

PRIMENA UČENJA POTKREPLJIVANJEM U PROTOKOLIMA RUTIRANJA ZA FANET MREŽE

Marija Malnar, Pavle Bugarčić
Univerzitet u Beogradu - Saobraćajni fakultet
m.malnar@sf.bg.ac.rs, p.bugarcic@sf.bg.ac.rs

Rezime: Tehnološki razvoj i sve šira upotreba bespilotnih letelica doveli su do povećane potrebe za obezbeđivanjem pouzdane komunikacije između njih. Bežične ad hoc mreže za letelice (*Flying Ad hoc Networks*, FANETs) predstavljaju odličnu soluciju za ostvarivanje ovog cilja. Najznačajnije karakteristike ovih mreža su česte promene u mrežnoj topologiji, velika brzina kretanja i mala gustina mrežnih čvorova. Protokoli rutiranja za FANET mreže treba da se prilagode njihovim karakteristikama, kako ne bi došlo do degradiranja ukupnih mrežnih performansi. Tradicionalni protokoli rutiranja ne uspevaju da isprate dinamičku prirodu FANET mreža, pa je neophodno uključiti nove tehnike koje bolje odgovaraju ovom izazovu. Uključivanje učenja potkrepljivanjem (*Reinforcement Learning*, RL) u protokole rutiranja može značajno unaprediti performanse FANET mreža, zbog mogućnosti stalnog praćenja i prilagođavanja promenama u mrežnom okruženju. U ovom radu biće predstavljeni najznačajniji protokoli rutiranja na bazi RL za FANET mreže, kao i njihova detaljna klasifikacija i komparativna analiza.

Ključne reči: FANET mreže, učenje potkrepljivanjem, protokoli rutiranja

1. Uvod

Poslednjih godina primetan je sve brži razvoj bespilotnih letelica koje se koriste u sve širem opsegu aplikacija i usluga. Povezivanje ovih letelica u organizovane grupe koje međusobno komuniciraju bežičnim putem moguće je formiranjem bežičnih ad hoc mreža za letelice (*Flying Ad hoc Networks*, FANETs). Ove mreže predstavljaju podskup mobilnih ad hoc mreža (*Mobile Ad hoc Networks*, MANETs), sa kojima dele osnovne karakteristike kao što su mobilnost čvorova, bežični komunikacioni medijum, decentralizovana kontrola i komunikacija preko višestrukih hopova. Karakteristike po kojima se FANET mreže izdvajaju su velika brzina kretanja mrežnih čvorova, česte promene u mrežnoj topologiji i mala gustina mreže. Uzevši u obzir sve ove karakteristike, dizajn efikasnog protokola rutiranja za FANET mreže predstavlja veoma značajan istraživački izazov. Protokoli rutiranja su zadušeni za izbor optimalne putanje za slanje podataka i veoma je važno da se prilagode stalnim promenama u mrežnoj topologiji, ali i da uzmu u obzir druga ograničenja koja su specifična za FANET mreže,

kao što su računarska snaga, energija i kašnjenje. Važno je napomenuti da se bespilotne letelice kreću u trodimenzionalnom prostoru, što takođe predstavlja veliki izazov pri kreiranju odgovarajućeg protokola rutiranja za FANET mreže.

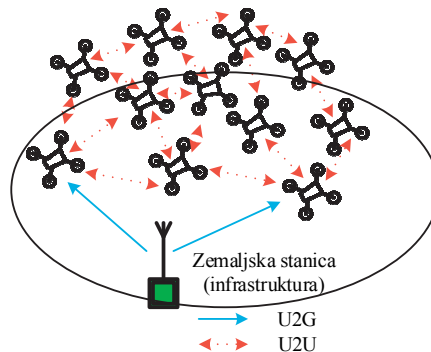
U prethodnim godinama mnogi istraživači pokušali su da poboljšaju performanse FANET mreža razvojem novih protokola rutiranja [1]. Uprkos velikim naporima, veoma je teško dizajnirati protokol koji garantuje efikasnu komunikaciju u ovim mrežama. Za rešavanje izazova sa kojima se susreću FANET mogu se koristiti tehnike mašinskog učenja, a najpogodniji tip učenja za ove mreže je učenje potkrepljivanjem (*Reinforcement Learning*, RL) [2]. Ovaj tip učenja podrazumeva stalnu interakciju agenta učenja sa okruženjem, tako da je moguće ispratiti i prilagoditi se čestim promenama u mrežnoj topologiji. Agent učenja na osnovu prikupljenih podataka iz okruženja treba da izabere najbolju akciju za postizanje definisanog cilja, kao što je kreiranje putanje sa maksimalnim procentom isporučenih paketa, minimalnim kašnjenjem, itd. S obzirom da protokoli rutiranja bazirani na RL pokazuju veliki potencijal za unapređenje performansi FANET mreža, neophodno je proučiti i izvršiti detaljan pregled ovih protokola kako bi se trasirao put za dalji razvoj ove oblasti. Upravo sa tim ciljem, u ovom radu je izvršen pregled najznačajnijih protokola rutiranja na bazi RL za FANET mreže, njihova klasifikacija i komparativna analiza.

Ostatak rada je organizovan na sledeći način. U drugom poglavlju su objašnjene osnovne karakteristike FANET mreža. U trećem poglavlju je izvršena klasifikacija protokola rutiranja baziranih na RL za FANET mreže na osnovu tipa učenja, agenta učenja i mrežne topologije, a ujedno su ukratko opisani predstavnici pojedinih klasa. Zatim je u četvrtom poglavlju data komparativna analiza ovih protokola na osnovu posmatranih mrežnih parametara koji su optimizovani, korišćenog mrežnog simulatora za evaluaciju mrežnih performansi i primenjenog modela mobilnosti mrežnih čvorova u simulacijama. Na kraju su u petom poglavlju sumirana zaključna razmatranja i predloženi pravci budućeg istraživanja.

2. FANET mreže

Nedavni napredak u bežičnim tehnologijama doveo je do novih mogućnosti povezivanja bespilotnih letelica (*Unmanned Aerial Vehicles*, UAVs), a jedan od najpopularnijih primera su FANET mreže. U suštini, FANET mreža se sastoji od grupe bespilotnih letelica koje međusobno saraduju i komuniciraju bez ikakve infrastrukture kako bi izvršile određeni zadatak bez ljudske intervencije. U ovim mrežama sve letelice mogu međusobno komunicirati (*UAV to UAV*, U2U), a samo manji broj njih može uspostaviti komunikaciju sa zemaljskom stanicom (*UAV to Ground station*, U2G). Na slici 1 je prikazan jednostavan primer FANET mreže, koja se sastoji iz više bespilotnih letelica i zemaljske stanice. Iako ove mreže karakteriše jednostavna implementacija i odsustvo složene mrežne infrastrukture, one se ipak susreću sa brojnim izazovima. Najveći izazov je održavanje povezanosti mrežnih čvorova, s obzirom da bespilotne letelice imaju veliku brzinu i malu gustinu pa su mogući česti prekidi mrežnih linkova putem kojih komuniciraju. Ovo dovodi do degradiranja brojnih mrežnih performansi, prvenstveno povećanja gubitaka paketa i kašnjenja paketa [3]. Još jedan bitan izazov u ovim mrežama je potrošnja energije, jer male bespilotne letelice korsište baterije niskog kapaciteta za snabdevanje potrebnom energijom za obradu podataka u realnom vremenu, komunikaciju i let. Ostali izazovi sa kojima se susreću FANET mreže su ograničeni

skladišni i procesorski kapaciteti bespilotnih letelica, velika kašnjenja paketa kao posledica komunikacije posredstvom više hopova, kompleksno upravljanje smetnjama zbog ograničenog propusnog opsega i dinamičke topologije, itd.



Slika 1. Primer FANET mreže

Zbog svega navedenog, veoma je kompleksno kreirati protokol rutiranja koji bi na efikasan način odgovorio svim izazovima koje nosi komunikacija unutar FANET mreža. Posebno značajan izazov u dizajniranju protokola rutiranja predstavlja činjenica da se bespilotne letelice kreću u trodimenzionalnom prostoru. Pored toga, ovi protokoli treba efikasno da koriste mrežne i energetske resurse i uzimaju u obzir mehanizme za sprečavanje petlji u rutiranju. U smislu efikasnosti, proces rutiranja treba da ima niske troškove rutiranja, visoku pouzdanost, niske gubitke paketa i prihvatljivo kašnjenje. Kako bi savladali sve ove izazove, brojni autori su razvili protokole rutiranja koji pomoću učenja potkrepljivanjem uzimaju u obzir bitne parametre FANET mreža pri izboru optimalne putanje i prilagođavaju proces rutiranja specifičnim zahtevima ovih mreža sa ciljem unapređenja ukupnih mrežnih performansi.

3. Klasifikacija protokola rutiranja na bazi RL za FANET mreže

U ovom poglavlju je izvršena klasifikacija aktuelnih protokola rutiranja za FANET mreže baziranih na RL, kao što je predstavljeno u tabeli 1. Protokoli su prvenstveno klasifikovani na osnovu tipa RL koji se koristi. Najčešće korišćeni tip je Q-učenje (*Q-Learning*, QL), dok neki autori koriste nešto složeniju tehniku dubokog učenja potkrepljivanjem (*Deep Reinforcement Learning*, DRL). Algoritmi bazirani na QL su jednostavniji, imaju manje procesorske zahteve i lakši su za implementaciju. Ovi algoritmi su pogodni za manje FANET mreže, gde je brzina učenja algoritma prihvatljiva zbog manjeg skupa mogućih stanja okruženja i akcija koje agent učenja može da preduzme. Međutim, kod većih FANET mreža brzina konvergencije QL algoritama nije na zadovoljavajućem nivou, pa se preporučuje primena nešto složenijih algoritama baziranih na DRL.

Protokole je zatim moguće klasifikovati prema organizaciji agenata učenja na algoritme kod kojih jedan agent (*Single Agent*, SA) samostalno uči i algoritme kod kojih grupa agenata (*Multi Agent*, MA) pokušava da nauči optimalno ponašanje kroz interakciju sa okruženjem. Algoritmi sa SA pristupom su lakši za implementaciju, jer je

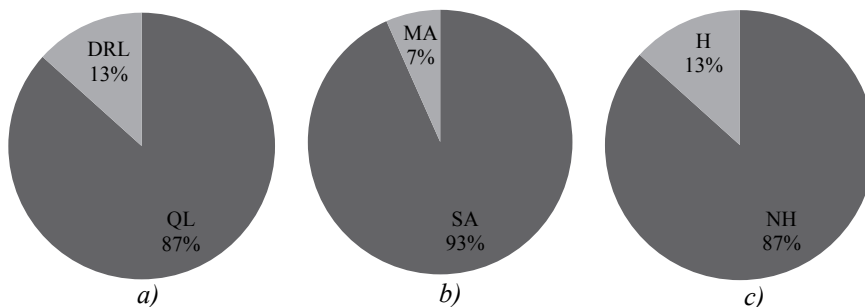
njihova računarska složenost manja od MA algoritama. Međutim, performanse SA algoritama nisu pogodne za velike mreže jer agentu treba dosta vremena da pronađe optimalnu putanju, što znači da je brzina konvergencije mala. Važan izazov kod MA algoritama je kako koordinisati i ostvariti optimalnu saradnju među agentima. Protokoli bazirani na MA algoritmima su otporniji na greške, što znači da ako jedan ili više agenata iz nekog razloga ne uspeju da nađu optimalnu putanju, ostali agenti nastavljaju sa optimizacijom rutiranja kako bi sprečili degradiranje ukupnih mrežnih performansi. Takođe, MA algoritmi imaju veću sposobnost učenja pa su pogodniji za veće mreže.

Poslednja klasifikacija protokola izvršena je na osnovu mrežne topologije, na nehijerarhijske (NH) i hijerarhijske (H) protokole. Kod nehijerarhijskih protokola svi čvorovi u mreži su ravnopravni i na osnovu algoritma rutiranja donose odluku o putanji kojom će proslediti pakete. Problem kod ovih protokola je mala skalabilnost i veliki overhed. Kod hijerarhijskih protokola mrežni čvorovi su podeljeni u nekoliko hijerarhijskih nivoa. Čvorovi na jednom nivou mogu se povezati jedni sa drugima, a takođe su povezani sa roditeljskim čvorom na višem nivou preko koga mogu da komuniciraju sa čvorovima na višim nivoima. Ovaj pristup odlikuje veća skalabilnost, efikasnija potrošnja mrežnih resursa i smanjenje overheda rutiranja. Glavni izazov kod ovih protokola je adekvatan izbor roditeljskih čvorova.

Tabela 1. Klasifikacija protokola rutiranja na osnovu tipa učenja, organizacije agenta učenja i mrežne topologije

Protokoli	Tip učenja		Agent učenja		Mrežna topologija	
	QL	DRL	SA	MA	NH	H
[4-14]	✓		✓		✓	
[15]	✓		✓			✓
[16]	✓			✓	✓	
[17]		✓	✓		✓	
[18]		✓	✓			✓

Na slici 2 je prikazana procentualna zastupljenost pojedinih tipova učenja, agenata učenja i mrežne topologije. Može se primetiti da najveći broj protokola koristi QL algoritam sa SA pristupom, uz nehijerarhijsku mrežnu topologiju. Po dva protokola koriste DRL algoritam i hijerarhijsku mrežnu topologiju, dok samo jedan protokol ima MA pristup učenju u procesu rutiranja.



Slika 2. Procentualna zastupljenost: a) tipa učenja, b) organizacije agenata učenja, c) tipa mrežne topologije

Primer protokola koji koristi QL sa SA pristupom i nehijerarhijskom mrežnom topologijom je *Q-learning-based topology-aware routing method* (QTAR) [4]. Ovaj protokol pokušava da otkrije najbolju putanju između izvora i odredišta na osnovu informacija od suseda koje uključuju poziciju, kašnjenje, brzinu i energiju suseda. QTAR pomaže u balansiranju mrežnog opterećenja jer se nivo energije čvorova uzima u obzir u procesu učenja. Takođe, ovaj protokol sprečava pojavljivanje petlji tako što prikuplja informacije i od suseda udaljenih dva hopa. Jedan od ključnih ciljeva je smanjenje dužine putanje (broja hopova). U procesu učenja svaki paket ima ulogu agenta, a susedni čvorovi predstavljaju skup mogućih stanja. Da bi se smanjila veličina skupa potencijalnih stanja, QTAR uzima u obzir samo čvorove koji su bliži odredištu u odnosu na tekući čvor. Ovo poboljšava brzinu konvergencije protokola. Izbor sledećeg hopa se smatra mogućom akcijom koju agent učenja može da preduzme. Nagrada za preduzetu akciju zavisi od energije, kašnjenja i brzine suseda.

Autori su u [15] razvili *intelligent clustering routing approach* (ICRA), koji koristi QL sa SA pristupom i hijerarhijskom topologijom. ICRA ima za cilj povećanje stabilnosti mrežne topologije i povećanje životnog veka mreže balansiranjem potrošnje energije. Čvorovi u mreži periodično razmenjuju kontrolne pakete koji sadrže informacije o lokaciji, brzini, pravcu kretanja i vremenu života. Ove informacije se koriste u procesu klasterovanja, koje služi za balansiranje potrošnje energije u mreži. ICRA algoritam se sastoji iz tri faze: klasterovanje, prilagođavanje strategije klasterovanja i rutiranje. U procesu klasterovanja, svaki čvor izračunava parametar korisnosti na osnovu četiri faktora korisnosti: preostale energije, pozicije u mreži, sličnosti brzine čvora i njegovih suseda i vremena zadržavanja linka. Čvorovi zatim razmenjuju parametre korisnosti kako bi odabrali čvor sa najvećom korisnošću za glavni čvor klastera (*Cluster Head*, CH). Ovi čvorovi su odgovorni za formiranje putanja između različitih klastera. Parametri korisnosti podležu procesu prilagođavanja strategiji klasterovanja zasnovanoj na QL. Ovaj proces učenja sledi centralizovanu strategiju i sprovodi ga zemaljska stanica (*Ground Station*, GS). U ovom procesu, GS igra ulogu agenta, a prostor akcija uključuje izbor četiri težinska koeficijenta koja odgovaraju faktorima korisnosti. Funkcija nagrade se izračunava na osnovu stabilnosti strukture klastera i stope promena energija čvorova. U procesu rutiranja, kada čvor koji nije odredišni primi paket od drugog čvora najpre proverava lokaciju odredišta. Ako je u pitanju njegov sused, direktno prosleđuje paket ka njemu. Ukoliko nije, šalje paket ka najbližem CH čvoru kako bi ga prosledio dalje ka odredištu.

U [16] je predložen *Q-learning geographic routing method* (QLGR), koji koristi QL sa MA pristupom i nehijerarhijskom topologijom. Ovaj protokol pomaže u smanjenju gubitka paketa i overheda u mreži. Informacije koje čvorovi dele u procesu učenja uključuju sekvencijalni broj i dužinu poruke, maksimalnu i zauzetu dužinu reda za čekanje kod slanja paketa. Svaki čvor predstavlja agenta učenja, a izbor susednog čvora za sledeći hop predstavlja moguću akciju koju agent može da preduzme. Nakon preduzimanja akcije, agent dobija lokalnu i globalnu nagradu od okruženja. Lokalne nagrade se računaju na osnovu kapaciteta opterećenja i kvaliteta veze. Ova nagrada ocenjuje kvalitet samo sledećeg čvora na putanji. Prenos podataka do odredišta garantuje se globalnom nagradom. Kada čvor bira sledeći hop za slanje podataka, uzima u obzir udaljenost tog čvora do odredišta i njegovu Q-vrednost.

Autori u [17] predložili su *traffic-aware Q-network geographic routing scheme based on greedy perimeter stateless routing* (TQNGPSR), koji koristi DRL sa SA

pristupom i nehijerarhijskom topologijom. Ovaj protokol uvodi strategiju balansiranja saobraćaja koja koristi informacije o zagušenju susednih čvorova za procenu kvaliteta veze. Na osnovu ovoga bira se najbolja putanja do odredišnog čvora kako bi se smanjila kašnjenja i gubici paketa. Svaki čvor izračunava nivo zagušenja na osnovu dužine reda za čekanje koju je dobio od svojih suseda, što je proporcionalno čekanju u redu za slanje paketa. U procesu rutiranja, Q-vrednosti se prvo računaju bez obzira na informacije o zagušenju, a zatim se ažuriraju na osnovu informacija o redu za čekanje. Ako je red za čekanje gotovo pun, čvor dobija visoku kaznu i samim tim se smanjuje njegova Q-vrednost kako bi se smanjila verovatnoća izbora ovog čvora za sledeći hop na putanji ka odredištu.

U [18] je predložen *deep Q-network-based vertical routing* (DQN-VR) za 5G FANET mreže, koji koristi DRL sa SA pristupom i hijerarhijskom topologijom. Ova metoda kombinuje distribuirane i centralizovane metode rutiranja. Podrazumeva se da 5G tehnologija podržava FANET mreže kako bi poboljšala skalabilnost i stabilnost mreže, kao i ravnotežu opterećenja u mreži. DQN-VR se sastoji od dve glavne faze: vertikalnog klasterovanja i vertikalnog rutiranja. U prvoj fazi, distribuirani kontroleri su odgovorni za implementaciju procesa vertikalnog klasterovanja kako bi formirali klustere na svakom nivou mreže, poboljšali stabilnost klastera i skalabilnost mreže, kao i smanjili kašnjenje između krajnjih tačaka. Svrha procesa klasterovanja je grupisanje čvorova na osnovu njihove prirode i ponašanja unutar klastera. U drugoj fazi centralni kontroler izvršava vertikalno rutiranje koristeći DRL kako bi odredio različite putanje na različitim nivoima mreže i stvorio puteve unutar nivoa i između njih. U ovom procesu, centralni kontroler igra ulogu agenta, a mreža se smatra okruženjem. Skup stanja predstavlja dvodimenzionalni niz koji uključuje informacije o kretanju i preostalu energiju čvorova. Skup akcija predstavlja izbor sledećeg čvora prema odredištu. Nagrada se računa na osnovu stepena uspešno isporučenih paketa i nivoa zagušenja u čvorovima.

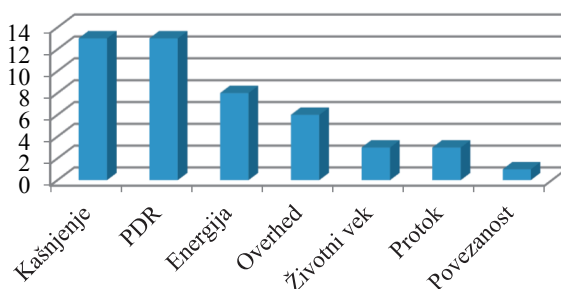
4. Komparativna analiza protokola rutiranja na bazi RL za FANET mreže

U ovom poglavlju je izvršena komparativna analiza protokola rutiranja baziranih na RL za FANET mreže na osnovu nekoliko kriterijuma, kao što je prikazano u tabeli 2. Prvi kriterijum su posmatrani parametri u procesu optimizacije procesa rutiranja, drugi kriterijum su simulacioni alati koji su korišćeni za testiranje i evaluaciju performansi ovih protokola, a treći kriterijum su modeli mobilnosti koji su primenjeni na mrežne čvorove u okviru simulacione analize.

Na slici 3 je prikazan odnos pojavljivanja pojedinih parametara u protokolima rutiranja. Može se primetiti da su najčešće posmatrani parametri kašnjenje paketa s kraja na kraj mreže [4-7], [9-17] i procenat uspešno isporučenih paketa (*Packet Delivery Ratio*, PDR) [4-11], [13-17]. Samim tim može se zaključiti da je cilj optimizacije većine protokola povećanje procenta isporučenih paketa i smanjenje kašnjenja paketa u mreži. U nekoliko protokola posmatran je veoma važan i kritičan parametar za FANET mreže, a to je potrošnja energije [4-5], [8], [12], [14-16], [18]. Sledeći parametar po zastupljenosti u protokolima je overhed rutiranja [4], [6], [8], [14], [17], gde je cilj protokola što manji overhed kako ne bi došlo do preopterećenja i zagušenja u mreži. Po tri protokola kao posmatrani parametar uzimaju životni vek mreže [4], [15], [18] i ostvareni aplikacioni protok [13], [16-17], dok je u jednom protokolu posmatrani parametar povezanost čvorova u mreži [18].

Tabela 2. Komparativna analiza protokola rutiranja na osnovu posmatranih mrežnih parametara, korišćenih simulacionih alata i modela mobilnosti

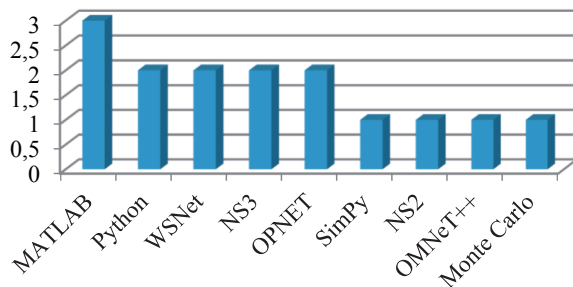
Protokol	Posmatrani parametri	Simulacioni alat	Model mobilnosti
[4]	potrošnja energije, kašnjenje, životni vek mreže, PDR, overhed	MATLAB	3D Gauss–Markov
[5]	potrošnja energije, kašnjenje, PDR	WSNet	Random Waypoint
[6]	kašnjenje, PDR, overhed	NS3	Gauss–Markov
[7]	kašnjenje, PDR	OPNET	Random Waypoint
[8]	potrošnja energije, PDR, overhed	nije naveden	nije naveden
[9]	kašnjenje, PDR	WSNet	Random Waypoint
[10]	kašnjenje, PDR	MATLAB, NS2	Random Waypoint
[11]	kašnjenje, PDR	OMNeT++	Random Waypoint
[12]	potrošnja energije, kašnjenje	nije naveden	nije naveden
[13]	kašnjenje, PDR, protok	nije naveden	Gauss–Markov
[14]	potrošnja energije, kašnjenje, PDR, overhed	Monte Carlo	3D Random Waypoint
[15]	potrošnja energije, kašnjenje, životni vek mreže, PDR	OPNET	Gauss–Markov
[16]	potrošnja energije, kašnjenje, PDR, protok	NS3	Gauss–Markov
[17]	kašnjenje, PDR, protok, overhed	Python, SimPy	Aircraft model
[18]	potrošnja energije, životni vek mreže, povezanost	MATLAB, Python	nije naveden



Slika 3. Posmatrani parametri u protokolima rutiranja

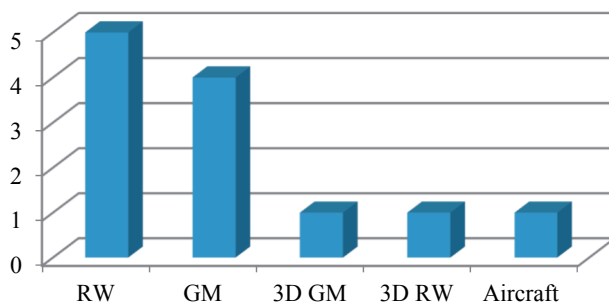
Na slici 4 je prikazana zastupljenost pojedinih simulacionih alata koji su korišćeni za testiranje predloženih protokola. Može se primetiti da najviše istraživača koristi MATLAB simulacioni alat [4], [10], [18], slede Python [17-18], WSNet [5], [9],

NS3 [6], [16] i OPNET [7], [15] sa po dve primene, dok se SimPy [17], NS2 [10], OMNET++ [11] i Monte Carlo [14] simulatori koriste za testiranje po jednog protokola rutiranja.



Slika 4. Simulacioni alati za testiranje protokola rutiranja

Konačno, na slici 5 je prikazana zastupljenost primene pojedinih modela mobilnosti u simulacionoj analizi. Najveći broj protokola je testiran uz primenu *Random Waypoint* (RW) [5], [7], [9-11] i *Gauss-Markov* (GM) [6], [13], [15-16] modela mobilnosti. Po jedan protokol testiran je uz primenu 3D GM [4], 3D RW [14] i *Aircraft* [17] modela mobilnosti. Iznenaduje činjenica da se i dalje 2D modeli koriste dosta češće od 3D modela.



Slika 5. Modeli mobilnosti u simulacionoj analizi

4. Zaključak

U ovom radu je predstavljen sistematičan pregled protokola rutiranja za FANET mreže, baziranih na RL. Najpre su opisane osnovne karakteristike i izazovi sa kojima se susreću FANET mreže, kako bi bilo jasnije koje zahteve moraju da zadovolje protokoli rutiranja za ove mreže. Zatim je izvršena klasifikacija protokola rutiranja za FANET mreže po nekoliko kriterijuma, nakon čega je detaljnije opisan po jedan predstavnik svake klase. Može se zaključiti da je u protokolima i dalje dominantan tip učenja QL i da se uglavnom koriste SA pristup učenju i nehijerarhijska mrežna topologija. Na kraju je izvršena komparativna analiza protokola prema posmatranim parametrima u procesu optimizacije rutiranja, korišćenim simulacionim alatima i modelima mobilnosti. Ovde je

primetno da se najčešće teži optimizaciji kašnenja i stepena isporučenih paketa, u svim istraživanjima je za evaluacija predloženih protokola korišćena simulaciona analiza, dok su najčešće korišćeni modeli mobilnosti i dalje RW i GM. U okviru budućih istraživanja moguće je izučiti dalje mogućnosti primene RL u protokolima rutiranja za dinamičke bežične *ad hoc* mreže. Takođe, ovaj pregled dosadašnjih istraživanja može da bude dobra osnova za razvoj novog protokola rutiranja koji bi pomogao u daljem unapređenju mrežnih performansi FANET mreža.

Literatura

- [1] O.S. Oubbati, A. Lakas, F. Zhou, M. Gunes, and M.B.Yagoubi, "A survey on position-based routing protocols for flying ad hoc networks (FANETs)", *Vehicular Communications*, vol. 10, pp. 29-56, October 2017. DOI: 10.1016/j.vehcom.2017.10.003
- [2] R. Sutton, A. Barto, *Reinforcement Learning: An Introduction, second edition*. Cambridge, Massachusetts, MIT Press, 2018.
- [3] F. Noor, M. A. Khan, A. Al-Zahrani, I. Ullah, K. A. Al-Dhlan, "A review on communications perspective of flying ad hoc networks: Key enabling wireless technologies, applications, challenges and open research topics", *Drones*, vol. 4, no. 5, pp. 65-78, September 2020. DOI: 10.3390/drones4040065
- [4] M. Y. Arafat and S. Moh, "A Q-learning-based topology-aware routing protocol for flying ad hoc networks", *IEEE Internet of Things Journal*, vol. 9, no. 3. pp. 1985-2000, February 2022. DOI: 10.1109/JIOT.2021.3089759
- [5] J. Liu, Q. Wang, C. He, K. Jaffrès-Runser, Y. Xu, Z. Li, and Y. J. Xu, "QMR: Q-learning based multi-objective optimization routing protocol for flying ad hoc networks", *Computer Communications*, vol. 150, pp. 304-316, January 2020. DOI: 10.1016/j.comcom.2019.11.011
- [6] W. S. Jung, J. Yim, and Y. B. Ko, "QGeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks", *IEEE Communications Letters*, vol. 21, no. 10, pp. 2258-2261, October 2017. DOI: 10.1109/LCOMM.2017.2656879
- [7] J. W. Lim and Y. B. Ko. "Q-learning based stepwise routing protocol for multi-uav networks", in *Proc. 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 307-309, April 2021. DOI: 10.1109/ICAIIIC51459.2021.9415265
- [8] A. Rovira-Sugranes, F. Afghah, J. Qu, and A. Razi, "Fully-echoed q-routing with simulated annealing inference for flying adhoc networks", *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2223-2234, July 2021. DOI: 10.1109/TNSE.2021.3085514
- [9] L. A. L. da Costa, R. Kunst, and E. P. de Freitas, "Q-FANET: Improved Q-learning based routing protocol for FANETs", *Computer Networks*, vol. 198, pp. 108379-108389, October 2021. DOI: 10.1016/j.comnet.2021.108379
- [10] Z. Zheng, A. K. Sangaiah, and T. Wang, "Adaptive communication protocols in flying ad hoc network", *IEEE Communications Magazine*, vol. 56, no. 1, pp. 136-142, January 2018. DOI: 10.1109/MCOM.2017.1700323
- [11] B. Sliwa, C. Schuler, M. Patchou, and C. Wietfeld, "PARRoT: Predictive ad hoc routing fueled by reinforcement learning and trajectory knowledge", in *Proc. 2021*

- IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1-7, April 2021. DOI: 10.1109/VTC2021-Spring51267.2021.9448959
- [12] Q. Yang, S. J. Jang, and S. J. Yoo, “Q-learning-based fuzzy logic for multi-objective routing algorithm in flying ad hoc networks”, *Wireless Personal Communications*, vol. 113, pp. 115-138, January 2020. DOI: 10.1007/s11277-020-07181-w
- [13] M. Zhang, C. Dong, S. Feng, X. Guan, H. Chen, and Q. Wu, “Adaptive 3D routing protocol for flying ad hoc networks based on prediction-driven Q-learning”, *China Communications*, vol. 19, no. 5, pp. 302-317, May 2022. DOI: 10.23919/JCC.2022.05.005
- [14] Y. Cui, Q. Zhang, Z. Feng, Z. Wei, C. Shi, and H. Yang, “Topology-Aware Resilient Routing Protocol for FANETs: An Adaptive Q-Learning Approach”, *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18632-18649, October 2022. DOI: 10.1109/JIOT.2022.3162849
- [15] J. Guo, H. Gao, Z. Liu, F. Huang, J. Zhang, X. Li, and J. Ma, “ICRA: An Intelligent Clustering Routing Approach for UAV Ad Hoc Networks”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 2, pp. 2447-2460, February 2023. DOI: 10.1109/TITS.2022.3145857
- [16] X. Qiu, Y. Xie, Y. Wang, L. Ye, and Y. Yang, “QLGR: A Q-learning-based geographic FANET routing algorithm based on multi-agent reinforcement learning”, *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 15, no. 11, November 2021. DOI: 10.3837/tiis.2021.11.020
- [17] Y. N. Chen, N. Q. Lyu, G. H. Song, B. W. Yang, and X. H. Jiang, “A traffic-aware Q-network enhanced routing protocol based on GPSR for unmanned aerial vehicle ad hoc networks”, *Frontiers of Information Technology & Electronic Engineering*, vol. 21, pp. 1308-1320, September 2020. DOI: 10.1631/FITEE.1900401
- [18] M. F. Khan, K. L. A. Yau, M. H. Ling, M. A. Imran, and Y. W. Chong, “An Intelligent Cluster-Based Routing Scheme in 5G Flying Ad Hoc Networks”, *Applied Sciences*, vol. 12, no. 7, pp. 3665-3699, April 2022. DOI: 10.3390/app12073665

Abstract: *Technological development and wider use of unmanned aerial vehicles increase the need to ensure reliable communication. Flying ad hoc networks (FANETs) represent an excellent solution for achieving this goal. The most significant characteristics of these networks are frequent network topology changes, and the high speed and low density of network nodes. Routing protocols for FANETs should be adapted to their characteristics, to improve the overall network performance. Traditional routing protocols fail to keep up with the dynamic nature of FANETs, so it is necessary to incorporate new techniques that better meet this challenge. The inclusion of reinforcement learning (RL) in routing protocols can significantly improve the performance of FANETs, due to its ability to continuously monitor and adapt to changes in the network environment. This paper will present the most important RL-based routing protocols for FANETs, as well as their detailed classification and comparative analysis.*

Keywords: *FANET, reinforcement learning, routing protocols*

APPLICATION OF REINFORCEMENT LEARNING IN ROUTING PROTOCOLS FOR FANETS

Marija Malnar, Pavle Bugarčić