

Mrežni promet i performanse

5. Laboratorijska Vježba

Jakša Fajs

Robert Kunštek

Borna Majerić

Fran Saganić

Anteo Vukasović

1. Ukratko **opišite svoju implementaciju** sustava i postupak razvoja.

Potrebno je napraviti generator prometa koji u određenom stanju Markovljeva lanca šalje promet negdje kroz mrežu. Odredili smo tri tipa prometa, umrežene igre, društvene mreže i strujanje videa koji su ujedno i stanja Markovljevog lanca. Svi ti podaci su obrađeni u 2. laboratorijskoj vježbi. Iz tih tipova prometa možemo dobiti distribuciju za veličinu prometa i vrijeme slanja. Naš generator prometa će u nekom stanju Markovljevog lanca generirati promet 1-30 sekundi. Promet će imati veličinu određenu prikladnom distribucijom i imati će međudolazno vrijeme paketa isto određeno prikladnom distribucijom. Generator prometa će slati *ping* poruke na server u lokalnoj mreži. Nakon isteka vremena generator će promijeniti stanje u lancu i ponavljati proces. *Ping* je napravljen u 4. laboratorijskoj vježbi. Tijekom rada generatora prometa će se poslani promet vidjeti u *Wiresharku*.

2. Navedite i opišite razvijene modele prometa odabranih usluga. Usporedite srednju vrijednost i medijan snimljenog prometa sa srednjom vrijednosti i medijanom distribucija kojima ste taj promet opisivali.

Odabrali smo sljedeće usluge:

- Umrežene video igre - npr. agar.io, steam igrice
- Društvene mreže - npr. facebook, twitter
- Servis video strujanja - npr. youtube, twitch

Podatke o najboljim distribucijama smo izvukli iz 2. laboratorijskih vježbi članova

Umrežene video igre:

Najbolja distribucija za veličinu paketa:

dweibull(c = 0.51, loc = 115.0, scale = 41.26)

Najbolja distribucija za međudolazna vremena:

pareto(b = 11.52, loc = -0.09, scale = 0.09)

Usporedba vrijednosti distribucije i podataka:

===VELIČINA PAKETA===

DATA

Medijan: 115.0

Prosječna vrijednost: 157.42186972662842

Maksimalna vrijednost1514.0

Minimalna vrijednost74.0

DISTRIBUCIJA

Medijan: 114.98185962615287

Prosječna vrijednost: 112.370556214154

Maksimalna vrijednost4608.281598272081

Minimalna vrijednost-4121.691930751631

===MEĐUDOLAZNO VRIJEME===

DATA

Medijan: 0.00717149999999922

Prosječna vrijednost: 0.008744238523852385

Maksimalna vrijednost0.4852170000000058

Minimalna vrijednost0.0

DISTRIBUCIJA

Medijan: 0.005722096575809653

Prosječna vrijednost: 0.008655079068619372

Maksimalna vrijednost0.15671945271609372

Minimalna vrijednost1.739697052152711e-06

Društvene mreže:

Najbolja distribucija za veličinu paketa:

halfgennorm(beta = 0.19, loc = 54.00, scale = 0.00)

Najbolja distribucija za međudolazna vremena:

kappa3(a = 2.11, loc = -0.0, scale = 0.02)

Usporedba vrijednosti distribucije i podataka:

===VELIČINA PAKETA===

DATA

Srednja vrijednost: 510.25349162011173

Medijan: 54.0

Minimum: 54.0

Maksimum: 14574.0

DISTRIBUCIJA

Srednja vrijednost: 108.46767469853057

Medijan: 61.827290866142285

Minimum: 54.00002917033475

Maksimum: 2241.3393721742113

===MEĐUDOLAZNO VRIJEME===

DATA

Srednja vrijednost: 0.09181417039106145

Medijan: 0.009791499999991515

Minimum: 0.0

Maksimum: 5.078023999999999

DISTRIBUCIJA

Srednja vrijednost: 0.02633194272076299

Medijan: 0.01616502132512545

Minimum: 4.323748279078275e-05

Maksimum: 1.1048906826039282

Servis video strujanja:

Najbolja distribucija za veličinu paketa:

foldcauchy(c = 118.91, loc = 0.2, scale = 10.86)

Najbolja distribucija za međudolazna vremena:

genhalflogistic(c = 0.0, loc = -0.04, scale = 7.0)

Usporedba vrijednosti distribucije i podataka:

===VELIČINA PAKETA===

DATA

Srednja vrijednost: 1167.4987850315893

Medijan: 1292.0

Minimum: 65.0

Maksimum: 1292.0

DISTRIBUCIJA

Srednja vrijednost: 1400.712972272204

Medijan: 1291.7565413330042

Minimum: 7.326758510066571

Maksimum: 556296.3441783073

===MEĐUDOLAZNO VRIJEME===

DATA

Srednja vrijednost: 1.755819669475049

Medijan: 0.0

Minimum: -0.039

Maksimum: 5169.37

DISTRIBUCIJA

Srednja vrijednost: 9.714773813663559

Medijan: 7.611447517219082

Minimum: -0.03497897535050311

Maksimum: 72.81083390455264

3. Navedite i opišite Markovljev lanac te distribucije trajanja usluga

Markovljev lanac se sastoji od 3 stanja, igranje umrežene igre, surfanje društvene mreže i gledanje youtube videa. Simuliramo korisnika koji u nekom trenu radi jednu aktivnost. Kada je gotov s jednom aktivnosti prelazi na drugu pomoću matrice prijelaza koja je proizvoljno definirana.

Matrica prijelaza je:

$\begin{bmatrix} 0.0 & 0.25 & 0.75 \\ 0.4 & 0.0 & 0.6 \\ 0.8 & 0.2 & 0.0 \end{bmatrix}$

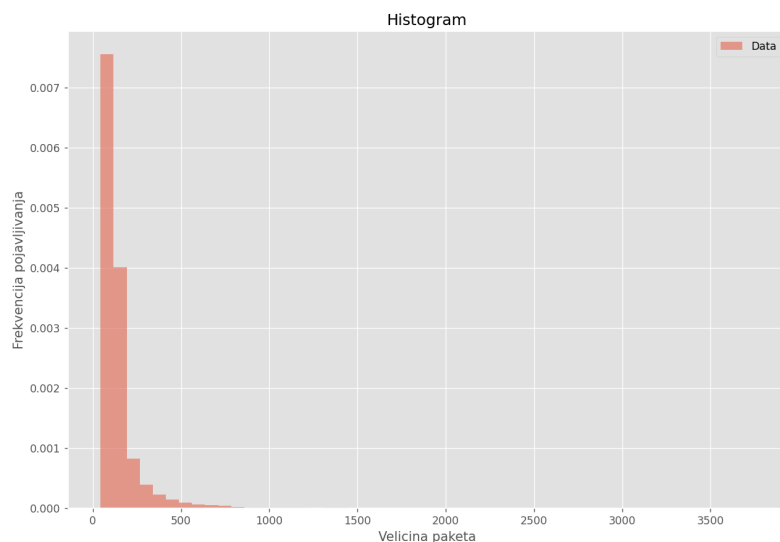
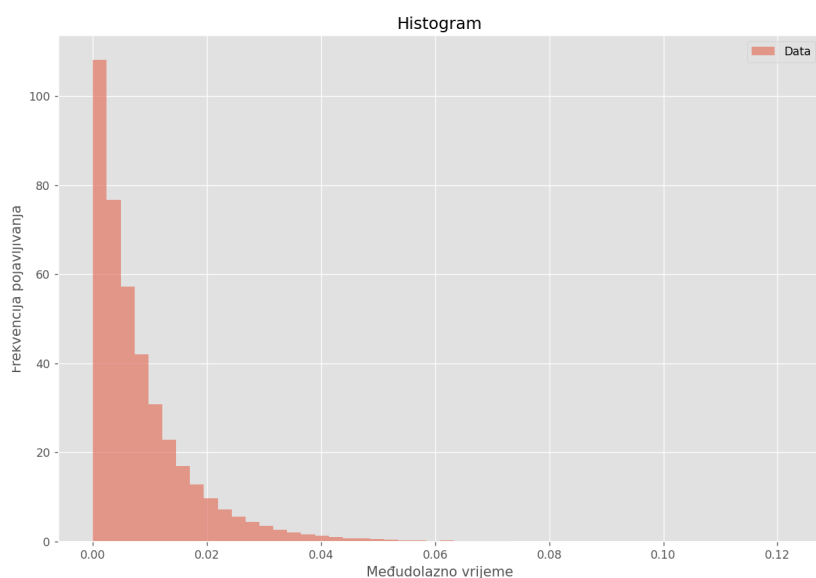
Vrijeme koje korisnik provodi u jednoj aktivnosti je definirana eksponencijalnom razdiobom sa λ (5, 10, 20).

4. Generirajte promet koristeći svoj generator te ga snimate koristeći Wireshark. Usporedite distribucije parametara (veličina paketa i međudolazna vremena) stvarnog snimljenog prometa usluge i generiranog prometa koristeći grafove funkcija razdiobe za oba modela.

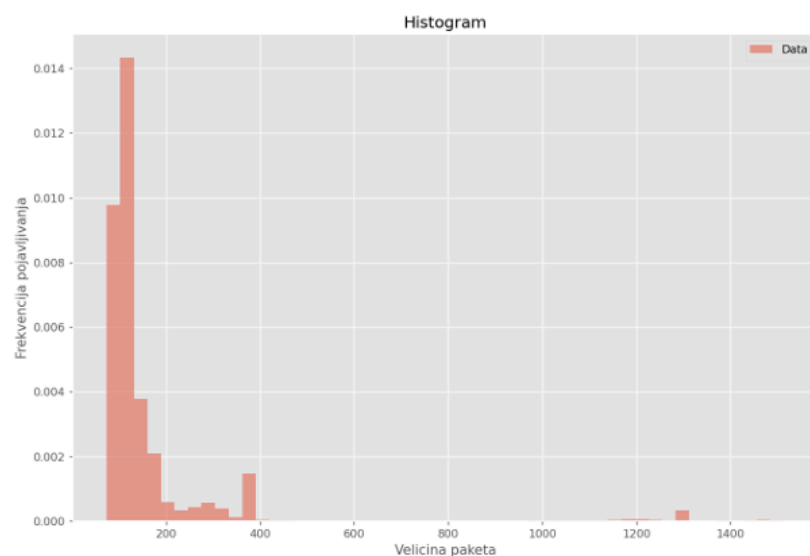
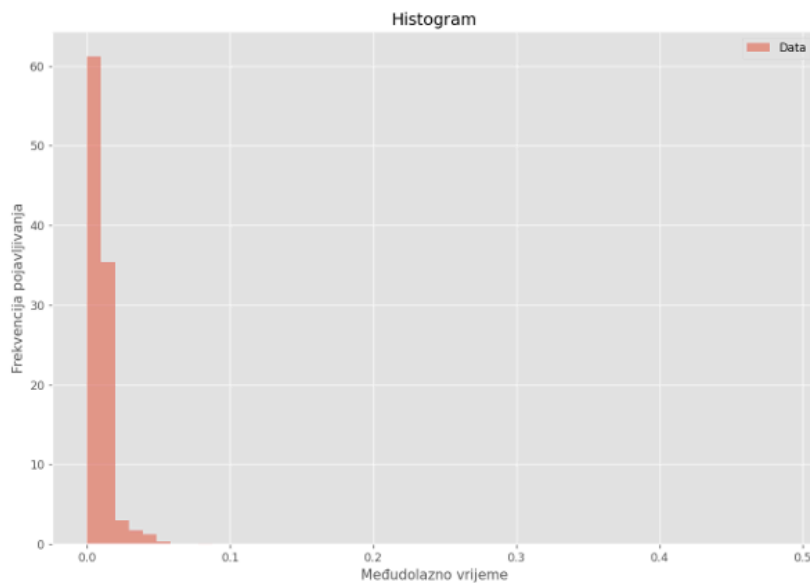
Komentirajte grafove

Na iduće 4 slike prikazani su grafovi za stvarni promet prilikom igranja video igre te promet generiran na temelju ranije dobivenih distribucija.

Prva dva grafa prikazuju međudolazna vremena i veličine paketa prometa generiranog u ovoj vježbi.



Iduća 2 grafa prikazuju stvarni promet snimljen Wireshark-om za vrijeme igranja video igre.

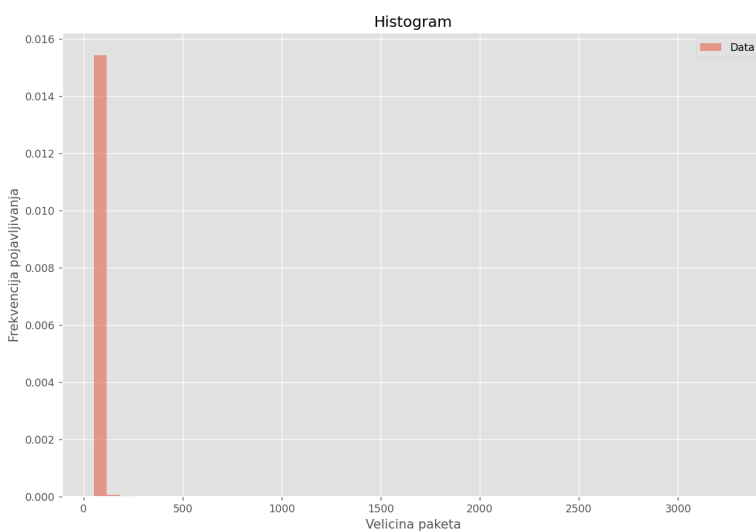
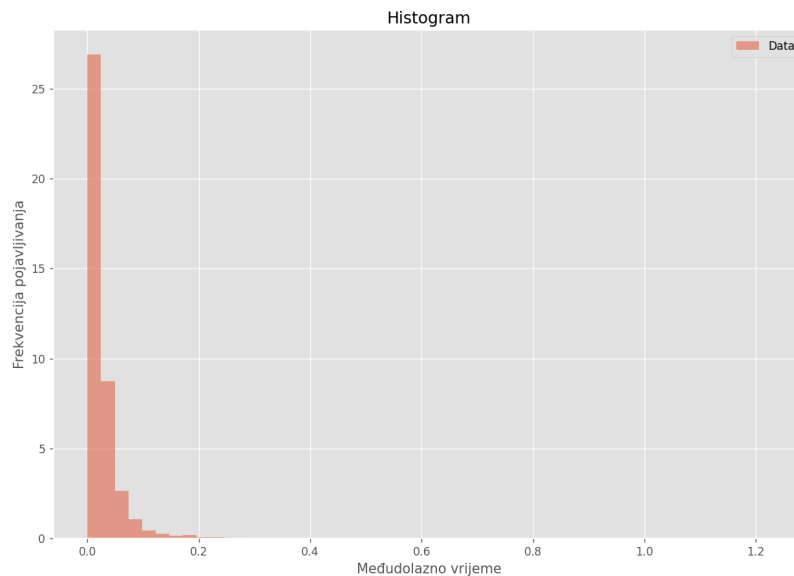


Na slikama je vidljivo kako se grafovi ne podudaraju u potpunosti. To se može pripisati nepreciznosti pri određivanju najbolje distribucije u drugom zadatku laboratorijske vježbe. Ovo je naročito vidljivo u grafovima za međudolazno vrijeme. Osim ispisanih grafova obavljena je i provjera distribucije koja najbolje opisuje generirani promet. Distribucija za veličinu paketa je $dweibull(c=0.52, loc=115.00, scale=25.66)$, a za međudolazno vrijeme je $pareto(b=11.89, loc=-0.09, scale=0.09)$. Razlike postoje, no distribucije su poprilično slične

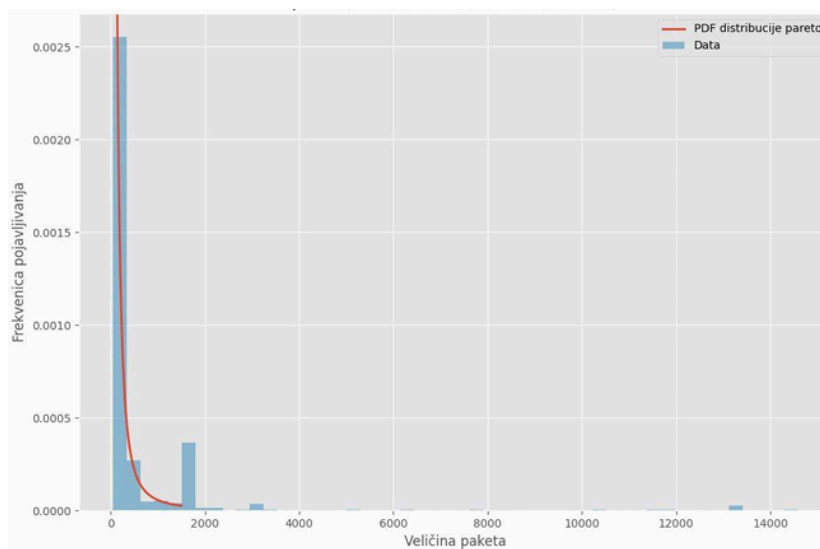
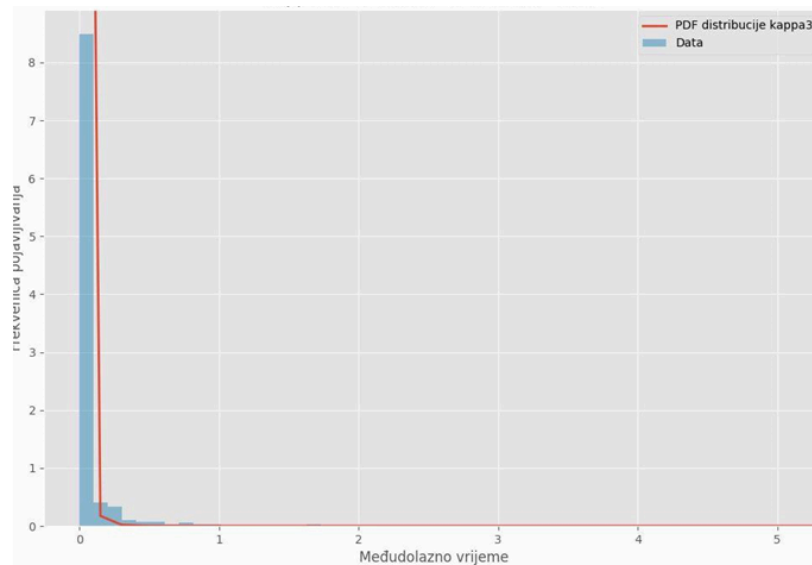
jedna drugoj, jedina veća razlika je scale parametar pareto distribucije.

Na iduće 4 slike prikazani su grafovi za stvarni promet prilikom korištenja društvenih mreža te promet generiran mrežnim generatorom na temelju ranije dobivenih distribucija.

Prva dva grafa prikazuju međudolazna vremena i veličine paketa prometa generiranog u ovoj vježbi.

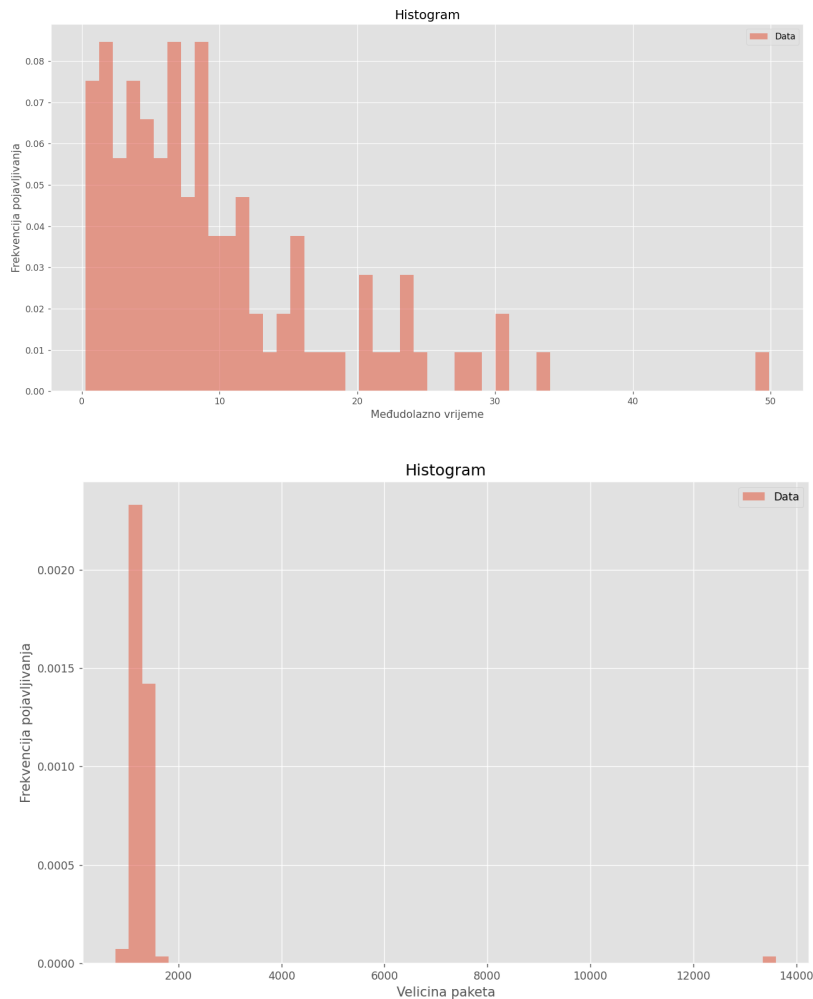


Iduća 2 grafa prikazuju stvarni promet snimljen Wireshark-om za vrijeme korištenja društvene mreže.

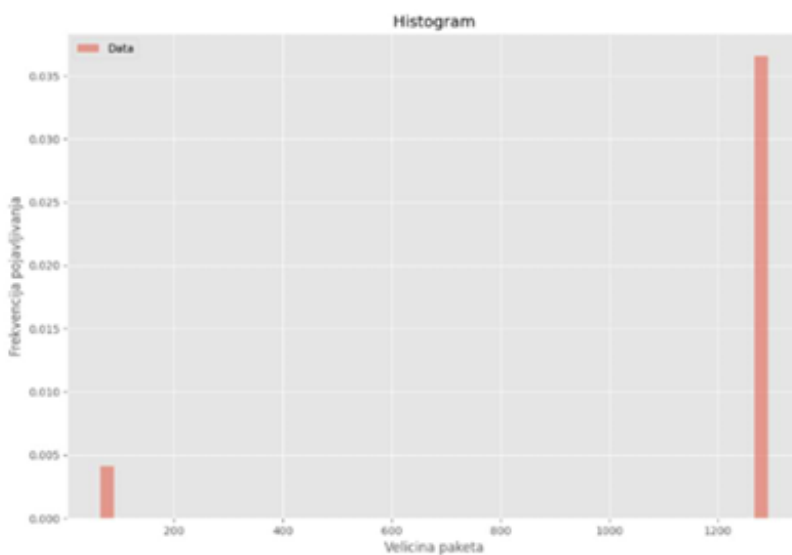
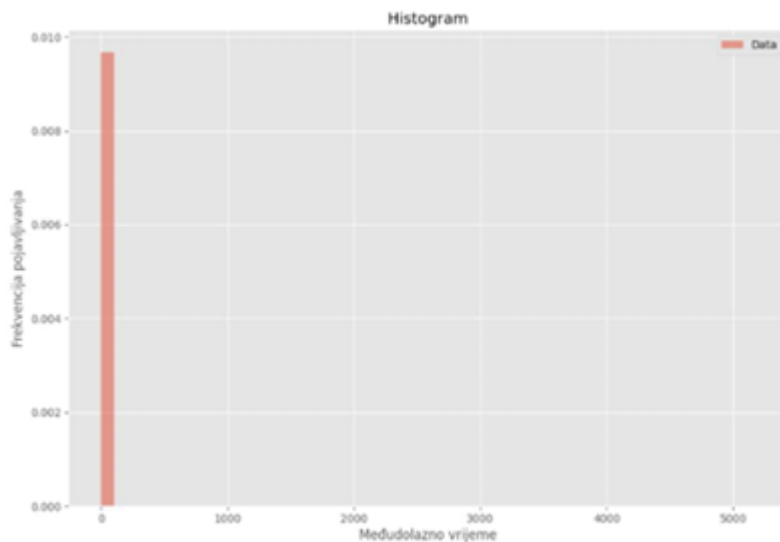


Grafovi za međudolazno vrijeme su jako slični jedan drugom, dok se u grafu za veličine paketa vide razlike. Generirani promet se sastojao uglavnom samo od paketa reda veličine ~ 55 , dok se u snimljenom prometu pojavljuje nešto više varijacije u veličinama paketa. Ovo može biti posljedica nepreciznog određivanja distribucije u drugom zadatku laboratorijske vježbe. Osim ispisanih grafova obavljena je i provjera distribucije koja najbolje opisuje generirani promet. Distribucija za veličinu paketa je $\text{expon}(\text{loc}=54.00, \text{scale}=3.52)$, a za međudolazno vrijeme je $\text{kappa3}(a=2.12, \text{loc}=0.00, \text{scale}=0.02)$. Distribucija za veličinu paketa generiranog

prometa je eksponencijalna što se razlikuje s obzirom na snimljeni stvarni promet gdje je distribucija bila halfgennorm. Na iduće 4 slike prikazani su grafovi za stvarni promet prilikom korištenja servisa video strujanja te promet generiran na temelju ranije dobivenih distribucija. Prva dva grafa prikazuju međudolazna vremena i veličine paketa prometa generiranog u ovoj vježbi.



Iduća 2 grafa prikazuju stvarni promet snimljen Wireshark-om za vrijeme korištenja servisa video strujanja.



U ovim grafovima su primjetne velike razlike, posebice na grafovima za međudolazne vrijednosti. Dok su na stvarnim podacima međudolazna vremena bila poprilično uniformna, unutar generiranog prometa se primjećuju veće razlike i odstupanja. Osim ispisanih grafova obavljena je i provjera distribucije koja najbolje opisuje generirani promet. Distribucija za veličinu paketa je $\text{foldcauchy}(c=95.34, \text{loc}=-0.20, \text{scale}=13.51)$, a za međudolazno vrijeme je $\text{genhalflogistic}(c=0.00, \text{loc}=0.26, \text{scale}=6.86)$. Distribucije po tipu su iste, no parametri se razlikuju.

5. Koristeći Kolmogorov-Smirnov test (iz scipy biblioteke) za dva uzorka (engl. two sample K-S test) usporedite distribucije parametara za oba modela (veličina paketa i međudolazna vremena) stvarnog snimljenog prometa usluge i generiranog prometa. Navedite vrijednosti dobivene testom i komentirajte ih.

Vrijednosti dobivene Kolmogorov-Smirnov testom:

==== VIDEO STREAMING ====

KS Test for Packet Lengths:

Statistic: 0.49380555357897143, P-value:
3.565963584670747e-24

KS Test for Interarrival Times:

Statistic: 0.944256848837984, P-value:
2.333707045079005e-128

==== VIDEO GAMES ====

KS Test for Packet Lengths:

Statistic: 0.254075083733078, P-value: 0.0

KS Test for Interarrival Times:

Statistic: 0.19915935893882702, P-value:
3.802965967662179e-268

==== SOCIAL NETWORKS ====

KS Test for Packet Lengths:

Statistic: 0.8290737743434449, P-value:
2.741192121578296e-74

KS Test for Interarrival Times:

Statistic: 0.9440296559285751, P-value:
1.6146452209589087e-113

Za KS test izvučene su dvije najvažnije vrijednosti kao rezultat:

Prva vrijednost koja nas zanima je statistička čija vrijednost označava koliko su dvije distribucije udaljene te se kreće u intervalu $[0, 1]$, 0 predstavljajući da su dvije distribucije identične, a 1 da su najviše moguće različite.

Druga nama zanimljiva vrijednost je p-value koja označava vjerojatnost da se razlika među distribucijama pojavila nasumično, odnosno da nije posljedica statističkih razlika među distribucijama. Visoku vrijednost p-value varijable većinom predstavljaju vrijednosti veće od 0.05, dok su male vrijednosti ispod 0.05.

Kao što se može vidjeti iz prikazanih rezultata, sve tri usluge prikazuju znatne razlike među distribucijama sa vrlo malim vrijednostima p-value što bi značilo da se podaci dobiveni generiranjem prometa i snimanjem istog znatno statistički razlikuju. Neki od razloga za takve razlike u podacima mogli bi biti to što smo snimanjem podataka komunicirali sa nekim postojećim serverom određene usluge, dok smo generiranjem prometa samo simulirali tu komunikaciju. Potrebno je uzeti u obzir i da su snimljeni podaci veoma teški za prikazati funkcijom distribucije te sama funkcija koju smo programski izračunali najvjerojatnije nije optimalan opis snimljenih podataka.

6. Iz simulacije izračunajte empirijsku vjerojatnost zadržavanja u pojedinom stanju (vrijeme koje je provedeno u pojedinom stanju podijeljeno s ukupnim vremenom) na temelju rezultata simulacije od barem 100 prelazaka iz stanja u stanje.

Izračunajte i teoretske stacionarne vjerojatnosti svakog stanja na temelju intenziteta prijelaza koje možete izračunati na temelju vjerojatnosti prijelaza i parametra eksponencijalne distribucije koja određuje vrijeme zadržavanja u pojedinom stanju. Usporedite dobivene vrijednosti te ih komentirajte.

Vektor stacionarnih vrijednosti dobiven računanjem tijekom simulacije iznosi

[0.388333333333333, 0.226666666666666, 0.384999999999999],
dok stvarni udio vremena proveden u pojedinom stanju iznosi
[0.2847684044097, 0.11713159135056, 0.5981000042397].

```
totalSimulationTime = 1744.6742040193976
totalSimulationPacketCount = 65278
Theoretical stationary probability vector:
[0.3883333333333336, 0.22666666666666666, 0.38499999999999995]

Empirical stationary probability vector:
[0.2847684044097233, 0.11713159135056461, 0.5981000042397121]
```

Vidimo da je svako stanje po “poretku” udjela vremena dobro prikazano izračunatim vektorom (2. stanje ima najmanji udio, 3. najveći), ali postoje određena odstupanja. U usporedbi s prethodnom laboratorijskom vježbom Markovljevog lanca, ova simulacija sastoji se od znatno manjeg broja koraka. Kao posljedica, stvarni udio vremena nije mogao doći do stvarnih stacionarnih vrijednosti (jer se za računanje limesa u ovakvim situacijama koristi broj koraka nekoliko redova veći od korištenog).