Synthesis
International Scientific Conference of IT and Business-Related Research

FLYBIT - GENERATOR SLUČAJNIH BROJEVA

FLYBIT - ONLINE RANDOM NUMBER GENERATOR

Milomir Tatović, Saša Adamović, Milan Milosavljević
Univerzitet Singidunum, Danijelova 32, Beograd, Srbija

Apstrakt:
U ovom radu razvili smo sopstveni izvor slučajnosti za generisanje slučajnih brojeva. Sinteza izvora slučajnosti zasniva se na javno dostupnim podacima civilnog avio saobraćaja. Na osnovu različitih tipova promenljivih podataka dobijenih u realnom vremenu od letilica, preko metoda za digitalnu obradu signala, generišemo sekvence slučajnih brojeva. Ovako dobijeni slučajni brojevi, na zahtev Internet korisnika distribuiraju se preko namenski razvijenog veb-servisa (FlyBit) po uzoru na već postojeće servise sa sličnim uslugama. Osim razvoja servisa za generisanje slučajnih brojeva na osnovu sopstvenog izvora slučajnosti, neophodno je sprovesti rigoroznu informacionu analizu, koja će potvrditi očekivane osobine slučajnih generatora. Osim standardne analize, data je i komparativna analiza u odnosu na druge izvore slučajnosti (atmosferi šum i radioaktivne materije). Kao dodatak servisu, obezbeđen je i dodatni modul za praćenje entropije izvora u realnom vremenu.

Ključne reči:
sopstveni izvor slučajnosti, slučajni brojevi, analiza podataka, NIST statistički testovi, avio saobraćaj.

Abstract:
In this paper, we have developed our own source of randomness to generate random numbers. Random source synthesis is based on publicly available civil air traffic data. Using digital signal processing of air traffic data in real time, we have obtained different types of variables, which are afterwards used to generate random number sequences. These sequences are distributed using web service (FlyBit), which has been created solely for this purpose. It was modelled on the already existing similar web services. Besides developing web service for generating random number sequences, it is necessary to conduct a thorough information analysis, which should confirm the expected characteristics of true random generators. Apart from the standard analysis, a comparative analysis of other sources of randomness (*i.e.* atmospheric noise, radioactive matter) is given. We have also provided a module used for real time tracking of source entropy.

Key words:
own source of randomness, random numbers, data analysis, NIST statistical tests, air traffic.

UVOD

Slučajnost i determinizam su pitanja kojima se bavi filozofija i teško je dati tačnu definiciju koja može biti korišćena u nauci. Klod Šenon (*Claude Shannon*) definiše 1948. godine pojam Entropije i time omogućava da se uvidi razlika između dva događaja, određenog i neodređenog. Entropija predstavlja meru za količinu neodređenosti ili prosečna količina informacije koju sadrže generisane poruke nekog informacionog izvora (Veinović & Adamović, 2013). Ukoliko se uzmu dva različita događaja, odnosno njihove verovatnoće, pomoću formule iz Šenonove definicije entropije može se odrediti vrednost entropije. Ako se kao informacioni izvor koristi impulsni generator, gde se kao jedan događaj uzima stanje kada impuls postoji i obeležava sa 1, a kao drugi događaj stanje kada impuls ne postoji i obeležava sa 0, tada su verovatnoće ova dva događaja P(1) i P(0). Imajući u vidu da impulsni generator ima samo dva stanja (0 ili 1) odnosno dva moguća događaja na izlazu, zbir vrednosti verovatnoće ta dva događaja mora biti jednak 1. Pa tako, ako se verovatnoća P(0) predstavi sa p a vrednost verovatnoće P(1) sa $1-p$ formula za entropiju binarnog izvora glasi:

$$H=-plogp-(1-p)log(1-p)$$

Ako je verovatnoća pojavljivanja nule ili jedinice na izlazu generatora jednaka, $P(0)=P(1)=0,5$ onda je entropija takvog generatora $H=1$. Za generator na čijem izlazu se dobija vrednost entropije $H\approx1$ kaže se da je generator slučajnih nizova.

Slučajni brojevi imaju širok spektar korišćenja, od simulacija, igara na sreću, estetike do sistema za donošenje odluka, ali značajnu primenu ima i u kriptografiji. Kako idealni kanal za prenos podataka ne postoji, prenos podataka obavlja se pomoću realnog kanala. Realni kanal je podložan smetnjama te tako nastaju gubici u prenosu. Ukoliko se podatak nadogradi slučajnim delovima gubici u prenosu se smanjuju. Prilikom testiranja sistema za donošenje odluka potrebno je simulirati stanja realnog sveta u kojima je potrebno doneti adekvatnu odluku. Kako bi sistem nesmetano funkcionisao u realnom svetu za testiranje se koriste slučajni događaji. Slučajne vrednosti su od vitalnog značaja i za igre na sreću. Nemogućnost pretpostavljanja vrednosti, odnosno konačnog ishoda, u ovoj oblasti omogućava ravnopravnost svih učesnika i samim tim veliku rasprostranjenost igara na sreću. Na kraju, ne i najmanje važna, kriptografija. Kriptografija je nauka ili umetnost koja pretvara smislene sekvence u na izgled nesmisleni slučajni šum koji se samo uz pomoć kriptološkog ključa može rekonstruisati u originalnu poruku (Kenny & Mosursk, 2005). Kriptografija se zasniva na matematičkim algoritmima i kriptološkim ključevima. Algoritmi se dele na simetrične i asimetrične, u zavisnosti da li koriste iste ili različite ključeve za šifrovanje i dešifrovanje. Kako su algoritmi javni, ceo sistem zavisi od kvaliteta samih ključeva. Proces generisanja kriptoloških ključeva zasniva se uvek na upotrebi generatora slučajnih brojeva i na taj način se obezbeđuje visoka sigurnost sistema usled nemogućnosti da se predvidi izlaz slučajnih generatora.

E-mail: milomir@tatovic.com
DOI: 10.15308/Synthesis-2015-107-111

Iz tog razloga u ovom radu pokušaćemo da obezbedimo generator kvalitetnih slučajnih sekvenci čiji izvor predstavljaju podacima koji su javno dostupni. Slučajne sekvence moći će da se preuzimaju putem Internet sajta.

2. PREGLED OBLASTI

Zbog sveobuhvatne potrebe za slučajnim događajima potrebno je obezbediti adekvatne generatore takvih događaja. Generatori slučajnih brojeva mogu biti deterministički i nedeterministički. Kod determinističkih generatora ulazni slučajni nizovi bivaju podvrgnuti matematičkim algoritmima kako bi dužina samih nizova bila uvećana ali uz očuvanje neodređenosti. Takvi generatori nazivaju se pseudo slučajni generatori (PRNG). Ovi generatori za svoj ulaz koriste najčešće izlaz nedeterminističkih generatora. Kod nedeterminističkih generatora ulaz je zasnovan na fizičkim pojavama koje se obično raznim fizičkim komponentama predstavljaju na izlazu pomoću nizova. Ovi generatori se nazivaju stvarno slučajni generatori (TRNG) i njihov izlaz je nemoguće ponoviti. Osim fizičkih, u koje spadaju generatori zasnovani na efektu elektronskog kola, fizičkih eksperimenata (radioaktivni raspad atoma) ili prirodnih pojava (atmosferski šum, cvetanje pupoljka) postoje i ne fizički zasnovani na događajima kao što su sistemsko vreme, vreme pristupa *hard* disku ili zasnovani na korisničkim interakcijama na računarima (npr. pokreti miša). Treba napomenuti da fizičke *TRN* generatore ne treba smatrati kao potpuno slučajne iako se zovu „stvarno" slučajni. Razlog za to je što generatori obično poseduju deo za post obradu signala koji sve nesavršenosti, kao što su dugi nizovi nula i jedinica odstranjuju, a sve u cilju konstantnog dobijanja izlaza većeg kvaliteta.

Realizacije *TRNG*-a se dosta razlikuju. Od jednostavnih strujnih kola sa samo par elemenata do realizacija za koje su potrebne kompleksne laboratorija. Kako se razlikuju same realizacije tako se razlikuju i dobijeni rezultati. Da bi se za neki generator reklo da daje slučajne sekvence na svom izlazu mora se prethodno podvrgnuti određenim kontrolnim testovima. Broj statističkih testova nije tačno poznat. Svi testovi se bave određenom anomalijom nego izvora ali i samog generator. Neki raniji testovi za koje se smatralo da su *de facto* standardni, danas se ne koriste jer ih i generatori za koje važi da su „loši" sa lakoćom prolaze. Dizajn testova, odnosno odabir na koje će se anomalije obratiti pažnja, deli statističke testove prema oblasti primene generatora. Skup statistički testovi koji je definisan od strane *Donald Knuth,* koristi se nad sekvencama koje se primenjuju za potrebe različitih simulacija, dok se *NIST*-ovi testovi primenjuju za slučajne sekvence koje će kasnije naći primenu u kriptografiji.

Svi *NIST*-ovi testovi su zasnovani na matematičkim formulama i primenjuju se na izlazne sekvence. Konačan rezultat svih testova predstavlja se vrednošću između nula i jedan. Rezultat testova predstavljen je sa vrednošću *P,* koja treba da ispuni sledeći uslov:

$$P \geq 0{,}01$$

Ukoliko je uslov ispunjen, generator se smatra slučajnim, što znači da daje slučajne sekvence na izlazu. Međutim, potrebno je da vrednost *P* bude što veća ($P \approx 1$) i na osnovu toga se generatori rangiraju po kvalitetu.

2.1. RANDOM.ORG

Veb-servis na adresi *www.random.org*, u vlasništvu firme *Randomness and Integrity Services Ltd* osnovan je 1997. godine. Od samog početka rada ovog servisa, kao izvor slučajnosti ovog generatora korišćen je atmosferski šum. U početku, prikupljan jeftinim radio prijemnikom, ali kako je servis postajao popularan, tako se i tehnologija prikupljanja unapređivala. Danas, slučajne sekvence generisane putem ovog servisa zasnovane su na atmosferskom šumu. Brzina generisanja je oko 3000 bitova u sekundi po radio prijemniku. Ukoliko želimo da preuzmemo binarnu slučajnu sekvencu, maksimalna vrednost koju je moguće preuzeti sa servisa tokom jednog upita je limitirana na 16 *KB*, odnosno 128 *Kbit*-a. Važno je napomenuti, da je komunikacioni kanal između korisnika i servisa zaštićen SSL protokolom. Ovo je značajno zbog potencijalne upotrebe slučajnih sekvenci u kriptografskim aplikacijama.

2.2. HOTBITS

HotBits je takođe servis na Internetu koji nudi slučajno generisane sekvence – brojeve (Walker, 2006). Vlasništvo je firme *Fourmilab* i postoji od 1996. godine. Kao izvor za generisanje slučajnih sekvenci koristi se proces radioaktivnog raspadanja atoma u jedinici vremena. Ovakav izvor slučajnosti smatra se izuzetno kvalitetnim, jer statistički testovi ukazuju na kvalitet izvora slučajnosti. Brzina generisanja slučajnih sekvenci primenom tehnologije koju koristi *HotBits* je oko 8000 bitova po sekundi. Ipak ukoliko želimo da preuzmemo binarnu sekvencu sa sajta (SAJT) možemo za jedan upit preuzeti maksimom od 2 *KB* odnosno 16 *Kbit*-a. Komunikacija klijent server je i kod ovog sajta realizovana pomoću *SSL*-a što u neku ruku garantuje da podaci koji putuju između servera i klijenta ne mogu biti kompromitovani.

3. METODOLOGIJA RADA

Istraživački radovi koji su za cilj imali pronalaženje adekvatnog izvora koji može biti iskorišćen za realizaciju sopstvenog generatora slučajnosti doveli su nas do podataka civilnog vazduhoplovstva. U ovom radu podaci civilnog vazduhoplovstva za razliku od rada „*One method for generating uniform random numbers via civil air traffic*" (Tatović et al, 2014) nisu prikupljani iz baza podataka koje se nalaze na internetu već se podaci prikupljaju direktno od aviona. Na ovaj način onemogućena je zloupotreba podataka.

3.1. PRIKUPLJANJE, OBRADA I BINARIZACIJA ADB-S PODATAKA

Instrumenti u avionima prikupljaju signal sa *GPS* satelita, računaju svoju geografsku poziciju i putem radio signala je distribuiraju dalje. Osim geografske širine i dužine, kao i nadmorske visine u okviru radio signala pakuje se i brzina kretanja, jedinstveni kod leta, kao i razni drugi podaci relevantni za kretanje aviona i komunikaciju kako sa kontrolom leta tako i sa drugim avionima u vazduhu. Taj sistem naziva se *ADS-B (*Hansman & Kunzi, 2011). Danas je usled bezbednosti avio saobraćaja procenat aviona koji poseduju *ADS-B* opremu dosta visok. *ADS-B* predajnik na avionu konstanto šalje radio signal koji se na zemlji prikuplja pomoću *OMNI* antena i kasnije pomoću dekodera pretvara iz radio signala u niz tekstualnih poruka.

U realizacije *FlyBit* generatora antene sa dekoderom su postavljene na dve tačke na teritoriji Srbije i međusobno su udaljene oko 300 *km* vazdušne linije. Obe antene šalju podatke ka serveru putem zaštićene veze SSL protokolom. Severna antena koja pokazuje bolje performanse, odnosno veći domet, je postavljena kako primarna, dok je južna antena sekundarna. Na server podaci dolaze sa obe antene i zatim se upoređuju. Uko-

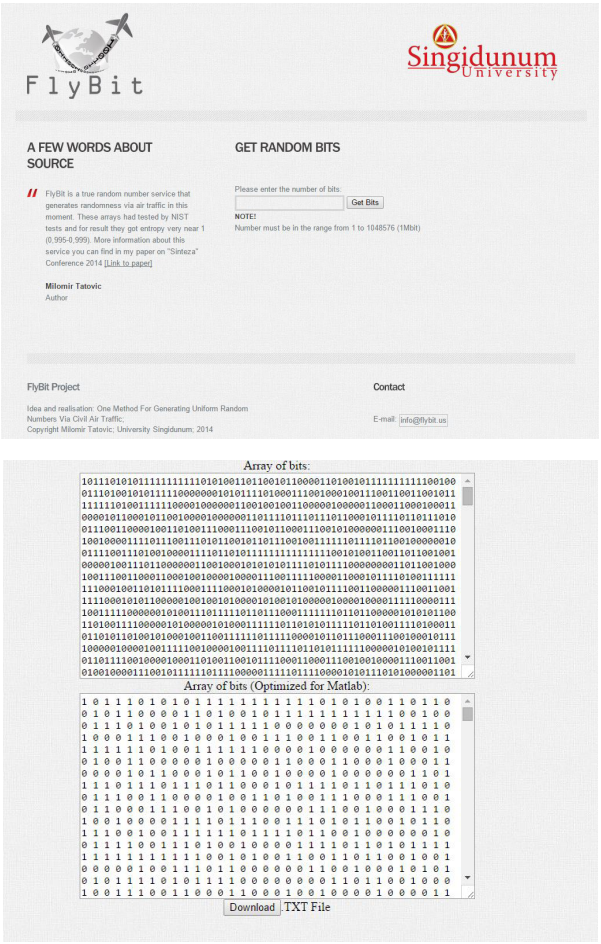
DOI: 10.15308/Synthesis-2015-107-111

liko su obe antene prikupile podatak istog aviona u vremenski bliskom trenutku (do 500*ms*) uzima se podatak primarne antene. Analizom je ustanovljeno da jedino podaci geografske širine i dužine jesu vremenski promenljivi, te se oni i koriste za binarizaciju. Svaki prikupljeni let koduje se sa 16 bitova, 8 bita predstavljaju geografsku širinu i 8 bita predstavljaju geografsku dužinu. Ovako postavljene antene omogućuju prikupljanje iznad teritorije Mađarske, Slovenije, teritorije Balkana, i dela Jadranskog i Jonskog mora. Broj letova u jedinici vremena dosta varira, od vremenski manje frekventnih do perioda visoke frekvencije saobraćaja.

3.2. DISTRIBUCIJA SLUČAJNIH BROJEVA – VEB SERVIS

Putem našeg servisa *www.flybit.us*, moguće je preuzeti slučajno generisane binarne sekvence preko izvora slučajnosti opisanog u prethodnom poglavlju. Binarne sekvence bi mogle biti generisane u trenutku zahteva korisnika sajta kada bi broj antena bio veći, ovako obezbeđena je sigurna baza koja se konstantno dopunjava novo generisanim binarnim sekvencama. Na taj način omogućeno je da korisnik interneta u bilo kom trenutku može da preuzme do 1 *Mbit* po upitu. Nakon preuzimanja binarna sekvenca se zauvek briše iz baze i na taj način osigurava da se podatak može preuzet samo jednom, što je izuzetno bitno ukoliko se slučajne sekvence koriste kao materijal za kriptografske ključeve. Princip upisa i ispisa iz baze je *FIFO*, prvi bit upisan u bazu, prvi će je i napustiti.

Na slici 1 prikazan je izgled veb-sajta našeg servisa. Korisnik servisa je potrebno da unese dužinu sekvence koju želi da preuzme (gore). Nakon toga na veb-sajtu se prikazuje generisana sekvenca (dole).



Slika 1. Izgled veb-sajta flybit.us

4. EKSPERIMENTALNA ANALIZA IZVORA SLUČAJNOSTI

Kako bi potvrdili polaznu pretpostavku da nizovi bitova generisani na ovaj način imaju visok stepen slučajnost, koriste se statistički testovi propisani od strane *NIST*-a (NIST SP 800-22) (Rukhin, 2001). Ovi testovi su razvijeni za testiranje slučajnosti binarne sekvence dobijene preko nekog izvora slučajnosti, tako što se fokusiraju na različite vrste ne slučajnosti koje bi mogle da postoje u testiranoj sekvenci. Testovi se kontinualno primenjuju nad generisanim binarnim sekvencama (približno 10.000 bitova) pre upisa u bazu i na taj način sekvence koje se pokažu kao ne slučajne se i ne upisuju u bazu. Ovakav vid obrade sekvenci je očekivan, jer kao što je već rečeno kod fizičkih generatora postoji mogućnost anomalija koje treba odstraniti. Na veb-sajtu *www.flybit.us* grafički su prikazane vrednost *P,* koje predstavljaju pojedinačni rezultat statističkih testova. Svaki od testova se primenjuje nad slučajnim vrednostima skladištenim u bazi podataka.

Prema preporuci *NIST*-a, potrebno je generisati onoliko uzoraka za testiranje koliko ima testova. Metodologija razvijena za testiranje izvora slučajnosti je u potpunosti primenjena u eksperimentalnom delu ovog rada.

Rezultati koji slede (Tabele 1, 2, 3, i 4), predstavljaju komparativnu analizu tri nezavisna izvora slučajnosti po kvalitetu generisanih slučajnih brojeva. Performanse navedenih izvora neće biti predmet ovog istraživanja. Analizom su obuhvaćeni *FlyBit* (avio saobraćaj), *Random.org* (atmosferski šum) i *Hotbits* (radioaktivni materijali).

Sa svakog izvora preuzeti su testirani uzorci u blokovima po 10,000 bitova, nad kojim će se obaviti pojedinačna testiranja sa: frekventnim testom, *runs* test, serijski test i entropijski test.

Baterija testova na koju smo se mi ograničili ne obuhvata sve testove predložene od strane *NIST*-a. Razlog tome je namena dizajniranog generatora. Naime, mi smo se ograničili na testove koji će garantovati kvalitet slučajnih brojeva sa aspekta kriptografske primene.

4.1. ISPITIVANJE UČESTANOSTI U NIZU (FREQUENCY TEST)

Preko frekventnog testa ispitujemo uravnoteženost jedinica i nula u generisanim binarnim sekvencama. Za uspešnost ovog testa poželjan je približno jednak odnos jedinica i nula. Na osnovu rezultata ovog testa, moguće je predvideti rezultat entropijskih testova.

	HotBits	random.org	FlyBit
P	0.7189	0.7507	0.8253

Tabela 1. Frekventni test

Na osnovu dobijenih rezultata, prikazanih u tabeli 1, zaključujemo da je naš izvor slučajnosti postigao odličan, bolji rezultat od preostala dva izvora slučajnosti. Međutim, vrednosti frekventnog testa nisu stacionarne i mogu da variraju u različitim vremenskim intervalima za svaki izvor slučajnosti.

4.2. ISPITIVANJE UZASTOPNIH PONAVLJANJA ISTIH BITOVA (RUNS TEST)

Preko *Runs* testa utvrđujemo broj uzastopnih ponavljanja nula ili jedinica. U tabeli 2, prikazani su dobijeni rezultati sa sve izvore slučajnosti.

DOI: 10.15308/Synthesis-2015-107-111

	HotBits	random.org	FlyBit
P	0.6536	0.9803	0.8836

Tabela 2. *Runs* test

Posle ovog testa, uslov za proglašenje testirane binarne sekvence slučajnom je ispunjen. Poređenjem dobijenih rezultata, *HotBits* je imao najlošiji rezultat, dok je *Random.org* imao najbolji. Rezultat našeg generatora se našao između ova dva rezultata.

4.3. SERIJSKI TEST (SERIAL TEST)

Preko serijskog određujemo ujednačenost raspodele preko trigrama. Za slučajnu binarnu sekvencu, očekuje se ujednačena distribucija svih trigrama.

	HotBits	Random.org	FlyBit
000	1323	1215	1303
001	1267	1250	1245
010	1263	1280	1251
011	1236	1195	1234
100	1268	1251	1246
101	1232	1225	1240
110	1237	1195	1235
111	1174	1245	1246
P_1	0.45514	0.3174	0.8223
P_2	0.8139	0.1088	0.6958

Tabela 3. Serijski test

Uspešnost serijskog testa se procenjuje preko dve *P* vrednosti. U oba slučaja *P* vrednost treba da ispunjava kriterijum za potvrdu slučajnosti testirane binarne sekvence. Uvidom u rezultate testa u tabeli 3, *Random.org* kao izvor slučajnosti dobio je najlošije rezultate, dok su *HotBits* i *FlyBit* postigli približne rezultate na ovom testiranju.

4.4. ENTROPIJSKI TEST

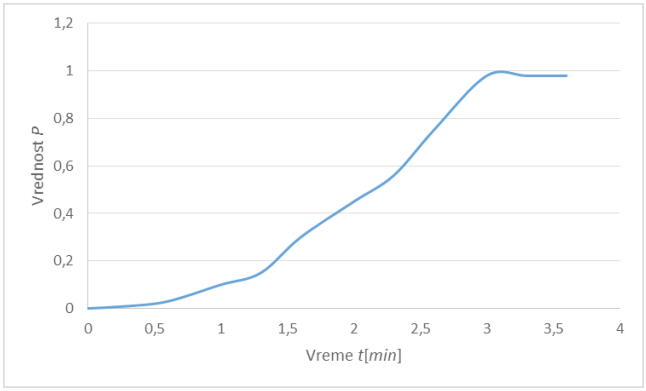
Posmatrana karakteristika u ovom testu je ukupan broj uzastopnih ponavljanja jedinice ili nule u čitavoj sekvenci, posmatrano prvo bit po bit, zatim preko bigrama, trigrama i na kraju matrica. Rezultat ovog testa je bio podjednako uspešan za sva tri izvora slučajnosti. Uvidom u rezultate u tabeli 4, postignuta je maksimalna količina informacije po jednom bitu ili prosečna količina informacije na izvoru generator je približno $H = 1$.

	HotBits	Random.org	FlyBit
Monobit	0.9997	0.9999	0.9999
Bigram	0.9997	0.9999	0.9999
Trigram	0.9997	0.9998	0.9999
4x4 Matrices	0.9995	0.9998	0.9999

Tabela 4. Entropijski test

4.5. REZULTATI STATISTIČKIH TESTOVA POSMATRANIH U REALNOM VREMENU

Statistički rezultati prikazani u realnom vremenu su dati za naš izvor slučajnosti. Na slici 2 prikazana je prosečna vrednost *P* navedenih statističkih testova i njihova zavisnost od vremena. Naime, prosečna vrednost *P* pokazala je zavisnost u odnosu na vreme ili na dužinu binarne sekvence koja zavisi direktno od gustine avio saobraćaja koji je doveden u naš generator slučajnosti. Početna vrednost za *P=0*, predstavlja momenat u kom je započet proces prikupljanja podataka (generator pušten u rad) i momenta u kada je otpočelo testiranje u realnom vremenu (testiranje dobijenih sekvenci čija je dužina manja od očekivane za testiranje). Nakon određenog vremena koje prošlo i uvida u stanje *P* vrednosti za naš izvor, potvrdili smo stacionarne osobine generatora slučajnih brojeva. Na ovaj način, sigurni smo da se rad generator, u vidu variranja kvaliteta neće menjati u jedinici vremena.



Slika 2. Rezultati statističkih testova u realnom vremenu (***P***)

5. ZAKLJUČAK I BUDUĆI RAD

U ovom radu, pokazali smo mogućnosti za sintezu sopstvenog izvora slučajnosti na osnovu stanja letilica civilnog saobraćaja. Distribuciju generisanih slučajnih brojeva obavljamo preko dodatno razvijenog veb-servisa na Internetu – *www.flybit.us*. Kvalitet našeg generator potvrđen je primenom adekvatnih statističkih testova i komparativnom analizom u koju su bili uključena još dva izvora slučajnosti (Random i HotBits). Dobijeni rezultati ukazuju na visok kvalitet slučajno generisanih brojeva preko našeg generatora. Takođe, statistikom u realnom vremenu potvrđena je njegova stacionarnost sa aspekta pouzdanog i stabilnog rada u eksploataciji. Za budući rad nameću se pitanja sigurne distribucije slučajno generisanih brojeva (SSL) i pitanja vezana za performanse generatora.

LITERATURA

Hansman, J., & Kunzi, F. (2011). *ADS-B benefits to general aviation and barriers to implementation*. Massachusetts Institute of Technology.

Kenny, C., & Mosursk, K. (2005). Random Number Generators: An Evaluation and Comparison of Random.org and Some Commonly Used Generators. Preuzeto sa https://www.random.org/analysis/Analysis2005.pdf

Random.org (2015). What's this fuss about true randomness? Preuzeto sa http:random.org.

DOI: 10.15308/Synthesis-2015-107-111

Rukhin, A., Soto, J., & Nechvatal, J. (2001). A Statistical Test Suite for the Validation of Random Number. Preuzeto sa http://csrc.nist.gov/groups/ST/toolkit/rng/documents/SP800-22b.pdf

Tatović, M., Adamović, S., Jevremović, A., & Milosavljević, M. (2014). One method for generating uniform random numbers via civil air traffic. Međunarodna naučna konferencija Univerziteta Singidunum Sinteza 2014, 25.-26.04.2014. Beograd: Univerzitet Singidunum.

Veinović , M., & Adamović, S. (2013). Kriptologija I : osnove za analizu i sintezu šifarskih sistema. Beograd: Univerzitet Singidunum.

Walker, J. (2006). Genuine random numbers, generated by radioactive decay. Preuzeto sa https://www.fourmilab.ch/hotbits/

DOI: 10.15308/Synthesis-2015-107-111