

MODERNE SOFTVERSKE TEHNIKE U PREPOZNAVANJU PROSKRIBOVANIH KOMPJUTERSKIH SADRŽAJA

Julijana Mirčevski, Informatička škola „Educon“, Beograd

julijana@afrodita.rcub.bg.ac.yu

Biljana Djokić, Informatička škola „Educon“, Beograd

djokicb@eunet.yu

Nikola Popović, Ministarstvo inostranih poslova

nikola.popovic@smip.sv.gov.yu

Abstrakt: Rad diskutuje mogućnost sakrivanja tajnog fajla u drugom računarski vidljivom fajlu kao i u digitalnoj formi slike. Moderna stegoanaliza obezbeđuje metode i postupke kojima je moguće detektovati ovakav fajl. Takve metode koriste softverske alate koji su realizovani kao programi na osnovu matematičkih modela i algoritama. Veoma efikasni programski paketi za analizu slike i njenu dekompoziciju na elementarne atome nastali su na wavelet transformaciji. U radu je izvršena procena efikasnosti programskih alata za otkrivanje nekorektnih kompjuterskih sadržaja.

Ključne reči: steganaliza, steganografija, wavelet transformacija, digitalna slika, vrednost piksela slike

MODERN SOFTWARE TECHNIQUE IN RECOGNITION OF A PROSCRIBE COMPUTER CONTENTS

Abstract: The paper discussed the ability to secrete file hiding within another computer readable file or within a digital image. Modern steganalysis provide the methods and approach which can be detected such file. Those methods use the software tools realised as a program on the base of mathematical models and algorithms. More efficient program packages to digital image analysing and decomposition in elementary atoms on the wavelet transformation are implemented. The evaluation of the program tools efficient for detection of uncorect computer contents was executed.

Key words: steganalysis, steganography, wavelet transformation, digital image, image's pixel value

1. UVOD

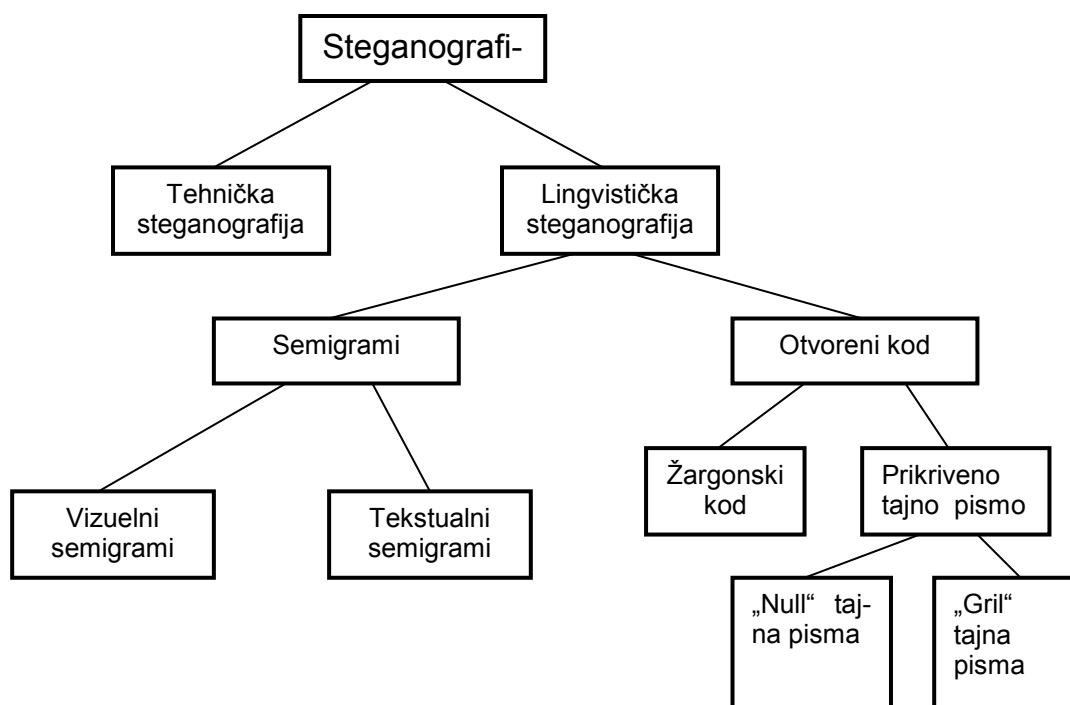
U kompleksnom digitalnom kodu fotografske slike može biti uvučen tajni fajl, ili nalog za ilegalnu finansijsku transakciju ili zabranjene informacije o računu korisnika ili kakva druga tajna poruka. Digitalna slika u formi **jpeg** fajla predstavlja idealan koverat za skrivanje tajne e-mail poruke, stenografske beleške ili uvučenog tajnog fajla. Digitalna slika u svojoj strukturi sadrži mnogo elementarnih prostora koji bi mogli biti potencijalni nosioci skrivenih podataka. Moguće je lagano pomeranje voxela na nivou bita a da se pritom ne vidi vizuelna razlika. Uobičajeni porodični snimak može da sadrži tajne informacije a ni istrenirano oko eksperta ne može da uoči razliku između ovakve i „normalne“ slike. Softverski alati su ovde neophodni za prepoznavanje i otkrivanje tajnog proskribovanog sadržaja. Steganografija ili stego tehnika [1], dozvoljava korisnicima da uvuku tajni fajl ili payload šifovanjem vrednosti boje samo neznatno prema redosledu bitova skrivenih podataka. Jednostavna stego tehnika se koristi da se nadje statistička mera koja može da se koristi za razlikovanje čiste slike od one koja sadrži skrivene podatke. Najjednostavniji nivo analize je razmatranje crno-bele ravanske slike gde svaki piksel ima skalu vrednosti sive od 0, što predstavlja crnu boju do vrednosti 255 što predstavlja belu boju. Fajl podataka za takvu sliku je dugačak niz brojeva - vrednosti skale sive koji reprezentuju svaki piksel.

Enkriptovani payload fajl [1], [2], može da se predstavi nizom nula i jedinica. Da bi se uvukao payload fajl, stego program upoređuje niz nula i jedinica koje pripadaju payload fajlu sa stringom vrednosti piksela u image fajlu. Stego program onda vrši promenu vrednosti piksela slike tako da vrednost neparnog piksela predstavlja nulu u payload stringu a vrednost parnog piksela predstavlja jedinicu. Primalac stego fajla, u nizu piksel vrednosti traži par-nepar da bi rekonstruisao niz podataka nula i jedinica koji posle toga treba da se dekriptuje da bi se pretražio tajni fajl.

Vizuelno, razlika u slici pre upisa i posle ne može da se uoči jer je pomeranje vrednosti piksela minorno. Međutim, šifovanje menja statističke osobine vrednosti piksela za sliku koju studiramo.

2. NEKI PRIMERI SKRIVANJA PORUKA U DIGITALNIM MEDIJIMA

Tajnu poruku je moguće sakriti na vrlo jednostavan način u Windows okruženju. U običnom Word fajlu upisati tekst u istoj boji kao i pozadina znači učiniti tekst vizuelno neuočivim. Takođe, u Power Point file, upisan tekst je moguće staviti u pozadinu slike i tako ga učiniti nevidljivim. Slične mogućnosti nudi i MS Excell, sadržaj ćelije se može učiniti nevidljivim.



Slika 1. Klasifikacija steganografskih tehnika prema [3]

Medjutim, ovakav tekst je pretraživ po sadržaju, ako se unapred zna ili pogodi niz karaktera koje tekst sadrži. Kod skrivenih poruka, najčešće se to ne zna.

Jedna od priznatih sistematizacija steganografskih tehnika prikazana je na slici 1. Ovu klasifikaciju napravio je Bauer 2002 godine a kasnije je u nekim podnivoima dopunjavana. Neke od sistematizovanih tehnika su vrlo stare ali su dobile novu značajnost pojavom internet komunikacije, eBay trgovinom, Web dizajnom ali i pojavom izlaganja i prenošenja pornografskih sadržaja putem interneta. Tako, na primer, „null chipers“ tip koristi vrlo jednostavan algoritam za skrivanje tajne poruke u vidljivoj poruci. Tajna poruka se formira tako što se iz javne poruke uzmu prva slova svake reči ili druga slova ili zadnja slova u reči. Korišćena još tokom Prvog svetskog rata u telegramima ova tehnika oživljava u spam porukama. Spam pošta je potencijalni nosilac ove vrste tajnih poruka.

U literaturi se često navodi primer sakrivanja karaktera šiftovanjem pojedinih bitova u osmobičnom zapisu u ASCII kodu. Ovde će biti dat primer skrivanja karaktera „G“ zamenom najmanje značajnog bita, prva pozicija s desna u levo, u primeru binarnog zapisa:

10010101 00001101 11001001 10010110 00001111 11001011 10011111 00010000
(1)

U gornjem zapisu najmanje značajan bit je podvučen.

Karakter „G“ u ASCII kodu je predstavljen binarnim nizom 01000111.

Ovih osam bitova, u čitanju najmanje značajnog bita može biti zapisano kao

10010100 00001101 11001000 10010110 00001110 11001011 10011111 00010001
(2)

pod uslovom da se prvobitni zapis (1) koristi kao nosilac informacije. Ovde je karakter „G“ utisnut u postojeći binarni zapis pomeranjem samo 4 bita najmanje značajnosti. Izmenjeni bitovi označeni su bold italic zapisom. Ova promena može biti

korišćena u promeni kodiranja nijanse RGB boje ili u zapisu koeficijenata u JPEG fajlu ili u zapisu vrednosti pokazivača u GIF ili BMP tipu fajla slike. Promena numeričke vrednosti bitova je vrlo mala i neuočljiva za ljudsko oko.

3. DETEKTOVANJE STEGANOGRAFSKIH FAJLOVA

Detektovanje steganografskog zapisa počiva na činjenici da steganografski proces prikriva tajnu poruku ali ne i interakciju dva računarska zapisa. Generalno uze-to, steganografski proces podrazumeva smeštanje tajne poruke u neki transportni medijum odnosno nosilac informacije. To bi moglo da se iskaže na sledeći način:

Steganografski fajl = skrivena poruka + nosilac informacije + steganografski ključ

Od vitalnog značaja za otkrivanje proskribovanog sadržaja je da se napravi veliki broj potencijalnih slika za pregled, varijacije i ispitivanje složenosti korišćenih algoritama za uvlačenje skrivenog fajla. Ove analize su poslužile za razvijanje jednostavnih i brzih tehnika kojima se vrši pregled i identifikacija slike koja sadrži skriveni fajl. Bergman i Davidson [1] su koristili programski sistem za prepoznavanje uzoraka u formi inteligentne neuralne mreže ANN (ANN – Artificial Neural Net) da ustanove razliku između „čiste slike“ i stego slike. Za obučavanje ANN sistema korišćeno je oko 1300 „čistih“ originalnih slika koje su na osam različitih načina promenjene nekom od metoda stego tehnike upisivanjem stranih fajlova korišćenjem software-a na bazi wavelet transformacije. Od ovakvih slika je iskreirana baza podataka od preko 10 000 slika koja omogućava da ANN prepozna novu sliku, kandidata za „čiste“ ili stego slike.

ANN uspostavlja jednu vrstu granične vrednosti odnosno vrednosti praga statističkih osobina slike. Preko te vrednosti fajl je sumnjiv i dalje se vrši identifikacija metoda korišćenog za uvlačenje stranog fajla, enkriptovanje i izdvajanje podataka tajne informacije. Za analizu sumnjive slike koriste se programski alati a software na bazi wavelet transformacije [3], [5], [7] je neizostavna komponenta takvih proizvoda.

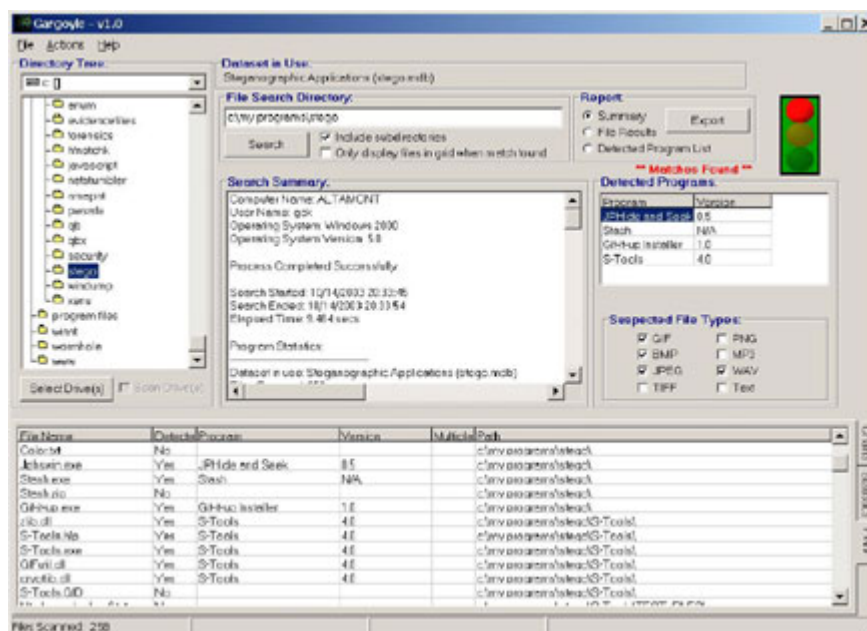
Tokom testiranja ANN sistem je prepoznao 92% stego slika a obeležio je samo 10% čistih slika. Istraživači sa Iowa State University, Department of Criminal Investigation rade na razvoju ovog programskog sistema u domenu unapredjenja korisnosti i grafičkog okruženja korisničkog interfejsa u cilju uvećanja „user friendly“ performansi.

4. SOFTVERSKI ALATI ZA STEGANOGRAFSKU DETEKCIJU

Evidentirano je preko sto steganografskih programa bilo da su komercijalno raspoloživi ili slobodno download-ovanih [2]. Oni se koriste za smeštanje, uvlačenje stranog fajla u fajl medijski nosilac. Medjutim za detektovanje steganografskih fajlova takodje se koriste softverski alati u vidu programa. Tekući raspoloživi programi za detekciju steganografskih fajlova raspolažu sa više različitih funkcija: otkrivaju prisustvo steganografskog programa, detektuju sumnjivi fajl nosilac i razaraju steganografski upisanu skrivenu poruku.

Analitičari kompjuterskog kriminala se zalažu za identifikovanje sumnjivog računara a to je računar koji raspolaže steganografskim softverom. Takodje, istraživanja pokazuju da programi za detektovanje proskribovanih sadržaja rade najefikasnije ako se u prvoj fazi koriste kao steganografsko upisivanje skrivenih poruka u fajl medijski nosilac. Ukoliko se nadje steganografski softver na sumnjivom računaru prema tipu tog softvera se odredjuje i metod stegoanalize. Na primer, S-Tools može direktno da ispituje GIF, BMP i WAV fajlove dok JP Hide & Seek omogućava analitičarima da traže sumnjivi fajl među JPEG fajlovima.

Bivši StegoDetect programski sistem od 2004. godine prerastao je u WetStone Technologies' Gargoyle i koristi se za otkrivanje prisustva steganografskog software-a. Gargoyle koristi sopstveni set podataka (hash set) sakupljen od svih poznatih steganografskih softverskih distribucija. Analizirajući fajlove upoređuje ih sa svojim hash setom. Slika 2 pokazuje izlaz u slučaju kada je Gargoyle našao proskribovani fajl sa steganografskim programom u direktorijumu na koji je bio usmeren.



Slika 2. Gargoyle izlaz u slučaju detektovanja sumnjivog fajla

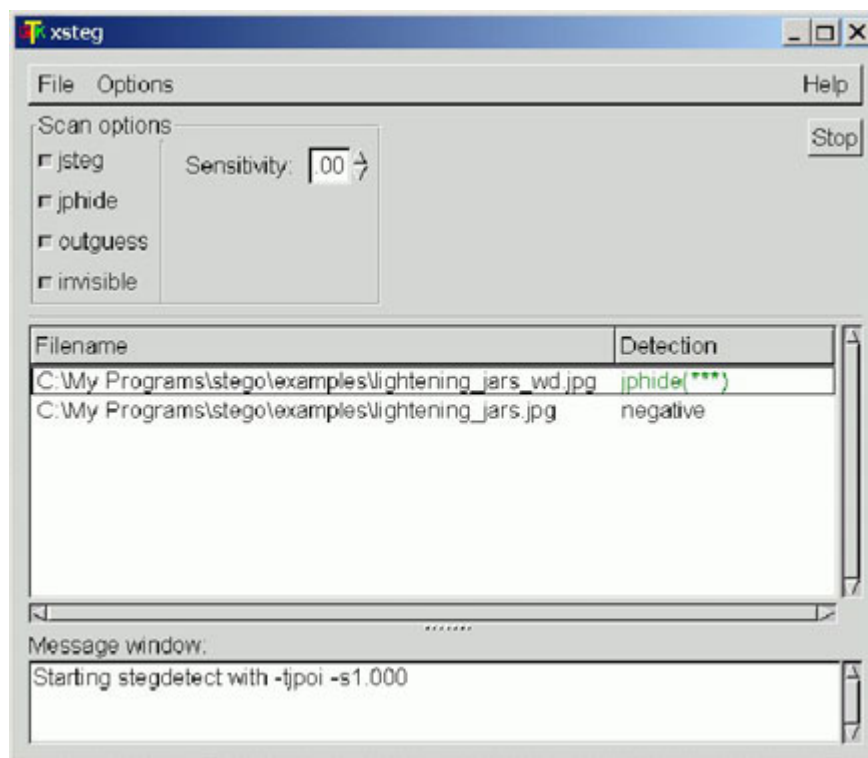
Ovaj program zajedno sa svojom hash setom se takodje koristi za otkrivanje prisustva kriptografije, instant poruka, Trojanskog konja, ključeva za logiranje i probije-nih šifara.

AccessData'sForensic Toolkit [5] koristi alat HashKeeper 2003 i *hash* setove za traženje vrlo različitih programa pri čemu treba obratiti pažnju da su steganografski programi dizajnirani tako da isključe *hash* setove poznatih „dobrih“ fajlova iz indeksa za pretraživanje. Gargoyle može da importuje ovakve *hash* setove.

Opšte prihvaćen i korišćen program za detekciju steganografskih fajlova je Niels Provos stegdetect. Ovaj program može da nadje skrivenu informaciju u JPEG sli-kama koristeći steganografske šeme kao što su F5, Invisible Secrets, JP Hide i Jsteg. Slika 3. pokazuje izlaz iz *xsteg*-a u funkciji grafičkog interfejsa za *stegdetect*.

Detekcija steganografskih zapisa postaje sve složeniji posao i zbog negativne korelacije izmedju sve manje veličine softverskih programa za stegoanalizu i povećane veličine memorijskog kapaciteta medija. Treba primetiti da S-Tools zahteva manje od 600 KB disk prostora i može da se izvršava direktno, sa pomoćnog medija (CD, floppy ili USB) bez dodatne instalacije. Pod tim uslovima ovaj program uopšte ne može biti registrovan na hard disku.

Današnji programi za detekciju uključuju inteligentnija rešenja i za detekciju steganografskog algoritma korišćenog za upisivanje tajne informacije tako da analitičar može odmah da pokrene proces rekonstrukcije tajne poruke.



Slika 3. Izlaz iz xsteg-a nakon ispitivanja dva sumnjiva JPEG fajla

Programski sistemi za analizu slike na bazi wavelet transformacije široko su primenjivani za konstituisanje forenzičkog softvera u domenu stegoanalize. Ovde će biti bliže opisana dva takva programa kojima se koristi veliki broj istraživača i analitičara. To su WaveLab i LastWave.

WaveLab je biblioteka MATLAB rutina za wavelet analize, wavelet-paket analize, kosinusne paketske analize i uparivanje. Biblioteka je besplatno raspoloživa na internetu i postoje verzije za mekintoš, unix i windows mašine. Wavelab se koristi u nastavnim kursevima u adaptiranim wavelet analizama na Stenfordu i na Berkliju. To je osnova za autorska wavelet istraživanja i može da se koristi za reprodukciju slika u njihovim objavljenim radovima i za ponovnu obradu tih slika variranjem različitih parametara. WaveLab sadrži preko 1100 fajlova koji su dokumentovani, indeksirani i ukršteno povezani na različite načine. MATLAB MEX fajlovi se opsežno koriste u cilju povećanja propusne moći izlaza.

U dodatku implementacije rutina osnovnih wavelet transformacija za konačne setove podataka (periodične transformacije i granično-korektivne transformacije), wavelet-paketne analize, kosinusne paketske analize i uparivanje, biblioteka sadrži skriptove za koje autori veruju da će pomoći u učenju praktičnih aspekata wavelet analize.

LastWave je programsko okruženje za Wavelet obradu signala i slike pisan u C jeziku za Unix i Macintosh računare. Ovaj programski paket je u kategoriji "free-ware" i "stand-alone", što uključuje samostalno funkcionisanje odnosno ne zahteva dodatne programe za izvršavanje. LastWave može da se pretražuje preko interneta na [8].

LastWave je kreirao i održavao Emanuel Bacry sa Ecole Polytechnique u Francuskoj. Kasnije je svoje doprinose u implementaciji ovog paketa u sadašnjem tekućem obliku dalo više istraživača uključujući i Stephane Mallat-a. Ovaj software objedinjuje komandni jezik kao objektno orijentisani grafički jezik visokog nivoa za pokazivanje jednostavnih objekata (dugmad, stringovi, ...) ali i više složenijih kao što su: signali, slike, wavelet transformacije, vremensko-frekventne ravni. Programi za računanje i komande su pregrupisani u nezavisne pakete. Takođe je raspoloživa i obi-

mna on-line dokumentacija. Nove komande, koje se dodaju implementirane su u komandnom jeziku ili kao C podrutina. LastWave paket se vrlo brzo razvijao baš poslednjih godina, koristeći kompatibilne pakete i rutine koje su dodavali internet korisnici širom sveta.

LastWave softversko okruženje ima izuzetno dobru dokumentaciju. Izvorni kod svake rutine je u zaglavlju dopunjen deklaracijom da se radi o „free ware“ software-u po principu GNU projekta na kome je ustvari pionirski sprovedena ideja „free-software“. Ovde će biti dati detalji o LastWave Kernel 1.6.1 odnosno jezgru komandnog jezika LastWave paketa za rad pod **unix** operativnim sistemom.

Ovaj program definiše sve osnovne komande koje su grupisane u različite tabele. Svaka komanda je opisana sa tri polja: ime komande, ime **C** funkcije i jedna **help** linija. Ime **C** funkcije korespondira sa stringom koji opisuje listu argumenata. Opis je strukturiran u tabelu komandnog interpretera.

5. ZAKLJUČAK

Steganografska istraživanja bi trebalo da se podrazumevaju i sprovode permanentno u cilju sprečavanja kompjuterskog kriminala slično kao DNK analiza. Korišćenje interneta kao medija [6,7] i kod nas je zastupljeno u meri koja zahteva da istina ne bude stvar slučaja. Stego analizu treba sprovoditi u cilju osnaživanja zakona u procesima koji predhode kriminalnom aktu i čine pripremu za izvršenje krivičnog ili kriminalnog dela a samo delo je izvan domera računarskih aktivnosti. U tom smislu, zalažemo se za dopunu zakonske regulative u cilju obuhvatanja softverske tehnologije na pronaženju proskribovanih sadržaja

Steganografija, međutim, obezbeđuje neke vrlo korisne i komercijalno značajne primene u digitalnom svetu. Najpoznatija takva funkcija je vodeni žig. Vodeni žig je izuzetno koristan u zaštiti intelektualne svojine nad proizvodima intelektualnog rada prisutnim na internetu.

Više istraživačkih timova i organizacija koje su sprovodile steganografska istraživanja na internet Web sajtovima u svojim radovima ne objavljuju imena i adrese tih sajtova ili publikuju adresu sajta bez navodjenja rezultata ispitivanja odnosno kakvih i koliko steganografskih fajlova je nadjeno.

6. LITERATURA

1. Ames Laboratory, Finding Computer Files Hidden In Plain Sight, May 24, 2006
2. Gary C. Kessler, An Overview of Steganography for the Computer Forensics Examiner, Forensic Science Communications, July 2004, Volume 6, Number 3, page 1-27
3. Bauer, F. L. Decrypted Secrets: Methods and Maxims of Cryptology, 3rd ed. Springer-Verlag, New York, 2002.
4. Artz, D. Digital Steganography: Hidding data within data, IEEE Internet Computing, 2001, Vol. 3, pag. 75-80
5. Farid, H. Detecting steganographic messages in Digital Images, Technical Report TR2001- 412, Dartmouth College, Computer Science Department, 2001, <http://www.cs.dartmouth.edu/~farid/publications>
6. Julijana Mirčevski, Biljana Djokić, Nikola Popović, The wavelet transform based software suitable for a digital contents analyze, V naučno-stručni skup "Nove tehnologije i standardi: digitalizacija nacionalne bastine ", 1-3. jun, Beograd

7. Nikola Popović, Julijana Mirčevski, Biljana Djokić, Elementi metodike implementacije sistema za upravljanje kolekcijama digitalizovanih dokumenata, V naučno-stručni skup "Nove tehnologije i standardi: digitalizacija nacionalne bastine ", 1-3. jun, Beograd
8. Željka Tomić, Julijana Mirčevski, Milesa Srećković, Miltenović Milan, Knežević Dragan, Programski alati za analizu slike nastale rasejanjem bazirani na wavelet algoritmima, Jahorina 2006, Srpsko Sarajevo, mart 2006