-\lambda nQ} \left( \frac{\lambda nQ + 1}{\lambda} \right)}{e^{-\lambda(n-1)Q} - e^{-\lambda nQ}} \end{aligned}$$

V števcu in imenovalcu izpostavimo in okrajšamo izraz  $e^{-\lambda nQ}$ . Zatem števec in imenovalec pomnožimo z  $\lambda$ , nekoliko preuredimo izraze v števcu in na koncu izraz razdelimo na vsoto enostavnih ulomkov.

$$\begin{aligned}
L_{1,n} &= \frac{e^{\lambda Q} \left( \frac{\lambda(n-1)Q+1}{\lambda} \right) - \left( \frac{\lambda n Q + 1}{\lambda} \right)}{e^{\lambda Q} - 1} = \frac{e^{\lambda Q}(\lambda n Q - \lambda Q + 1) - (\lambda n Q + 1)}{\lambda(e^{\lambda Q} - 1)} \\
&= \frac{(e^{\lambda Q} - 1)(\lambda n Q + 1) - \lambda Q e^{\lambda Q}}{\lambda(e^{\lambda Q} - 1)} = \frac{\lambda n Q + 1}{\lambda} - \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} \\
&= nQ + \frac{1}{\lambda} - \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} \tag{C.89}
\end{aligned}$$

Ob upoštevanju enačbe (6.54) ter robnih pogojev  $DC_0 = 0$  in  $n = 1, 2, \dots, \infty$ , lahko za naš primer zapišemo:

$$DC_1 = \sum_{n=1}^{\infty} p_{1,n}(nQ - L_{1,n}) \tag{C.90}$$

Če sedaj v zgornji izraz vstavimo za naš primer izračunane izraze za  $L_{1,n}$ , dobimo:

$$\begin{aligned}
DC_1 &= \sum_{n=1}^{\infty} p_{1,n} \left( nQ - nQ - \frac{1}{\lambda} + \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} \right) \\
&= \sum_{n=1}^{\infty} p_{1,n} \left( \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \right)
\end{aligned}$$

Ker je izraz v oklepaju konstanta, ga lahko iz vsote izpostavimo, za izraz  $\sum_{n=1}^{\infty} p_{1,n}$  pa smo že z (C.84) dokazali, da je enak ena. Zato za  $DC_1$  dobimo preprost izraz (C.91).

$$DC_1 = \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \tag{C.91}$$

Za drugi in vse nadaljnje pakete v čakalni vrsti moramo upoštevati, da je  $DC \geq 0$ . Če za osnovo zopet vzamemo enačbe splošnega primera ter jih priredimo eksponentni verjetnosti porazdelitve dolžin paketov, dobimo izraze za  $p_{m,n}$ ,  $L_{m,n}$ ,  $DC_m$  in  $O_m$ . Najprej izrazimo vrednosti  $p_{m,0}$ ,  $L_{m,0}$ , ki veljajo za  $m > 1$ . Pri  $m = 1$  namreč ti dve vrednosti nista definirani.

$$p_{m,0} = P(l \leq DC_{m-1}) = F(DC_{m-1}) = 1 - e^{-\lambda DC_{m-1}} \quad \text{pri } m > 1 \tag{C.92}$$

$$\begin{aligned}
L_{m,0} &= \frac{\int_0^{DC_{m-1}} l \lambda e^{-\lambda l} dl}{\int_0^{DC_{m-1}} \lambda e^{-\lambda l} dl} = \frac{-\lambda e^{-\lambda l} \left( \frac{\lambda l + 1}{\lambda^2} \right) \Big|_0^{DC_{m-1}}}{\frac{-\lambda e^{-\lambda l}}{\lambda} \Big|_0^{DC_{m-1}}} \\
&= \frac{\frac{1}{\lambda} - e^{-\lambda DC_{m-1}} \left( \frac{\lambda DC_{m-1} + 1}{\lambda} \right)}{1 - e^{-\lambda DC_{m-1}}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\lambda} \frac{\left(1 - e^{-\lambda DC_{m-1}}\right) - DC_{m-1} e^{-\lambda DC_{m-1}}}{1 - e^{-\lambda DC_{m-1}}} \\
&= \frac{1}{\lambda} - \frac{DC_{m-1} e^{-\lambda DC_{m-1}}}{1 - e^{-\lambda DC_{m-1}}} \\
&= \frac{1}{\lambda} + DC_{m-1} - \frac{DC_{m-1}}{1 - e^{-\lambda DC_{m-1}}} \quad \text{pri } m > 1 \quad (\text{C.93})
\end{aligned}$$

Vidimo, da sta izraza  $p_{m,0}$ ,  $L_{m,0}$  na moč podobna izrazoma  $p_{1,1}$ ,  $L_{1,1}$ , le spremenljivko  $Q$  nadomešča spremenljivka  $DC_{m-1}$ . Sedaj pa zapišimo še  $p_{m,n}$ ,  $L_{m,n}$  pri  $m = 1, 2, 3, \dots, M$  in  $n = 1, 2, 3, \dots, \infty$ .

$$\begin{aligned}
p_{m,n} &= P\left((n-1)Q + DC_{m-1} < l \leq nQ + DC_{m-1}\right) \\
&= F(nQ + DC_{m-1}) - F\left((n-1)Q + DC_{m-1}\right) \\
&= e^{-\lambda nQ} e^{-\lambda DC_{m-1}} \left(e^{\lambda Q} - 1\right) \quad \text{pri } n > 0 \quad (\text{C.94})
\end{aligned}$$

$$\begin{aligned}
L_{m,n} &= \frac{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} l f(l) dl}{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} f(l) dl} \\
&= nQ + DC_{m-1} + \frac{1}{\lambda} - \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} \quad \text{pri } n > 0 \quad (\text{C.95})
\end{aligned}$$

Preden nadaljujemo, preverimo še vrednost gotovega dogodka  $S_m$ , ki mora biti enak ena. Z uporabo enačb C.82 in C.83 dobimo:

$$\begin{aligned}
S_m &= \sum_{n=0}^{\infty} p_{m,n} = p_{m,0} + \sum_{n=1}^{\infty} p_{m,n} \\
&= 1 - e^{-\lambda DC_{m-1}} + \sum_{n=1}^{\infty} e^{-\lambda DC_{m-1}} e^{-\lambda nQ} \left(e^{\lambda Q} - 1\right) \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} \left(e^{\lambda Q} - 1\right) \sum_{n=1}^{\infty} e^{-\lambda nQ} \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} \left(e^{\lambda Q} - 1\right) \frac{e^{-\lambda Q}}{1 - e^{-\lambda Q}} \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} \frac{1 - e^{-\lambda Q}}{1 - e^{-\lambda Q}} \\
&= 1 \quad (\text{C.96})
\end{aligned}$$

Pri izračunu povprečnega števila operacij za oddajo  $m$ -tega paketa moramo upoštevati vrednosti  $DC_{m-1}$  in  $p_{m,0}$ , tako da se izpeljava enačbe (C.88), ki velja za  $O_1$ , nekoliko spremeni. Če upoštevamo omenjene spremembe, splošni izraz za izračun  $O_m$  in izraz (C.86), dobimo:

$$\begin{aligned}
O_m &= p_{m,0} + \sum_{n=1}^{\infty} n p_{m,n} \\
&= 1 - e^{-\lambda DC_{m-1}} + \sum_{n=1}^{\infty} n e^{-\lambda DC_{m-1}} e^{-\lambda n Q} (e^{\lambda Q} - 1) \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} (e^{\lambda Q} - 1) \sum_{n=1}^{\infty} n e^{-\lambda n Q} \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} (e^{\lambda Q} - 1) \frac{e^{-\lambda Q}}{(e^{-\lambda Q} - 1)^2} \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} (e^{\lambda Q} - 1) \frac{e^{-\lambda Q}}{e^{-2\lambda Q} (e^{\lambda Q} - 1)^2} \\
&= 1 - e^{-\lambda DC_{m-1}} + e^{-\lambda DC_{m-1}} \frac{e^{\lambda Q}}{e^{\lambda Q} - 1} \\
&= \frac{e^{-\lambda DC_{m-1}} + e^{\lambda Q} - 1}{e^{\lambda Q} - 1} \\
&= 1 + \frac{e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \tag{C.97}
\end{aligned}$$

Vidimo, da je predzadnji izraz v (C.97) podoben izrazu (C.88). Razlikuje se samo v členu  $e^{-\lambda DC_{m-1}} - 1$  v števcu. Če vzamemo  $DC_{m-1} = 0$ , pa izraza postaneta enaka. Izračun  $O_m$  torej velja tudi pri  $m = 1$ , saj vemo, da je po definiciji  $DC_0 = 0$ . Na koncu izračunajmo še vrednosti  $DC_m$ . V splošno enačbo za  $DC_m$  (6.54) vnesemo že izračunane izraze za eksponentno porazdelitev. Tako pri  $m = 1, 2, 3, \dots, M$  in  $n = 0, 1, 2, \dots, \infty$  dobimo:

$$dc_{m,n} = nQ + DC_{m-1} - L_{m,n} \tag{C.98}$$

$$\begin{aligned}
DC_m &= p_{m,0} dc_{m,0} + p_{m,1} dc_{m,1} + \dots + p_{m,n} dc_{m,n} + \dots \\
&= p_{m,0} (DC_{m-1} - L_{m,0}) + \sum_{n=1}^{\infty} p_{m,n} (nQ + DC_{m-1} - L_{m,n}) \\
&= (1 - e^{-\lambda DC_{m-1}}) \left( DC_{m-1} - \frac{1}{\lambda} - DC_{m-1} + \frac{DC_{m-1}}{1 - e^{-\lambda DC_{m-1}}} \right) \\
&+ \sum_{n=1}^{\infty} p_{m,n} \left( nQ + DC_{m-1} - nQ - DC_{m-1} - \frac{1}{\lambda} + \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} \right) \\
&= (1 - e^{-\lambda DC_{m-1}}) \left( \frac{DC_{m-1}}{1 - e^{-\lambda DC_{m-1}}} - \frac{1}{\lambda} \right) + \sum_{n=1}^{\infty} p_{m,n} \left( \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \right) \\
&= DC_{m-1} - \frac{1}{\lambda} + \frac{1}{\lambda} e^{-\lambda DC_{m-1}} + \left( \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \right) \sum_{n=1}^{\infty} p_{m,n} \tag{C.99}
\end{aligned}$$

Ker velja  $\sum_{n=0}^{\infty} p_{m,n} = 1$ , lahko zadnjo vsoto  $\sum_{n=1}^{\infty} p_{m,n}$  v zgornjem izrazu izrazimo tudi kot:

$$\sum_{n=1}^{\infty} p_{m,n} = \sum_{n=0}^{\infty} p_{m,n} - p_{m,0} = 1 - \left(1 - e^{-\lambda DC_{m-1}}\right) = e^{-\lambda DC_{m-1}} \quad (\text{C.100})$$

tako za  $DC_m$  dobimo izraz:

$$\begin{aligned} DC_m &= DC_{m-1} - \frac{1}{\lambda} + \frac{1}{\lambda} e^{-\lambda DC_{m-1}} + \left( \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \right) e^{-\lambda DC_{m-1}} \\ &= DC_{m-1} - \frac{1}{\lambda} + \frac{1}{\lambda} e^{-\lambda DC_{m-1}} + \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} e^{-\lambda DC_{m-1}} - \frac{1}{\lambda} e^{-\lambda DC_{m-1}} \\ &= DC_{m-1} - \frac{1}{\lambda} + \frac{Q e^{\lambda Q} e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \end{aligned} \quad (\text{C.101})$$

Kot vidimo, je potrebno vrednosti  $DC_m$  računati iterativno, saj so odvisne od  $DC_{m-1}$ . Ker so tudi vse ostale spremenljivke odvisne od  $DC_{m-1}$ , tudi njih računamo iterativno.

### C.4.3 Zbir rezultatov za eksponentno porazdelitev dolžin paketov

Zapišimo še končne rezultate za eksponentno porazdelitev dolžin paketov. Pri tem moramo upoštevati definicije spremenljivk iz razdelka C.4.1, začetni pogoj  $DC_0 = 0$  ter zalogo vrednosti indeksov  $m = 1, 2, 3, \dots, M$  in  $n = 0, 1, 2, \dots, \infty$ . Tako dobimo:

$$DC_m = DC_{m-1} - \frac{1}{\lambda} + \frac{Q e^{\lambda Q} e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \quad (\text{C.102})$$

$$p_{m,n} = \begin{cases} 1 - e^{-\lambda DC_{m-1}} & n = 0 \\ e^{-\lambda n Q} e^{-\lambda DC_{m-1}} (e^{\lambda Q} - 1) & n > 0 \end{cases} \quad (\text{C.103})$$

$$L_{m,n} = \begin{cases} 0 & m = 1, n = 0 \\ \frac{1}{\lambda} + DC_{m-1} - \frac{DC_{m-1}}{1 - e^{-\lambda DC_{m-1}}} & m > 1, n = 0 \\ nQ + DC_{m-1} + \frac{1}{\lambda} - \frac{Q e^{\lambda Q}}{e^{\lambda Q} - 1} & m > 1, n > 0 \end{cases} \quad (\text{C.104})$$

$$O_m = 1 + \frac{e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1} \quad (\text{C.105})$$

### C.4.4 Primer

Poglejmo si primer izračuna nekaterih izpeljanih vrednosti za eksponentno porazdelitev dolžin paketov pri naslednjih pogojih:

$$Q = 1000 \qquad \lambda = 0.001 \qquad \bar{L} = \frac{1}{\lambda} = 1000$$

V našem primeru je vrednost kvanta  $Q$  kar enaka povprečni dolžini paketov. S tem želimo doseči, da bo večina paketov v čakalni vrsti za odpravo potrebovala samo

eno operacijo. Pri eksponentni porazdelitvi je namreč verjetnost, da bo paket krajši od povprečne dolžine, večja, kot da bo paket daljši od povprečne dolžine. Najprej si pogledjmo izračunane vrednosti  $p_{m,n}$  in  $L_{m,n}$ , ki so zapisane v tabelah C.1 in C.2.

$m \setminus n$	0	1	2	3	4	5
1	0	0,6321	0,2325	0,0855	0,0315	0,0116
2	0,4412	0,3532	0,1299	0,0478	0,0176	0,0065
3	0,3725	0,3967	0,1459	0,0537	0,0197	0,0073
4	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
5	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
6	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
7	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
8	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
9	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073
10	0,3679	0,3996	0,1470	0,0541	0,0199	0,0073

Tabela C.1 – Izračunane vrednosti  $p_{m,n}$  pri  $Q = 1000$

V tabelah C.1 in C.2 opazimo, da gledano po indeksu  $m$  pride do prehodnega pojava. To je neposredno povezano s spreminjanjem vrednosti  $DC_m$ , kar komentiramo pozneje. Vrednosti  $p_{m,n}$  in  $L_{m,n}$  se namreč obnašajo na podoben način kot  $DC_m$ , saj je njihov izračun neposredno odvisen od  $DC_m$ .

V tabeli C.1 opazimo, da se z naraščanjem indeksa  $n$  hitro manjša verjetnost, da bo paket oddan v  $n$ -tem obhodu. To je posledica lastnosti eksponentne verjetnosti porazdelitve, pri kateri verjetnost, da bo paket imel dolžino  $l$ , pada eksponentno z  $l$ .

$m \setminus n$	0	1	2	3	4	5
1	0	418,0233	1418,0233	2418,0233	3418,0233	4418,0233
2	262,9217	1000,0000	2000,0000	3000,0000	4000,0000	5000,0000
3	214,9577	883,9970	1883,9970	2883,9970	3883,9970	4883,9970
4	211,8781	876,7250	1876,7250	2876,7250	3876,7250	4876,7250
5	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984
6	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984
7	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984
8	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984
9	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984
10	211,8668	876,6984	1876,6984	2876,6984	3876,6984	4876,6984

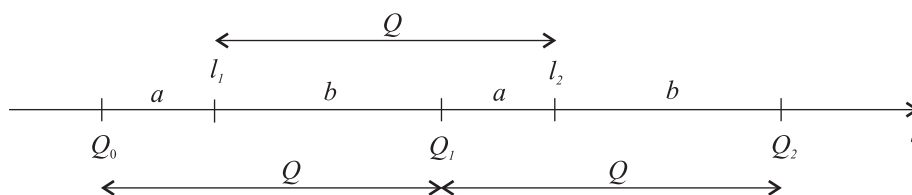
Tabela C.2 – Izračunane vrednosti  $L_{m,n}$  pri  $Q = 1000$

V tabeli C.2 pa opazimo, da se z naraščanjem indeksa  $n$  povprečna dolžina odpravljenega paketa povečuje za vrednost  $Q$ . Tudi to je posledica lastnosti eksponentne verjetnosti porazdelitve, ki ima na kateremkoli odseku enake dolžine, enako obliko. Na primer, če računamo povprečno dolžino paketa  $\bar{l}$  na kateremkoli odseku dolžine  $Q$ , ga



bo le ta vedno razdelila na dva dela v enakih razmerjih. Ker pri izračunu  $L_{m,n}$  obravnavamo odseke dolžine  $Q$ , je tudi povprečna dolžina odpravljenega paketa  $\bar{l}$  na istem mestu vsakega odseka. Če si ti odseki med seboj stikajo (so premaknjeni za vrednost  $Q$ ), so tudi zaporedne povprečne dolžine odpravljenih paketov med seboj razmaknjene za  $Q$ . To pa je razvidno tudi iz tabele C.2.

Zgornjo trditev ponazorimo s sliko C.3. Če vzamemo stikajoča se odseka  $(Q_0, Q_1)$  in  $(Q_1, Q_2)$  širine  $Q$ , potem ju njuni povprečni vrednosti  $l_1$  in  $l_2$  razdelita v enakih razmerjih  $a : b$ . In ker je  $a + b = Q$ , vidimo, da velja  $l_2 - l_1 = Q$ . To pa se sklada z rezultati v tabeli C.2.

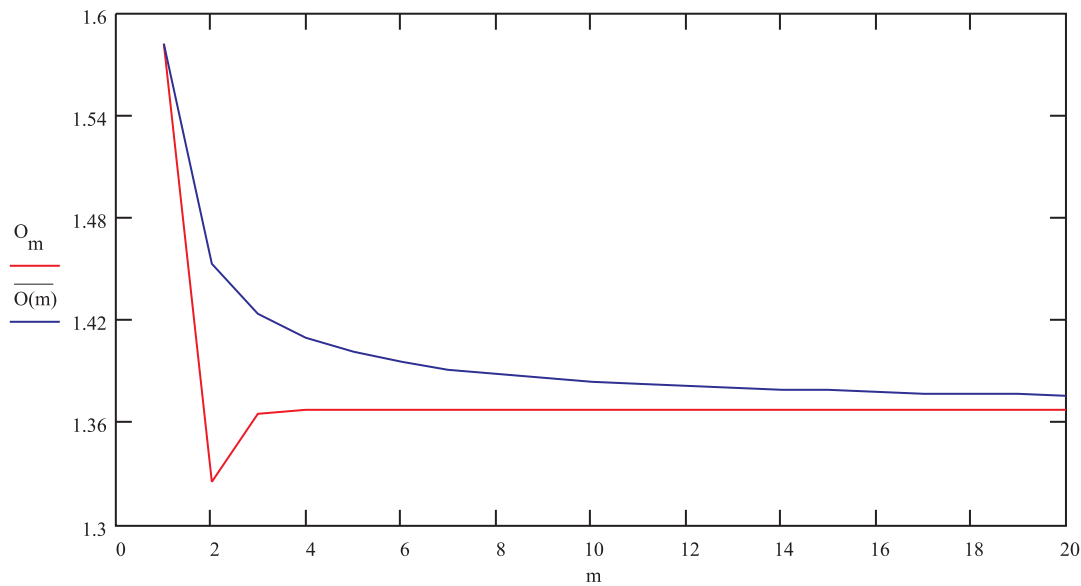


Slika C.3 – Lastnosti eksponentne porazdelitve

$m$	$DC_m$	$O_m$	$\overline{O(m)}$
1	581,9767	1,582	1,5820
2	465,9737	1,3252	1,4536
3	458,7017	1,3652	1,4241
4	458,6751	1,3679	1,4101
5	458,6751	1,3679	1,4016
6	458,6751	1,3679	1,3960
7	458,6751	1,3679	1,3920
8	458,6751	1,3679	1,3890
9	458,6751	1,3679	1,3866
10	458,6751	1,3679	1,3848

Tabela C.3 – Izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$  pri  $Q = 1000$ 

Poglejmo si še izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$ . Slednji sta prikazani tudi v grafu na sliki C.4. Za  $DC_m$  vidimo, da ima najvišjo vrednost pri  $m = 1$ , potem pa hitro pada proti svoji stacionarni vrednosti. Kot posledica tega prehodnega pojava se pojavijo prehodni pojavi tudi pri ostalih spremenljivkah.  $O_m$  na primer zaniha in se potem hitro ustali pri svoji stacionarni vrednosti.  $O_m$  ima najvišjo vrednost pri  $m = 1$ , saj je takrat na voljo najmanjši trenutni DC. Vzrok leži v dejstvu, da prvi paket, ki čakalno vrsto spravi v aktivno stanje, naleti na  $DC_{m-1} = DC_0 = 0$  in posledično je največji paket, ki se lahko postreže v prvem obhodu, enak vrednosti kvanta  $Q$  ( $l \leq Q$ ). Vsi naslednji paketi imajo na voljo še preostanek  $DC_{m-1} \geq 0$  iz prejšnjega obhoda. V najboljšem položaju je ravno drugi paket, ker mu pripade največji ostanek DC, saj je iz tabele C.4 razvidno, da velja  $DC_1 \geq DC_2 \geq DC_3 \dots \geq 0$ . Podobna razlaga je možna tudi za nihanja po indeksu  $m$  pri vrednostih v tabelah C.1 in C.2.  $\overline{O(m)}$  predstavlja

Slika C.4 – Primerjava povprečnega števila operacij pri  $Q = 1000$ 

tekoče povprečje števila operacij, potrebnih za odpravo enega paketa, računano do  $m$ -tega paketa. Če nekoliko priredimo izraz (6.56) in za zgornjo mejo vsote namesto  $M$  vstavimo  $m$ , dobimo naslednji izraz:

$$\overline{O(m)} = \frac{1}{m} \sum_{i=1}^m O_i \quad (\text{C.106})$$

#### C.4.5 Poenostavljen izračun $DC_m$

Iterativno računanje vrednosti  $DC_m$ , in preko njih tudi ostalih vrednosti, je zamudno in nepraktično. Zato bi želeli za posebne primere te izračune poenostaviti. Iz zgornjega primera vidimo, da prehodni pojav sčasoma izzveni in čakalna vrsta, v smislu povprečnega števila operacij, potrebnih za odpravo enega paketa, preide v stacionarno stanje. Izračun vseh ostalih vrednosti je odvisen od  $DC_m$ . Ko v stacionarno stanje, glede na indeks  $m$ , preide  $DC_m$ , vanj preidejo tudi ostale vrednosti. O tem se lahko prepričamo v tabelah C.1, C.2 in C.3.

Prva poenostavitev je izračun asimptote, h kateri teži vrednost  $\overline{O(m)}$ , ko gre  $m \rightarrow \infty$ . Izhajamo iz enačbe (C.102).

$$DC_m = DC_{m-1} - \frac{1}{\lambda} + \frac{Qe^{\lambda Q} e^{-\lambda DC_{m-1}}}{e^{\lambda Q} - 1}$$

Stacionarno stanje nastopi, ko velja  $DC_m = DC_{m-1}$ . Zgornjo enačbo tako zapišemo kot

$$\frac{Qe^{\lambda Q} e^{-\lambda DC_m}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} = 0 \quad (\text{C.107})$$

Sedaj je potrebno iz zgornje enačbe izraziti vrednost  $DC_m$ , ki bo predstavljala stacionarno vrednost  $DC$ , ko gre  $m \rightarrow \infty$ .

$$\begin{aligned}
\frac{Qe^{\lambda Q}e^{-\lambda DC_m}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} &= 0 \\
\frac{Qe^{\lambda Q}e^{-\lambda DC_m}}{e^{\lambda Q} - 1} &= \frac{1}{\lambda} \\
e^{\lambda Q}e^{-\lambda DC_m} &= \frac{1}{\lambda Q}(e^{\lambda Q} - 1) \\
e^{-\lambda DC_m} &= \frac{1}{\lambda Q}(e^{\lambda Q} - 1)e^{-\lambda Q} \\
-\lambda DC_m &= \ln \left[ \frac{1}{\lambda Q}(e^{\lambda Q} - 1)e^{-\lambda Q} \right] \\
DC_m &= -\frac{1}{\lambda} \ln \left[ \frac{1}{\lambda Q}(1 - e^{-\lambda Q}) \right] \tag{C.108}
\end{aligned}$$

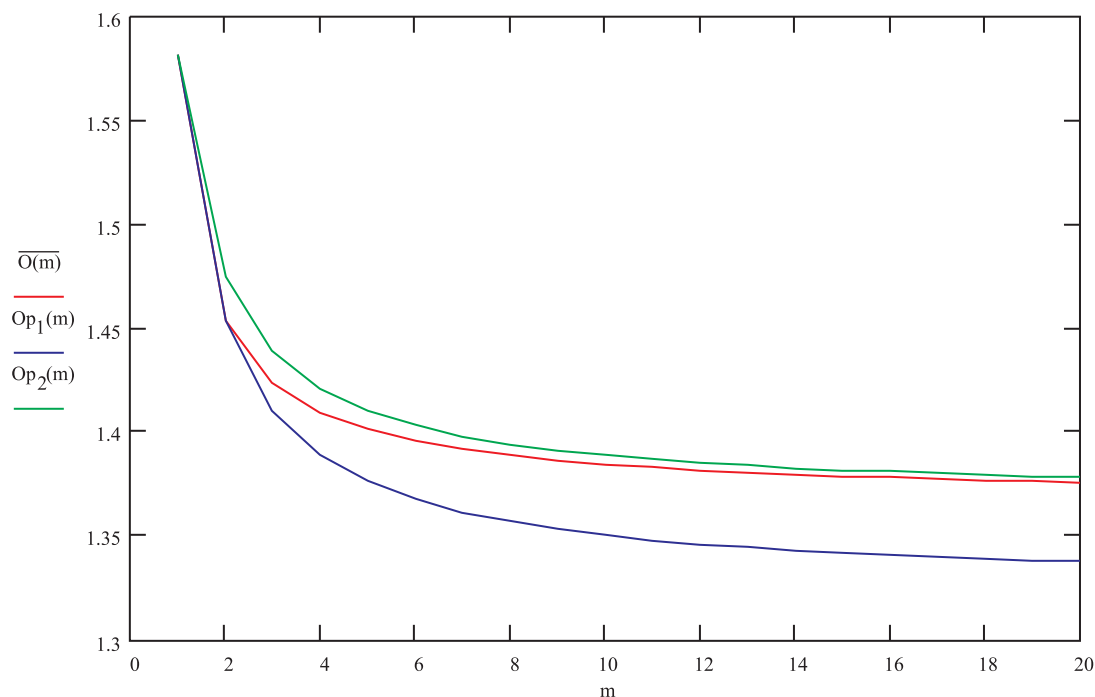
Za majhne vrednosti  $m$  pa lahko uporabimo kar izraz  $DC_1$ , ki smo ga izračunali po odpravi prvega paketa. Če pri izračunih  $DC_m$  pri  $m > 1$  zanemarimo verjetnosti  $p_{m,0}$ , postanejo vrednosti  $DC_m$  enake vrednosti  $DC_1$ .

$$DC_m = DC_1 = \frac{Qe^{\lambda Q}}{e^{\lambda Q} - 1} - \frac{1}{\lambda} \tag{C.109}$$

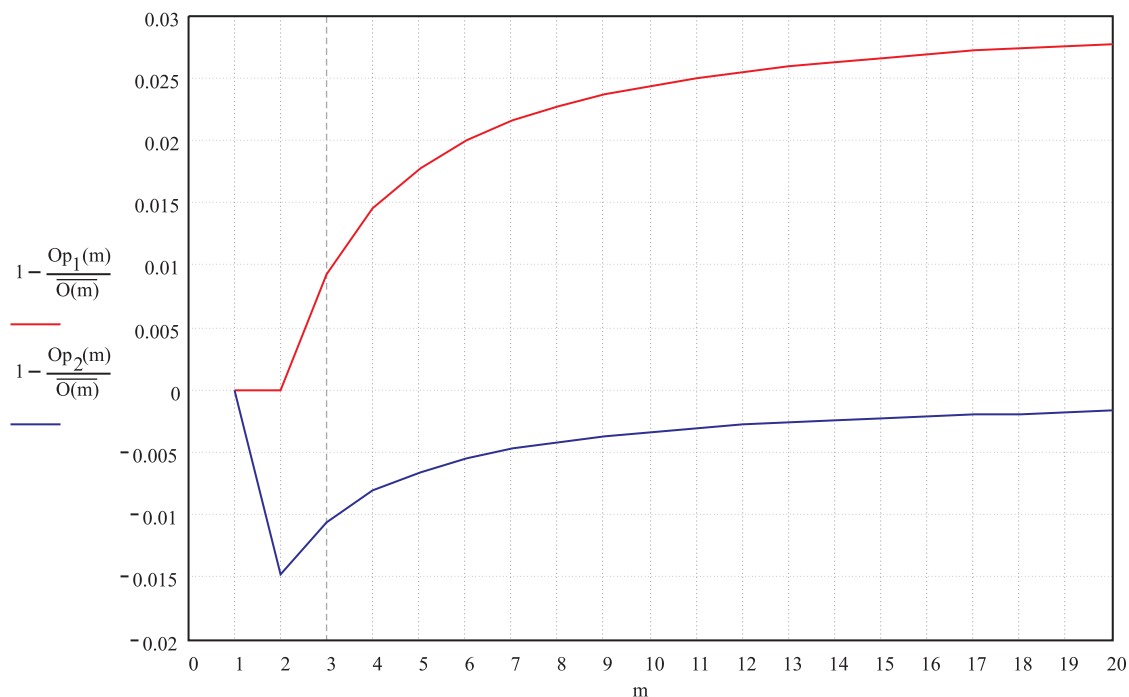
Na sliki C.5 vidimo iterativno izračunane poteke povprečnega števila operacij  $\overline{O(m)}$  in poteke obeh poenostavitvev  $Op_1(m)$  in  $Op_2(m)$ , vse izračunano pri  $Q = 1000$ . Pri tem krivulja  $Op_1(m)$  predstavlja poenostavitev za majhne vrednosti,  $Op_2(m)$  pa za velike vrednosti indeksa  $m$ . Kot je bilo pričakovati, krivulja  $Op_1(m)$  z večanjem  $m$  vedno bolj odstopa od iterativnega izračuna, krivulja  $Op_2(m)$  pa najbolj odstopa ravno na začetku, kasneje pa se krivulji  $\overline{O(m)}$  asimptotično približuje. Taki rezultati so tudi pričakovani, saj smo pravkar ugotovljena dejstva vzeli kot osnovo pri izpeljavi obeh poenostavitvev.

Poskusimo sedaj ugotoviti območja vrednosti indeksa  $m$ , pri katerih je smiselno uporabiti posamezno poenostavitev. Na sliki C.6 je prikazan potek relativnega odstopanja poenostavljenih izračunov povprečnega števila operacij od analitičnega izračuna pri  $Q = 1000$ . Pri tem krivulja  $1 - \frac{Op_1(m)}{\overline{O(m)}}$  predstavlja relativno odstopanje glede na iterativno izračunane vrednosti povprečnega števila operacij  $\overline{O(m)}$  za majhne vrednosti indeksa  $m$ , krivulja  $1 - \frac{Op_2(m)}{\overline{O(m)}}$  pa za velike vrednosti indeksa  $m$ . Izkaže se, da je meja nekje pri tretjem paketu v čakalni vrsti. Pod to mejo je bolj smiselno uporabiti prvo, nad njo pa drugo poenostavitev.

Seveda nas zanima tudi, v katerih načinih delovanja čakalne vrste bo sploh prišlo do priložnosti uporabe posameznih poenostavitvev. Najprej si zamislimo delovanje čakalne vrste v malo obremenjenem omrežju. Takrat stanje čakalne vrste stalno alternira med aktivnim in pasivnim. To pomeni, da bo v večini primerov v posameznem stanju aktivnosti postreženih le nekaj paketov. V tem primeru je bolje uporabiti poenostavitev za majhne vrednosti indeksa  $m$ . V primeru zelo obremenjenega omrežja pa je v posameznem stanju aktivnosti čakalne vrste postreženih veliko paketov in je



Slika C.5 – Primerjava poenostavljenih in analitičnega izračuna povprečnega števila operacij pri  $Q = 1000$



Slika C.6 – Relativno odstopanje poenostavljenih izračunov povprečnega števila operacij od analitičnega pri  $Q = 1000$

bolje uporabiti poenostavitev za velike vrednosti indeksa  $m$ . V srednje obremenjenem omrežju je ravno tako bolje uporabiti slednjo poenostavitev, saj ta v povprečju manj odstopa od pravih, iterativno izračunanih vrednosti.

## C.5 Zahtevnost SCDRR pri enakomerni porazdelitvi dolžin paketov

Lotimo se še analize računske zahtevnosti pri enakomerni verjetnostni porazdelitvi dolžin paketov.

### C.5.1 Definicije

Pri analizi, tako kot pri eksponentni verjetnostni porazdelitvi, uporabljamo spremenljivke in definicije iz 6.5.1, pri čemer lahko za nekatere izmed njih zapišemo izraze in vrednosti, ki ustrezajo enakomerni verjetnostni porazdelitvi dolžin paketov.

$$\begin{aligned} 0 < L_{min} \leq L_{max} & & \bar{L} &= \frac{1}{2}(L_{max} + L_{min}) \\ 0 < L_{max} \leq \infty & & L_{min} \leq Q < L_{max} \\ N_m \leq \infty & & & \end{aligned}$$

Zapišimo še izraze, ki veljajo za verjetnostno porazdelitev in kumulativno verjetnostno porazdelitev dolžin paketov.

$$\begin{aligned} f(l) &= \begin{cases} \frac{1}{L_{max} - L_{min}} & L_{min} \leq l \leq L_{max} \\ 0 & \text{drugje} \end{cases} \\ F(l) &= \begin{cases} 0 & l < L_{min} \\ \frac{l - L_{min}}{L_{max} - L_{min}} & L_{min} \leq l < L_{max} \\ 1 & l \geq L_{max} \end{cases} \end{aligned}$$

Upoštevajmo še robne pogoje iz 6.5.2 in lahko se lotimo izračuna števila operacij pri enakomerni verjetnostni porazdelitvi dolžin paketov.

### C.5.2 Izračun števila potrebnih operacij

Za izračun uporabimo kar enačbe od (6.48) do (6.54), pri tem pa upoštevamo zgornje lastnosti in enačbe za enakomerno verjetnostno porazdelitev dolžin paketov.

Analizo pričnemo pri prvem paketu v čakalni vrsti, torej pri  $m = 1$ . Veljajo enaki robni pogoji kot pri splošni analizi. Začetna vrednost  $DC$  ob času  $t_0$  je enaka nič  $DC = DC_0 = 0$ , vrednost kvanta  $Q$  pa se nahaja med najmanjšo in največjo možno dolžino paketa  $L_{min} \leq Q < L_{max}$ . Vrednosti manjše od  $L_{min}$  niso smiselne, saj v tem primeru noben paket ne bi mogel biti odpravljen v prvem obhodu. Vrednosti večje ali enake  $L_{max}$  pa ne predstavljajo več podkritične vrednosti kvanta  $Q$ . Najprej izračunajmo  $N_1$ .

$$N_1 = \left\lceil \frac{L_{max}}{Q} \right\rceil \quad (\text{C.110})$$

Sedaj izrazimo  $p_{1,n}$  pri  $n = 1, 2, \dots, N_1$ .

$$p_{1,1} = P(l \leq Q) = F(Q) = \frac{Q - L_{min}}{L_{max} - L_{min}} \quad (C.111)$$

$$\begin{aligned} p_{1,n} &= P((n-1)Q < l \leq nQ) = F(nQ) - F((n-1)Q) \\ &= \frac{Q}{L_{max} - L_{min}} \quad \text{pri } 1 < n < N_1 \end{aligned} \quad (C.112)$$

$$\begin{aligned} p_{1,N_1} &= P((N_1-1)Q < l \leq N_1Q) = F(N_1Q) - F((N_1-1)Q) \\ &= 1 - \frac{(N_1-1)Q - L_{min}}{L_{max} - L_{min}} = \frac{L_{max} - (N_1-1)Q}{L_{max} - L_{min}} \end{aligned} \quad (C.113)$$

pri čemer upoštevamo, da je  $N_1Q \geq L_{max}$ . Preden nadaljujemo, preverimo, ali je verjetnost gotovega dogodka  $S_1$  za podane vrednosti  $p_{1,n}$  res enaka 1. Če uporabimo izraz (6.28), lahko zapišemo:

$$\begin{aligned} S_1 &= \sum_{n=1}^{N_1} p_{1,n} = \frac{Q - L_{min}}{L_{max} - L_{min}} + \sum_{n=2}^{N_1-1} \frac{Q}{L_{max} - L_{min}} + \frac{L_{max} - (N_1-1)Q}{L_{max} - L_{min}} \\ &= \frac{Q - L_{min} + (N_1-2)Q + L_{max} - (N_1-1)Q}{L_{max} - L_{min}} = 1 \end{aligned} \quad (C.114)$$

Sedaj lahko izračunamo povprečno število operacij  $O_1$  potrebnih za oddajo prvega paketa:

$$\begin{aligned} O_1 &= \sum_{n=1}^{N_1} np_{1,n} \\ &= \frac{Q - L_{min}}{L_{max} - L_{min}} + \sum_{n=2}^{N_1-1} n \frac{Q}{L_{max} - L_{min}} + N_1 \frac{L_{max} - (N_1-1)Q}{L_{max} - L_{min}} \\ &= \frac{1}{L_{max} - L_{min}} \left[ Q - L_{min} + \frac{1}{2}(N_1-2)(N_1+1)Q + N_1(L_{max} - (N_1-1)Q) \right] \\ &= \frac{1}{L_{max} - L_{min}} \left[ Q - L_{min} + \frac{N_1^2Q - N_1Q}{2} - Q + N_1L_{max} - N_1^2Q + N_1Q \right] \\ &= \frac{1}{L_{max} - L_{min}} \left[ N_1L_{max} - L_{min} + (1 - N_1) \frac{N_1Q}{2} \right] \end{aligned} \quad (C.115)$$

Za izračun števila operacij  $O_2$ , potrebnih za odpravo drugega paketa, zopet potrebujemo povprečno vrednost  $DC$ , ki ostane po odpravi prvega paketa in jo označimo z  $DC_1$ . Za njen izračun moramo najprej izračunati vrednosti  $L_{1,n}$  pri  $n = 1, 2, \dots, N_1$ . V enačbo (6.52) za splošni primer vnesemo izraze, ki veljajo za enakomerno verjetnostno porazdelitev dolžin paketov ter upoštevamo še  $m = 1$  in  $n = 1, 2, \dots, N_1$ .

$$L_{1,1} = \frac{\int_{L_{min}}^Q lf(l)dl}{\int_{L_{min}}^Q f(l)dl} = \frac{\int_{L_{min}}^Q \frac{l}{L_{max} - L_{min}} dl}{\int_{L_{min}}^Q \frac{1}{L_{max} - L_{min}} dl} = \frac{\frac{l^2}{2} \Big|_{L_{min}}^Q}{\frac{l}{L_{max} - L_{min}} \Big|_{L_{min}}^Q}$$

$$= \frac{Q^2 - L_{min}^2}{2(Q - L_{min})} = \frac{Q + L_{min}}{2} \quad (\text{C.116})$$

$$\begin{aligned} L_{1,n} &= \frac{\int_{(n-1)Q}^{nQ} \frac{l}{L_{max} - L_{min}} dl}{\int_{(n-1)Q}^{nQ} \frac{1}{L_{max} - L_{min}} dl} = \frac{\left. \frac{l^2}{2(L_{max} - L_{min})} \right|_{(n-1)Q}^{nQ}}{\left. \frac{l}{L_{max} - L_{min}} \right|_{(n-1)Q}^{nQ}} \\ &= \frac{(nQ)^2 - ((n-1)Q)^2}{2(nQ - (n-1)Q)} = \frac{(nQ - (n-1)Q)(nQ + (n-1)Q)}{2(nQ - (n-1)Q)} \\ &= \frac{(2n-1)Q}{2} \quad \text{pri } 1 < n < N_1 \end{aligned} \quad (\text{C.117})$$

$$\begin{aligned} L_{1,N_1} &= \frac{\int_{(N_1-1)Q}^{L_{max}} \frac{l}{L_{max} - L_{min}} dl}{\int_{(N_1-1)Q}^{L_{max}} \frac{1}{L_{max} - L_{min}} dl} = \frac{\left. \frac{l^2}{2(L_{max} - L_{min})} \right|_{(N_1-1)Q}^{L_{max}}}{\left. \frac{l}{L_{max} - L_{min}} \right|_{(N_1-1)Q}^{L_{max}}} \\ &= \frac{L_{max}^2 - ((N_1-1)Q)^2}{2(L_{max} - (N_1-1)Q)} = \frac{L_{max} + (N_1-1)Q}{2} \end{aligned} \quad (\text{C.118})$$

Ob upoštevanju enačbe (6.54) ter robnih pogojev  $DC_0 = 0$  in  $n = 1, 2, \dots, N_1$ , lahko za ta primer zapišemo:

$$DC_1 = \sum_{n=1}^{N_1} p_{1,n}(nQ - L_{1,n}) \quad (\text{C.119})$$

Če sedaj v zgornji izraz vstavimo pravkar izračunane vrednosti  $L_{1,n}$ , dobimo:

$$\begin{aligned} DC_1 &= p_{1,1}(Q - L_{1,1}) + \sum_{n=2}^{N_1-1} p_{1,n}(nQ - L_{1,n}) + p_{1,N_1}(N_1Q - L_{1,N_1}) \\ &= \frac{Q - L_{min}}{L_{max} - L_{min}} \left[ Q - \frac{Q + L_{min}}{2} \right] \\ &+ \sum_{n=2}^{N_1-1} \frac{Q}{L_{max} - L_{min}} \left[ nQ - \frac{(2n-1)Q}{2} \right] \\ &+ \frac{L_{max} - (N_1-1)Q}{L_{max} - L_{min}} \left[ N_1Q - \frac{L_{max} + (N_1-1)Q}{2} \right] \\ &= \frac{(Q - L_{min})(Q - L_{min})}{2(L_{max} - L_{min})} + \sum_{n=2}^{N_1-1} \frac{Q^2}{2(L_{max} - L_{min})} \\ &+ \frac{(L_{max} - (N_1-1)Q)(N_1Q - L_{max} + Q)}{2(L_{max} - L_{min})} \end{aligned}$$

$$= \frac{L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_1QL_{max} + N_1Q^2(1 - N_1)}{2(L_{max} - L_{min})} \quad (C.120)$$

Za drugi in vse nadaljnje pakete v čakalni vrsti moramo upoštevati, da je  $DC \geq 0$ . Zato moramo izračunati  $N_m$  ali največje možno število obhodov, potrebnih za oddajo  $m$ -tega paketa po enačbi

$$N_m = \left\lceil \frac{L_{max} - DC_{m-1}}{Q} \right\rceil \quad (C.121)$$

kar pomeni, da se število potrebnih obhodov spreminja tudi v odvisnosti od  $DC_{m-1}$ .  $N_m$  ima največjo vrednost ravno pri  $m = 1$ , saj je takrat vrednost  $DC_{m-1} = DC_0 = 0$  po definiciji enaka nič.

Poleg tega pa pri enakomerni porazdelitvi dobimo dva različna primera izračuna vrednosti za  $m > 1$ . Prvi primer nastopi pri izpolnjenem pogoju  $DC_{m-1} < L_{min}$ . V tem primeru  $m$ -ti paket ne more biti postrežen v istem obhodu kot  $m - 1$  paket, zato velja  $n = 1, 2, \dots, N_m$ . V drugem primeru, ko je izpolnjen pogoj  $DC_{m-1} \geq L_{min}$ , pa je lahko  $m$ -ti paket postrežen v istem obhodu kot  $m - 1$  paket. Zato izračun velja za  $n = 0, 1, 2, \dots, N_m$ .

### C.5.2.1 Izračun pri izpolnjenem pogoju $DC_{m-1} < L_{min}$

Lotimo se prvega primera pri pogoju  $DC_{m-1} < L_{min}$ . Izrazimo vrednosti  $p_{m,n}$ ,  $L_{m,n}$ ,  $DC_m$  in  $O_m$  pri  $m > 1$  in  $n = 1, 2, \dots, N_m$ .

$$\begin{aligned} p_{m,1} &= P(l \leq Q + DC_{m-1}) = F(Q + DC_{m-1}) \\ &= \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} \end{aligned} \quad (C.122)$$

$$\begin{aligned} p_{m,n} &= P((n-1)Q + DC_{m-1} < l \leq nQ + DC_{m-1}) \\ &= F(nQ + DC_{m-1}) - F((n-1)Q + DC_{m-1}) \\ &= \frac{Q}{L_{max} - L_{min}} \quad \text{pri } 1 < n < N_m \end{aligned} \quad (C.123)$$

$$\begin{aligned} p_{m,N_m} &= P((N_m - 1)Q + DC_{m-1} < l \leq N_mQ + DC_{m-1}) \\ &= F(N_mQ + DC_{m-1}) - F((N_m - 1)Q + DC_{m-1}) \\ &= \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \end{aligned} \quad (C.124)$$

Preverimo še ali je verjetnost gotovega dogodka  $S_m$  za podane vrednosti  $p_{m,n}$  res enaka 1. Če uporabimo izraz (6.28), lahko zapišemo:

$$S_m = \sum_{n=1}^{N_m} p_{m,n}$$



$$\begin{aligned}
&= \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} + \sum_{n=2}^{N_m-1} \frac{Q}{L_{max} - L_{min}} + \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \\
&= \frac{Q + DC_{m-1} - L_{min} + (N_m - 2)Q + L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \\
&= 1
\end{aligned} \tag{C.125}$$

Sedaj lahko izračunamo povprečno število operacij  $O_m$ , potrebnih za oddajo  $m$ -tega paketa:

$$\begin{aligned}
O_m &= \sum_{n=1}^{N_m} np_{m,n} \\
&= \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} + \sum_{n=2}^{N_m-1} n \frac{Q}{L_{max} - L_{min}} \\
&+ N_m \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \\
&= \frac{1}{L_{max} - L_{min}} \left[ Q + DC_{m-1} - L_{min} + \frac{1}{2}(N_m - 2)(N_m + 1)Q \right. \\
&+ \left. N_m (L_{max} - (N_m - 1)Q - DC_{m-1}) \right] \\
&= \frac{1}{L_{max} - L_{min}} \left[ Q + DC_{m-1} - L_{min} + \frac{N_m^2 Q - N_m Q}{2} - Q \right. \\
&+ \left. N_m L_{max} - N_m^2 Q + N_m Q - N_m DC_{m-1} \right] \\
&= \frac{1}{L_{max} - L_{min}} \left[ N_m L_{max} - L_{min} + (1 - N_m) \left( DC_{m-1} + \frac{N_m Q}{2} \right) \right]
\end{aligned} \tag{C.126}$$

Izračun vrednosti  $L_{m,n}$  je skoraj enak kot pri prvem paketu. Razlikuje se le v nekaterih izrazih po dodani konstanti  $DC_{m-1}$ .

$$\begin{aligned}
L_{m,1} &= \frac{\int_{L_{min}}^{Q+DC_{m-1}} l f(l) dl}{\int_{L_{min}}^{Q+DC_{m-1}} f(l) dl} = \frac{\int_{L_{min}}^{Q+DC_{m-1}} \frac{l}{L_{max} - L_{min}} dl}{\int_{L_{min}}^{Q+DC_{m-1}} \frac{1}{L_{max} - L_{min}} dl} = \frac{\frac{l^2}{2(L_{max} - L_{min})} \Big|_{L_{min}}^{Q+DC_{m-1}}}{\frac{l}{L_{max} - L_{min}} \Big|_{L_{min}}^{Q+DC_{m-1}}} \\
&= \frac{(Q + DC_{m-1})^2 - L_{min}^2}{2(Q + DC_{m-1} - L_{min})} = \frac{Q + DC_{m-1} + L_{min}}{2}
\end{aligned} \tag{C.127}$$

$$\begin{aligned}
L_{m,n} &= \frac{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} \frac{l}{L_{max} - L_{min}} dl}{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} \frac{1}{L_{max} - L_{min}} dl} = \frac{\frac{l^2}{2(L_{max} - L_{min})} \Big|_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}}}{\frac{l}{L_{max} - L_{min}} \Big|_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}}} \\
&= \frac{(nQ + DC_{m-1})^2 - ((n-1)Q + DC_{m-1})^2}{2(nQ + DC_{m-1} - ((n-1)Q + DC_{m-1}))}
\end{aligned}$$

$$= \frac{(2n-1)Q}{2} + DC_{m-1} \quad \text{pri } 1 < n < N_m \quad (\text{C.128})$$

$$\begin{aligned} L_{m,N_m} &= \frac{\int_{(N_m-1)Q+DC_{m-1}}^{L_{max}} \frac{l}{L_{max}-L_{min}} dl}{\int_{(N_m-1)Q+DC_{m-1}}^{L_{max}} \frac{1}{L_{max}-L_{min}} dl} = \frac{\left. \frac{l^2}{2(L_{max}-L_{min})} \right|_{(N_m-1)Q+DC_{m-1}}^{L_{max}}}{\left. \frac{l}{L_{max}-L_{min}} \right|_{(N_m-1)Q+DC_{m-1}}^{L_{max}}} \\ &= \frac{L_{max}^2 - ((N_m-1)Q + DC_{m-1})^2}{2(L_{max} - ((N_m-1)Q + DC_{m-1}))} \\ &= \frac{L_{max} + (N_m-1)Q + DC_{m-1}}{2} \end{aligned} \quad (\text{C.129})$$

Sedaj lahko izračunamo  $DC_m$

$$\begin{aligned} DC_m &= p_{m,1}(Q + DC_{m-1} - L_{m,1}) + \sum_{n=2}^{N_m-1} p_{m,n}(nQ + DC_{m-1} - L_{m,n}) \\ &+ p_{m,N_m}(N_mQ + DC_{m-1} - L_{m,N_m}) \\ &= \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} \left[ Q + DC_{m-1} - \frac{Q + DC_{m-1} + L_{min}}{2} \right] \\ &+ \sum_{n=2}^{N_m-1} \frac{Q}{L_{max} - L_{min}} \left[ nQ + DC_{m-1} - \frac{(2n-1)Q}{2} - DC_{m-1} \right] \\ &+ \frac{L_{max} - (N_m-1)Q - DC_{m-1}}{L_{max} - L_{min}} \left[ N_mQ + DC_{m-1} - \frac{L_{max} + (N_m-1)Q + DC_{m-1}}{2} \right] \\ &= \frac{(Q + DC_{m-1} - L_{min})(Q + DC_{m-1} - L_{min})}{2(L_{max} - L_{min})} + \sum_{n=2}^{N_m-1} \frac{Q^2}{2(L_{max} - L_{min})} \\ &+ \frac{(L_{max} - (N_m-1)Q - DC_{m-1})(N_mQ + DC_{m-1} - L_{max} + Q)}{2(L_{max} - L_{min})} \\ &= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_mQL_{max} \right. \\ &\left. + N_mQ^2(1 - N_m) + 2DC_{m-1}(L_{max} - L_{min} + Q(1 - N_m)) \right] \end{aligned} \quad (\text{C.130})$$

Opazimo, da se vrednost  $DC_m$  za  $m > 1$  od vrednosti  $DC_1$  razlikuje samo v členu:

$$DC_{m-1} \left[ 1 + \frac{Q(1 - N_m)}{L_{max} - L_{min}} \right] \quad (\text{C.131})$$

### C.5.2.2 Izračun pri izpolnjenem pogoju $DC_{m-1} \geq L_{min}$

Drugi primer se od prvega razlikuje po možnosti odprave paketa  $m$  v istem obhodu, kot je bil odpravljen  $m-1$  paket. To se lahko zgodi pri pogoju  $DC_{m-1} \geq L_{min}$ . Ob upoštevanju slednjega lahko za  $m > 1$  in  $n = 0, 1, 2, \dots, N_m$  zapišemo:

$$\begin{aligned}
p_{m,0} &= P(L_{min} \leq l \leq DC_{m-1}) = F(DC_{m-1}) - F(L_{min}) \\
&= \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}}
\end{aligned} \tag{C.132}$$

$$\begin{aligned}
p_{m,n} &= P((n-1)Q + DC_{m-1} < l \leq nQ + DC_{m-1}) \\
&= F(nQ + DC_{m-1}) - F((n-1)Q + DC_{m-1}) \\
&= \frac{Q}{L_{max} - L_{min}} \quad \text{pri } 0 < n < N_m
\end{aligned} \tag{C.133}$$

$$\begin{aligned}
p_{m,N_m} &= P((N_m - 1)Q + DC_{m-1} < l \leq N_m Q + DC_{m-1}) \\
&= F(N_m Q + DC_{m-1}) - F((N_m - 1)Q + DC_{m-1}) \\
&= \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}}
\end{aligned} \tag{C.134}$$

Preverimo še, ali je verjetnost gotovega dogodka  $S_m$  za podane vrednosti  $p_{m,n}$  res enaka 1. Če upoštevamo dejstvo, da je vsota verjetnost  $p_{m,0} + p_{m,1}$  za drugi primer ( $DC_m \geq L_{min}$ ) enaka verjetnosti  $p_{m,1}$  za prvi primer ( $DC_m < L_{min}$ ) in da so ostale vrednosti  $p_{m,n}$  za  $n > 1$  v obeh primerih enake, morata biti tudi vsoti  $S_m$  za oba primera enaki in imata vrednost ena. Sedaj lahko izračunamo povprečno število operacij  $O_m$ , potrebnih za oddajo  $m$ -tega paketa:

$$\begin{aligned}
O_m &= p_{m,0} + \sum_{n=1}^{N_m} n p_{m,n} \\
&= \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}} + \sum_{n=1}^{N_m-1} n \frac{Q}{L_{max} - L_{min}} \\
&\quad + N_m \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \\
&= \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}} + \frac{Q}{L_{max} - L_{min}} + \sum_{n=2}^{N_m-1} n \frac{Q}{L_{max} - L_{min}} \\
&\quad + N_m \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \\
&= \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} + \sum_{n=2}^{N_m-1} n \frac{Q}{L_{max} - L_{min}} \\
&\quad + N_m \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}}
\end{aligned} \tag{C.135}$$

Vidimo, da je rezultat (C.135) enak drugi vrstici izraza (C.126). To pomeni, da morata biti tudi končna rezultata obeh izrazov enaka. Zato velja:

$$O_m = \frac{1}{L_{max} - L_{min}} \left[ N_m L_{max} - L_{min} + (1 - N_m) \left( DC_{m-1} + \frac{N_m Q}{2} \right) \right] \tag{C.136}$$

Za izračun vrednosti  $L_{m,n}$  velja enako kot pri  $L_{1,n}$ . Ob upoštevanju splošnega obrazca (6.52) ter relacij  $m > 1$  in  $n = 0, 1, 2 \dots N_m$  zapišemo:

$$\begin{aligned} L_{m,0} &= \frac{\int_{L_{min}}^{DC_{m-1}} \frac{l}{L_{max} - L_{min}} dl}{\int_{L_{min}}^{DC_{m-1}} \frac{1}{L_{max} - L_{min}} dl} = \frac{l^2}{2(L_{max}L_{min})} \Big|_{L_{min}}^{DC_{m-1}} \\ &= \frac{DC_{m-1} + L_{min}}{2} \end{aligned} \quad (C.137)$$

$$\begin{aligned} L_{m,n} &= \frac{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} \frac{l}{L_{max} - L_{min}} dl}{\int_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} \frac{1}{L_{max} - L_{min}} dl} = \frac{l^2}{2(L_{max} - L_{min})} \Big|_{(n-1)Q+DC_{m-1}}^{nQ+DC_{m-1}} \\ &= \frac{(2n-1)Q}{2} + DC_{m-1} \end{aligned} \quad (C.138)$$

$$\begin{aligned} L_{m,N_m} &= \frac{\int_{(N_m-1)Q+DC_{m-1}}^{L_{max}} \frac{l}{L_{max} - L_{min}} dl}{\int_{(N_m-1)Q+DC_{m-1}}^{L_{max}} \frac{1}{L_{max} - L_{min}} dl} = \frac{l^2}{2(L_{max} - L_{min})} \Big|_{(N_m-1)Q+DC_{m-1}}^{L_{max}} \\ &= \frac{L_{max} + (N_m - 1)Q + DC_{m-1}}{2} \end{aligned} \quad (C.139)$$

Sedaj lahko izračunamo  $DC_m$

$$\begin{aligned} DC_m &= p_{m,0}(DC_{m-1} - L_{m,1}) + \sum_{n=1}^{N_m-1} p_{m,n}(nQ + DC_{m-1} - L_{m,n}) \\ &+ p_{m,N_m}(N_mQ + DC_{m-1} - L_{m,N_m}) \\ &= \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}} \left[ DC_{m-1} - \frac{DC_{m-1} + L_{min}}{2} \right] \\ &+ \sum_{n=1}^{N_m-1} \frac{Q}{L_{max} - L_{min}} \left[ nQ + DC_{m-1} - \frac{(2n-1)Q}{2} - DC_{m-1} \right] \\ &+ \frac{L_{max} - (N_m - 1)Q - DC_{m-1}}{L_{max} - L_{min}} \left[ N_mQ + DC_{m-1} \right. \\ &\left. - \frac{L_{max} + (N_m - 1)Q + DC_{m-1}}{2} \right] \\ &= \frac{(DC_{m-1} - L_{min})(DC_{m-1} - L_{min})}{2(L_{max} - L_{min})} + \sum_{n=1}^{N_m-1} \frac{Q^2}{2(L_{max} - L_{min})} \\ &+ \frac{(L_{max} - (N_m - 1)Q - DC_{m-1})(N_mQ + DC_{m-1} - L_{max} + Q)}{2(L_{max} - L_{min})} \\ &= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right] \end{aligned}$$

$$+ 2DC_{m-1}(L_{max} - L_{min} - QN_m) \quad (C.140)$$

Opazimo, da se vrednost  $DC_m$  v drugem primeru ( $DC_m \geq L_{min}$ ) od vrednosti  $DC_m$  v prvem primeru ( $DC_m < L_{min}$ ) razlikuje samo v členu

$$\frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}}Q = p_{m,0}Q \quad (C.141)$$

ki predstavlja ravno verjetnost  $p_{m,0}$ , pomnoženo z velikostjo kvanta  $Q$ .

### C.5.3 Zbir rezultatov za enakomerno porazdelitev dolžin paketov

Zapišimo še končne rezultate za enakomerno porazdelitev dolžin paketov. Pri tem moramo upoštevati definicije spremenljivk iz razdelka C.5.1, začetni pogoj  $DC_0 = 0$  ter zalogo vrednosti indeksov  $m = 1, 2, 3, \dots, M$  in  $n = 0, 1, 2, \dots, N_m$ . Pri enakomerni porazdelitvi moramo upoštevati dva različna primera izračuna vrednosti spremenljivk glede na vrednost  $DC_{m-1}$ .

$$N_m = \left\lceil \frac{L_{max} - DC_{m-1}}{Q} \right\rceil \quad (C.142)$$

$$DC_m = \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) + 2DC_{m-1}(L_{max} - L_{min} + Q(1 - N_m)) \right] \quad DC_{m-1} < L_{min} \quad (C.143)$$

$$DC_m = \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) + 2DC_{m-1}(L_{max} - L_{min} - QN_m) \right] \quad DC_{m-1} \geq L_{min} \quad (C.144)$$

$$p_{m,n} = \begin{cases} 0 & n = 0, DC_{m-1} < L_{min} \\ \frac{DC_{m-1} - L_{min}}{L_{max} - L_{min}} & n = 0, DC_{m-1} \geq L_{min} \\ \frac{Q + DC_{m-1} - L_{min}}{L_{max} - L_{min}} & n = 1, DC_{m-1} < L_{min} \\ \frac{Q}{L_{max} - L_{min}} & n = 1, DC_{m-1} \geq L_{min} \\ \frac{L_{max} - L_{min}}{Q} & 1 < n < N_m \\ \frac{L_{max} - L_{min}}{L_{max} - (N_m - 1)Q - DC_{m-1}} & n = N_m \end{cases} \quad (C.145)$$

$$L_{m,n} = \begin{cases} 0 & n = 0, DC_{m-1} < L_{min} \\ \frac{DC_{m-1} + L_{min}}{2} & n = 0, DC_{m-1} \geq L_{min} \\ \frac{Q + DC_{m-1} + L_{min}}{2} & n = 1, DC_{m-1} < L_{min} \\ \frac{(2n-1)Q}{2} + DC_{m-1} & n = 1, DC_{m-1} \geq L_{min} \\ \frac{(2n-1)Q}{2} + DC_{m-1} & 1 < n < N_m \\ \frac{L_{max} + (N_m - 1)Q + DC_{m-1}}{2} & n = N_m \end{cases} \quad (C.146)$$

$$O_m = \frac{1}{L_{max} - L_{min}} \left[ N_m L_{max} - L_{min} + (1 - N_m) \left( DC_{m-1} + \frac{N_m Q}{2} \right) \right] \quad (C.147)$$

### C.5.4 Primer 1

Poglejmo si primer izračuna nekaterih izpeljanih vrednosti za enakomerno porazdelitev dolžin paketov pri naslednjih pogojih:

$$Q = 1000 \qquad L_{min} = 0 \qquad L_{max} = 2000, 4000$$

V prvem primeru pri  $L_{max} = 2000$  je vrednost kvanta  $Q$  kar enaka povprečni dolžini paketov  $\bar{L} = 1000$ . S tem želimo doseči, da bo večina paketov v čakalni vrsti za odpravo potrebovala samo eno operacijo, v drugem primeru pri  $L_{max} = 4000$  pa dve operaciji. Pri tem seveda upoštevamo, da je pri  $m > 1$  prisoten tudi ostanek  $DC_{m-1}$ , ki prispeva k hitrejši odpravi posameznega paketa. Najprej si pogledjmo izračunane vrednosti  $p_{m,n}$  in  $L_{m,n}$  za oba primera, ki so zapisane v tabelah C.4 in C.5. Pri tem je potrebno za oba primera navesti še vrednosti spremenljivke  $N_m$ , ki določa območje vrednosti indeksa  $n$ . Pri  $L_{max} = 2000$  je vrednost  $N_m = 2$  za vsak  $m$ , zato velja  $n = 0, 1, 2$ . Pri  $L_{max} = 4000$  pa je vrednost  $N_m = 4$  za vsak  $m$ , zato velja  $n = 0, 1, \dots, 4$ .

m \ n	0	1	2
1	0	0,5	0,5
2	0,25	0,5	0,25
3	0,25	0,5	0,25
4	0,25	0,5	0,25
5	0,25	0,5	0,25
6	0,25	0,5	0,25
7	0,25	0,5	0,25
8	0,25	0,5	0,25
9	0,25	0,5	0,25
10	0,25	0,5	0,25

$L_{max} = 2000$

m \ n	0	1	2	3	4
1	0	0,25	0,25	0,25	0,25
2	0,125	0,25	0,25	0,25	0,125
3	0,125	0,25	0,25	0,25	0,125
4	0,125	0,25	0,25	0,25	0,125
5	0,125	0,25	0,25	0,25	0,125
6	0,125	0,25	0,25	0,25	0,125
7	0,125	0,25	0,25	0,25	0,125
8	0,125	0,25	0,25	0,25	0,125
9	0,125	0,25	0,25	0,25	0,125
10	0,125	0,25	0,25	0,25	0,125

$L_{max} = 4000$

Tabela C.4 – Izračunane vrednosti  $p_{m,n}$  pri  $Q = 1000$  in  $L_{max} = 2000, 4000$

Poglejmo si še izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$ , ki so naštetni v tabeli C.6. Za  $DC_m$  vidimo, da ima kar konstantno vrednost. To ob podanih pogojih izhaja iz enačb v razdelku C.5.3. Posledica tega so tudi nezanimivi rezultati za  $p_{m,n}$ ,  $L_{m,n}$  in  $O_m$ , ki so različni samo pri  $m = 1$ , za  $m > 1$  pa so spet konstantni. Vzrok za nezanimive rezultate so izbrane vrednosti  $L_{max}$  in  $Q$ , katerih deljenje da rezultat brez ostanka. Posledica tega je, ob upoštevanju vrednosti  $L_{min} = 0$ , da se enačba za  $DC_m$  poenostavi do te mere, da nam da konstanten rezultat. Za bolj splošen rezultat si moramo izbrati drugačne pogoje.

m \ n	0	1	2
1	0	500	1500
2	250	1000	1750
3	250	1000	1750
4	250	1000	1750
5	250	1000	1750
6	250	1000	1750
7	250	1000	1750
8	250	1000	1750
9	250	1000	1750
10	250	1000	1750

$L_{max} = 2000$

m \ n	0	1	2	3	4
1	0	500	1500	2500	3500
2	250	1000	2000	3000	3750
3	250	1000	2000	3000	3750
4	250	1000	2000	3000	3750
5	250	1000	2000	3000	3750
6	250	1000	2000	3000	3750
7	250	1000	2000	3000	3750
8	250	1000	2000	3000	3750
9	250	1000	2000	3000	3750
10	250	1000	2000	3000	3750

$L_{max} = 4000$

Tabela C.5 – Izračunane vrednosti  $L_{m,n}$  pri  $Q = 1000$  in  $L_{max} = 2000, 4000$ 

m \ n	$DC_m$	$O_m$	$\overline{O(m)}$
1	500	1,50	1,500
2	500	1,25	1,375
3	500	1,25	1,333
4	500	1,25	1,313
5	500	1,25	1,300
6	500	1,25	1,292
7	500	1,25	1,286
8	500	1,25	1,281
9	500	1,25	1,278
10	500	1,25	1,275

$L_{max} = 2000$

m \ n	$DC_m$	$O_m$	$\overline{O(m)}$
1	500	2,500	2,500
2	500	2,125	2,313
3	500	2,125	2,250
4	500	2,125	2,219
5	500	2,125	2,200
6	500	2,125	2,188
7	500	2,125	2,179
8	500	2,125	2,172
9	500	2,125	2,167
10	500	2,125	2,163

$L_{max} = 4000$

Tabela C.6 – Izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$  pri  $Q = 1000$  in  $L_{max} = 2000, 4000$ 

### C.5.5 Primer 2

Poglejmo si primer izračuna nekaterih izpeljanih vrednosti za enakomerno porazdelitev dolžin paketov pri naslednjih pogojih:

$$Q = 1000 \qquad L_{min} = 100 \qquad L_{max} = 1600, 3600$$

S tako izbiro pogojev dosežemo, da  $L_{max}$  ni več mnogokratnik  $Q$  in da  $L_{min} \neq 0$ . Tako se enačbe iz razdelka C.5.3 ne poenostavijo in  $DC_m$  ni več konstanten. Najprej si pogledjmo izračunane vrednosti  $p_{m,n}$  in  $L_{m,n}$  za oba primera, ki so zapisane v tabelah C.7 in C.8. Pri tem je potrebno za oba primera navesti še vrednosti spremenljivke  $N_m$ , ki določa območje vrednosti indeksa  $n$ . Pri  $L_{max} = 1600$  je vrednost  $N_m = 2$  za vsak  $m$ , zato velja  $n = 0, 1, 2$ . Pri  $L_{max} = 3600$  pa je vrednost  $N_m = 4$  za vsak  $m$ , zato velja  $n = 0, 1, \dots, 4$ .

m \ n	0	1	2
1	0	0,600	0,400
2	0,300	0,667	0,033
3	0,222	0,667	0,111
4	0,248	0,667	0,085
5	0,240	0,667	0,094
6	0,242	0,667	0,091
7	0,241	0,667	0,092
8	0,242	0,667	0,092
9	0,242	0,667	0,092
10	0,242	0,667	0,092

$$L_{max} = 1600$$

m \ n	0	1	2	3	4
1	0	0,257	0,286	0,286	0,171
2	0,120	0,286	0,286	0,286	0,022
3	0,107	0,286	0,286	0,286	0,036
4	0,109	0,286	0,286	0,286	0,034
5	0,109	0,286	0,286	0,286	0,034
6	0,109	0,286	0,286	0,286	0,034
7	0,109	0,286	0,286	0,286	0,034
8	0,109	0,286	0,286	0,286	0,034
9	0,109	0,286	0,286	0,286	0,034
10	0,109	0,286	0,286	0,286	0,034

$$L_{max} = 3600$$

Tabela C.7 – Izračunane vrednosti  $p_{m,n}$  pri  $Q = 1000$  in  $L_{max} = 1600, 3600$

m \ n	0	1	2
1	0	550	1300
2	325	1050	1575
3	267	933	1517
4	286	972	1536
5	280	959	1530
6	282	964	1532
7	281	962	1531
8	281	963	1531
9	281	963	1531
10	281	963	1531

$$L_{max} = 1600$$

m \ n	0	1	2	3	4
1	0	550	1500	2500	3300
2	311	1021	2021	3021	3561
3	288	976	1976	2976	3538
4	291	982	1982	2982	3541
5	291	981	1981	2981	3541
6	291	981	1981	2981	3541
7	291	981	1981	2981	3541
8	291	981	1981	2981	3541
9	291	981	1981	2981	3541
10	291	981	1981	2981	3541

$$L_{max} = 3600$$

Tabela C.8 – Izračunane vrednosti  $L_{m,n}$  pri  $Q = 1000$  in  $L_{max} = 1600, 3600$

V tabelah C.7 in C.8 opazimo, da se, gledano po indeksu  $m$ , izvrši prehodni pojav. To je neposredno povezano s spreminjanjem vrednosti  $DC_m$ , kar komentiramo pozneje. Vrednosti  $p_{m,n}$  in  $L_{m,n}$  se namreč obnašajo na podoben način kot  $DC_m$ , saj je njihov izračun neposredno odvisen od  $DC_m$ .

V tabelah C.7 opazimo, da se z naraščanjem indeksa  $n$  verjetnost, da bo paket oddan v tem obhodu, ne spreminja bistveno. To je posledica lastnosti enakomerne verjetnosti porazdelitve, pri kateri je verjetnost, da bo paket imel dolžino  $l$ , enakomerna v celotnem razponu njenega definicijskega območja med  $L_{min}$  in  $L_{max}$ .

V tabelah C.8 pa opazimo, da se z naraščanjem indeksa  $n$  povprečna dolžina odpravljenega paketa povečuje za vrednost  $Q$ . Tudi to je posledica lastnosti enakomerne verjetnostne porazdelitve.

Poglejmo si še izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$ , ki so našteje v tabeli C.9. Vrednosti  $O_m$  in  $\overline{O(m)}$ , za vrednost  $L_{max} = 1600$ , sta prikazani tudi v grafu na sliki C.7. Za  $DC_m$  vidimo, da ima najvišjo vrednost pri  $m = 1$ , potem pa se izniha proti

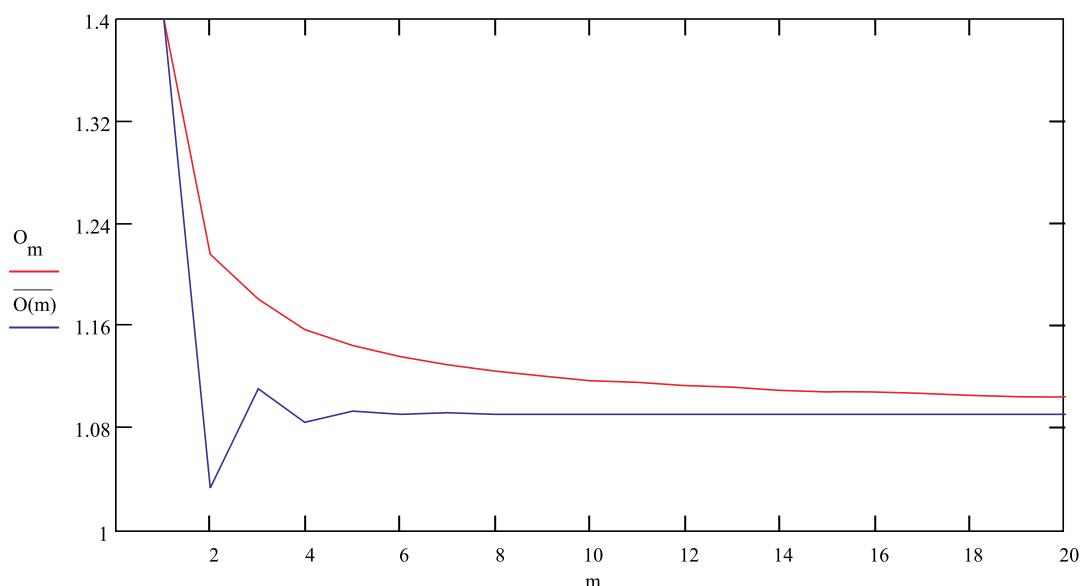


m \ n	$DC_m$	$O_m$	$\overline{O(m)}$
1	550,00	1,400	1,400
2	433,33	1,033	1,217
3	472,22	1,111	1,181
4	459,25	1,085	1,157
5	463,58	1,094	1,145
6	462,14	1,091	1,136
7	462,62	1,092	1,129
8	462,46	1,092	1,125
9	462,51	1,092	1,121
10	462,49	1,092	1,118

$L_{max} = 1600$

m \ n	$DC_m$	$O_m$	$\overline{O(m)}$
1	521,43	2,371	2,371
2	475,51	1,924	2,148
3	482,07	1,964	2,087
4	481,13	1,958	2,054
5	481,27	1,959	2,035
6	481,25	1,959	2,023
7	481,25	1,959	2,014
8	481,25	1,959	2,007
9	481,25	1,959	2,001
10	481,25	1,959	1,997

$L_{max} = 3600$

Tabela C.9 – Izračunane vrednosti  $DC_m$ ,  $O_m$  in  $\overline{O(m)}$  pri  $Q = 1000$  in  $L_{max} = 1600, 3600$ Slika C.7 – Primerjava povprečnega števila operacij pri  $Q = 1000$ ,  $L_{min} = 100$ ,  $L_{max} = 1600$ 

svoji stacionarni vrednosti. Kot posledica tega prehodnega pojava se pojavijo prehodni pojavi tudi pri ostalih spremenljivkah.  $O_m$ , na primer, ravno tako zaniha in se nato hitro ustali pri svoji stacionarni vrednosti.  $O_m$  ima najvišjo vrednost pri  $m = 1$ , saj je takrat na voljo najmanjši trenutni  $DC$ . Vzrok leži v dejstvu, da prvi paket, ki čakalno vrsto spravi v aktivno stanje, naleti na  $DC_{m-1} = DC_0 = 0$ , torej najmanjšo možno vrednost. Posledično je največji paket, ki se lahko postreže v prvem obhodu, enak vrednosti kvanta  $Q$  ( $l \leq Q$ ). Vsi naslednji paketi imajo na voljo še preostanek  $DC_{m-1} \geq 0$  iz prejšnjega obhoda. Podobna razlaga je možna tudi za nihanja po indeksu  $m$  pri vrednostih v tabelah C.7 in C.8.  $\overline{O(m)}$  predstavlja tekoče povprečje števila operacij, potrebnih za odpravo enega paketa, računano do  $m$ -tega paketa, ki se izračuna po izrazu (C.106).

### C.5.6 Poenostavljen izračun $DC_m$

Iterativno računanje vrednosti  $DC_m$ , in preko njih tudi ostalih vrednosti, je zamudno. Zato bi želeli za posebne primere te izračune poenostaviti. Iz zgornjega primera vidimo, da prehodni pojav sčasoma izzveni in čakalna vrsta, v smislu povprečnega števila operacij, potrebnih za odpravo enega paketa, preide v stacionarno stanje. Izračun vseh ostalih vrednosti je odvisen od  $DC_m$ . Ko v stacionarno stanje, glede na indeks  $m$ , preide  $DC_m$ , vanj preidejo tudi ostale vrednosti. O tem se lahko prepričamo v tabelah C.7, C.8 in C.9.

Lotimo se poenostavljenega izračuna vrednosti  $DC_m$ . Tu ločimo dva primera, ki se nanašata na dve možni varianti izračuna  $DC_m$  v enačbah v razdelku C.5.3.

Najprej izračunajmo asimptoto, h kateri teži  $DC_m$ , ko gre  $m \rightarrow \infty$ , pri pogoju  $DC_m < L_{min}$ . Stacionarno stanje nastopi, ko velja  $DC_m = DC_{m-1}$ . Takrat lahko zapišemo:

$$\begin{aligned}
DC_m &= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right. \\
&\quad \left. + 2DC_m(L_{max} - L_{min} + Q(1 - N_m)) \right] \\
&= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right] \\
&\quad + \frac{2DC_m(L_{max} - L_{min} + Q(1 - N_m))}{2(L_{max} - L_{min})} \\
&= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - 2QL_{min} - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right] \\
&\quad + DC_m + \frac{2DC_mQ(1 - N_m)}{2(L_{max} - L_{min})} \tag{C.148}
\end{aligned}$$

če okrajšamo člen  $DC_m$  na levi in desni strani zgornje enačbe, obe strani pomnožimo s členom  $2(L_{max} - L_{min})$  ter preuredimo. Končno dobimo poenostavljen rezultat za  $DC_m$ , ki velja pri pogoju  $DC_m < L_{min}$  in ga označimo z  $DC1_m$ :

$$DC1_m = \frac{L_{max}^2 - L_{min}^2 + 2QL_{min} - 2N_mQL_{max} - N_mQ^2(1 - N_m)}{2Q(1 - N_m)} \tag{C.149}$$

Izračunajmo še asimptoto, h kateri teži  $DC_m$ , ko gre  $m \rightarrow \infty$ , pri pogoju  $DC_m \geq L_{min}$ . Stacionarno stanje nastopi, ko velja  $DC_m = DC_{m-1}$ . Takrat lahko zapišemo:

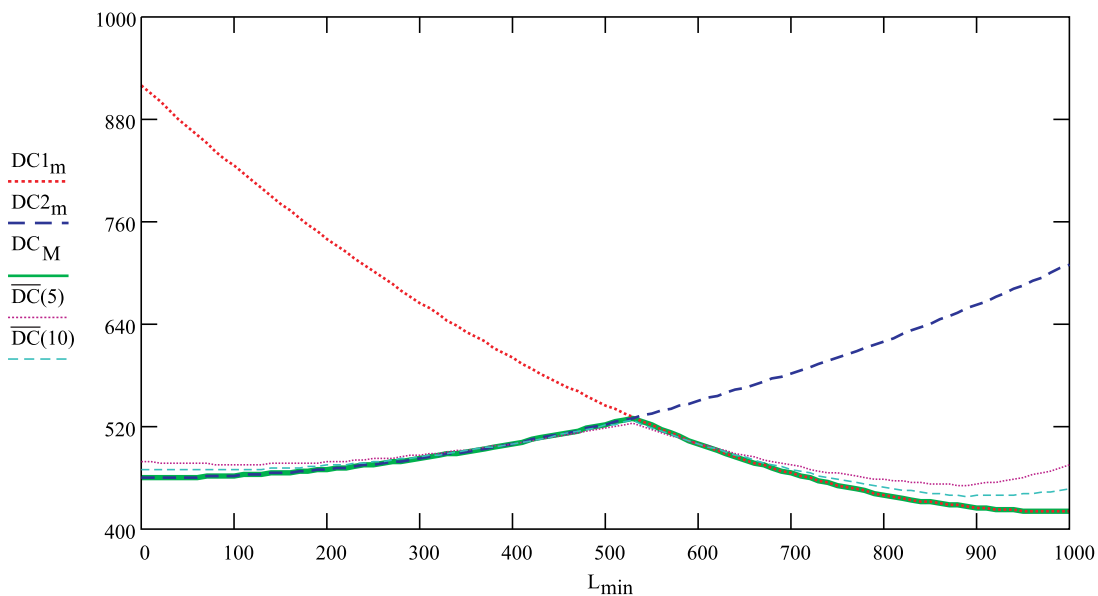
$$\begin{aligned}
DC_m &= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right. \\
&\quad \left. + 2DC_{m-1}(L_{max} - L_{min} - QN_m) \right] \\
&= \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right]
\end{aligned}$$

$$\begin{aligned}
& + \frac{2DC_{m-1}(L_{max} - L_{min} - QN_m)}{2(L_{max} - L_{min})} \\
& = \frac{1}{2(L_{max} - L_{min})} \left[ L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) \right] \\
& + DC_{m-1} - \frac{2DC_{m-1}QN_m}{2(L_{max} - L_{min})} \tag{C.150}
\end{aligned}$$

če okrajšamo člen  $DC_m$  na levi in desni strani zgornje enačbe, obe strani pomnožimo s členom  $2(L_{max} - L_{min})$  ter preuredimo. Končno dobimo poenostavljen rezultat za  $DC_m$ , ki velja pri pogoju  $DC_m \geq L_{min}$  in ga označimo z  $DC2_m$ :

$$DC2_m = \frac{L_{min}^2 - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m)}{2QN_m} \tag{C.151}$$

Sedaj se postavi vprašanje, katero izmed obeh poenostavitvev  $DC1_m$  in  $DC2_m$  uporabiti. Glede na to, da je vsakokratna izbira poenostavitve odvisna od  $L_{min}$  in od trenutne vrednosti  $DC_m$ , bi bilo potrebno za vsak paket v čakalni vrsti ugotoviti, katera poenostavitev izračuna  $DC_m$  mu glede na vrednost  $DC_{m-1}$  bolj ustreza. To je seveda nepraktično in v bistvi ne izniči iterativnosti. Naša želja je, da bi v celotnem času aktivnosti čakalne vrste uporabljali isto poenostavitev. Najprimernejšo lahko najdemo s pomočjo slike C.8.



Slika C.8 – Poenostavitve izračuna  $DC_m$  pri  $Q = 1000$  in  $L_{max} = 1600$

Na sliki C.8 so vse krivulje izrisane glede na vrednosti  $L_{min}$ , ki se nahajajo med 0 in  $Q$ . Poenostavitvi  $DC1_m$  in  $DC2_m$  primerjamo z iterativno izračunano asimptoto  $DC_M$ , pri čemer  $M$  predstavlja visoko vrednost indeksa  $m$ . Za popolnoma točno vrednost asimptote bi morala biti vrednost  $M = \infty$ , česar pa z iterativnim postopkom seveda ne moremo izračunati. V našem izračunu je vrednost  $M = 100$ , kar za potrebe grafične primerjave popolnoma zadošča.

Najprej primerjajmo iterativno izračunano asimptoto  $DC_M$  s poenostavljenima izračunoma  $DC1_m$  in  $DC2_m$ . Opazimo, da je za visoke vrednosti  $L_{min}$  boljša poenostavitev  $DC1_m$ , ki se v tem delu popolnoma prilega asimptoti  $DC_M$ . Enako velja za poenostavitev  $DC2_m$  in nizke vrednosti  $L_{min}$ . Meja med obema je malo nad polovično vrednostjo  $Q$ . Tak rezultat je bil pričakovan. Poenostavitev  $DC1_m$  velja pri pogoju  $DC_m < L_{min}$ , zato je verjetnost, da bo ta pogoj izpolnjen, večja pri višjih vrednostih in manjša pri nižjih vrednostih  $L_{min}$ . Uporaba poenostavitve  $DC1_m$  je zato bolj verjetna pri višjih vrednostih  $L_{min}$ . Ravno obratno pa velja za poenostavitev  $DC2_m$ , katere uporaba je zaradi pogoja  $DC_m \geq L_{min}$  bolj verjetna pri nižjih vrednostih  $L_{min}$ .

Vse tri do sedaj primerjane vrednosti so asimptote, izračunane pri pogoju  $m \rightarrow \infty$  in zato veljajo za velike vrednosti indeksa  $m$ , ki lahko nastopi le, če je v času aktivnosti čakalne vrste v njej veliko število paketov (visoko obremenjeno omrežje). Kaj pa se dogaja pri nižjih vrednostih  $m$ , ko je v času aktivnosti čakalne vrste v njej le nekaj paketov (nizko obremenjeno omrežje)? Katera poenostavitev je bolj primerna v takem primeru? V ta namen smo izračunali še povprečne iterativno izračunane vrednosti  $DC_m$ , ki so v grafu na sliki C.8 označene kot  $\overline{DC}(x)$ , pri čemer argument  $x = 5, 10$  označuje število paketov, ki so bili postreženi v času aktivnosti čakalne vrste. Povprečje se izračuna po enačbi (C.106). Primerjajmo rezultate  $\overline{DC}(5)$  in  $\overline{DC}(10)$  z iterativno izračunanimi rezultati  $DC_M$ . Vidimo, da se v sredini zelo dobro prilegajo, pri skrajnih vrednostih  $L_{min}$ , pa se začnejo razhajati. Razhajanje je večje pri manjšem številu paketov. To lahko pojasnimo s prehodnim pojavom, ki izzveni šele po nekaj iteracijah in v tem času močno vpliva na izbiro veljavne poenostavitve.

Problem pravilne izbire poenostavitve je težje obvladljiv pri majhnem številu paketov v čakalni vrsti. Taki pogoji pa so značilni za nizko obremenjeno omrežje, ki navadno ni zanimivo za analizo in tudi ne predstavlja nobenih problemov. Bolj nas zanimajo pogoji in delovanje pri visokih obremenitvah omrežja. V tem primeru pa imamo navadno veliko število paketov v čakalni vrsti in izbira poenostavitve je enostavna.

Izračunajmo še presečišče krivulj obeh poenostavitev. Če izenačimo enačbi za  $DC1_m$  (C.149) in  $DC2_m$  (C.151) ter jih preuredimo, dobimo kvadratno enačbo za spremenljivko  $L_{min}$ :

$$L_{min}^2 - 2QN_mL_{min} - L_{max}^2 + 2N_mQL_{max} + N_mQ^2(1 - N_m) = 0 \quad (\text{C.152})$$

Korena te kvadratne enačbe izračunamo po obrazcu

$$L_{min_{1,2}} = QN_m \pm \sqrt{L_{max}^2 - 2N_mQL_{max} - N_mQ^2(1 - 2N_m)} \quad (\text{C.153})$$

Izračunajmo korena  $L_{min_{1,2}}$  za podane vrednosti  $Q = 1000$ ,  $L_{max} = 1600$  in  $N_m = 2$ , ki so bile uporabljene tudi za izris grafov na sliki C.8. Dobljena rezultata sta  $L_{min_1} = 3469,7$  in  $L_{min_2} = 530,3$ . Prva rešitev v našem primeru ne velja, saj za  $L_{min}$  velja pogoj  $L_{min} \leq Q$ . Druga rešitev pa se lepo sklada z grafom na sliki C.8.

Mejno vrednost  $L_{min}$ , ki določa uporabo posamezne poenostavitve, lahko torej enostavno določimo s pomočjo enačbe (C.152). Glede na to, da je v večini omrežij vrednost  $L_{min}$  dokaj nizka (najmanjše dovoljene dolžine paketov so nekaj deset oktetov), je zato večinoma uporabljena kar poenostavitev  $DC2_m$ .

### C.5.7 Poenostavljen izračun $O_m$

Zanima nas še, kako poenostavljen izračun za  $DC_m$  vpliva na izračun števila operacij  $O_m$ , potrebnih za odpravo  $m$ -tega paketa. Ker v enačbi (C.136) za  $O_m$  nastopa člen  $DC_{m-1}$ , moramo zopet upoštevati dva različna primera.

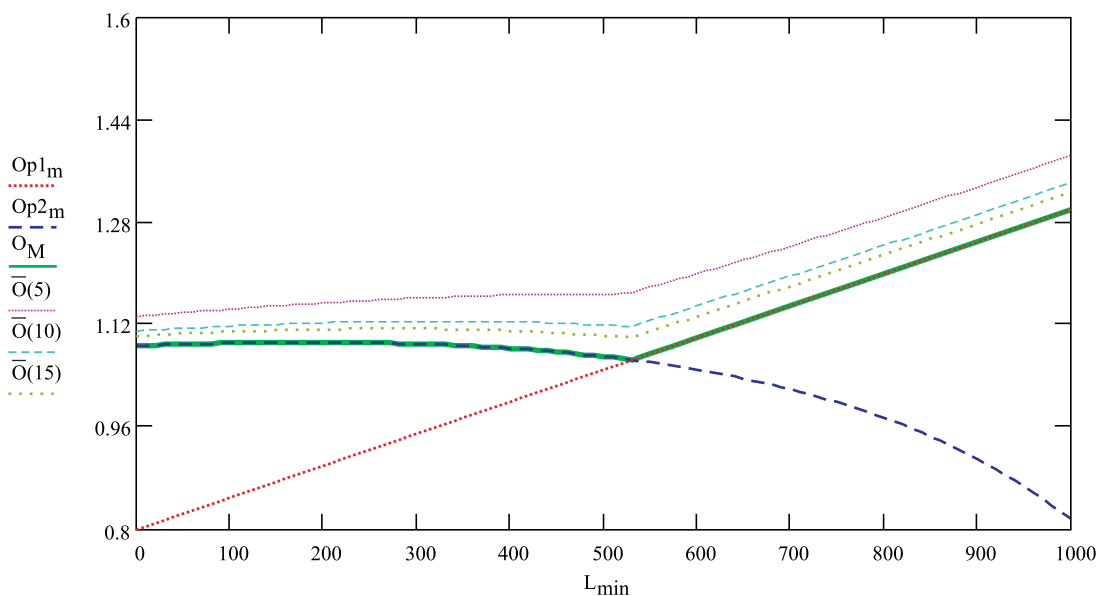
Prvi primer velja, ko je izpolnjen pogoj  $DC_m < L_{min}$ . Uporabimo poenostavljen izračun  $DC1_m$  (C.149), ga vstavimo v enačbo za  $O_m$  (C.136), izraz poenostavimo in dobimo poenostavljen izračun  $Op1_m$

$$Op1_m = \frac{L_{min} + L_{max}}{2Q} \quad (C.154)$$

Drugi primer velja, ko je izpolnjen pogoj  $DC_m \geq L_{min}$ . Uporabimo poenostavljen izračun  $DC2_m$  (C.151), ga vstavimo v enačbo za  $O_m$  (C.136), izraz poenostavimo in dobimo poenostavljen izračun  $Op2_m$

$$Op2_m = 1 - \frac{(1 - N_m)(L_{max}^2 - L_{min}^2 - N_m Q^2)}{2N_m Q(L_{max} - L_{min})} \quad (C.155)$$

Na sliki C.9 so vse krivulje izrisane glede na vrednosti  $L_{min}$ , ki se nahajajo med 0 in  $Q$ . Poenostavitvi  $Op1_m$  in  $Op2_m$  najprej primerjamo z iterativno izračunano asimptoto  $O_M$ , pri čemer  $M$  predstavlja visoko vrednost indeksa  $m$ . Za popolnoma točno vrednost asimptote bi morala biti vrednost  $M = \infty$ , česar pa z iterativnim postopkom seveda ne moremo izračunati. V našem izračunu je vrednost  $M = 100$ , kar za potrebe grafične primerjave popolnoma zadošča.



Slika C.9 – Poenostavitve izračuna  $O_m$  pri  $Q = 1000$  in  $L_{max} = 1600$

Vidimo, da se obe poenostavitvi odsekoma lepo prilegata iterativno izračunani asimptoti. Razmere so torej zelo podobne tistim pri poenostavljenem izračunu  $DC1_m$  in  $DC2_m$ , kar je bilo pričakovano. Poenostavitev  $Op1_m$  se bolje obnese pri visokih,  $Op2_m$  pa pri nizkih vrednostih  $L_{min}$ . Razlogi za to so enaki kot pri  $DC1_m$  in  $DC2_m$  in so navedeni v prejšnjem razdelku.

Zopet se postavi vprašanje, katero izmed obeh poenostavitvev  $Op1_m$  ali  $Op2_m$  uporabiti. Glede na to, da je vsakokratna izbira poenostavitve odvisna od  $L_{min}$  in od trenutne vrednosti  $O_m$ , bi bilo potrebno za vsak paket v čakalni vrsti ugotoviti, katera poenostavitvev mu glede na vrednost  $DC_{m-1}$  bolj ustreza. To je seveda nepraktično in v bistvi ne izniči iterativnosti. Naša želja je, da bi v celotnem času aktivnosti čakalne vrste uporabljali isto poenostavitvev. Najprimernejšo lahko najdemo s pomočjo slike C.9.

Vse tri do sedaj primerjane vrednosti so asimptote, izračunane pri pogoju  $m \rightarrow \infty$ , in zato veljajo za velike vrednosti indeksa  $m$ , ki lahko nastopi le, če je v času aktivnosti čakalne vrste v njej veliko število paketov (visoko obremenjeno omrežje). Kaj pa se dogaja pri nižjih vrednostih  $m$ , ko je v času aktivnosti čakalne vrste v njej le nekaj paketov (nizko obremenjeno omrežje)? V ta namen smo izračunali še povprečne iterativno izračunane vrednosti  $O_m$ , ki so v grafu na sliki C.9 označene kot  $\bar{O}(x)$ , pri čemer argument  $x = 5, 10, 15$  označuje število paketov, ki so bili postreženi v času aktivnosti čakalne vrste. Povprečje se izračuna po enačbi (C.106). Primerjajmo rezultate  $\bar{O}(5)$ ,  $\bar{O}(10)$  in  $\bar{O}(15)$  z iterativno izračunanimi rezultati  $O_M$ . Vidimo, da se z večanjem števila upoštevanih paketov vse bolj približujemo asimptoti. Izbira prave poenostavitve pa je zopet odvisna le od vrednosti  $L_{min}$ , meja se nahaja v presečišču krivulj  $Op1_m$  ali  $Op2_m$ , ki ga izračunamo z izenačenjem obeh poenostavitvev  $Op1_m = Op2_m$ . Dobimo enak rezultat kot pri izračunu presečišča  $DC1_m$  in  $DC2_m$ , ki je zapisan z enačbo (C.152).

Mejno vrednost  $L_{min}$ , ki določa uporabo posamezne poenostavitve, lahko torej enostavno določimo s pomočjo enačbe (C.152). Glede na to, da je v večini omrežij vrednost  $L_{min}$  dokaj nizka (najmanjše dovoljene dolžine paketov so nekaj deset oktetov), je zato večinoma uporabljena kar poenostavitvev  $Op2_m$ .

## Literatura

- [1] J. M. Arco, D. Meziat, B. Alarcos: *Deficit Round Robin Alternated - A New Scheduling Algorithm*, Conference on Protocols For Multimedia Systems - PROMS2000, Cracow, Poland, October 22-25, 2000
- [2] Jon C.R. Bennett and H. Zhang: *Why WFQ Is Not Good Enough for Integrated Services Networks*, Proceedings of NOSSDAV'96, April, 1996
- [3] Jon C.R. Bennett and H. Zhang: *Hierarchical Packet Fair Queueing Algorithms*, IEEE/ACM Transactions on Networking, 5(5):675-689, October 1997
- [4] Jon C.R. Bennett and H. Zhang: *WF2Q: Worst-case Fair Weighted Fair Queueing*, INFOCOM'96, March, 1996
- [5] Yoram Bernet: *The Complementary Roles of RSVP and Differentiated Services in the Full-Service QoS Network*, IEEE Communications Magazine, February 2000
- [6] Uyless Black: *Emerging Communications Technologies*, Prentice Hall, 1997
- [7] W. Chan, J. Nieh: *Group Ratio Round-Robin: An  $O(1)$  Proportional Share Scheduler*, Tech.Rep. CUCS 012-03, Department of Computer Science, Columbia University, April 2003
- [8] A. Demers, S. Keshav, S. Shenker: *Analysis and Simulation of a Fair Queueing Algorithm*, ACM SIGCOMM Computer Communication Review, Volume 19, Issue 4, September 1989
- [9] Constantinos Dovrolis, Parameswaran Ramanathan: *A Case for Relative Differentiated Services and the Proportional Differentiation Model*, IEEE Network, September 1999
- [10] Tim Godfrey: <http://www.us.design-reuse.com/articles/article6865.html>, December, 2003
- [11] S. Golestani: *A self-clocked fair queueing scheme for broadband applications*, Proc. IEEE INFOCOM, Toronto, ON, Canada, June 1994
- [12] R. Guérin, V. Peris: *Quality-of-Service in Packet Networks Basic Mechanisms and Directions*, Computer Networks, 31(3):169-189, February 1999
- [13] R. Guérin, S. Blake, S. Herzog: *Aggregating RSVP-based QoS Requests - IETF draft*, IETF, November 1997
- [14] Go Hasegawa, Takahiro Matsuo, Masayuki Murata, Hideo Miyahara: *Comparisons of packet scheduling algorithms for fair service among connections on the internet*, Journal of High Speed Networks, Volume 12, Issue 1,2, 2003
- [15] David Hayes, Michael Rumsewicz, Andrew Lachlan: *Quality of Service Driven Packet Scheduling Disciplines for Real-Time Applications: Looking Beyond Fairness*, Proceedings of the Conference on Computer Communications (IEEE Infocom), New York, March 1999

- 
- [16] Peter Homan: *Konfiguracija in krmiljenje čakalnih vrst za zagotavljanje kakovosti telekomunikacijskih storitev v paketnih stikalih*, doktorska disertacija, Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, 2001
- [17] ITU E.430: *Telephone Network and ISDN - Quality of Service, Network Management and Traffic Engineering*, International Telecommunication Union, Geneva, 1992
- [18] ITU E.800: *Terms and definitions related to quality of service and network performance including dependability*, International Telecommunication Union, Geneva, August 1995
- [19] ITU G.114: *One-way transmission time*, International Telecommunication Union, Geneva, February 1996
- [20] ITU G.131: *Control of talker echo*, International Telecommunication Union, Geneva, August 1996
- [21] ITU X.902: *Information technology - Open distributed processing - Reference Model*, International Telecommunication Union, Geneva, November 1995
- [22] Jayaraman Iyer: *Integrated Services on the Internet*,  
[http://www.cse.ohio-state.edu/~jain/cis788-97/ftp/integrated\\_services/](http://www.cse.ohio-state.edu/~jain/cis788-97/ftp/integrated_services/)
- [23] Anton Kos: *Prenos govora preko podatkovnih omrežij*, magistrsko delo, Univerza v Ljubljani, Fakulteta za elektrotehniko, april 1998
- [24] Anton Kos, Sašo Tomažič: *Izohroni Ethernet*, ERK '97, strani 91-94, Portorož 1997
- [25] Anton Kos, Sašo Tomažič: *Existing LAN Technologies as a Part of Future Communication Networks*, VITEL, D21-D26, Ljubljana 1998
- [26] Anton Kos, Sašo Tomažič: *Prenos podatkov v realnem času preko javnega Internet omrežja*, ERK '99, strani 165-168, Portorož 1999
- [27] Anton Kos, Sašo Tomažič: *Internet and Transmission in Real Time*, IEEE Middle East Workshop on Networking, Beirut, 2.-4. November 1999
- [28] Anton Kos, Sašo Tomažič: *A Two-bit Scheme for Service Differentiation in IP Networks*, Melecon 2000, Vol. 1, str. 148-151, May 29-31, 2000, Cyprus.
- [29] Borut Klepec, Anton Kos, Sašo Tomažič: *Vpliv diferenciacije IP prometa na delovanje aplikacij v realnem času*, ERK 2000, 123-126, Portorož, september 2000
- [30] Anton Kos, Borut Klepec: *Supporting Real-time Applications in Differentiated Services Network*, Simpozij VITEL, Ljubljana, October 2000
- [31] Anton Kos, Borut Klepec, Sašo Tomažič: *Real-time Application Performance in Differentiated Services Network*, TENCON 2001, August 19-22, 2001, Singapore
- [32] Borut Klepec, Anton Kos: *Performance of VoIP Applications in a Simple Differentiated Services Network Architecture*, EUROCON 2001, July 5-7, 2001, Bratislava, Slovakia
- [33] Anton Kos: *Prenos podatkov v realnem času in zagotavljanje kakovosti storitev v IP omrežjih*, Raziskovalno delo podiplomskih študentov Slovenije - Novo tisočletje: Naravoslovje in tehnika, Društvo mladih raziskovalcev Slovenije, 2001
- [34] Anton Kos: *Prenos podatkov v realnem času in zagotavljanje kakovosti storitev v IP omrežjih*, Raziskovalno delo podiplomskih študentov v Sloveniji - novo tisočletje, Naravoslovje in tehnika : jesen - zima. Ljubljana: Društvo mladih raziskovalcev Slovenije, 2001, str. 209-221



- 
- [35] Anton Kos, Borut Klepec, Sašo Tomažič; *Techniques for Performance Improvement of VoIP Applications*, Melecon 2002, str. 250-254, May 7-9, 2002, Cairo, Egipt
- [36] Anton Kos, Sašo Tomažič; *Kakovost storitve*, ERK 2003, Zvezek AB, strani 103-106, Ljubljana, september 2003
- [37] Anton Kos, Robert Verlič, Sašo Tomažič; *Kakovost storitve v paketnih omrežjih*, Elektrotehniški vestnik, vol. 71(3), strani 103-108, 2004
- [38] Anton Kos, Sašo Tomažič; *Združitev integriranih in diferenciranih storitev*, ERK 2004, Zvezek A, strani 195-198, Portorož, september 2004
- [39] Anton Kos; *Paketna omrežja in kakovost storitve*, ZNANSTVENO delo podiplomskih studentov v Sloveniji - "publish or perish!", e-knjiga, Društvo mladih raziskovalcev Slovenije, 2004
- [40] Anton Kos, Sašo Tomažič; *Nov način združevanja RSVP pretokov*, ERK 2006, Portorož, september 2005
- [41] Andrej Kos: *Prometni inženiring v omrežjih MPLS*
- [42] Salil S. Kanhere, Harish Sethu: *On the Latency Bound of Deficit Round Robin*, Proceedings of the International Conference on Computer Communications and Networks, Miami, Florida, USA, October 14-16, 2002
- [43] Salil S. Kanhere, Harish Sethu: *Fair, Efficient and Low-Latency Packet Scheduling Using Nested Deficit Round Robin*, Proceedings of the IEEE Workshop on High Performance Switching and Routing, Dallas, Texas, USA, May 29-31, 2001
- [44] Salil S. Kanhere, Harish Sethu, Alpa B. Parekh: *Fair and Efficient Packet Scheduling Using Elastic Round Robin*, IEEE Transactions on Parallel and Distributed Systems, Volume 13, Issue 3, March 2002
- [45] Salil S. Kanhere, Harish Sethu, Alpa B. Parekh: *Low-latency guaranteed-rate scheduling using elastic round robin*, Computer Communications, Volume 25, no. 14, pp. 1315-1322, September 2002
- [46] Salil S. Kanhere, Harish Sethu: *On the Latency Bound of Pre-Order Deficit Round Robin*, Proceedings of the IEEE Conference on Local Computer Networks, Tampa, Florida, USA, November 6-8, 2002
- [47] Srinivasan Keshav: *An Engineering Approach to Computer Networking: ATM networks, the internet, and the telephone network*, Addison-Wesley, Reading, Massachusetts, 1997
- [48] Peter J.B. King: *Computer and Communication Systems Performance Modelling*, Prentice Hall, 1990
- [49] Leonard Kleinrock: *Queueing Systems, Volume I: Theory*, John Wiley & Sons, 1975
- [50] Leonard Kleinrock: *Queueing Systems, Volume II: Computer Applications*, John Wiley & Sons, 1976
- [51] Jung-Shian Li: *An Evaluation of Deficit Round Robin Fair Queuing Applied in Router Congestion Control*, Journal of Information Science and Engineering 18, 333-339, 2002
- [52] M.H. MacGregor, W. Shi: *Deficits for Bursty Latency-critical Flows: DRR++*, Proceedings of IEEE International Conference on Networks (ICON 2000), Singapore, September 2000
- [53] M.H. MacGregor, W. Shi: *Scheduling Latency-Critical Traffic: A Measurement Study of DRR+ and DRR++*, Workshop on High Performance Switching and Routing (HPSR 2002), Kobe, Japan, May 26-29, 2002

- 
- [54] James Martin, Kathleen Kavanagh Chapman, Joe Leben: *Enterprise Networking: data link subnetworks*, Prentice Hall, 1996
  - [55] John Nagle: *On Packet Switches with Infinite Storage*, IEEE Transactions on Communications, Volume 35, pp 435-438, 1987
  - [56] Anindya Neogi, Tzi-cher Chiueh, Paul Stirpe: *Performance Analysis of an RSVP-Capable Router*, IEEE Network, September 1999
  - [57] Athanasios Papoulis; *Probability, random variables, and stochastic processes*, McGraw-Hill, 1991
  - [58] Abhay K. Parekh, Robert G. Gallager: *A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case*, IEEE/ACM Transactions on Networking, Volume 1, Issue 3, June 1993
  - [59] Abhay K. Parekh, Robert G. Gallager: *A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple-Node Case*, IEEE/ACM Transactions on Networking, Volume 2, Issue 2, April 1994
  - [60] Nihal Pekergin: *Stochastic Bounds on Delays of Fair Queueing Algorithms*, Proceedings of the Conference on Computer Communications (IEEE Infocom), New York, March 1999
  - [61] RFC 1349: *Type of Service in the Internet Protocol Suite*, IETF, July 1992
  - [62] RFC 1633: *Integrated Services in the Internet Architecture: an Overview*, IETF, July 1994
  - [63] RFC 1932: *IP over ATM: A Framework Document*, IETF, April 1996
  - [64] RFC 1946: *Native ATM Support for ST2+*, IETF, May 1996
  - [65] RFC 2113: *IP Router Alert Option*, IETF, February 1997
  - [66] RFC 2205: *Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification*, IETF, September 1997
  - [67] RFC 2208: *Resource ReSerVation Protocol (RSVP) - Version 1 Applicability Statement Some Guidelines on Deployment*, IETF, September 1997
  - [68] RFC 2209: *Resource ReSerVation Protocol (RSVP) - Version 1 Message Processing Rules*, IETF, September 1997
  - [69] RFC 2210: *The Use of RSVP with IETF Integrated Services*, IETF, September 1997
  - [70] RFC 2211: *Specification of the Controlled-Load Network Element Service*, IETF, September 1997
  - [71] RFC 2212: *Specification of Guaranteed Quality of Service*, IETF, September 1997
  - [72] RFC 2474: *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, IETF, December 1998
  - [73] RFC 2475: *An Architecture for Differentiated Services*, IETF, December 1998
  - [74] RFC 2597: *Assured Forwarding PHB Group*, IETF, June 1999
  - [75] RFC 2598: *An Expedited Forwarding PHB*, IETF, June 1999
  - [76] RFC 2711: *IPv6 Router Alert Option*, IETF, October 1999
  - [77] RFC 2746: *RSVP Operation Over IP Tunnels*, IETF, January 2000
  - [78] RFC 2814: *SBM (Subnet Bandwidth Manager): A Protocol for RSVP-based Admission Control over IEEE 802-style networks*, IETF, May 2000
  - [79] RFC 2815: *Integrated Service Mappings on IEEE 802 Networks*, IETF, May 2000
  - [80] RFC 2996: *Format of the RSVP DCLASS Object*, IETF, November 2000
  - [81] RFC 2998: *A Framework for Integrated Services Operation over Diffserv Networks*, IETF, November 2000

- 
- [82] RFC 3031: *Multiprotocol Label Switching Architecture*, IETF, January 2001
  - [83] RFC 3168: *The Addition of Explicit Congestion Notification (ECN) to IP*, IETF, September 2001
  - [84] RFC 3246: *An Expedited Forwarding PHB (Per-Hop Behavior)*, IETF, March 2002
  - [85] RFC 3247: *Supplemental Information for the New Definition of the EF PHB*, IETF, March 2002
  - [86] RFC 791: *Internet Protocol*, IETF, September 1981
  - [87] Kiran Rege et al: *QoS Management in Trunk-and-Branch Switched Ethernet Networks*, IEEE Communications Magazine, December 2002
  - [88] Debanjan Saha, Sarit Mukherjee, Satish K. Tripathi: *Carry-Over Round Robin: A Simple Cell Scheduling Mechanism for ATM Networks*, IEEE/ACM Transactions on Networking, Vol. 6, No. 6, December 1998
  - [89] Chuck Semeria: *Supporting Differentiated Service Classes - Queue Scheduling Disciplines*, White Paper, Juniper Networks, Inc., December 2001
  - [90] Chuck Semeria: *Supporting Differentiated Service Classes - Active Queue Memory Management*, White Paper, Juniper Networks, Inc., February 2002
  - [91] Chuck Semeria, John W. Stewart: *Supporting Differentiated Service Classes in Large IP Networks*, White Paper, Juniper Networks, Inc., December 2001
  - [92] Syed Ijlal Ali Shah: *Bringing Comprehensive Quality of Service Capabilities to Next-Generation Networks*, White Paper, Motorola, Inc., October 2001
  - [93] M. Shreedhar, George Varghese: *Efficient Fair Queuing Using Deficit Round Robin*, IEEE/ACM Transactions on Networking, Volume 4, Issue 3, June 1996
  - [94] Slovar ITkT, <http://ltfe.org/>
  - [95] Tomi Solala: *A Framework for Integrated Services over Diffserv Network*, University of Helsinki, November 2000
  - [96] Donpaul Stephens, Jon Bennett, Hui Zhang: *Implementing Scheduling Algorithms in High Speed Networks*, IEEE JSAC Special Issue on High Performance Switches/Routers, September 1999
  - [97] Donpaul Stephens, Hui Zhang: *Implementing Scheduling Algorithms in High Speed Networks*, IEEE JSAC Special Issue on High Performance Switches/Routers, September 1999
  - [98] D. Stiliadis, A.Varma: *Latency-Rate Servers - A General Model for Analysis of Traffic Scheduling Algorithms*, IEEE/ACM Transactions on Networking, Volume 6, Issue 5, October 1998
  - [99] D. Stiliadis, A.Varma: *Rate-Proportional Servers: A Design Methodology for Fair Queueing Algorithms*, IEEE/ACM Transactions on Networking, Volume 6, Issue 1, February 1998
  - [100] Ion Stoica, Hui Zhang: *Implementing Distributed Packet Fair Queueing in a Scalable Switch Architecture*, IEEE INFOCOM'98 Conference, San Francisco, CA, March 1998
  - [101] Ion Stoica, Scott Shenker, Hui Zhang: *Core-Stateless Fair Queueing: A Scalable Architecture to Approximate Fair Bandwidth Allocations in High Speed Networks*, IEEE/ACM Transactions on Networking, Volume 11, Issue 1, February 2003
  - [102] Yeali S. Sun, Y.C. Tu and W.K. Shih: *Frame-based Worst-Case Weighted Fair Queueing with Jitter Control*, IEICE Transactions on Communications, Vol.E84-B, No.8, pp.2266-2278, August 2001

- 
- [103] TTI Vanguard: *ATM Lexicon*, <http://www.tticom.com/atmglosy/atmlex.htm#qos>
- [104] *The Token Bucket(Leaky Bucket) Model*,  
<http://qbone.internet2.edu/bb/Bucket.doc/>
- [105] P. F. Chimento *Standard Token Bucket Terminology*,  
<http://qbone.internet2.edu/bb/Traffic.pdf/>
- [106] Shih-Chiang Tsao, Ying-Dar Lin: *Pre-order Deficit Round Robin: a new scheduling algorithm for packet-switched networks*, *Computer Networks: The International Journal of Computer and Telecommunications Networking*, Volume 35, Issue 2-3, February 2001
- [107] Srinivas Vegesna: *IP Quality of Service*, Cisco Press, Indianapolis, 2001
- [108] Rolland Vida, Adrian Sergiu Darabant: *Current Technologies for QoS and Traffic Control*, Babes-Bolyai University, Faculty of Mathematics and Computer Science Research Seminars, Seminar on Computer Science, Preprint, no. 2, 1998
- [109] Andreas Vogel: *Distributed Multimedia and Quality of Service: A Survey*, <http://www.dstc.edu.au/AU/staff/andreas-vogel/papers/ieee-mm.ps>
- [110] Lixia Zhang, et al: *RSVP: A New Resource ReSerVation Protocol*, *IEEE Network*, September 1993
- [111] H. Zhang and D. Ferrari: *Rate-Controlled Static-Priority Queueing*, *Proceedings of INFOCOM'93*, San Francisco, CA, April 1993
- [112] H. Zhang: *Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks*, *Proceedings of the IEEE*, 83(10), Oct 1995
- [113] H. Zhang and D. Ferrari: *Rate-Controlled Service Disciplines*, *Journal of High Speed Networks: Special Issue on Quality of Service*, 3(4), 1994
- [114] Jian Zhang, Timo Hamalainen, Jyrki Joutsensalo: *Stochastic Analysis of Upper Delay Bound of GPS-Based Packetized Fair Queueing Algorithms*, Dept. of Mathematical Information Technology, University of Jyvaskyla, Finland
- [115] Yunkai Zhou, Harish Sethu: *On the Relationship Between Absolute and Relative Fairness Bounds*, *IEEE Communications Letters*, Vol. 6, No. 1, January 2002

## Seznam kratic

<b>ADREP</b>	Admission Reply	Odgovor na zahtevo za dostop
<b>ADREQ</b>	Admission Request	Zahteva za dostop
<b>ADSL</b>	Asymmetric Digital Subscriber Line	Asimetrična digitalna naročniška linija
<b>AF</b>	Assured Forwarding	Zagotovljeno posredovanje
<b>AFM</b>	Absolute Fairness Measure	Absolutna mera pravičnosti
<b>ARCnet</b>	Attached Resource Computer network	Računalniško omrežje priključenih virov
<b>ATM</b>	Asynchronous Transfer Mode	Asinhroni način prenosa
<b>BE</b>	Best Effort	Po najboljših zmožnostih
<b>BECN</b>	Backward Explicit Congestion Notification	Neposredno povratno obvestilo o zamašitvi
<b>BRR</b>	Bit-by-bit Round Robin	Bitno krožno odpravljanje
<b>CDDI</b>	Copper Distributed Data Interface	Žično porazdeljeni podatkovni vmesnik
<b>CIR</b>	Committed Information Rate	Zajamčena informacijska hitrost
<b>CLS</b>	Controlled Load Service	Storitev nadzorovanega bremena
<b>DC</b>	Deficit Counter	Števec primanjkljaja
<b>DIS</b>	Differentiated and Integrated Services	Diferencirane in integrirane storitve
<b>DQDB</b>	Distributed Queue Dual Bus	Dvojno vodilo s porazdeljeno čakalno vrsto
<b>DRR</b>	Deficit Round Robin	Deficitarno krožno odpravljanje
<b>DS</b>	Differentiated Services (Diffserv)	Diferencirane storitve
<b>DSCP</b>	Differentiated Services Point Code	Kodna oznaka diferenciranih storitev
<b>ECN</b>	Explicit Congestion Notification	Neposredno opozorilo o zamašitvi
<b>EPD</b>	Early Packet Discard	Zgodnje odmetavanje paketov
<b>EF</b>	Expedited Forwarding	Ekspeditivno posredovanje
<b>FCFS</b>	First Come First Serve	Prvi prispe prvi postrežen
<b>FDDI</b>	Fiber Distributed Data Interface	Optični porazdeljeni podatkovni vmesnik
<b>FEC</b>	Forwarding Equivalence Class	Ekvivalentni posredovalni razred
<b>FECN</b>	Forward Explicit Congestion Notification	Neposredno vnaprejšnje obvestilo o zamašitvi
<b>FIFO</b>	First In First Out	Prvi noter prvi ven
<b>FQ</b>	Fair Queueing	Pravično odpravljanje
<b>FR</b>	Frame Relay	Blokovno posredovanje

<b>FRED</b>	Fair Random Early Detection	Pravično naključno zgodnje odmetavanje paketov
<b>FTP</b>	File Transfer Protocol	Protokol za prenos datotek
<b>GPS</b>	Generalized Processor Sharing	Posplošena souporaba procesnih zmogljivosti
<b>GS</b>	Guaranteed Service	Zagotovljena storitev
<b>HRR</b>	Hierarchical Round Robin	Hierarhično krožno odpravljanje
<b>ICMP</b>	Internet Control Message Protocol	Internetni protokol za izmenjavo kontrolnih sporočil
<b>IEEE</b>	Institute of Electrical and Electronics Engineers	Inštitut inženirjev elektrotehnike in elektronike
<b>IETF</b>	Internet Engineering Task Force	Skupina za razvoj Interneta
<b>IP</b>	Internet Protocol	Internetni protokol
<b>IS</b>	Integrated Services	Integrirane storitve
<b>ISDN</b>	Integrated Services Digital Network	Digitalno omrežje z integriranimi storitvami
<b>ITU</b>	International Telecommunications Union	Mednarodna telekomunikacijska unija
<b>LAN</b>	Local Area Network	Lokalno omrežje
<b>LIFO</b>	Last In First Out	Zadnji noter prvi ven
<b>LIRO</b>	Last In Random Out	Zadnji noter naključno ven
<b>LR</b>	Latency-Rate	Zakasnitev - Bitna hitrost
<b>MAN</b>	Metropolitan Area Network	Mestno omrežje
<b>MBZ</b>	Must Be Zero	Vrednost mora biti enaka nič
<b>MPLS</b>	Multiprotocol Label Switching	Multiprotokolna komutacija na osnovi label
<b>PGPS</b>	Packet-By-Packet Generalized Processor Sharing	Paketna posplošena souporaba procesnih zmogljivosti
<b>PHB</b>	Per-Hop Behaviour	Obnašanje vedenjske skupine paketov
<b>PPP</b>	Point-To-Point Protocol	Protokol točka točka
<b>PPD</b>	Partial Packet Discard	Delno odmetavanje paketov
<b>PQ</b>	Priority Queueing	Prednostno odpravljanje
<b>PS</b>	Processor Sharing	Souporaba procesnih zmogljivosti
<b>PSTN</b>	Public Switched Telephone Network	Javno komutirano telefonsko omrežje
<b>QoE</b>	Quality Of Experience	Kakovost izkušnje
<b>QoS</b>	Quality Of Service	Kakovost storitve
<b>RED</b>	Random Early Detection	Naključno zgodnje odmetavanje paketov
<b>RFC</b>	Request For Comments	Poziv za komentarje
<b>RF</b>	Relative Fairness	Relativna pravičnost
<b>RFM</b>	Relative Fairness Measure	Relativna mera pravičnosti
<b>RIFO</b>	Random In First Out	Naključno noter prvi ven
<b>RIRO</b>	Random In Random Out	Naključno noter naključno ven
<b>RR</b>	Round Robin	Krožno odpravljanje
<b>RSVP</b>	Resource Reservation Protocol	Protokol za rezervacijo virov

---

<b>RTP</b>	Real-time Transport Protocol	Protokol za prenos v realnem času
<b>SCDRR</b>	Sub-Critical Deficit Round Robin	Podkritično deficitarno krožno odpravljanje
<b>SCFQ</b>	Self-Clocked Fair Queueing	Pravično odpravljanje z lastnim proženjem
<b>SCTP</b>	Stream Control Transmission Protocol	Protokol za krmiljenje prenosa pretokov
<b>SDH</b>	Synchronous Digital Hierarchy	Sinhrona digitalna hierarhija
<b>SLA</b>	Service Level Agreement	Dogovor o nivoju storitev
<b>TCP</b>	Transmission Control Protocol	Protokol za nadzor prenosa
<b>ToS</b>	Type Of Service	Tip storitve
<b>UDP</b>	User Datagram Protocol	Uporabniški datagramski protokol
<b>VC</b>	Virtual Clock	Navidezna ura
<b>WAN</b>	Wide Area Network	Prostrano omrežje
<b>WBRR</b>	Weighted Bit-By-Bit Round Robin	Uteženo bitno krožno odpravljanje
<b>WF<sup>2</sup>Q</b>	Worst Case Fair Weighted Fair Queueing	Uteženo pravično odpravljanje za najslabši primer
<b>WFQ</b>	Weighted Fair Queueing	Uteženo pravično odpravljanje
<b>WRED</b>	Weighted Random Early Detection	Uteženo naključno zgodnje odmetavanje paketov
<b>WRR</b>	Weighted Round Robin	Uteženo krožno odpravljanje
<b>WWW</b>	World Wide Web	Svetovni splet
<b>xDSL</b>	Digital Subscriber Line	Digitalna naročniška linija





## Izjava

Izjavljam, da sem doktorsko disertacijo samostojno izdelal pod vodstvom mentorja prof. dr. Saša Tomažiča, rednega profesorja na Fakulteti za elektrotehniko v Ljubljani. Izkazano pomoč drugih sodelavcev sem v celoti navedel v zahvali.

Anton Kos