UNIVERZITET U BEOGRADU

ELEKTROTEHNIČKI FAKULTET

Jelica P. Radomirović

Jedna nova klasa sistema za destilaciju apsolutno tajnih kriptografskih ključeva zasnovanih na zajedničkoj slučajnosti

doktorska disertacija

Beograd, 2026.

UNIVERSITY OF BELGRADE

SCHOOL OF ELECTRICAL ENGINEERING

Jelica P. Radomirović

A novel class of systems for distilling absolutely secret cryptographic keys based on common randomness

Doctoral Dissertation

Belgrade, 2026.

Mentor:

dr Branko Kovačević, profesor emeritus
Univerzitet u Beogradu, Elektrotehnički fakultet

Članovi komisije:

dr Goran Kvaščev, redovni profesor
Univerzitet u Beogradu, Elektrotehnički fakultet

dr Zoran Banjac, viši naučni saradnik
Institut Vlatakom, Beograd

dr Pavle Vuletić, redovni profesor
Univerzitet u Beogradu, Elektrotehnički fakultet

Datum odbrane

_____

Zahvalnica

Ova doktorska disertacija nastala je uz nesebičnu podršku moje porodice, koja mi je tokom čitavog života najveći oslonac i motivacija. Mojoj sestri Milici, tati Predragu, mami Gordani i teti Vesni hvala na bezuslovnoj ljubavi, strpljenju, veri i vetru u leđa koji mi pružaju na svakom koraku. Disertaciju posvećujem vama.

Veliku zahvalnost dugujem svojim mentorima prof. dr Branku Kovačeviću i prof. dr Milanu Milosavljeviću, koji su bili inspiracija i uzor tokom doktorskih studija i koji su svojim znanjem, strpljenjem, posvećenošću i nesebičnom pomoći značajno doprineli nastanku ove disertacije. Na izdvojenom vremenu, stručnim savetima i pomoći tokom izrade disertacije zahvaljujem se svom komentoru dr Zoranu Banjcu.

Zahvaljujem se i Institutu Vlatakom koji mi je omogućio profesionalno i naučno usavršavanje, kao i saradnju sa vrhunskim stručnjacima i profesorima iz koje je proistekla ova disertacija.

Na savetima, bodrenju i svakoj vrsti pomoći koju su mi pružili tokom doktorskih studija zahvaljujem kolegama dr Milošu Pavloviću i Tamari Parojčić. Neizostavan deo procesa izrade disertacije činili su i prijatelji, čije su reči podrške bila dodatna motivacija. Hvala vam na strpljenju, razumevanju i ohrabrenju.

Naslov teze: Jedna nova klasa sistema za destilaciju apsolutno tajnih kriptografskih ključeva zasnovanih na zajedničkoj slučajnosti

Rezime:

Informaciono-teorijski pristup bezbednosti informacija ponovo dobija na značaju usled ubrzanog razvoja kvantnih računarskih tehnologija, koje dovode u pitanje bezbednost klasičnih računarski zasnovanih kriptografskih sistema. Za razliku od računarski bezbednih metoda, informaciono-teorijski pristup omogućava ostvarivanje apsolutne tajnosti komunikacije, pri čemu prisluškivač, nezavisno od raspoloživih računarskih resursa, ne može steći nikakvu informaciju o sadržaju poruke. Ovakav nivo bezbednosti zahteva generisanje i distribuciju velike količine tajnih ključeva maksimalne entropije, što predstavlja jedan od ključnih izazova savremene kriptografije.

Predmet ove disertacije je analiza, razvoj i eksperimentalna evaluacija nove klase sistema za generisanje i distribuciju apsolutno tajnih kriptografskih ključeva zasnovanih na različitim izvorima zajedničke slučajnosti. Istraživanje je usmereno na klasične informaciono-teorijske metode koje predstavljaju praktičnu, ekonomičnu i kvantno-otpornu alternativu kvantnoj distribuciji ključeva, uz mogućnost implementacije u postojećoj komunikacionoj infrastrukturi.

Centralni doprinos rada predstavlja uvođenje i primena Renijeve entropije drugog reda prisluškivača kao ključne informaciono-teorijske mere za analizu bezbednosti i optimizaciju procesa pojačanja privatnosti u sistemima za destilaciju tajnih ključeva. Razvijena je teorijsko-empirijska metodologija koja omogućava pouzdanu procenu bezbednosne margine, curenja informacija i efikasnosti iskorišćenja zajedničke slučajnosti. Poseban doprinos čini integracija metoda mašinskog učenja i intervala predviđanja dubokih neuronskih mreža u dizajn adaptivnih strategija pojačanja privatnosti, koje funkcionišu isključivo na osnovu lokalno dostupnih informacija.

U disertaciji je demonstrirano da biometrijski signali, posebno govorni i elektroencefalografski (EEG) signali, mogu predstavljati efikasne izvore zajedničke slučajnosti za sisteme destilacije tajnih ključeva. Na osnovu predloženih metoda projektovan je i eksperimentalno verifikovan autonomni sistem apsolutno tajne govorne komunikacije zasnovan na MELPe vokoderu i Vernamovoj šifri, sposoban za rad u realnom vremenu pri brzinama prenosa do 8,4 kb/s. Eksperimentalni rezultati potvrđuju visoku efikasnost, zanemarljivo curenje informacija i otpornost sistema na kvantne napade, čime se predloženi pristup nameće kao ozbiljna alternativa savremenim kvantnim i hibridnim kriptografskim rešenjima.

Ključne reči: Tajni ključevi, informaciono-teorijska bezbednost, pojačanje privatnosti, apsolutna tajnost, biometrijski signali

Naučna oblast: Elektrotehnika i računarstvo

Uža naučna oblast: Obrada signala i bezbednost informacija

Dissertation title: A novel class of systems for distilling absolutely secret cryptographic keys based on common randomness

Abstract:

The information-theoretic approach to information security has regained significant attention due to the rapid development of quantum computing technologies, which threaten the security of classical computationally secure cryptographic systems. Unlike computational security methods, the information-theoretic approach enables perfect secrecy of communication, whereby an eavesdropper, regardless of the available computational resources, cannot obtain any information about the message content. Achieving such a level of security requires the generation and distribution of large quantities of secret keys with maximum entropy, which represents one of the key challenges of modern cryptography.

The subject of this dissertation is the analysis, development, and experimental evaluation of a new class of systems for the generation and distribution of perfectly secret cryptographic keys based on various sources of common randomness (SCR). The research focuses on classical information-theoretic methods that provide a practical, cost-effective, and quantum-resistant alternative to quantum key distribution, with the ability to be implemented within existing communication infrastructures.

The central contribution of this work is the introduction and application of the eavesdropper's conditional Rényi entropy of order two (ECRE2) as a key information-theoretic measure for security analysis and optimization of the privacy amplification process in secret key distillation systems. A theoretical–empirical methodology is developed that enables reliable estimation of security margins, information leakage, and the efficiency of common randomness utilization. A further significant contribution is the integration of machine learning methods and prediction interval deep neural networks (PIDNN) into the design of adaptive privacy amplification strategies that operate exclusively on locally available information.

The dissertation demonstrates that biometric signals, particularly speech and electroencephalographic (EEG) signals, can serve as efficient sources of common randomness for secret key distillation systems. Based on the proposed methods, an autonomous system for perfectly secure speech communication was designed and experimentally validated, employing a MELPe vocoder and the Vernam cipher, and capable of real-time operation at transmission rates of up to 8.4 kb/s. Experimental results confirm high efficiency, negligible information leakage, and resistance to quantum attacks, positioning the proposed approach as a viable alternative to contemporary quantum and hybrid cryptographic solutions.

Key words: Secret keys, information-theoretic security, privacy amplification, perfect secrecy, biometric signals

Scientific field: Electrical Engineering and Computer Science

Scientific subfield: Signal processing and information security

Sadržaj

Lista skraćenica

AD	*Advantage Distillation*
APS-VCS	*Autonomous Perfectly Secure Low-Bit-Rate Voice Communication System*
BER	*Bit Error Rate*
BP	*Bit Parity*
BSC	*Binary Symmetric Channel*
DMS	*Discrete Memoryless Source*
ECC	*Error-Correcting Code*
ECRE2	*Eavesdropper Conditional Renyi Entropy of order 2*
EEG	*Elektroencefalogram*
EKG	*Elektrokardiogram*
FIFO	*First In First Out*
HC	*Huffman coder*
IR	*Information Reconciliation*
KAR	*Key Acceptance Rate*
KR	*Key Rate*
KSG	*Keystream Generator*
LDPC	*Low-Density Parity-Check*
LHL	*Leftover Hash Lemma*
LP	*Linear prediction*
LR	*Leakage Rate*
LSP	*Linear Spectral Pairs*
MELPe	*Mixed-Excitation Linear Prediction voice coder*
ML	*Machine Learning*
MPIW	*Mean Prediction Interval Width*
NIST	*National Institute of Standards and Technology*
OTP	*One-Time-Pad*
PA	*Privacy Amplification*
PCM	*Pulse code modulation*
PICP	*Prediction Interval Coverage Probability*
PIDNN	*Prediction Interval Deep Neural Network*
PKI	*Public Key Infrastructure*
QKD	*Quantum Key Distribution*
SCR	*Source of Common Randomness*
SHAP	*SHapley Additive exPlanations*
SKD	*Sequential Key Distillation*
SNR	*Signal-to-Noise Ratio*
VoIP	*Voice over Internet Protocol*
vPCP-V	*Vlatacom Personal Crypto Platform for Voice encryption*
vTRNG	*Vlatacom True Random Number Generator*

1 Uvod

1.1 Predmet i motivacija istraživanja

Informaciono-teorijski pristup bezbednosti informacija ponovo je privukao značajnu pažnju usled napredaka u kvantnim računarskim tehnologijama. Suštinski princip ovog pristupa lako je izraziti: kriptografski sistem obezbeđuje apsolutnu tajnost (eng. *perfect secrecy*) poruka, odnosno informaciono-teorijsku tajnost, ako i samo ako entropija njegovog tajnog ključa nije manja od entropije samih poruka. U takvim uslovima, apriorna neodređenost sadržaja prisluškivanih poruka ne može biti smanjena od strane napadača, bez obzira na računarske resurse koje poseduje, uključujući i resurse kvantnih računara [1], [2]. Cena koju je potrebno platiti za apsolutnu tajnost je potreba za enormnom količinom generisanih i distribuisanih tajnih kriptografskih ključeva. Shodno tome, savremena kriptografija ulazi u eru u kojoj prikupljanje i ekstrakcija slučajnosti, nezavisno od njenog tipa, porekla ili mesta nastanka, predstavljaju prioritetan zadatak u procesu generisanja i distribucije kriptografskih ključeva maksimalne entropije [3], [4], [5].

Savremeni pristupi u generisanju i distribuciji tajnih kriptografskih ključeva obuhvataju kako klasične informaciono-teorijske metode, tako i napredne kvantne protokole i hibridne sisteme. Najintenzivnija oblast istraživanja je kvantna distribucija ključeva (eng. *Quantum Key Distribution* - QKD) zasnovana na teoremi o nemogućnosti kloniranja iz kvantne fizike, koja tvrdi da se ne može napraviti identična kopija nepoznatog kvantnog stanja, što omogućava detekciju prisluškivanja [6]. Međutim, praktična primena QKD-a suočava se sa višestrukim izazovima koji ograničavaju njegov domet i efikasnost u realnim komunikacionim mrežama, uključujući ograničenja u brzini generisanja ključeva, dometu prenosa i troškovima implementacije [7], [8].

Razvoj oblasti informaciono-teorijskih pristupa generisanju i distribuciji tajnih kriptografskih ključeva započinje radovima *Shannon*-a [1] i *Wyner*-a [9], a zatim je kroz radove *Ahlswede* i *Csiszár*-a i *Maurer*-a evoluirala u sofisticirane modele destilacije ključeva zasnovane na zajedničkoj slučajnosti i interaktivnoj javnoj diskusiji, sa formalnom karakterizacijom kapaciteta tajnih ključeva (eng. *secret key capacity*) i praktičnim protokolima koji omogućavaju njihovo pouzdano generisanje u postojećim komunikacionim sistemima [3], [10], [11], [12]. Svaki od ovih pristupa donosi specifične prednosti i izazove vezane za bezbednost, efikasnost implementacije i otpornost na napade, što ih čini predmetom intenzivnog teorijskog i eksperimentalnog istraživanja u savremenoj kriptografiji [1], [5].

Predmet ove disertacije je analiza, razvoj i eksperimentalna evaluacija nove klase sistema za destilaciju apsolutno tajnih kriptografskih ključeva zasnovanih na različitim izvorima zajedničke slučajnosti. Prednosti predloženih sistema u odnosu na kvantni pristup su:

- jednostavna implementacija u postojećoj komunikacionoj infrastukturi,
- dostupnost svuda gde postoji komunikacioni kanal i neosetljivost na spoljna ometanja,
- manji troškovi implementacije (koristi postojeću infrastrukturu),
- informaciono-teorijska bezbednost koja je otporna na kvantne napade,
- adaptivnost na različite kanale i uslove,
- korišćenje različitih izvora entropije, kako iz samog komunikacionog kanala tako i iz nekog eksternog izvora slučajnosti,
- veće brzine generisanja ključeva i mogućnost generisanja u realnom vremenu [13], [14].

Usavršavanje i razvijanje informaciono-teorijski bezbednih klasičnih metoda za generisanje i distribuciju tajnih ključeva predstavlja efikasan, ekonomičan i teorijski utemeljen pristup. Ove

metode komplementarne su kvantnim metodama i nude praktičnu alternativu u scenarijima gde je kvantna infrastruktura nedostupna ili neisplativa.

1.2 Cilj disertacije

Problemi sa kojim se suočavaju trenutno korišćeni algoritmi u sistemima zaštite informacija predstavljaju osnovnu motivaciju za nastanak ove disertacije. Primarni ciljevi disertacije se mogu sažeti u sledeća četiri podcilja:

1. Sinteza nove klase sistema za generisanje i destilaciju tajnih ključeva (eng. *Secret Key Distillation* - SKD) zasnovanih na različitim klasama izvora zajedničke slučajnosti (eng. *Source of Common Randomness* - SCR). U okviru ovog cilja teži se karakterizaciji zadatog SCR praktično pogodnim pokazateljem koji meri njegovu maksimalnu količinu destilovanih apsolutno tajnih ključeva za zadati protokol usaglašavanja informacija (eng. *Information Reconciliation* - IR) i optimalnu strategiju prisluškivača kanala za javnu diskusiju. Istraživanje se dalje fokusira na ispitivanje mogućnosti uključivanja ovog pokazatelja u optimizaciju bloka za pojačanje privatnosti (eng. *Privacy Amplification* - PA), sa namerom da se dobijene optimalne strategije automatski nauče uz pomoć metoda mašinskog učenja i dubokih neuronskih mreža.
2. Posebna integracija ove nove klase SKD sistema u postojeće vokoderske sisteme malih brzina prenosa, sa težnjom da se ostvare istovremeno tri osnovna svojstva idealnog apsolutno tajnog autonomnog šifarskog sistema za kriptozaštitu govora.
    - Prvo svojstvo podrazumeva interni SCR koji obezbeđuje potpunu autonomiju generisanja i distribucije apsolutno tajnih kriptografskih ključeva, nezavisno od korišćenog telekomunikacionog kanala.
    - Drugo svojstvo se odnosi na postizanje dovoljne brzine destilacije i distribucije ključeva za robusnu implementaciju apsolutno tajnog Vernamovog šifarskog sistema govora u realnom vremenu.
    - Treće svojstvo obuhvata potpunu kontrolu svih kritičnih parametara sistema, kao što su neodređenost generisanih ključeva, brzina oticanja informacija ka potencijalnom prisluškivaču sa neograničenim računarskim resursima, kao i obezbeđivanje rada u realnom vremenu za zadatu klasu vokodera brzina prenosa do 8 kb/s.
3. Hibridizacija prirodnih SCR dodavanjem determinističke komponente, koja transformiše ove izvore u hibridne SCR, kontrolisanih entropijskih svojstava i brzine destilacije tajnih ključeva.
4. Identifikacija i eksperimentalna evaluacija novih SCR pogodnih za praktičnu primenu u SKD protokolima sa javnom diskusijom. Posebna pažnja posvećena je detaljnijoj analizi biometrijskih signala, kao što su elektroencefalografski (EEG) i govorni signali, sa ciljem ispitivanja njihovih mogućnosti kao izvora zajedničke slučajnosti u sistemima zaštite informacija.

1.3 Struktura rada

Rad je podeljen na 9 poglavlja, uključujući uvod i spisak korišćene literature. Pojedini delovi disertacije uključuju adaptirane slike i rezultate iz prethodno publikovanih radova autora disertacije koji su navedeni u spisku korišćene literature. U uvodnom poglavlju izložen je predmet, motivacija i cilj sprovedenih istraživanja.

Nakon poglavlja vezanog za pregled literature u kome se analizira trenutno stanje razvijenih teorijskih modela i praktičnih pristupa u realizaciji SKD protokola, sledi treće poglavlje koje se odnosi na teorijski okvir istraživanja. Usvajajući informaciono-teorijski pristup analizi i sintezi sistema za zaštitu informacija, ovo poglavlje daje kraći pregled korišćenih informacionih mera i kriterijuma koji postuliraju pojmove apsolutne tajnosti. Zatim sledi pregled postojećih modela i metoda generisanja i distribucije tajnih ključeva zasnovanih na SCR. Ostatak poglavlja posvećen je

detaljnom opisu SKD protokola i njegovih komponenti: izvora zajedničke slučajnosti, destilacije prednosti, usklađivanja informacija i pojačanja privatnosti.

U četvrtom poglavlju se uvodi ključna informaciono-teorijska mera – *Renyi*-jeva entropija drugog reda prisluškivača (eng. *Eavesdropper Conditional Renyi Entropy of order 2* - ECRE2), koja predstavlja bazičnu veličinu za sve teorijske i praktične rezultate prezentovane u ovoj disertaciji. Sledi pregled novih efikasnih PA strategija dobijenih na osnovu poznavanja ECRE2. Završni deo poglavlja odgovara na pitanje kako sintetisati efikasne estimatore ECRE2 zasnovane na mašinskom učenju i dubokim neuronskim mrežama iz klase PIDNN (eng. *Predicting Interval Deep Neural Networks*). U okviru procesa sinteze posebna pažnja je posvećena inženjerstvu ulaznih obeležja. Demonstrirana je efikasnost stilometrijskih obeležja izdvojenih iz prethodno transformisanih signala dostupnih u toku sprovođenja SKD protokola.

U petom poglavlju analizirani su SCR zasnovani na biometrijskim signalima, sa posebnim akcentom na EEG i govorne signale.

U šestom poglavlju prezentovana je eksperimentalna platforma za evaluaciju predloženih SKD sistema sa različitim SCR i različitim PA strategijama, uključujući i strategiju zasnovanu na PIDNN. Data je detaljna analiza svih eksperimentalnih rezultata uz poređenje ostvarenih performansi sa performansama postojećih sistema referisanih u dostupnoj literaturi.

U sedmom poglavlju prezentovan je praktično realizovan apsolutno tajni autonomni sistem zaštite govora na niskim brzinama prenosa. U okviru već realizovanog sistema za prenos govora (eng. *Vlatacom Personal Crypto Platform for Voice encryption* - vPCP-V), zasnovanog na MELP-e vokoderu (eng. *Mixed-Excitation Linear Prediction voice coder*), identifikovan je SCR koga čine sintetisani govor na osnovu vokoderskih parametara na predajnoj i prijemnoj strani. Sprovedena je rigorozna informaciono-teorijska i eksperimentalna analiza koja pokazuje da izvori ove klase mogu podržavati apsolutno tajnu zaštitu govora zasnovanu na Vernamovoj šifri, za brzine prenosa do 8.4 kb/s.

U osmom poglavlju su sumirani ključni rezultati i doprinosi disertacije, ograničenja predloženih rešenja i mogući pravci za dalja istraživanja.

2 Pregled literature

2.1 Informaciono-teorijski pristup u kriptografiji

Informaciono-teorijski pristup jedan je od najranije razvijenijih pristupa u kriptografiji, a odlikuje se sposobnošću da obezbedi zaštitu informacija nezavisno od računarskih moći napadača pružajući time bezuslovnu bezbednost štićenih informacija. Razvoj kvantnih računara dovodi do značajno naprednije i ubrzane kriptoanalize čime se ugrožavaju široko korišćeni mehanizmi zaštite informacija zasnovani na računskoj bezbednosti i nameće informaciono-teorijski pristup kao adekvatno rešenje.

Temelje ovog pristupa postavio je *Shannon* svojim radom [1] u kojem definiše i daje matematičku formulaciju konceptu apsolutne tajnosti. Uvođenjem informacionih mera, poput entropije, ekvivokacije i međusobne informacije, *Shannon* je omogućio kvantifikaciju informacione bezbednosti. Entropija (ili neodređenost) $H(M)$ je informaciona mera koja govori o tome koliko neodređenosti poseduje poruka M. Uslovna entropija (ekvivokacija) $H(M|C)$ meri neodređenost poruke ukoliko je poznat šifrat C, dok $I(M;C)$ pokazuje koliko informacija o poruci je dostupno ukoliko se poznaje šifrat C. Za sistem sa apsolutnom tajnošću važi da je $I(M;C) = 0$. Apsolutnu tajnost moguće je postići samo ukoliko je entropija ključa veća ili jednaka od entropije poruka, što se matematički formuliše kao:

$$H(|K|) \geq H(|M|) \tag{2.1}$$

gde je sa M je označena poruka koja se šifruje, a sa K označen ključ koji mora biti čisto slučajan niz generisan uniformno iz skupa svih mogućih ključeva zadate dužine. Dodatno, mora biti ispunjen uslov jedinstvenosti, tj., za svaki par poruke i šifrata mora postojati tačno jedan ključ koji ih povezuje. Ukoliko su prethodni zahtevi postignuti kriptografski sistem ne otkriva nikakve informacije o poruci. Realni kriptografski sistemi iako teže apsolutnoj tajnosti moraju uzeti u obzir da dužine ključeva zahtevane u jednom takvom sistemu nisu primenljive u praksi. Stoga su razvijeni i implementirani algoritmi koji se fokusiraju na „računsku bezbednost“.

2.2 Modeli i metodi generisanja i distribucije tajnih ključeva zasnovanih na izvorima zajedničke slučajnosti

Wyner-ov model *wiretap* kanala predstavlja značajno proširenje *Shannon*-ove teorije tajnog komuniciranja. Potpuno kontraintuitivno, u svom radu [9] pokazao je da je moguća apsolutno tajna komunikacija između dva legitimna učesnika pod uslovom da je kanal prisluškivača degradiran (zašumljeniji) u odnosu na glavni kanala. Njegova postavka iako revolucionarna sa aspekta praktične implementacije i realizacije, uslovljena je ograničenjima u pogledu pouzdanosti i bezbednosti. Odredio je gornju granicu za maksimalnu brzinu sigurne komunikacije koju je nazvao kapacitet tajnosti C_s a koja se definiše kao

$$C_s = max(I(X;Y) - I(X;Z)) \tag{2.2}$$

gde je sa $I(X;Y)$ označena međusobna informacija između ulaza i izlaza glavnog kanala dok je sa $I(X;Z)$ označena međusobna informacija između ulaza i izlaza kanala prisluškivača.

Nadogradnju *Shannon*-ove teorije apsolutne tajnosti i *Wyner*-ovog pristupa sa kanalima sa prisluškivanjem, omogućili su *Csiszár*, *Narayan* i *Ahlswede*, koji su u svojim radovima povezali teoriju informacija sa savremenim zahtevima praktične kriptografije. Između ostalog, u tim radovima postavili su matematičke temelje za generisanje ključeva koji ne zahtevaju prethodnu razmenu tajnih informacija. U radu [10] razmatrana je mogućnost uspostave tajnih ključeva na udaljenim lokacijama ukoliko dve legitimne strane u komunikaciji na početku poseduju korelisane opservacije. Razmena informacija kojom bi se izdvojila zajednička slučajnost i dobio tajni ključ odvija se putem javnog kanala koji može biti posmatran i od strane pasivnog napadača ali koji na kraju ne sme imati nikakvu informaciju o ključu. Predložen je dvofazni protokol u kojem se u prvoj fazi usaglašavaju informacije, sa ciljem dobijanja identične sekvence, dok se u drugoj fazi otklanjaju sve informacije koje su potencijalno procurele do prisluškivača. Potreba za eliminacijom potencijalno kompromitovanih informacija posledica je javne komunikacije, a obavlja se lokalno kod legitimnih učesnika komunikacije primenom iste heš funkcije (eng. *hash*). Dodatno, razmatraju i dva modela za izvor zajedničke slučajnosti. Model kanala pretpostavlja da legitimne strane u komunikaciji primaju korelisane signale preko odvojenih izlaza kanala. Model izvora podrazumeva da legitimni učesnici u komunikaciji imaju pristup nekom diskretnom izvoru slučajnosti koji generiše korelisane sekvence. Pretpostavlja se da prisluškivači imaju pristup korelisanim signalima istog izvora.

Ovim radom predstavljen je pristup koji postavlja teorijske osnove za savremene protokole informaciono-teorijske bezbednosti i distribucije tajnih ključeva, omogućavajući praktičnu implementaciju sigurne komunikacije u scenarijima gde tradicionalna kriptografija sa prethodno deljenim ključevima nije izvodljiva.

U nastavku svog istraživanja, *Ahlswede* i *Csiszár* u [15] uvode definiciju kapaciteta zajedničke slučajnosti i kvantifikaciju odnosa između brzine generisanja ključa i složenosti komunikacije. Uvode i osnovne granice koje protokoli za generisanje zajedničkih ključeva moraju da ostvare kako bi ostali bezbedni.

Csiszár i *Narayan* u radu [11] proširuju informaciono-teorijski pristup kriptografiji uvođenjem modela generisanja tajnih ključeva koji omogućava da više učesnika istovremeno učestvuju u procesu. Definišu tri kategorije bezbednosnih ograničenja na osnovu toga od koga se ključ krije: samo od javnih poruka, od spoljašnjih prisluškivača, ili od određenih učesnika u komunikaciji. Pored toga, uvode koncept komunikacione složenosti kao minimalnu količinu javne komunikacije potrebne da se postigne optimalna brzina generisanja tajnog ključa, što predstavlja kompromis između efikasnosti i bezbednosti u distribuiranim kriptografskim protokolima.

Istovremeno sa ovim radovima *Maurer* je u svojim radovima formalizovao koncept uslovne entropije u kontekstu kriptografskih sistema i uveo precizne matematičke alate za kvantifikovanje bezbednosti. Njegova teorija se zasniva na ideji da se bezbednost kriptografskog sistema može meriti kroz količinu informacija koju protivnik može da izvuče o tajnim podacima, pri čemu se ova količina kvantifikuje pomoću teorijsko-informacijskih mera poput entropije i međusobnih informacija. Ovaj pristup omogućava dokazivanje bezbednosti koja ne zavisi od ograničenih računskih resursa protivnika, već predstavlja apsolutnu bezbednost u matematičkom smislu. Razvio je model za usaglašavanje ključeva baziranim na prirodnim izvorima slučajnosti (poput satelitskog signala) i pokazao da je moguće generisati ključeve bez prethodno podeljene zajedničke tajnosti. U radu [3] demonstrirao je da je moguće izvući zajedničku informaciju iz slučajnih promenljivih dobijenih pomoću binarnog simetričnog izvora. One se nakon generisanja distribuiraju putem zašumljenih binarnih simetričnih kanala do korisnika uz pretpostavku nezavisnosti šuma u tim kanalima. Pokazao je i da nije neophodno da legitimni učesnici u komunikaciji imaju prednost u odnosu na potencijalnog napadača. Eliminisao je potrebu za unapred deljenom tajnošću i za posebnim tajnim kanalima putem kojeg bi se informacije razmenjivale, već je pretpostavio scenario u kom svi imaju pristup informacijama koje se razmenjuju putem javnog kanala. Definisao je i teorijske granice brzine generisanja tajnog ključa tako da napadač ima zanemarljivu količinu informacija o konačno usaglašenom tajnom ključu. Ovaj rad predstavlja teorijski osnov u praktičnim dizajnima kriptografskih sistema u kojima zajednički izvor slučajnosti igra važnu ulogu u generisanju tajnih ključeva.

Rad *Bennett, Brassard, Crépeau* i *Maurer* [16] predstavlja značajan teorijski napredak u oblasti informaciono-teorijske bezbednosti uvođenjem formalizacije i generalizacije procesa pojačavanja privatnosti putem javne diskusije. Osnovni cilj rada je uspostavljanje opšteg teorijskog okvira koji omogućava dvema stranama da izdvoje tajni ključ iz zajedničke slučajne promenljive o kojoj napadač ima delimične informacije, proširujući time originalne rezultate na scenarije gde strane ne poseduju potpune informacije o saznanjima napadača. Definisali su precizne matematičke granice za količinu tajnosti koja može biti izdvojena iz delimično kompromitovanih podataka. Destilovanje tajnog ključa iz zajedničke slučajne promenljive o kojoj prisluškivač ima delimične informacije omogućili su korišćenjem teorije univerzalnih heš funkcija. Rad takođe uvodi konstruktivan pristup koji omogućava praktičnu implementaciju protokola, za razliku od prethodnih pristupa koji su često bili ograničeni na asimptotske rezultate.

Bloch i *Barros* u svojoj knjizi [14] proširuju *Maurer*-ove teorijske osnove na praktične implementacije bezbednosti na fizičkom sloju. Njihov pristup se zasniva na korišćenju već prisutnih šumova na koriščenim komunikacionim kanalima kao izvora zajedničke slučajnosti. Umesto da se šumovi tretiraju kao prepreka komunikaciji koju treba eliminisati koristi se kao resurs za postizanje informaciono-teorijske bezbednosti. Posebna pažnja je posvećena postojećim stvarnim bežičnim mrežama, gde su prisutni fenomeni kao što su slabljenja signala (eng. *fading*), višestruki antenski sistemi i promena svojstava kanala u vremenu. Ovim radom pokazuju da fizički sloj može biti važan resurs u praktičnoj realizaciji bezbednosnih mehanizama u savremenim komunikacijama.

2.3 Opravdanje istraživanja klasične kriptografije u eri kvantnog računarstva

Istraživanje klasične kriptografije bazirane na informaciono-teorijskoj bezbednosti ne predstavlja zastareo pristup, već ključan element u razvoju buduće kriptografske arhitekture. Savremeni trendovi u kriptografiji ukazuju na hibridne pristupe koji kombinuju različite tipove kriptografskih algoritma kako bi se postignula maksimalna bezbednost i otpornost na napade.

Rad [17] demonstrira da hibridna kvantna razmena ključeva omogućava zadržavanje sigurnosnih garancija pre-kvantnih šema dok istovremeno postiže kvantnu otpornost. Ovaj pristup implementira konkretan sistem kombinovanja tri tipa ključeva, što jasno pokazuje da se klasični algoritmi ne napuštaju, već se integrišu sa kvantnim i post-kvantnim pristupima u jedinstvene sigurnosne arhitekture. Slično tome, rad [18] ilustruje kako se klasična kriptografija zasnovana na javnom ključu može kombinovati sa sigurnošću na fizičkom sloju. Ovaj hibridni protokol koristi ograničeno vreme računanja dostupno napadaču za probijanje kriptografskih algoritama zasnovanih na javnom ključu. Na taj način legitimne strane mogu da iskoriste fizički sloj komunikacije za uspostavljanje zajedničkih tajnih ključeva. Ovakav pristup pokazuje da iako kvantni računari predstavljaju pretnju tradicionalnim algoritmima, oblast se i dalje razvija u svim smerovima kroz inovativna hibridna rešenja.

Hibridizacija se zasniva na principu komplementarnih prednosti i nedostatka koje imaju različiti kriptografski pristupi. Klasična kriptografija donosi efikasnost i zrelost implementacije, kvantna kriptografija pruža fundamentalnu bezbednost zasnovanu na zakonima fizike, dok post-kvantna kriptografija nudi otpornost na napade kvantnim računarima. Kombinovanjem ovih pristupa se dobijaju sistemi koji su otporni na kompromitovanje bilo kog pojedinačnog algoritma.

Dodatno, informaciono-teorijska bezbednost, predstavlja fundamentalan teorijski okvir koji se koristi i za analizu kvantnih protokola. Koncepti kao što su entropija, uslovne verovatnoće i teorija informacija su jednako relevantni u kvantnoj kriptografiji, što znači da znanje o klasičnim pristupima direktno doprinosi razumevanju i razvoju kvantnih sistema. Stoga, istraživanje klasične kriptografije bazirane na informaciono-teorijskoj bezbednosti predstavlja kritičnu komponentu u širem spektru kriptografskih istraživanja. Umesto da bude u suprotnosti sa kvantnom kriptografijom, ova oblast pruža fundamentalne alate, teorijske osnove i praktična rešenja koja se integrišu sa naprednim kvantnim tehnologijama u kreiranje robusnih, višeslojnih sigurnosnih sistema koji mogu da odgovore na izazove kvantne ere.

3 Teorijski okvir istraživanja

3.1 Osnovne informacione mere

U ovom poglavlju će biti definisane osnovne informacione mere koje opisuju zavisnosti između slučajnih promenljivih, količinu neodređenosti i zajedničkih informacija među njima.
Shannon-ova entropija predstavlja osnovnu meru neodređenosti slučajnih promenljivih. Neka je X slučajna promenljiva definisana na konačnom skupu vrednosti $\mathcal{X}$ i neka je raspodela slučajne promenljive označena sa $p(x) = P\{X = x\}$. Entropija $H(X)$ definiše se kao:

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x). \tag{3.1}$$

Osnovno svojstvo entropije je da je uvek nenegativna, odnosno važi:

$$H(X) \geq 0, \tag{3.2}$$

pri čemu jednakost označava da je promenljiva X deterministička. Maksimalna vrednost entropije zavisi od kardinalnosti skupa mogućih vrednosti $|\mathcal{X}|$. Za binarne slučajne promenljive maksimalna entropija iznosi 1 bit.

Uslovna entropija za dve slučajne promenljive X i Y sa zajedničkom raspodelom $p(x, y)$ meri neodređenost promenljive X kada je poznata vrednost Y i definiše se kao

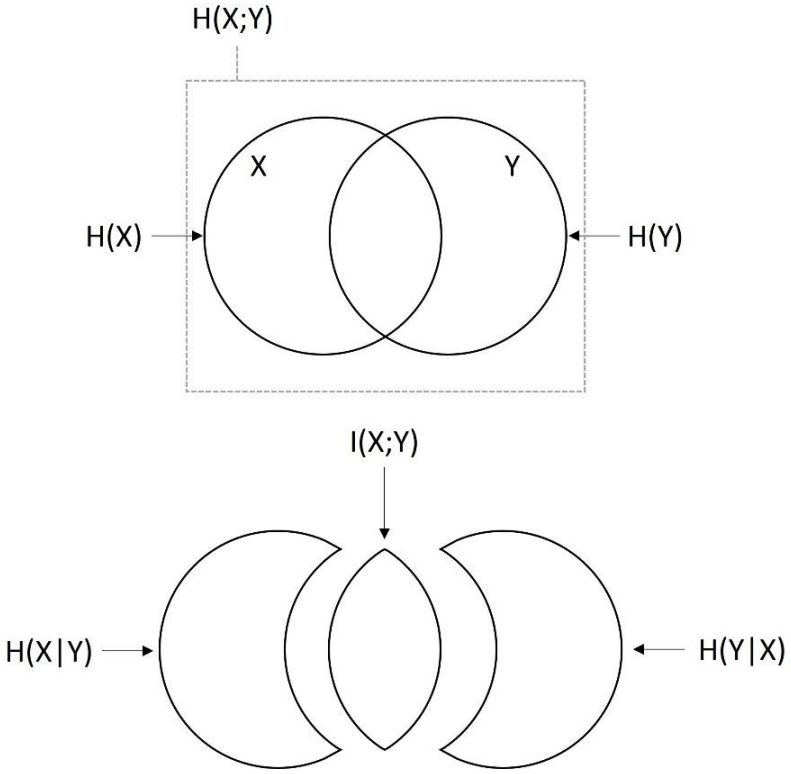
$$H(X|Y) = -\sum_{x,y} p(x,y)\log_2 p(x|y) = H(X,Y) - H(Y), \tag{3.3}$$

gde je $H(X, Y)$ združena entropija slučajnih promenljivih X i Y.
Međusobna informacija dve slučajne promenljive X i Y kvantifikuje količinu informacija koje jedna promenljiva sadrži o drugoj i definiše se kao

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X). \tag{3.4}$$

U slučaju nezavisnih promenljivih važi $I(X;Y) = 0$.

Grafička reprezentacija veza između osnovnih informacionih mera prikazana je na slici 1.



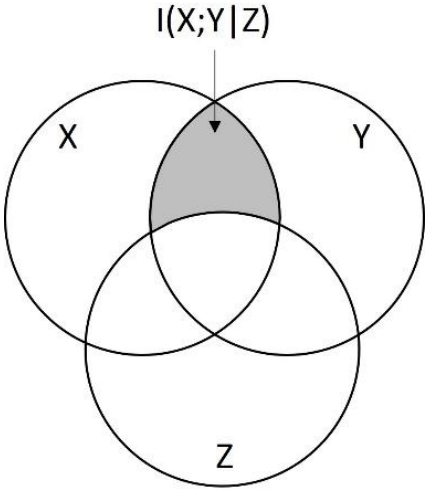
Slika 1. Veza između osnovnih informacionih mera.

Kapacitet tajnog ključa C_k predstavlja najveću stopu generisanja zajedničkog tajnog ključa koja se može ostvariti u prisustvu napadača uz očuvanje bezbednosti [3], pri čemu se stopa meri brojem bitova ključa po jednoj upotrebi izvora slučajnosti, i definiše se kao

$$C_k = min\{I(X;Y), I(X;Y|Z)\}, \tag{3.5}$$

gde $I(X;Y)$ predstavlja međusobnu informaciju između sekvenci legitimnih korisnika, Z sekvencu napadača, a $I(X;Y|Z)$ uslovnu međusobnu informaciju u prisustvu napadača, kao što je prikazano na slici 2, a koja se definiše kao

$$I(X;Y|Z) = H(X|Z) - H(X \mid Y,Z). \tag{3.6}$$



Slika 2. Uslovna međusobna informacija dve promenljive.

Maksimalni kapacitet tajnosti dostiže se u slučaju nezavisnosti sekvence napadača u odnosu na sekvence legitimnih učesnika protokola, i dat je sa

$$C_{k\,max} = I(X;Y) = I(X;Y|Z). \tag{3.7}$$

Ove veličine čine osnovu teorijskog aparata koji će se u narednim poglavljima koristi za analizu i projektovanje sistema za destilaciju kriptografskih ključeva.

3.2 Modeli izvora i kanala za destilaciju kriptografskih ključeva

Glavni izazov informaciono-teorijskog pristupa je naći kvalitetan izvor slučajnosti dostupan na udaljenim lokacijama. Oslanjajući se na ideju *Wyner*-a [9] i *Maurer*-a [3], *Ahlswede* i *Csiszár* [10] su formalizovali podelu na dva pristupa za generisanje zajedničke slučajnosti u odnosu na lokaciju izvora neodređenosti:

- model izvora, koji podrazumeva ekstrakciju slučajnosti iz izvora nezavisnih od komunikacionih kanala (eng. *source model*),
- model kanala, koji podrazumeva ekstrakciju slučajnosti iz postojećih komunikacionih kanala (eng. *channel model*).

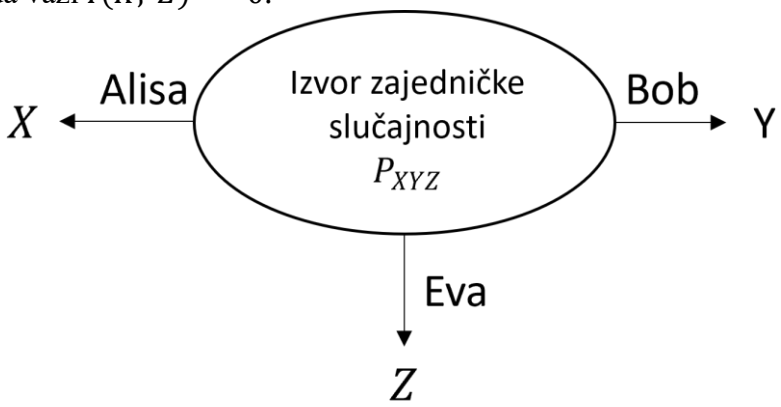
Model izvora u kriptografskim sistemima podrazumeva postojanje zajedničkog izvora za generisanje korelisanih opservacija, koje se koriste za dobijanje tajnih ključeva. Najčešće proučavani model u ovom domenu je diskretan izvor bez memorije (eng. *Discrete Memoryless Source* - DMS). Njegovo osnovno svojstvo podrazumeva nezavisnost uzastopno generisanih uzoraka iz zajedničke raspodele, što implicira da trenutni izlaz ne zavisi od prethodno generisanih vrednosti. Model kanala predstavlja pristup u kojem komunikacione strane koriste različita svojstva zajedničkog komunikacionog kanala za generisanje tajnih ključeva. Na primer, legitimni korisnici, mogu iskoristiti reciprocitet impulsnog odziva kanala i prostorno-vremenske karakteristike komunikacionog okruženja za kreiranje korelisanih signala.

Model izvora poseduje nekoliko važnih karakteristika koje su relevantne u teorijskom i praktičnom kontekstu:

- Nezavisnost izvora slučajnosti od komunikacione infrastrukture. Ovo je značajan bezbednosni benefit jer potencijalno kompromitovanje izvora slučajnosti ne utiče na bezbednost ostatka sistema.
- Energetska efikasnost. Početni signali generišu se lokalno i nije potrebno trošiti energiju za njihovu transmisiju.
- Pasivnost rada sistema. Odsustvo aktivne emisije signala čime se sprečava početno kompromitovanje i presretanje.
- Robusnost na eksterne smetnje. Otpornost na spoljnje elektromagnetno ometanje i ciljane napade na komunikacione kanale.
- Jednostavnost implementacije u višenamenskim sistemima. Zajednički izvor slučajnosti omogućava korišćenje za više korisnika bez degradacije performansi.
- Kontinuirana dostupnost. Kod modela kanala izvori slučajnosti nisu dostupni u svakoj situaciji. Na primer u situacijama kada protivnik može da kontroliše ili manipuliše komunikacionim kanalom, kada kanal fizički nije dostupan, kada postoje energetska ograničenja za transmisiju, ili kada bezbednosni zahtevi zabranjuju aktivnu komunikaciju između korisnika.
- Sinhronizacija podataka. Kod modela izvora sinhronizacija je inherentna jer signali dele ista fizička stanja dok je kod modela kanala neophodna fazna sinhronizacija.

Zbog svih pobrojanih prednosti, ovaj rad će se fokusirati na razvoj novih SKD protokola zasnovanih isključivo na modelu izvora.

U oblasti informaciono-teorijske kriptografije, koncept DMS izvora zauzima centralno mesto kada je reč o protokolima za generisanje i distribuciju kriptografskih ključeva zasnovanim na zajedničkoj slučajnosti. *Maurer*-ova fundamentalna postavka [3] razmatra situaciju u kojoj tri učesnika, dva legitimna (Alisa i Bob) i jedan pasivni napadač (Eva), primaju podatke koji predstavljaju izlaz iz istog stohastičkog procesa sa tri komponente. Ovaj proces generiše sekvence X, Y i Z prema zajedničkoj raspodeli verovatnoće P_{XYZ}, pri čemu se pretpostavlja da nijedan učesnik ne može uticati na generisanje ovih sekvenci, ali im je statistička struktura izvora u potpunosti poznata (slika 3). Legitimni učesnici nastoje da se iz svojih međusobno korelisanih opservacija X i Y usaglase oko identičnog tajnog ključa K, uz bezbednosni zahtev da napadač ne poseduje nikakvu informaciju o ključu, odnosno da važi $I(K;\ Z) = 0$.



Slika 3. Diskretni izvor bez memorije.

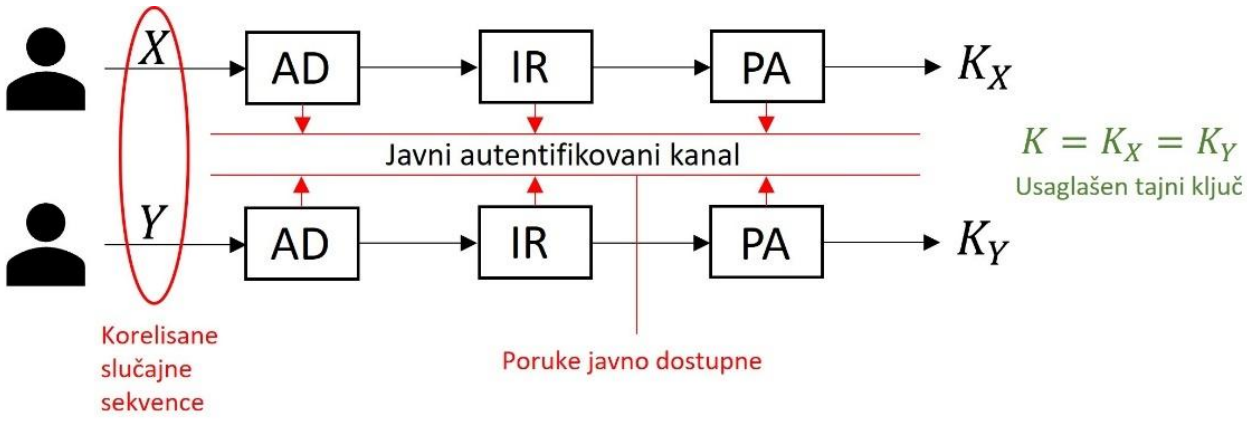
U originalnoj postavci satelitskog scenarija, korelacija nastaje tehničkim putem. Satelit emituje referentnu sekvencu koju sve tri strane primaju sa različitim nivoima degradacije usled karakteristika ekvivalentnih prenosnih kanala i antenskih prijemnika. Ovakva postavka je više teorijska nego praktično ostvariva, budući da zahteva složenu infrastrukturu. Biometrijski signali predstavljaju alternativni pristup gde zajednička slučajnost prirodno proizilazi iz merenja istog fiziološkog procesa. Digitalizovane opservacije biometrijskih signala poput elektroencefalografskih, govornih, elektrokardiografskih i drugih mogu se posmatrati kao realizacije virtuelnog DMS-a čija korelaciona struktura nastaje prirodno. U kontekstu govornih signala, ista izgovorena fraza od strane različitih subjekata generiše sekvence X, Y i Z čija korelacija proizilazi iz zajedničke lingvističke strukture, dok individualnost govornog aparata pojedinca uvodi potreban stepen neizvesnosti. Slično, kod EEG signala, merenje moždane aktivnosti tokom istog kognitivnog zadatka rezultuje korelisanim opservacijama, dok senzorski šum i individualne fiziološke varijacije obezbeđuju neophodnu neodređenost. Primenom odgovarajućih tehnika predobrade, biometrijski signali mogu se transformisati u sekvence koje zadovoljavaju svojstva slična satelitskom scenariju, čime se omogućava direktna primena postojećih protokola i bezbednosnih mehanizama za generisanje kvalitetnih kriptografskih ključeva.

3.3 Sekvencijalni protokol destilovanja ključeva (*Sequential Key Distillation* - SKD)

Sekvencijalni protokol destilovanja tajnih ključeva predstavlja najrasprostranjeniju tehniku koja legitimnim učesnicima protokola omogućava uspostavljanje zajedničkog tajnog ključa, uz uklanjanje informacija koje bi omogućile prisluškivaču njegovu parcijalnu ili potpunu restauraciju. Formalizacija ovog protokola se zasniva na *Shannon*-ovom konceptu apsolutne tajnosti uz minimizaciju međusobne informacije između sekvenci legitimnih korisnika i potencijalnog prisluškivača.

Blok dijagram predstavljen na slici 4 ilustruje arhitekturu tipičnog SKD sistema. Osnovni cilj ovog sistema je postizanje identičnosti finalnih kriptografskih ključeva koji se generišu kod legitimnih učesnika protokola, Alise i Boba, tako da važi jednakost $K_X = K_Y$. Istovremeno, sistem mora garantovati da ključ koji generiše protivnik, Eva, ne sadrži nikakve korisne informacije o legitimnim ključevima.

Sistem funkcioniše pod pretpostavkom potpune transparentnosti prema protivniku, što je u skladu sa osnovnim Kerckhoffsovim principom koji nalaže da bezbednost kriptografskog sistema ne sme zavisiti od skrivanja algoritma ili sistemske arhitekture [19]. Shodno tome pretpostavlja se da je Eva potpuno informisana o svim sistemskim elementima, uključujući parametre i funkcionalne karakteristike pojedinačnih blokova. Analiza predstavljena u radu [3] demonstrira da je optimalna strategija za Evu da ponavlja identične radnje koje legitimni učesnici izvršavaju kroz komunikaciju na javnom kanalu.



Slika 4. Arhitektura osnovnog SKD sistema.

Početni koraci u radu sistema obuhvataju predobradu raspoloživih korelisanih signala, kao što je npr. analogno-digitalna konverzija, normalizacija, kvantizacija, serijalizacija i slično. Zatim slede tri osnovne faze svakog SKD zasnovanog na SCR.

Prvi blok, destilacija prednosti (eng. *Advantage Distillation - AD*), koja ima za cilj povećanje statističkih korelacija između Alisine i Bobove sekvence u odnosu na Evinu.

Drugi blok, usaglašavanje informacija, (eng. *Information Reconciliation – IR*), kroz koji se eliminišu razlike između sekvenci legitimnih učesnika.

Treći blok, pojačavanje privatnosti, (eng. *Privacy Amplification - PA*), realizuje se implementacijom pažljivo odabrane familije univerzalnih heš funkcija, što rezultuje kompresijom usaglašenih podataka u kraći, ali sigurniji ključ.

Na kraju ciklusa rada sistema, Alisa i Bob dele identičan tajni ključ, $K_X = K_Y$. Istovremeno, sistem garantuje da Evin ključ sadrži proizvoljno malu količinu informacija o uspostavljenom tajnom ključu između Alise i Boba, čime se postiže fundamentalni cilj informaciono-teorijske bezbednosti. Ova arhitektura predstavlja praktičnu implementaciju teorijskih principa destilovanja ključeva u kontrolisanom okruženju gde protivnik poseduje delimične informacije o komunikaciji između legitimnih strana.

3.3 Komponente SKD protokola

3.3.1 Izvori zajedničke slučajnosti

Izvori zajedničke slučajnosti mogu se posmatrati kroz nekoliko grupa, pri čemu se tri posebno izdvajaju i predstavljaju različite mehanizme generisanja čisto slučajnih signala:

- kvantni izvori slučajnosti eksploatišu slučajnost i nepredvidivost kvantnih čestica i smatraju se pravim prirodnim izvorom jer su zasnovani na kvantnim fenomenima koji su po svojoj prirodi nedeterministički,
- izvori slučajnosti koriste slučajnost pogodno odabranog fizičkog procesa. Najčešće prisutan u telekomunikacionim uređajima je termalni šum koji nastaje kao posledica slučajnog kretanja elektrona u provodnicima na temperaturama iznad apsolutne nule,.
- biometrijski izvori slučajnosti kod kojih slučajnost proizilazi iz specifičnih i jedinstvenih ljudskih karakteristika. Prirodna varijabilnost bioloških procesa i individualne razlike između pojedinaca omogućavaju generisanje čisto slučajnih signala na osnovu fizičkih parametara i parametara ponašanja koji su vezani za pojedinca.

Izbor SCR mora biti takav da poseduje veliku entropiju, dovoljno veliku korelaciju između generisanih sekvenci, da je moguće izvršiti akviziciju u realnom vremenu i da poseduju robusnost na šum i druge smetnje. Biometrijski izvori slučajnosti ispunjavaju sve navedene zahteve i stoga su odabrani za eksperimentalnu evaluaciju predložene nove klase SKD sistema. Dodatno, biometrijski SCR ne zahtevaju specijalizovane uređaje poput kvantnih generatora i deo su fizičkog integriteta legitimnih učesnika protokola. U sistemima gde je potrebno postići ravnotežu između bezbednosti, dostupnosti i implementacione izvodljivosti, ovi izvori slučajnosti se nameću kao racionalan izbor.

3.3.2 Destilacija prednosti (Advantage distillaiton - AD)

U opštem slučaju, mora se pretpostaviti najnepovoljni scenario koji se karakteriše Evinom početnom prednošću u pogledu količine informacije koju poseduje o sekvenci bilo kog legitimnog učesnika u odnosu na međusobnu količinu informacija između sekvenci legitimnih učesnika. Konkretno, pretpostavlja se da je normalizovano Hamingovo rastojanje (eng. *Hamming distance*) između Evine sekvence i bilo koje sekvence legitimnih učesnika (Alise ili Boba) manje od normalizovanog Hamingovog rastojanja koje postoji između Alisine i Bobove sekvence. Faza destilacije prednosti, dizajnirana je za prevazilaženje ovog inicijalnog nedostatka. Kroz ovu fazu, Alisa i Bob uspostavljaju komunikaciju putem javnog autentifikovanog kanala i razmenjuju poruke čiji je cilj uspostavljanje veće međusobne informacije između legitimnih učesnika protokola da bi se konačnom rezultatu promenila prednost u njihovu korist. Protokol parnosti bitova (eng. *Bit Parity Advantage Distillation* - BP AD) predložen u [20], pokazao je da je uz adekvatnu strategiju moguće generisati tajne ključeve čak i kada je kanal napadača pouzdaniji, odnosno kada je korelacija između legitimnih sekvenci učesnika protokola manja od korelacije prema sekvenci napadača. U [21] je demonstrirano da iterativni protokol postiže optimalnu efikasnost kada se koriste blokovi od dva bita za proveru parnosti, čime se postiže maksimalna brzina generisanja ključa uz eliminaciju početne prednosti napadača. Detaljni koraci ovog algoritma opisani su u Algoritmu 1. U okviru Algoritma 1, n_{AD0} predstavlja početnu dužinu binarnih sekvenci pre početka AD protokola, dok X_k označava k-ti bit sekvenci učesnika protokola.

Algoritam 1. *Bit Parity AD* protokol

1: Alisa i Bob grupišu n_{AD0} bite u blokove od 2 bita.
2: Alisa i Bob računaju bite parnosti ovih blokova, $\left\{X_{2i+1} \oplus X_{2i+2} \mid i=0,1,...,\lfloor \frac{n_{AD0}}{2} \rfloor -1\right\}$.
3: Alisa šalje $\lfloor \frac{n_{AD0}}{2} \rfloor$ bita parnosti Bobu preko javnog autentifikovanog kanala. Ukoliko se parnosti bloka poklapaju, Bob šalje potvrdnu poruku na javni kanal.
4: Blokovi čija se parnost ne poklapa se odbacuju, dok kod blokova čija je parnost ista i Alisa i Bob zadržavaju prvi bit bloka formirajući tako kraću bit sekvencu koja služi kao ulaz u sledeću rundu.

U okviru faze destilacije prednosti dolazi do značajne redukcije informacija zarad očuvanja bezbednosti. Svaki poslati bit o parnosti bloka rezultuje odbacivanju jednog bita iz sekvenci legitimnih učesnika, pri čemu optimalna veličina bloka od dva bita predstavlja kompromis između efikasnosti destilacije i minimizacije informacionih gubitaka [21]. Ova faza je neophodna jer algoritmi za usaglašavanje informacija u sledećoj fazi postižu optimalne performanse samo kada je brzina bitskih grešaka (eng. *Bit Error Rate - BER*) u sekvencama ispod određenog praga, što direktno utiče na efikasnost celokupnog protokola i konačnu dužinu generisanog ključa.

Efikasnost generisanja ključeva i praktičnost primene opisanog BP AD algoritma u realnim komunikacionim sistemima direktno zavisi od njene računske složenosti. Smatrajući da je n dužina ulaznih signala, BP AD algoritam postiže linearnu računsku složenost, $O(n)$, pri čemu iterativni karakter algoritma ne narušava složenost koja ostaje linearna. Ovo svojstvo algoritam čini praktičnim čak i za obradu velikih sekvenci podataka, što je od krucijalnog značaja za SKD implementacije u modernim komunikacionim sistemima.

3.3.3 Usklađivanje informacija (Information Reconciliation - IR)

Osnovni cilj IR faze svakog SKD protokola je usklađivanje informacionog sadržaja sekvenci legitimnih korisnika uz minimizaciju informacija koje tom prilikom otiču prisluškivaču. IR stoga predstavlja značajnu fazu u SKD sistemima zasnovanih na SCR i ima za cilj da ukloni, odnosno ispravi, sve razlike koje do te faze još uvek postoje u sekvencama legitimnih korisnika. IR se odvija u formi diskusije preko javnog autentifikovanog kanala, što znači da napadač ima pristup svim razmenjenim porukama, ali ih ne može modifikovati jer je autentifikacijom onemogućeno lažno predstavljanje i neovlašćeno menjanje sadržaja. Zbog toga se efikasnost algoritama za usklađivanje ne meri se samo njihovom sposobnošću da isprave greške, već i minimalnom količinom informacija koju moraju razmeniti da bi to postigli.

Osnovni koncepti za rešavanje ovog problema mogu se pronaći u radu [22] u kome je predstavljen BBBSS protokol, jedan od najranijih protokola IR, čija se osnovna ideja zasniva na podeli sekvence na manje segmente, blokove, nakon čega se pronalaze i otklanjaju neusaglašeni bitovi korišćenjem binarne pretrage. Koncepti koje je ovaj protokol uveo poslužili su kao osnova za razvoj brojnih algoritama, uključujući kaskadni algoritam (eng. *cascade*) [23]. On je prevazišao ograničenja BBBSS-a kroz implementaciju adaptivnih strategija za upravljanje blokovima i optimizovanu komunikacionu efikasnost. Na ovaj način je proširen opseg početnih grešaka, BER, za koje je faza usklađivanja informacija efikasno izvršena.

Kaskadni algoritam jedan je od najčešće korišćenih algoritma zahvaljujući svojoj jednostavnosti, robusnosti, praktičnosti implementacije i komunikacionoj efikasnosti. Naziv algoritma potiče od njegovog svojstva da korekcijom jedne greške u bloku mogu biti otkrivene greške u prethodno proverenim blokovima. Ovaj efekat garantuje potpunu korekciju svih grešaka, što je veoma značajno za primenu u kriptografskim sistemima. Osnovna varijanta protokola podrazumeva rast veličine blokova kroz uzastopne iteracije što utiče na krajnju efikasnost sistema. Postoje različite optimizovane varijante koje zahtevaju procenu početnog BER-a između sekvenci i na taj način prilagođavaju veličinu blokova. Međutim, u praktičnim sistemima za generisanje i destilaciju kriptografskih ključeva ovaj parametar često nije dostupan unapred. Zbog toga osnovna varijanta protokola ostaje najčešći izbor.

Teorijska analiza protokola za usklađivanje informacija u radu [23] pokazuje da optimalan protokol za usklađivanje informacija ima efikasnost jednaku jedinici. Efikasnost protokola za usklađivanje informacija definiše se kao odnos stvarne količine informacije koja se razmeni tokom procesa korekcije grešaka i teorijski minimalne količine informacije određene entropijom izvora (*Shannon*-ovom granicom). Efikasnost jednaka 1 označava da protokol dostiže ovu donju granicu, odnosno da ne postoji dodatno curenje informacije u odnosu na optimalni slučaj, već se razmena informacija poklapa sa minimalno potrebnom količinom za uspešnu rekonstrukciju sekvence [14]. U radu se ističe da se kaskadni protokol jednostavno može implementirati, pri čemu količina informacija

koja se otkriva kroz komunikaciju ostaje bliska teorijskoj donjoj granici za slučaj prenosa informacija preko binarnog simetričnog kanala, ali pod uslovom da greške između sekvenci čije se informacije usklađuju ne prelaze 15%. U literaturi se ističe [91] da kaskadni protokol ostvaruje malu količinu otkrivenih informacija tokom procesa usaglašavanja sekvenci, dok protokoli zasnovani na kodovima za korekciju grešaka, poput *low-density parity-check* (LDPC) pristupa, zahtevaju veoma duge kodove i pažljivo prilagođavanje parametara kako bi dostigli sličnu efikasnost, zbog čega se kaskadni protokol često razmatra kao pogodniji u praktičnim scenarijima sa promenljivom stopom greške.

LDPC kodovi za korekciju grešaka su za razliku od kaskadnog protokola probabilističke prirode i ne garantuju potpunu korekciju grešaka, što zahteva uvođenje dodatnih mehanizama provere potpunog poklapanja. Pored toga, LDPC kodovi su računski složeniji i kompromituju više informacionih bitova u odnosu na kaskadni protokol, što rezultuje kraćim finalnim ključevima..

Pokušaj da se iskoriste prednosti kaskadnog algoritma rezultovali su razvijanjem *Winnowog* algoritma koji zadržava deterministička prirodu i garantovanu korekciju grešaka. Ovaj algoritam baziran je na podeli sekvence na blokove i računanju sindroma korišćenjem Hamingovih matrica. Njegova glavna prednost u odnosu na kaskadni algoritam je brzina, jer ne zahteva binarnu pretragu, jednostavan je za implementaciju, ali se mora biti obazriv pri izboru dužine blokova jer ovaj algoritam omogućava ispravljanje samo jedne grešku po bloku. Komunikaciona složenost *Winnowog* algoritma zavisi od dužine bloka, broja grešaka i njihove rasprostranjenosti kroz sekvencu kao i broja potrebnih iteracija.

Kriptografski sistemi zahtevaju identične tajne ključeve. Greška u jednom bitu bi onemogućila ispravno funkcionisanje šifrovanja i dešifrovanja. Ovaj strogi zahtev za perfektnom sinhronizacijom ključeva predstavlja osnovu za potrebu determinističkih algoritama u fazama usklađivanja informacija. Za razliku od drugih komunikacionih scenarija gde se greške mogu tolerisati ili naknadno ispraviti, generisanje kriptografskih ključeva ne dozvoljava kompromise u pogledu tačnosti, što čini potpunu korekciju grešaka obaveznim svojstvom IR faze protokola.

Izbor IR algoritma zavisi od specifičnih zahteva primene, pri čemu se mora balansirati između komunikacione efikasnosti, implementacijske složenosti i praktičnosti. Kaskadni algoritam je teorijski optimalan u pogledu komunikacione efikasnosti, dok praktični zahtevi favorizuju alternativne pristupe poput *Winnow* algoritma. Komparativna analiza ova dva algoritma aktuelna je u oblasti razvoja SKD sistema [24]. Zbog kompletnosti eksperimentalne evaluacije SKD sistema predloženih u ovoj disertaciji, u IR fazi su korišćeni i kaskadni i *Winnow* algoritam. Kao što je napomenuto, osnovna varijanta Kaskadnog protokola najčešći je izbor i biće implementirana u ovom radu. Koraci algoritma predstavljeni su Algoritmom 2, dok je Algoritmom 3 opisan pomoćni Binarni algoritam [92].

Algoritam 2. Kaskadni protokol

Ulaz: A, B - binarne sekvence Alise i Boba
Izlaz: K - usaglašeni ključ

1: Alisa i Bob dele svoje sekvence na blokove unapred dogovorene dužine, a zatim Alisa računa parnosti svojih blokova i šalje ih Bobu.
2: Bob računa svoje parnosti i nastavlja sa binarnim algoritmom (Algoritam 3).
3: Na početku svake sledeće iteracije Bob promeša bite svoje sekvence i ponavlja korake 1 i 2 ali povećavajući veličinu bloka, veličina novog bloka = 2 * veličina starog bloka.
4: Ispravljeni biti će uzrokovati kaskadni efekat na promešanim blokovima iz prethodnih iteracija, nakon čega se vrši povratak na te blokove i primena binarnog algoritma.
5: Ponavljati korake 3 i 4 sve dok se ne dostigne unapred zadat broj iteracija.

Algoritam 3. Binarni algoritam

U slučaju kada su blokovi sekvenci A i B različite parnosti:

1: Alisa deli blok A_i na dva jednaka dela i šalje Bobu parnost prve polovine A_i.
2: Bob deli B_i na isti način i poredi parnost sa Alisom da bi odredio u kojoj se polovini nalazi neparan broj grešaka.
3: Koraci 1 i 2 se ponavljaju sve dok se greška ne pronađe.

Detaljan opis *Winnow*-og protokola kao i pomoćnog Hamingovog algoritma dati su Algoritmom 4 i Algoritmom 5 [59].

Algoritam 4. *Winnow* protokol

1: Alisa i Bob permutuju svoje binarne sekvence na unapred dogovoren način a zatim ih dele na blokove dužine $N = 2^r, r \geq 3$.
2: Alisa i Bob računaju parnost svakog N-bitnog bloka.
3: Alisa i Bob porede parnosti svojih blokova.
Ako su parnosti bloka usaglašene operacija ispravljanja grešaka se ne primenjuje ali se prvi bit bloka odbacuje zarad očuvanja tajnosti. U suprotnom, razlika u parnosti blokova označava da postoji neparan broj grešaka u N-bitnom bloku i primenjuje se operacija ispravljanja grešaka korišćenjem Hamingovog algoritma (Algoritam 5). I u ovom slučaju se prvi bit bloka odbacuje pa se ispravljanje grešaka obavlja na preostalih $N - 1$ bita.
4: Koraci 1-3 se izvršavaju iterativno sve dok broj iteracija ne dostigne unapred zadatu granicu ili dok se ne isprave sve greške.

Algoritam 5. Hamingov algoritam ispravljanja grešaka

1: Neka je I_a Alisin blok od $(N - 1)$ bit, a I_b Bobov blok od $(N - 1)$ bit. Alisa i Bob računaju sindrome dužine r bita, S_a i S_b, $S_a = I_a H^T$, a $S_b = I_b H^T$ na svojim $(N - 1)$-bitnim blokovima, respektivno, gde je H matrica koja predstavlja kontrolnu matricu Hamingovog koda.
2: Alisa prenosi S_a ka Bobu, a greške se otkrivaju u slučaju da razlika između sindroma, S_d, nije jednaka nuli, gde je $S_d = S_a \oplus S_b$. Bob zatim ispravlja otkrivene greške.
3: Na kraju, r bita biva odbačeno iz svakog bloka da bi se sprečilo potencijalno oticanje informacija ka Evi zbog Alisinog javnog slanja sindroma

Pomenuta H matrica rezultat je jednih od najranijih radova za detekciju i korekciju pojedinačnih grešaka u binarnim sekvencama [25]. Ograničena je mogućnošću da ispravi najviše jednu grešku u bloku zbog čega je od važnosti veličina bloka koja se posmatra kao i rasprostranjenost i broj grešaka binarne sekvence. Veličina matrice određena je veličinom bloka pa je tako za blok dužine 2^r broj redova jednak r, dok je broj kolona $2^r - 1$. Matrica omogućava jednoznačno mapiranje pozicije bita na osnovu koje se detektuje pozicija greške. Izgled matrice za blokove dužine 8 je:

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{3.8}$$

Složenost faze usaglašavanja informacija razlikuje se u zavisnosti od izabranog algoritma. Ukoliko se smatra da je ulazna dužina sekvenci u IR fazu n, može se reći da kaskadni algoritam

karakteriše složenost $O(n\log n)$, zbog svoje hijerarhijske strukture i kaskadnog svojstva. Iako efikasan u pogledu eliminacije pogrešnih bita i korišćenju javnog kanala za komunikaciju, njegova složenost može predstavljati ograničavajući faktor pri obradi dugih sekvenci, posebno u aplikacijama koje zahtevaju rad u realnom vremenu.

Za razliku od kaskadnog algoritma, *Winnow* algoritam postiže značajno bolju, linearnu složenost $O(n)$. Podelom na blokove, korišćenjem Hamingove matrica za dobijanje sindroma koji identifikuju i koriguju greške omogućava se efikasnija obrada podataka. Linearna složenost čini *Winnow* algoritam posebno atraktivnim za primenu u sistemima koji zahtevaju obradu dugih sekvenci podataka i za rad u realnom vremenu.

Dalje istraživanje će pokazati kako se teorijski principi opisanih IR algoritama prenose u praktičnu implementaciju i koje su stvarne performanse različitih pristupa u kontrolisanim uslovima.

3.3.4 Pojačanje privatnosti (Privacy amplification - PA)

Pojačanje privatnosti je koncept uveden u radu [26] i predstavlja poslednju i ključnu fazu u SKD protokolima. Njegova osnovna uloga je eliminacija svih informacija koje je napadač stekao tokom izvršavanja AD i IR faze protokola. Dobijena usaglašena binarna sekvenca koju poseduju legitimni učesnici nakon IR faze, se u okviru PA faze transformiše u kraći niz u kojem je preostala informacija napadača o finalnom destilovanom ključu proizvoljno mala, ili tačnije, manja od unapred zadatog nivoa.

Teoretski, PA se može implementirati kroz dva glavna pristupa: primenom odgovarajućih heš funkcija iz klase univerzalnih heš funkcija ili korišćenjem klase funkcija poznate kao ekstraktori [27]. U teorijskom smislu, oba pristupa su ekvivalentna, međutim, nedavne studije pokazuju da upotreba ekstraktora postaje superiorna u odnosu na heš funkcije tek kod izuzetno velikih dužina ključeva koji su reda većeg od 10^5 bitova [28]. Imajući na umu tipičnu praksu u implementaciji SKD sistema, u radu se dalje razmatra PA zasnovana na heš funkcijama iz univerzalne klase heš funkcija.

Definicija 1 [29]. Za data dva konačna skupa A i B, familija $\mathcal{G}$ funkcija g: $\mathcal{A} \to \mathcal{B}$ je 2-univerzalna (skraćeno univerzalna) ako je

$$\forall\, x_1, x_2 \in \mathcal{A}\ \ x_1 \neq x_2 \Longrightarrow P_G[G(x_1) = G(x_2)] \leq \frac{1}{|\mathcal{B}|}\ , \tag{3.9}$$

gde je G slučajna promenljiva koja predstavlja slučajan i uniforman izbor funkcije g$\in \mathcal{G}$.

U radu [30] a zatim i preciznije u radu [31] dat je odgovor, upotrebom *Leftover Hash Lemma* (LHL), na pitanje da li se iz sekvence dužine n o kojoj napadač poseduje informaciju o t bitova može izdvojiti bezbedan tajni ključ. Naime, LHL tvrdi da se odgovarajućom transformacijom, korišćenjem univerzalne klase heš funkcija, može izvući ključ dužine $k = n - t$ bitova o kojima napadač ima zanemarljivo malu količinu informacija. Stoga, LHL predstavlja teorijski osnov za korišćenje univerzalne klase heš funkcija kao PA mehanizma.

U kontekstu praktične implementacije tranformacija se može odraditi bilo kojom funkcijom $g: \{0,1\}^n \to \{0,1\}^k$ iz univerzalne klase heš funkcija. Tipična realizacija takve funkcije zasnovana je na linearnim tranformacijama u polju Galoa sa dva elementa, $GF(2)$ i definisana je kao

$$G = \{g_{M_T}: M_T \in GF(2)^{kxn}\}, \tag{3.10}$$

$$g_{M_T}(x) = M_T \cdot x, \tag{3.11}$$

gde je M_T binarna matrica dimenzija kxn a sve operacije definisane su na $GF(2)$.

Elementi matrice M_T slučajno se biraju slučajno. Ako su k i n veliki, generisanje slučajnih binarnih matrica je teško ostvariti na efikasan način jer je računska složenost u tom slučaju kvadratna,

$O(n^2)$. Problem se može prevazići korišćenjem *Toeplitz* matrica zbog njenih povoljnih karakteristika složenosti. *Toeplitz* matrice omogućavaju složenost $O(n\log n)$ zbog svoje specifične strukture koje doprinose bržem matričnom množenju. Matrice su određene prvom vrstom i prvom kolonom i zahtevaju svega $k+n-1$ bitova za opis, zadržavajući 2-univerzalnost [32]. Izbor *Toeplitz* matrica predstavlja optimalan kompromis između računske efikasnosti i bezbednosnih zahteva, omogućavajući praktičnu implementaciju čak i za dugačke ključeve.
Toeplitz matrica dimenzija kxn definiše se

$$\mathrm{M}_T = \begin{bmatrix} m_0 & m_{-1} & m_{-2} & \dots & m_{-(n-2)} & m_{-(n-1)} \\ m_1 & m_0 & m_{-1} & \dots & m_{-(n-3)} & m_{-(n-2)} \\ m_2 & m_1 & m_0 & \dots & m_{-(n-4)} & m_{-(n-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ m_{k-1} & m_{k-2} & m_{k-3} & \dots & m_{k-n-1} & m_{k-n} \end{bmatrix} \tag{3.12}$$

pri čemu važi $m_i \in \{0,1\}$. Kao i čisto slučajne matrice i *Toeplitz* matrice obezbeđuju istu bezbednost destilovanih ključeva, uz znatno manju kompleksnost generisanja.

U tipičnoj realizaciji PA faze, matrica transformacije M_T se generiše slučajno od strane jednog legitimnog učesnika komunikacije (na primer Alise) ili se deterministički izvodi iz prethodno dogovorenog zajedničkog inicijalnog parametra (eng. *shared random seed*). Opis matrice zatim se javno objavljuje Bobu preko javnog autentifikovanog kanala. Bezbednost postupka ne oslanja se na tajnost samog opisa matrice, već na svojstvo univerzalnosti odgovarajuće heš familije i na preostalu entropiju izvora posle IR faze. Iako je M poznata Evi i primenjuje istu transformaciju na sopstvenu, delimično korelisanu sekvencu, prema LHL, ona ne dolazi do značajnije informacije o konačnom destilovanom ključu. Ipak, neophodno je da se matrica ne koristi u više nezavisnih sesija jer bi se time narušila pretpostavka o stohastičkoj prirodi ove transformacije, što bi dovelo do odstupanja od teorijskog modela PA i garantovanja unapred zadatog nivoa curenja informacija o destilovanom ključu ka Evi.

4 Pojačanje privatnosti zasnovano na estimaciji ECRE2

4.1 Klasične PA metode i njihovi nedostaci

Klasične metode pojačanja privatnosti najčešće se oslanjaju na matematičke funkcije koje transformišu izvorni niz u kraći uz maksimizaciju entropije, pri čemu protivnikova informacija postaje zanemarljiva.

Za primenu PA Alisa i Bob moraju imati neku procenu količine informacije koju je protivnik (Eva) mogao steći o njihovom izvornom nizu. Ukoliko se protivnikova informacija potceni, transformisani niz može biti nedovoljno smanjen i Eva će i dalje imati nezanemarljivu informaciju o konačnom ključu, čime se narušava primarna namena SKD. S druge strane, preterano konzervativno skraćivanje ključa značajno utiče na smanjenje efikasnosti i brzine generisanja destilovanih ključeva. Dakle, pogrešna pretpostavka o Evinoj količini informacija o nizu ključa koji ulazi u PA blok može kompromitovati ceo protokol. U ovoj fazi protokola nije moguće dodati novu količinu neodređenosti, već samo koncentrisati već postojeću entropiju usaglašenih sekvenci. Ukoliko Eva raspolaže značajnom količinom informacija o ključu, to će rezultovati veoma kratkim ključem. U krajnjem slučaju, može se desiti da do formiranja konačnog ključa uopšte i ne dođe zbog manjka neodređenosti u odnosu na sekvencu napadača. U tim slučajevima protokol se prekida čime se sistem osigurava od potencijalne kompromitacije. Stoga povećanje bezbednosti vodi ka kraćim ključevima i manjoj brzini destilacije.

Iako klasične metode pojačanja privatnosti obezbeđuju dokazivu informaciono-teorijsku bezbednost, njihova praktična primena ograničena je zahtevima preciznog modelovanja izvora zajedničke slučajnosti i procene količine informacija koje su procurele da napadača. U realnim uslovima, kada se kao izvori zajedničke slučajnosti koriste biometrijski signali, procena količine informacija dostupnih napadaču predstavlja izuzetno složen zadatak. Cilj ovog rada je analiza različitih strategija pojačanja privatnosti i razvoj adaptivnog modela koji omogućava dinamičko određivanje optimalnog broja bitova koje je potrebno odbaciti u PA fazi, na osnovu procene uslovne entropije (ekvivalentno, međusobne informacije) između sekvenci legitimnih učesnika i napadača. Takav pristup omogućava balans između bezbednost i efikasnosti, jer se dužina skraćenja ne određuje unapred fiksno, već se prilagođava stepenu korelacije i procenjenom curenju informacija u konkretnom komunikacionom scenariju. Time se ostvaruje znatno efikasniji i robusniji okvir za projektovanje SKD sistema koji može pouzdano raditi i u uslovima promenljivih verovatnosnih svojstava SCR.

4.2 Uloga *Renyi*-jeve uslovne entropije drugog reda u proceni količine kompromitovanih informacija dostupnih u javnom kanalu

Za kvantitativnu procenu nivoa bezbednosti u fazi pojačanja privatnosti entropija kolizije koristi se kao odgovarajuća informaciona mera. Ona predstavlja entropijsku meru koncentracije raspodele verovatnoće koja posebno naglašava prisustvo dominantnih ishoda. U kriptografskom kontekstu, može se interpretirati kao mera verovatnoće kolizije između dva nezavisno izabrana uzorka iz iste raspodele, što odgovara verovatnoći uspešnog pogotka napadača koji vrši nasumično pogađanje. Veće vrednosti entropije kolizije odgovaraju manjoj verovatnoći ovakvog poklapanja.

Ova mera pruža konzervativniju i informacijski bogatiju procenu neizvesnosti u odnosu na klasičnu *Shannon*-ovu entropiju. Zbog toga se koristi kao osnovni parametar u analitičkim modelima za procenu broja kompromitovanih bitova i maksimalne dužine tajnog ključa koji se može bezbedno destilovati iz delimično kompromitovane sekvence. U radu [16] pokazano je da ova entropijska mera daje donju granicu neizvesnosti (iz ugla Eve) i samim tim gornju granicu dužine bezbednog ključa, što čini osnovu kvantitativne analize PA.

Definicije i svojstva ključnih informacionih mera za PA fazu navedeni su u nastavku.

Definicija 2. Entropija kolizije diskretne slučajne promenljive $X \in X$ je

$$H_c(X) \triangleq -\log E[p_X(x)] = -\log P_c(x) = -\log(\sum_{x\in X} p_X(x)^2), \tag{4.1}$$

gde je

$$P_c(x) = \sum_{x\in X} p_X(x)^2 \tag{4.2}$$

verovatnoća kolizije.

Za dve diskretne slučajne varijable, $X \in \mathcal{X}$ i $Y \in \mathcal{Y}$, uslovna entropija kolizije od X ukoliko je poznato Y je

$$H_c(X|Y) \triangleq \sum_{y\in \mathcal{Y}} p_Y(y) H_c(X|Y = y). \tag{4.3}$$

Za bilo koje diskretne slučajne promenljive $X \in \mathcal{X}$, entropije kolizije zadovoljava $H(X) \geq H_c(X) \geq 0$. Ako je X uniformno raspodeljenja u $\mathcal{X}$, tada je $H(X) = H_c(X) = log\,|\mathcal{X}|$, gde je $H(X)$ *Shannon*-ova entropija.

Naziv entropija kolizije potiče od činjenice da je to funkcija verovatnoće kolizije (4.1) dobijanja iste realizacije slučajne promenljive dva puta u dva nezavisna eksperimenta. Za diskretnu slučajnu promenljivu X, *Renyi*-jeva entropija reda α je

$$R_\alpha(X) = \frac{1}{1-\alpha} log(\sum_{x\in X} p_X(x)^\alpha). \tag{4.4}$$

Stoga, entropija kolizije je identična *Renyi*-jevoj entropiji reda 2, odnosno, $H_c(X) = R_2(X)$.

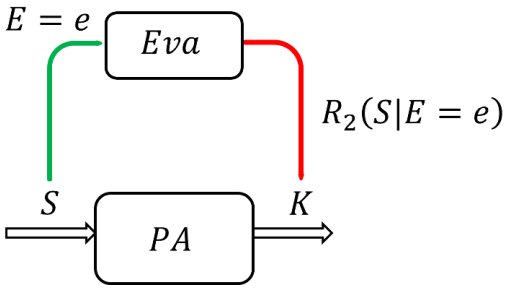
Nadovezujući se na rad [16], *Cachin* i *Maurer* u radu [33] uspostavljaju eksplicitnu vezu između IR i PA faze, pokazujući da količina informacija koja se razmeni tokom ispravljanja grešaka direktno utiče na smanjenje *Renyi*-jeve entropije izvora, a samim tim i na broj bitova koji se mogu bezbedno destilovati iz usaglašene sekvence. Time su obezbedili kvantitativan okvir koji povezuje informacije prenesene javnim kanalom sa preostalom dostupnom entropijom čime je omogućena preciznija analiza ukupne bezbednosti SKD sistema.

Veza između *Renyi*-jeve entropije i PA zasnovane na univerzalnoj familiji heš funkcija formulisana je u sledećoj teoremi [16]:

Teorema 1. [16] Neka je $S \in \{0,1\}^n$ slučajna promenljiva koja predstavlja zajedničku sekvencu Alise i Boba, a neka je E slučajna promenljiva koja predstavlja ukupno znanje o S dostupno Evi. Neka je e posebna relizacija od E. Ako Alisa i Bob znaju da je ECRE2, $R_2(S|E = e)$ najmanje neka konstanta c, i ako biraju $K = G(S)$ kao svoj tajni ključ, gde je G heš funkcija izabrana uniformno slučajno iz univerzalne familije heš funkcija $\mathcal{G}$: $\{0,1\}^n \longrightarrow \{0,1\}^k$, tada

$$H(K|G, E = e) \geq k - \frac{2^{k-c}}{ln\,2}. \tag{4.5}$$

Na slici je prikazan konceptualni tok faze pojačanja privatnosti. Znanja napadača o izvornoj sekvenci S izražena su uslovnom *Renyi*-jevom entropijom drugog reda pomoću koje se određuje koliko bitova treba odbaciti u PA bloku tako da napad sadrži zanemarljivo malo informacija o izlaznom ključu K.



Slika 5. Faza pojačanja privatnosti u prisustvu napadača. Adaptirano iz rada [34], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Napomena 1. Teorema 1 tvrdi da Alisa i Bob mogu da generišu deljeni tajni ključ dužine $k < c$, ako znaju donju granicu c za ECRE2. Kombinovanjem (4.5) i činjenice da binarna sekvenca dužine k ne može imati *Shannon*-ovu entropiju veću od k, dobijamo

$$k \geq H(K|G, E = e) \geq k - \frac{2^{k-c}}{ln\,2}. \tag{4.6}$$

Ovo dalje znači da ako Alisa i Bob biraju dužinu deljenog tajnog ključa

$$k_{GLB_c}(e) = c - s, \tag{4.7}$$

gde je s sigurnosni parametar $s > 0$, koji predstavlja marginu sigurnosti, odnosno razliku između dostupne količine slučajnosti u sistemu i količine informacije koja je iskorišćena za formiranje finalnog ključa. Kako je s izbor dizajnera sistema, generisani ključevi će se razlikovati eksponencijalno malo po s od sekvenci maksimalne entropije, dok će Evina ukupna informacija o tom tajnom ključu biti eksponencijalno mala po s. U nastavku rada, ova PA strategija biće označena kao strategija globalne donje granice.

Strategija globalne donje granice dominira savremenom praksom primene PA u SKD sistemima zasnovanim na modelu izvora. Pošto konstanta c ne zavisi od Evine konkretne sekvence $E = e$, jasno je da za svaku pojedinačnu sekvencu e postoji manje ili veće odstupanje od uspostavljene fiksne donje granice c, što vodi ka nepotrebnom gubitku u dužini generisanih ključeva, za iste radne uslove SKD protokola i isti sigurnosni parametar $s > 0$.

Sledeća teorema obezbeđuje osnovu za strategiju zasnovanu na lokalnoj donjoj granici za ECRE2.

Teorema 2. [34] Neka je $S \in \{0,1\}^n$ slučajna promenljiva koja predstavlja zajedničku sekvencu Alise i Boba, a neka je E slučajna promenljiva koja predstavlja ukupno znanje o S dostupno Evi. Neka je e posebna realizacija od E. Ako Alisa i Bob biraju $K = G(S)$ kao svoj tajni ključ, gde je G heš funkcija izabrana uniformno slučajno iz univerzalne familije heš funkcija $\mathcal{G}$: $\{0,1\}^n \longrightarrow \{0,1\}^k$, tada važi

$$k \geq H(K|G, E = e) \geq k - \frac{2^{k - R_2(S|\,E = e)}}{\ln 2}. \tag{4.8}$$

Dokaz. Dokaz u potpunosti sledi korake dokaza Teoreme 3 iz [16], ako su sve relevantne verovatnosne mere proširene dodatnim uslovom $E = e$.

Obzirom da važi $H(K|G,E=e) \geq R_2(K|G,E=e)$, dovoljno je utvrditi donju granicu za $R_2(K|G,E=e)$, tj. ECRE2. Važi da

$$\begin{aligned} R_2(K|G,E=e) &= \sum_{g\in\mathcal{G}} p_G(g) R_2(K|G=g,E=e) \\ &= \sum_{g\in G} p_G(g)\left(-\log E_{K|G=g,E=e}\left[p_{K|GE}(K|g,e)\right]\right) \\ &\geq -\log\left(\sum_{g\in G} p_G(g) E_{K|G=g,E=e}\left[p_{K|GE}(K|g,e)\right]\right) \end{aligned} \tag{4.9}$$

Gde poslednja nejednakost sledi iz konveksnosti funkcije $x \mapsto -\log x$ i Jensenove nejednakosti. Sada, neka su $S_1 \in \{0,1\}^n$ i $S_2 \in \{0,1\}^n$ dve slučajne promenljive koje su međusobno nezavisne i nezavisne od G, pri čemu su raspodeljene po $p_{S|E=e}$. Tada važi,

$$\begin{aligned} P[G(S_1)=G(S_2)|G=g] &= \sum_{kk\in\{0,1\}^k} p_{G(S)|GE}(kk|g,e)\, p_{G(S)|GE}(kk|g,e) \\ &= E_{K|G=g,E=e}\left[p_{K|GE}(K|g,e)\right], \end{aligned}$$

gde je $k \in K$, a jednačina (4.9) može se napisati u obliku

$$R_2(K|G,E=e) \geq -\log P\,[G(S_1)=G(S_2)]. \tag{4.10}$$

Dalje se razmatra formiranje gornje granice za $P[G(S_1)=G(S_2)]$. Po zakonu totalne verovatnoće važi,

$$\begin{aligned} &P[G(S_1)=G(S_2)] = \\ &= P[G(S_1)=G(S_2), S_1=S_2]P[S_1=S_2] + P[G(S_1)=G(S_2), S_1\neq S_2]P[S_1\neq S_2]. \end{aligned} \tag{4.11}$$

Pri čemu je $P[G(S_1)=G(S_2)|E=e, S_1=S_2] \leq 1$ i $P[S_1\neq S_2|E=e] \leq 1$. Pored toga, na osnovu definicije entropije kolizije važi

$$P[S_1=S_2] = \textstyle\sum_{s\in\{0,1\}^n} p_{S|E=e}(s|e)^2 = 2^{-R_2(S|E=e)}.$$

Konačno, imajući u vidu da je heš funkcija $\mathcal{G}$ izabrana iz familije univerzalnih funkcija, važi da je

$$P[G(S_1)=G(S_2)|S_1\neq S_2] \leq 2^{-k}.$$

Zamenom ovih nejednakosti u (4.11), dobija se

$$P[G(S_1)=G(S_2)] \leq 2^{-R_2(S|E=e)} + 2^{-k} = 2^{-k}\left(1+2^{k-R_2(S|E=e)}\right). \tag{4.12}$$

Zamenom jednačine (4.12) u (4.10) i uzimajući u obzir činjenicu da je $\ln(1+x) \leq x$ for all $x > -1$, dobija se

$$R_2(K|G,E=e) \geq k - \frac{2^{k-R_2(S|E=e)}}{\ln 2}.$$

Konačno, imajući u vidu gornju granicu *Shannon*-ove entropije, dobija se

$$k \geq H(K|G,E=e) \geq k - \frac{2^{k-R_2(S|E=e)}}{\ln 2} \quad \square$$

U određenim praktičnim implementacijama, estimirana vrednost donje granice za ECRE2 važi sa određenom verovatnoćom. Teorema 3 definiše precizne uslove pod kojima PA blok može i u ovoj situaciji obezbediti maksimalnu neizvesnost Evinog znanja o uspostavljenim tajnim ključevima.

Teorema 3. [34] Neka je $S \in \{0,1\}^n$ slučajna promenljiva koja predstavlja zajedničku sekvencu koju dele Alisa i Bob, a neka je E slučajna promenljiva koja predstavlja ukupno znanje o S dostupno Evi. Neka je e posebna realizacija od E. Neka je verovatnoća da je e tačno određena realizacija od E koja uzima vrednost koja zadovoljava $R_2(S|E=e) \geq R_{2\delta}$ najmanje $1-\delta$. Neka je s proizvoljni sigurnosni parametar. Ako Alisa i Bob biraju $k(e) = R_{2\delta} - s$ kao svoj tajni ključ, gde je G heš funkcija izabrana uniformno slučajno iz univerzalne familije heš funkcija $G: \{0,1\}^n \to \{0,1\}^k$, tada je ekvivokacija ključa sa Evine strane

$$H(\mathrm{K}|\mathrm{G,E}) \geq (1\text{-}\delta)\left(\mathrm{k} - \tfrac{1}{\ln 2}2^{-\mathrm{s}}\right). \tag{4.13}$$

Dokaz. Direktnom primenom Teoreme 2, dobijamo

$$H(K|G,E) = \textstyle\sum_{all\ e} H(K|G,E=e)p(e) \geq \sum_{all\ e}\left[k - \frac{1}{ln\,2}2^{k\,-\,R_2(S|E\,=\,e)}\right]p(e).$$

Neka se skup svih sekvenci e podeli u dva skupa

$$E_+ = \{e|R_2(S|E=e) \geq R_{2\delta}\},$$

$$E_- = \{e|R_2(S|E=e) < R_{2\delta}\}.$$

Tada važi

$$\sum_{all\ e}\left[k - \frac{1}{\ln 2}2^{k\,-\,R_2(S|E\,=\,e)}\right]p(e) =$$

$$= \sum_{e\in E_-}\left[k - \frac{1}{\ln 2}2^{k\,-\,R_2(S|E\,=\,e\,)}\right]p(e) + \sum_{e\in E_+}\left[k\ - \frac{1}{\ln 2}2^{k\,-\,R_2(S|E\,=\,e\,)}\right]p(e)$$

$$\geq \sum_{e\in E_+}\left[k - \frac{1}{\ln 2}2^{k\,-\,R_2(S|E\,=\,e\,)}\right]p(e)$$

$$\geq \sum_{e\in E_+}\left[k - \frac{1}{\ln 2}2^{k\,-\,R_{2\delta}}\right]p(e)$$

$$= \sum_{e\in E_+}\left[k - \frac{1}{\ln 2}\cdot 2^{\,-s}\right]p(e)$$

$$= (1-\delta)\cdot\left(k - \frac{1}{\ln 2}2^{\,-s}\right)$$

što je i trebalo dokazati. □

Napomena 2. Iz Teoreme 3, i činjenice da maksimalna vrednost *Shannon*-ove entropije binarne sekvence dužine k ne može biti veća od k, za malo δ, K ima skoro maksimalnu entropiju za Evu:

$$k \geq H(K|G,E) \geq (1-\delta)\left(k - \frac{1}{\ln 2}2^{-s}\right) \tag{4.14}$$

Napomena 3. Ako je, tokom izvršavanja IR faze protokola, razmenjeno n_b bita parnosti putem javnog kanala, prema Lemi 4 [16], potrebno je izvršiti dodatnu kompresiju za istu količinu od n_b bitova u PA fazi.

Na osnovu Teoreme 2 i Napomena 1 i 3, može se tvrditi da je *optimalna PA strategija* zasnovana na ECRE2 data sa

$$k_{Opt}(e) = R_2(S|E=e) - n_b - s. \tag{4.15}$$

Glavna prepreka za primenu ove strategije je činjenica da ECRE2 nije poznata Alisi i Bobu jer je uslovljena poznavanjem Evine sekvence e, koja im u opštem slučaju nije dostupna. Ova fundamentalna neizvesnost predstavlja ključni problem u praktičnoj implementaciji optimalne PA strategije. Za prevazilaženje ove prepreke, u ovoj disertaciji je predloženo rešenje zasnovano na naprednim tehnikama mašinskog učenja, što otvara novi pravac istraživanja u oblasti SKD sistema. Uprkos svojim ograničenjima, klasične PA tehnike predstavljaju osnov informaciono-teorijski bezbednih SKD protokola, osiguravajući da svaki ostatak čiste slučajnosti u korist legitimnih strana bude pretvoren u perfektno tajne ključeve, uz neminovno žrtvovanje efikasnosti. U nastavku teksta biće razmotrena alternativa za estimaciju donje granice zasnovane na fenomenu *Spoiling knowledge* [16].

Neka je S sekvenca koju dele Alisa i Bob neposredno pre primene PA faze, kao što je prikazano na slici 5. Neka je E sekvenca koju dobija Eva prisluškivanjem sekvence S preko binarnog simetričnog kanala (eng. *Binary Symmetric Channel* - BSC) sa verovatnoćom greške ε. Pretpostavlja se da Eva, pored konkretne prisluškivane sekvence $E = e$, raspolaže dodatnom informacijom (eng. *side information*) u obliku slučajne promenljive $u = D_H(S,e)$, koja predstavlja Hamingovo rastojanje između S i posmatrane sekvence e.

Za dato $U = u$, sve sekvence s, dužine n, koje se nalaze na rastojanju u od e jednako su verovatni kandidati za S. Stoga se ECRE2 definiše kao [16]:

$$R_2(S|\,U=u, E=e) = \log_2\binom{n}{u}. \tag{4.16}$$

U [35], str. 309, Lema 7 pokazuje da važi:

$$\binom{n}{\lambda u} \geq \frac{2^{h(\lambda)}}{\sqrt{2n}}, \tag{4.17}$$

za svako $\lambda \in (0,1)$, gde je $h(x) = -x\cdot\log(x) - (1-x)\cdot\log(1-x)$, $0 < x < 1$, tzv. binarna entropijska funkcija.

Neka LB_u označava donju granicu za u, takvu da:

$$P\{LB_u \leq u\} \geq 1-\delta, \tag{4.18}$$

za svako $\delta > 0$. Tada važi:

$$\binom{n}{u} > \binom{n}{LB_u} = \binom{n}{n\frac{LB_u}{n}}, \quad \text{sa verovatnoćom najmanje } 1-\delta, \tag{4.19}$$

jer $\frac{LB_u}{n} \in (0,1)$. Uzimajući u obzir (4.17), sledi:

$$\binom{n}{u} \geq \frac{2^{nh\left(\frac{LB_u}{n}\right)}}{\sqrt{2n}}, \quad \text{sa verovatnoćom najmanje } 1-\delta. \tag{4.20}$$

Zamenom (4.20) u (4.16) dobija se donja granica za ECRE2:

$$R_2(S|\, U=u, E=e) \geq nh\left(\frac{LB_u}{n}\right) - \log_2\sqrt{2n}, \text{ sa verovatnoćom najmanje } 1-\delta. \tag{4.21}$$

Napomena 4. Donja granica (4.21) predstavlja eksplicitnu formu donje granice za ECRE2, koja se koristi u objašnjenju formulacije Teoreme 8 u [16] u kontekstu analize fenomena „Spoiling Knowledge". Stoga se ova granica naziva *Spoiling Knowledge donja granica za ECRE2* i označava se sa $R_{2Spoil}(LB_u, \delta)$.

Teorema 4. [36] Neka je $S \in \{0,1\}^n$ slučajna promenljiva koja predstavlja sekvencu koju dele Alisa i Bob neposredno pre PA faze nekog SKD protokola. Neka E označava slučajnu promenljivu koja obuhvata sve Evine informacije o S. Neka je e jedna realizacija E. Pored toga, Eva poseduje slučajnu promenljivu U, koja je zajednički distribuirana sa S i E prema nekoj raspodeli P_{UES}, pri čemu marginalna raspodela $[E,S]$ odgovara P_{ES}. Neka Alisa i Bob formiraju zajednički tajni ključ na osnovu preslikavanja $K = G(S)$, gde je G heš funkcija izabrana uniformno iz univerzalne klase heš funkcija $G: \{0,1\}^n \to \{0,1\}^r$. Tada neizvesnost generisanog ključa, posmatrana sa Evine strane, sa verovatnoćom najmanje $1-\delta$, zadovoljava nejednakost:

$$H(\mathrm{K}|\mathrm{G,E}) \geq \mathrm{r} - \textstyle\sum_{U,E} P(u,e)\log_2\left(1+2^{r-R_{2Spoil}(LB_u,\delta)}\right), \tag{4.22}$$

gde je

$$R_{2Spoil}(LB_u, \delta) = nh\left(\frac{LB_u}{n}\right) - \log_2\sqrt{2n}\,. \tag{4.23}$$

Dokaz. Dokaz neposredno sledi iz Posledice 7 [16] i činjenice da donja granica $R_{2Spoil}(LB_u, \delta)$ važi sa verovatnoćom najmanje $1-\delta$. □

Teorema 5. [36] Neka su veličine S, E, G, K i U definisane kao u Teoremi 4, i neka je $t \geq 0$ proizvoljan bezbednosni parametar. Ako Alisa i Bob odaberu dužinu k generisanog tajnog ključa $K = G(S)$, za svaku sekvencu S, tj. funkciju F takvu da važi:

$$k(F) = R_{2Spoil}(LB_u(F), \delta) - t, \tag{4.24}$$

tada neizvesnost generisanog tajnog ključa, posmatrana sa Evine strane, zadovoljava nejednakost:

$$H(K|G,E) \;\geq\; k(F) - log_2(1+2^{-t}), \quad \text{sa verovatnoćom najmanje } 1-\delta. \tag{4.25}$$

Dokaz. Prema Teoremi 4, neposredno dobijamo, nakon zamene (4.24) u (4.22):

$$H(K|G,E) \geq \mathrm{r} - \sum_{U,E} P(u,e)\, log_2(1+2^{r-R_{2Spoil}(LB_u(F),\delta)})$$

$$= k(F) - \sum_{U,E} P(u,e)\log_2(1+2^{-t})$$

$$= k(F) - \log_2(1+2^{-t})\text{ , sa verovatnoćom najmanje } 1-\delta.$$

Napomena 5. S obzirom da je *Shannon*-ova entropija $H(K)$ proizvoljne binarne sekvence K odozgo ograničena sa $H(K) \leq \mid K \mid = k$, iz Teoreme 5 neposredno sledi:

$$k(F) \geq H(K|G,E) \geq k(F) - \log_2(1+2^{-t}) \text{ , sa verovatnoćom najmanje } 1-\delta. \quad (4.26)$$

Dakle, destilovani tajni ključ K poseduje maksimalnu neizvesnost sa Evine strane.

Važno je istaći da, za razliku od *Shannon*-ove entropije *Renyi*-jeva entropija može da se poveća ukoliko je uslovljena slučajnom promenljivom [16] o čemu govori ovaj rezultat.

4.3 Razvoj nove PA strategije zasnovane na proceni ECRE2

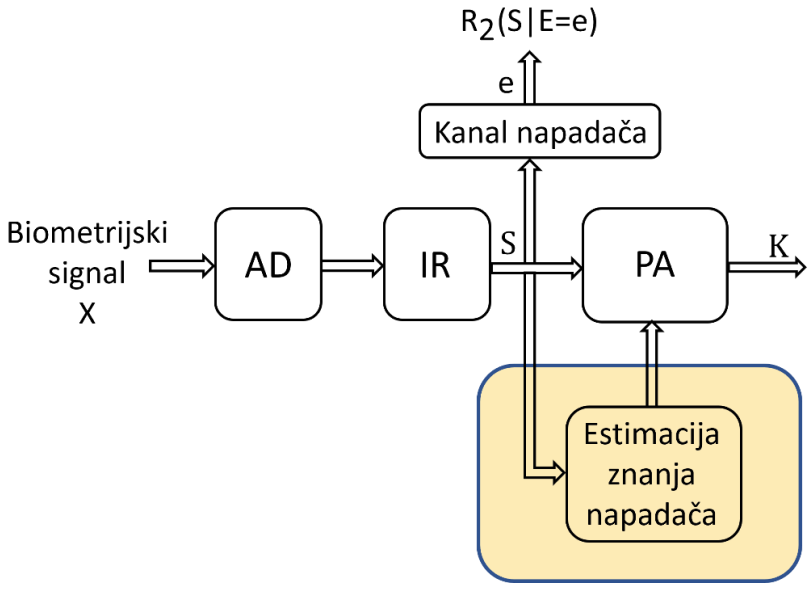
U realnim radnim uslovima datog SKD-a, budući da je verovatnosna struktura izabranog SCR poznata, moguće je formirati obučavajuće skupove podatka sledeće strukture, videti sliku 6:

$$\{[X_{ij}],[Y_{ij}],[Z_{ij}],S_i,e_i\}, i = 1,\dots,M, j = 1,\dots,N, \quad (4.27)$$

gde je N dužina pojedinačnih DMS sekvenci, X_{ij}, Y_{ij} i Z_{ij}, koje učestvuju u protokolu, a M predstavlja ukupan broj takvih sekvenci u skupu za obuku. Konkretno X_{ij} i Y_{ij} predstavljaju sekvence legitimnih korisnika, dok Z_{ij} predstavlja sekvencu prisluškivača. Sa S_i je označena usaglašena sekvenca legitimnih korisnika nakon IR faze, dok je sa e_i označena sekvenca prisluškivača nakon iste faze koja je dobijena na osnovu Z_{ij}, i svih informacija razmenjenih po javnom kanalu u toku realizacija AD i IR faze protokola. Pošto parametri S_i i e_i jednoznačno određuju uslovnu *Renyi*-jevu entropiju drugog reda, $R_{2i}(S_i|E = e_i)$, dobija se konačna forma obučavajućeg skupa $\{[X_{ij}],[Y_{ij}],R_{2i}\}$ koja se zapisuje kao

$$\{F_i, R_{2i}\}, i = 1,\dots,M. \quad (4.28)$$

Vektori obeležja F_i formiraju se tokom izvršavanja SKD-a na osnovu informacija koje poseduje jedan od legitimnih učesnika protokola, Alisa, tako da su u potpunosti izračunljivi i ne zahtevaju dodatnu komunikaciju. Jedan primer skupova obeležja koji zadovoljavaju ovaj uslov dati su u nastavu u Tabelama 1 i 2.



Slika 6. Generička arhitektura nove klase SKD sistema koja eliminiše estimirano znanje napadača iz konačnog ključa. Osenčeni blok predstavlja dodatak standardnoj teorijskoj SKD postavci.

Kao što je simbolično prikazano na slici 6, ove karakteristike se obično formiraju na osnovu informacija sa Alisine strane nakon izvršavanja pojedinačnih faza SKD protokola. Treba napomenuti da se razmatranje ograničava na direktni SKD, u kome Alisa pokreće protokol i određuje konačnu dužinu ključa [14]. Ista procedura važi i za inverzni protokol, u kome Alisa i Bob menjaju uloge.

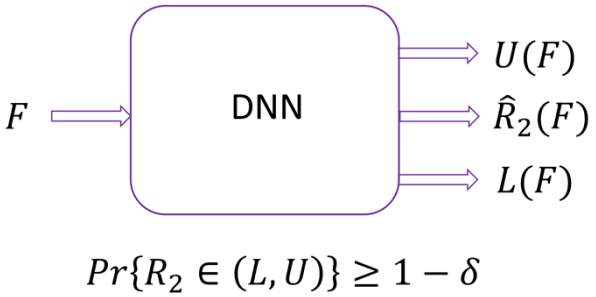
Napomena 6. Prelazak obučavajućeg skupa (4.27) na konačni oblik (4.28) zahteva izračunavanje ECRE2 za sve parove (S_i, e_i). To se može uraditi na osnovu izraza

$$R_2(S_i|E = e_i) = -log_2(\varepsilon^2 + (1-\varepsilon)^2), \tag{4.29}$$

gde je ε verovatnoća greške bita ekvivalentnog BSC-a, čiji je ulaz S_i a izlaz e_i [16]. Dobra procena za ε, je normalizovano Hamingovo rastojanje $D_h(S_i, e_i)$ između S_i i e_i. Normalizovano Hamingovo rastojanje između dve binarne sekvence X i Y istih dužina je dato sa

$$D_h(X,Y) = \frac{broj\ bita\ koji\ se\ ne\ poklapaju}{ukupan\ broj\ bita\ koji\ se\ uporedjuju}. \tag{4.30}$$

Ako bi blok mašinskog učenja (eng. *Machine Learning* - ML) na izlazu davao samo procenu $\widehat{R_2}$ za ECRE2, koja bi se zatim koristila u (4.15) za izračunavanje dužine destilovanog tajnog ključa, ne postoji garancija da će vrednost $\widehat{R_2}$ biti manja od prave vrednosti ECRE2. Prema Teoremi 2, tajni ključevi generisani na ovaj način ne bi imali poželjna kriptografska svojstva neizvesnosti i zanemarljivog curenja informacija Evi.Zato ML blok treba da na izlazu da interval u kome se, sa datom visokom verovatnoćom $1-\delta$, prava vrednost ECRE2 nalazi, videti sliku 7.



$$Pr\{R_2 \in (L,U)\} \geq 1-\delta$$

Slika 7. Blok mreže dubokog učenja sa predikcionim intervalom za procenu ECRE2 (PIDNN). Preuzeto iz rada [34], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Tada bi bilo moguće koristiti donju granicu $L(F)$ tog intervala kao procenu za ECRE2 u (4.15). Prema Teoremi 3, u tom slučaju postoji garancija poželjnih kriptografskih svojstava za dobijene ključeve.

4.4 Dizajn PIDNN

U oblasti mašinskog učenja, regresioni blok prikazan na slici 7 poznat je kao model duboke neuronske mreže (DNN) za interval predviđanja (PI), dizajniran da proizvede PI za svaki uzorak [37-39]. Uobičajen pristup za obuku PIDNN zasnovan je na dve kriterijumske funkcije: pokrivanje i srednja širina intervala predviđanja [40].

Pokrivanje je odnos uzoraka iz skupa podataka koji padaju u svoje odgovarajuće PI, mereno korišćenjem metrike verovatnoće pokrivanja intervala predviđanja (eng. *prediction interval coverage probability* - PICP)

$$PICP = \frac{1}{n}\sum_{i=1}^{n} m_i, \tag{4.31}$$

gde n označava broj uzoraka i $m_i = 1$ ako $R_{2i} \in (L(F_i), U(F_i))$, inače $m_i = 0$. Očigledno je da PICP teži ka $1 - \delta$.

Srednja širina intervala predviđanja (eng. *mean prediction interval width* - MPIW) je metrika kvaliteta za generisane PI čiji je cilj proizvođenje što je moguće uže granice:

$$MPIW = \frac{1}{n}\sum_{i=1}^{n} U(F_i) - L(F_i), \tag{4.32}$$

Treniranje PIDNN se izvršava postupkom optimizacije minimizacije MPIW uz zadržavanje predefinisane PICP. Kombinovanjem u jedinstven kriterijum, dobija se neograničena funkcija gubitka

$$J_{PI} = MPIW_\theta + \lambda\Psi(1 - \delta - PICP_\theta), \tag{4.33}$$

$$\Psi(x) = max(0, x)^2, \tag{4.34}$$

Gde je Ψ kvadratna penalizujuća funkcija a λ je hiperparametar koji kontroliše relativnu važnost širine u odnosu na pokrivanje. Algoritam korišćen u ovom radu zasnovan je na optimizaciji opisanoj u [39] i softverskom paketu dostupnom na odgovarajućem GitHub repozitorijumu [41].

PA strategija zasnovana na PIDNN u nastavku teksta označava se kao strategija mašinskog učenja. Može se formulisati kao

$$k_{ML}(F_i) = L(F_i) - n_b - s. \tag{4.35}$$

U PA sistemu, zasnovanom na strategiji (4.35), može se desiti situacija gde je $L(F_i) < c$, a gde je c globalna donja granica ECRE2. Tada je bolja strategija globalne donje granice (4.7), što opravdava uvođenje sledeće strategije, koja će u nastavku biti označena kao *hibridna PA strategija*

$$k_{Hyb}(F_i) = L_{Hyb}(F_i) - n_b - s. \tag{4.36}$$

gde je

$$L_{Hyb}(F_i) = max(L(F_i), R_{2\delta}), \tag{4.37}$$

dok je $R_{2\delta}$ vrednost ECRE2 koja zadovoljava uslov

$$P\{R_2 \geq R_{2\delta}\} \geq 1 - \delta, \delta = \frac{1-PICP}{2}, \tag{4.38}$$

PICP je vrednost pokrivanja (4.31) PIDNN obučene na datom DMS.

Algoritam 6 prikazuje korake predložene PA metodologije dizajna zasnovane na mašinskom učenju.

Algoritam 6. Metodološki tok eksperimenta

1: Za dati DMS generisati reprezentativnu populaciju (X_i, Y_i, Z_i),
$$|X_i| = |Y_i| = |Z_i| = n_i, \quad i = 1,2, \dots, M$$
2: Formirati skup $\{ S_i, e_i, u_i = D_H(S_i, e_i)\}, \quad i = 1, \dots, M.$
3: Formirati skup $\{F_i, D_H(S_i, e_i)\}, \quad i = 1,2, \dots, M.$
4: Započeti 10-tostruku kros-validaciju
5: Trenirati PIDNN za predikciju $D_H(S_i, e_i)$ i njegove gornje i donje granice, $\{UB_u(F_i), LB_u(F_i)\}$, na trening delu trenutnog podskupa
6: Primeniti globalnu strategiju i ML PA strategiju na sve sekvence test dela trenutnog podskupa. Minimalna vrednost ECRE2 nalazi se na trening delu trenutnog podskupa. Vrednost ECRE2 dobija se izračunavanjem sledećeg izraza
$$R_2(S_i|E = e_i) = -\log_2\left(D_{hi}{}^2 + \left(1 - D_{hi}{}^2\right)^2\right).$$
7: Izračunati sve parametre performansi trenutnog podskupa
8: Kraj kros-validacije
9: Računanje statistika, srednje vrednosti i standardne devijacije, svih parametara

U Tabeli 1 su prikazana informaciono-teorijska obeležja koja se mogu izračunati u različitim fazama SKD protokola, pre primene faze pojačanja privatnosti. Ova obeležja razmatrana su za procenu uslovne *Renyi*-jeve entropije drugog reda u prethodno opisanom sistemu.

Treba napomenuti da su obeležja 12* i 13* korišćena isključivo u sistemu zasnovanom na kaskadnom algoritmu. Razlog za to leži u činjenici da kaskadni algoritam ne odbacuje sigurnosne bitove tokom procesa usaglašavanja informacija, već se svi bitovi koriste do faze pojačanja privatnosti. Nasuprot tome, kod *Winnow* algoritma dolazi do odbacivanja sigurnosnih bitova već u toku same procedure usaglašavanja.

Tabela 1. Informaciono-teorijska obeležja od interesa za sintezu PIDNN bloka za predikciju donje granice ECRE2

Oznaka	Opis obeležja
IT 1	Dužina početne sekvence
IT 2	Dužina sekvence nakon prve iteracije AD algoritma
IT 3	Dužina sekvence nakon druge iteracije AD algoritma
IT 4	Normalizovana blok entropija na početku računata sa blokom dužine 8
IT 5	Normalizovana blok entropija računata sa blokom dužine 8 nakon prve iteracije AD algoritma
IT 6	Normalizovana blok entropija računata sa blokom dužine 8 nakon druge iteracije AD algoritma
IT 7	Broj razmenjenih poruka o parnosti u okviru AD faze
IT 8	Dužina sekvence nakon IR faze
IT 9	Normalizovana blok entropija računata sa blokom dužine 8 nakon IR algoritma
IT 10	Dužina sekvence nakon primene kompresionog algoritma bez gubitaka
IT 11	Normalizovana blok entropija računata sa blokom dužine 8 nakon primene kompresionog algoritma bez gubitaka
IT 12*	Broj poruka parnosti razmenjenih u okviru IR algoritma
IT 13*	Broj bita koje prisluškivač tačno zna nakon IR algoritma

Pored prethodno opisanih informaciono-teorijskih obeležja, koja su se ispostavila izuzetno kvalitetna za procenu donje granice uslovne *Renyi*-jeve entropije drugog reda u slučaju eeg izvora, izabranom skupu su pridodata i stilometrijska obeležja, originalno razvijena za analizu iris biometrijskih podataka [42]. Stilometrija, iako tradicionalno vezana za analizu pisanog teksta, zasniva

se na fundamentalnom principu identifikacije stabilnih, karakterističnih statističkih obrazaca koji opisuju način generisanja podataka. Ovaj pristup se može uspešno proširiti na različite tipove signala, uključujući biometrijske sekvence, gde omogućava kvantifikaciju individualnih statističkih varijacija bez potrebe za složenom strukturnom ili semantičkom analizom.

U kontekstu SKD sistema, stilometrijska obeležja pružaju dodatnu prednost jer se mogu izračunati lokalno od strane legitimnih učesnika Alise i Boba, bez potrebe za komunikacijom preko javnog kanala. Konkretno, razmatran je skup od 22 stilometrijska obeležja (ST 1-ST 22) detaljno opisanih u [42], pri čemu je izostavljena poslednja karakteristika iz originalnog skupa, Tabela 2. Ova obeležja su dobijena prethodnom transformacijom kvantizovane sekvence korišćenjem Base64 enkodera, čime se sekvenca konvertuje u "rečenice" jezika definisanog nad alfabetom veličine 64 [42]. Ovakav pristup omogućava da se numeričke sekvence posmatraju kroz okvir stilometrijske analize, čime se dobija robustan i informativan opis signala pogodan za dalju klasifikaciju

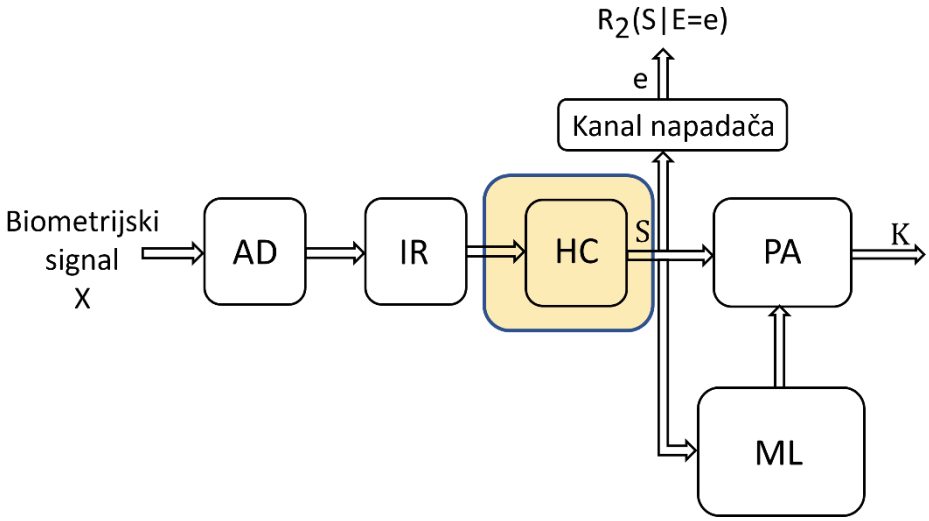
Tabela 2. Stilometrijska obeležja od interesa za sintezu PIDNN bloka za predikciju donje granice ECRE2

Oznaka	Opis obeležja
ST 1	Prosečna dužina rečenice u tekstu
ST 2	Procenat velikih slova u odnosu na mala slova u tekstu
ST 3	Procenat malih slova u odnosu na ukupan broj znakova u tekstu
ST 4	Procenat interpunkcijskih znakova u odnosu na ukupan broj razmaka u tekstu
ST 5	Procenat numeričkih znakova u odnosu na ukupan broj slovnih znakova u tekstu
ST 6	Prosečna dužina reči u rečenici u tekstu
ST 7	Frekvencija najčešće pojavljivane reči na kraju rečenice u tekstu
ST 8	Frekvencija najčešće pojavljivane reči na početku rečenice u tekstu
ST 9	Frekvencija početnog slova koje se najčešće pojavljuje u početnoj reči
ST 10	Frekvencija početnog slova koje se najčešće pojavljuje u reči na kraju
ST 11	Broj reči koje se samo jednom pojavljuju u tekstu
ST 12	Broj reči koje se dva puta pojavljuju u tekstu
ST 13-ST 17	Broj reči u tekstu sa određenom dužinom (dužine 3, 4, 5, 6, 7)
ST 18-ST 20	Broj reči čije su dužine u odgovarajućem opsegu. (opseg [3-3], [3-9], [3-15])
ST 21	Broj samoglasnika u tekstu
ST 22	Deflate algoritam - najbolja kompresija teksta

4.5 *Huffman*-ov koder

Huffman-ovo kodovanje je optimalan algoritam kodovanja izvora bez gubitaka, koji bira dužine kodnih reči obrnuto proporcionalno frekvenciji simbola koje emituje dati izvor [43]. Zbog mogućnosti da transformiše niz i redistribuira informaciju bez gubitka, *Huffman* se nameće kao prirodna lokalna transformacija u kriptografskim protokolima. Računarski je efikasan, lako se integriše u sekvencijalne protokole i ne zahteva iterativnu komunikaciju niti razmenu pomoćnih podataka. U okviru SKD-a može se koristiti za predobradu pre PA faze, videti sliku 8, gde kompresija može da transformiše usaglašenu sekvencu u kraću, ali informacijski ekvivalentnu reprezentaciju sa smanjenom statističkom pristrasnošću, što predstavlja poželjan ulaz za konačno pojačanje privatnosti.

Huffman-ovo kodovanje predstavlja jedan od najefikasnijih algoritama kompresije podataka bez gubitaka zasnovanom na promenljivoj dužini kodnih reči. Predložen je u radu [43] i standardno se koristi za različite potrebe digitalne komunikacije i skladištenje podataka. Glavna prednost primene *Huffman*-ovog kodovanja u kontekstu destilovanja tajnih ključeva ogleda se u lokalnom generisanju kodnog rečnika zasnovano na statističkim karakteristikama sekvenci koje se koduju. S obzirom da legitimni učesnici poseduju usaglašene sekvence nakon IR faze, oni generišu identične kodne rečnike i primenjuju istovetan proces kodovanja. Nasuprot tome, čak i minimalna razlika u sekvenci potencijalnog napadača rezultuje generisanjem potpuno različitog kodnog rečnika. Ova razlika u rečnicima dovodi do toga da komprimovane sekvence između legitimnih učesnika i napadača mogu značajno divergirati kako po sadržaju, tako i po dužini. Na taj način, *Huffman*-ovo kodovanje ne doprinosi samo smanjenju redundanse, već dodatno uvećava statističku distancu između sekvenci legitimnih učesnika i potencijalnog napadača, čime se posredno povećava bezbednost protokola. Lokalnim generisanjem rečnika eliminiše se potreba za razmenom informacija putem kanala, čime izostaje dodatno opterećenje komunikacionog kanala, kao i dodatno oticanje informacija.



Slika 8. Proširena arhitektura osnovnog SKD sistema. Osenčeni blok označava *Huffman*-ov optimalni koder (eng. *Huffman coder* - HC).

Osnovna ideja algoritma zasnovana je na činjenici da se simboli koji se češće pojavljuju mogu predstaviti kraćim kodnim rečima dok se ređe pojavljivani simboli, suprotno ovome, kodiraju dužim kodnim rečima. Na taj način se ukupna prosečna dužina kodirane sekvence minimizuje i postaje približno jednaka entropiji izvora. Koraci *Huffman*-ovog kodiranja predstavljeni su algoritmom 7.

Algoritam 7. *Huffman*-ov koder

1: Određivanje verovatnoće simbola zadate dužine
2: Sortiranje simbola po rastućim verovatnoćama koji predstavljaju čvorove stabla
3: Izdvajaju se dva čvora sa najmanjim verovatnoća i spajaju tako da kreiraju novi, unutrašnji čvor čija je verovatnoća jednaka zbiru verovatnoća ta dva čvora
4: Novi čvor zamenjuje stare čvorove
5: Koraci 2-4 se ponavljaju sve dok ne ostane samo jedan čvor koji predstavlja koren *Huffman*-ovog stabla
6: Pri svakom prelazu od jednog nivoa stabla ka drugom dodeljuje se bit 0, za levi čvor, bit 1 za desni čvor
7: Za svaki simbol sekvence na osnovu stabla određuje se njegov kod

Ukoliko se smatra da je dužina ulazne sekvence n, složenost *Huffman*-ovog algoritma za kompresiju sekvenci iznosi $O(n\log n)$, jer algoritam nakon određivanja učestalosti simbola mora da izgradi Hafmanovo stablo spajanjem čvorova. Tokom izgradnje stabla izvršavaju se operacije spajanja, pri čemu svako spajanje zahteva pronalaženje i uklanjanje dva čvora sa najmanjim učestalostima iz prioritetnog reda, što rezultuje složenosti $O(n\log n)$. Ovaj rezultat ukazuje na mogućnost efikasne primene kompresionog algoritma u realnim komunikacionim sistemima bez obzira na dužinu ulazne sekvence, jer broj potrebnih operacija raste umereno sa njenom dužinom.

5 Biometrijski signali kao izvori zajedničke slučajnosti

5.1. Pregled i izbor biometrijskih signala pogodnih za SKD

Značajan potencijal za generisanje kriptografskih ključeva uočava se u biometrijskim signalima, koji predstavljaju prirodan izbor za savremene protokole čiji je imperativ nalaženje efikasnih izvora čiste slučajnosti.

U okviru SKD klase protokola zasnovanih na modelu izvora [14], biometrijski signali korišćeni su kao izvor zajedničke slučajnosti, uključujući hod [44,45], elektrokardiogram (EKG) [46], pokrete oka i miša [47], kao i EEG signale [34,48]. Ostvarena brzina generisanja tajnih ključeva kreće se u opsegu od 2 do 26 b/s za takozvani model izvora bez dodatnih informacija dostupnih napadaču [10]. Ovaj model obuhvata hod, EKG i druge izvore pogodne za generisanje tajnih kriptografskih ključeva radi bezbedne komunikacije između različitih uređaja lociranih na ljudskom telu [44]. U ovim uslovima, kapacitet tajnih ključeva određen je stopom međusobne informacije između terminalnih signala dostupnih na početku protokola.

U slučaju takozvanog modela izvora sa dodatnim informacijama dostupnim napadaču [10], razmatranog u radovima [34,48], ostvarene su brzine generisanja ključeva od oko 10 b/s za tzv. EEG signale metrika, kao i do 1200 b/s za sirove EEG signale sa 14 kanala. Za ovaj tip modela, kapacitet tajnih ključeva određen je brzinom međusobne informacije između terminalnih signala, uslovljene odgovarajućim signalom prisluškivača dostupnim na početku protokola.

U radovima [49,50], rastojanje između legitimnih čvorova u mobilnim bežičnim mrežama posmatra se kao izvor zajedničke slučajnosti. Eksperimentalni rezultati ukazuju na brzinu generisanja ključeva u opsegu od 0,1 do 0,6 b/s, u zavisnosti od brzine kretanja terminala i položaja prisluškivača.

Svi navedeni rezultati ukazuju na to da, do sada, ne postoje sistemi zasnovani na modelu izvora sa dodatnim informacijama dostupnim pasivnom napadaču koji ostvaruju visoke brzine generisanja tajnih ključeva, izuzev sistema zasnovanog na sirovim EEG signalima predstavljenog u radu [34].

Odlučeno je da se analiziraju dva različita tipa izvora zajedničke slučajnosti, EEG signali i govorni signali zbog toga što predstavljaju biometrijske signale praktično neograničenog trajanja što omogućava izdvajanje proizvoljne, unapred određene, količine kriptografskih ključeva.

EEG se teže kompromituje od tradicionalnih biometrijskih metoda jer odražava trenutnu moždanu aktivnost. Stres ili prinuda pri prikupljanju EEG signala menjaju karakteristike signala, što omogućava detekciju pokušaja zloupotrebe i čini kompromitovane podatke neupotrebljivim. Takođe, EEG signal pojedinca u različitim vremenskim trenucima pri istom mentalnom zadatku nikada nije identičan, iako se karakteriše velikom korelacijom, te ga nemogućnost identične reprodukcije čini idealnim entropijskim izvorom. Ova osobina pruža i dugoročnu bezbednost sistema, jer omogućava da dva učesnika u komunikaciji tokom različitih sesija generišu potpuno novi ključ bez rizika od kompromitovanja prethodno generisanih ključeva. Za potrebe akvizicije EEG signala koriste se neinvazivne metode, a za potrebe ovog istraživanja izabran je bežični EEG uređaj EMOTIV EPOC+ zbog svoje pristupačnosti i univerzalne dostupnosti na tržištu. Govorni signal izabran je kao drugi izvor zbog jednostavnosti generisanja i izostanka potrebe za specijalnim uređajima za akviziciju signala. Način izgovora, akcentovanje, ritam, tempo, pa i osnovna (eng. *pitch*) frekvencija pružaju jedinstvenost i dobar su izvor neodređenosti. U radovima [51,52] ova dva tipa biometrijskih signala korišćena su za generisanje kriptografskih ključeva izdvajanjem specifičnih karakteristika. U ovom radu navedeni biometrijski signali korišćeni su u izvornom obliku budući da se već u toj formi pojavljuje korelaciona struktura adekvatna za SCR.

5.1.1 Elektroencefalografski (EEG) signali

Zbog svoje velike količine prirodne stohastičke neodređenosti, EEG signali privlače sve veću pažnju istraživača u oblasti kriptografije posebno imajući u vidu savremeni imperativ za nalaženjem adekvatnih prirodnih izvora slučajnosti. U radu [51], autori pokazuju da se iz EEG signala može ekstrahovati dovoljna količina entropije za generisanje 192-bitnih kriptografskih ključeva sa procentom uspešnosti od 99%. Slično tome u radu [53] naglašena je mogućnost generisanja ključeva koji zavise od trenutnog stanja moždane aktivnosti, čime se otežava njihova reprodukcija i predikcija od strane napadača. Iako ovi radovi potvrđuju da EEG signali poseduju značajan potencijal kao izvor slučajnosti za kriptografske primene, u dostupnoj literaturi nije zabeležena njihova primena u SKD sistemima što pruža prostor za ispitivanje upotrebe EEG signala u okviru SKD sistema koje će biti razmatrano u nastavku.

Izvor EEG-a biće analiziran kroz dva pristupa. Prvi pristup, označen kao sirovi EEG, formiran je serijalizacijom podataka snimljenih bežičnim uređajem EMOTIV EPOC+ pomoću četrnaest elektroda koje detektuju promene električnog potencijala koje se registruju na površini glave. Time je omogućeno dobijanje signala koji zadržava izvorne informacije o električnoj aktivnosti mozga u vremenu, bez naknadne obrade i filtriranja, odnosno ispitivanje entropijskih kvaliteta neprocesiranih informacija. Drugi pristup, označen kao metrike EEG-a, formiran je serijalizacijom šestodimenzionalnih metrika performansi koje kvantitativno opisuju emocionalne i kognitivne komponente ponašanja ispitanika. Ove metrike obuhvataju interesovanje (privlačnost ili odbojnost zadatka), angažovanost (ili dosadu u negativnom naboju), uzbuđenost (emocionalni intenzitet), stres (frustraciju), opuštenost (meditaciju), i fokus (pažnju). Njihove vrednosti dobijeni su u okviru EMOTIV EPOC+ software-a. Svaka metrika predstavlja relativni udeo određene psihofiziološke komponente u EEG signalu koji se procentualno izražava. Viši procenat označava snažniju prisutnost odgovarajućeg stanja u određenom trenutku.

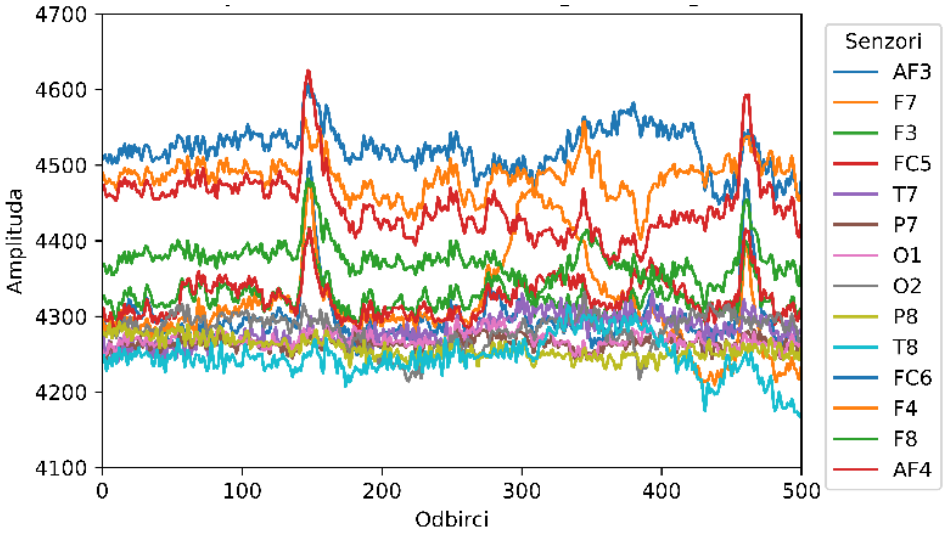
Snimanje signala realizovano je u asinhronom režimu. Posebno je značajno što su korelacione strukture EEG signala invarijantne na vreme i mesto testiranja, kao i na različite testirane subjekte, što omogućava pouzdanu asinhronu akviziciju podataka. Ovo svojstvo predstavlja značajnu praktičnu prednost u situacijama kada je kompleksno uspostaviti preciznu sinhronizaciju koja bi zahtevala dodatne resurse i koja bi povećala sistemsku kompleksnost SKD implementacije.

Podaci su prikupljani od 50 učesnika uzrasta između 20 i 65 godina, slučajno odabranih među zaposlenima Instituta za visoke tehnologije Vlatacom u Beogradu. Učesnici su bili potpuno informisani o postupku istraživanja, uključujući proces postavljanja senzora, i dobrovoljno su pristali na učešće u testiranju. Institucionalni etički komitet je formalno odobrio ovo istraživanje u skladu sa principima Helsinške deklaracije. Na slici 9 je prikazana akvizicija EEG signala na jednom od ispitanika.

Slika 9. Eksperimentalno prikupljanje EEG signala pomoću EMOTIV EPOC+ uređaja.

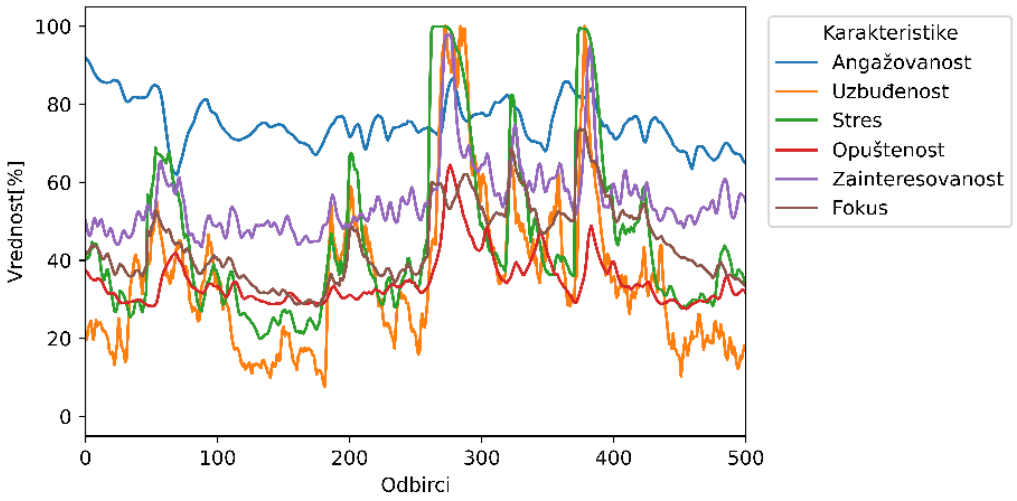
Svaka sesija snimanja EEG signala trajala je 20 minuta po učesniku, tokom čega su učesnici imali slobodu da obavljaju aktivnosti po sopstvenom izboru. Tipične aktivnosti uključivale su čitanje internet sadržaja, igranje video igara, rad na ličnim projektima ili meditaciju. Za svakog učesnika, slučajno su selektovana dva uzorka od 2 sekunde za sirovi EEG izvor čija je frekvencija odabiranja 128Hz, odnosno dva uzorka od 300 sekundi za izvor EEG metrika čija je frekvencija odabiranja 2Hz. Ovim pristupom je formiran skup od 100 statistički reprezentativnih uzoraka za svaki izvor na osnovu kojih će biti izvršena analiza performansi predloženih protokola. Izbor dva vremenski odvojena EEG uzorka istog ispitanika uveden je radi analize uticaja vremenskih varijacija EEG signala jednog ispitanika na performanse procesa generisanja tajnog ključa. Iako signali istog ispitanika pokazuju visok stepen međusobne korelacije i zadržavaju karakteristične individualne obrasce, njihove statističke i entropijske karakteristike razlikuju se u različitim vremenskim trenucima usled nestacionarne prirode EEG aktivnosti, promena mentalnog stanja ispitanika i uslova akvizicije signala. Dodatno, u cilju analize otpornosti sistema na pasivnog napadača, sekvenca napadača birana je nasumično iz skupa preostalih EEG sekvenci, odnosno iz skupa koji ne sadrži sekvence korišćene za formiranje para legitimnih učesnika protokola. Ovakav model omogućava razmatranje scenarija u kojem napadač raspolaže prethodno snimljenim EEG signalima drugih korisnika sistema, uključujući i slučaj potencijalne kompromitacije ranije snimljene biometrije jednog od učesnika. Time se modeluje insajderski napadač koji poseduje bazu EEG signala populacije korisnika sistema i pokušava da na osnovu dostupnih biometrijskih podataka proceni informacije o generisanom tajnom ključu.

Na slici 10 prikazan je isečak EEG signala dobijenog promenama električnog potencijala izmerenog na senzorima koji je predstavljen kvantizovanim jedinicama amplituda moždanih aktivnosti.



Slika 10. Isečak EEG signala merenog na senzorima.

Na slici 11 prikazan je isečak EEG signala kojim su interpretirane metrike koje procenjuju emocionalne i kognitivne parametre ispitanika. Vrednosti metrika procentualno su predstavljene.



Slika 11. Isečak EEG signala sa procentualno izraženim karakteristikama emocionalnih i kognitivnih parametara.

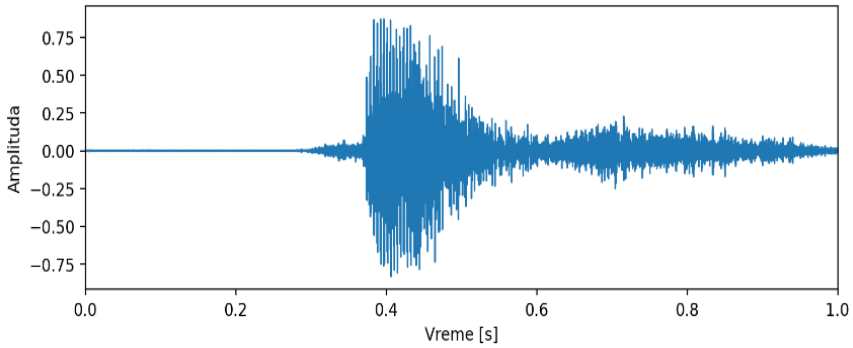
5.1.2 Govorni signali

Govorni signal predstavlja kompleksan biološki i akustički fenomen koji nastaje kroz koordinisanu aktivnost različitih fizioloških podsistema (pluća, glasnih žica, usne šupljine i nosa) koji zajedno formiraju sistem za generisanje i modulaciju zvuka. Tokom artikulacije, protok vazduha iz pluća prolazi kroz vokalni trakt, gde vibracije glasnih žica stvaraju osnovni ton, dok oblik i položaj rezonantnih šupljina određuju spektar i karakteristike proizvedenog zvučnog talasa [54]. U fizičkom smislu, govor je akustički talas koji se prostire u vazduhu, dok se u tehničkom smislu posmatra kao vremenski kontinuiran analogni signal koji može biti digitalizovan i analiziran kao diskretno-vremenska sekvenca uzoraka. Sa stanovišta teorije informacija govor može biti posmatran kao realizacija slučajnog procesa [55]. Preliminarnom analizom vrednosti govornog signala i njihovih raspodela uočava se da govorni signal poseduje visoku entropiju i lokalnu zavisnost.

Kao i kod EEG signala dinamička priroda govornog signala i njegove jedinstvene karakteristike vezane za pojedinca kandiduje ovaj tip biometrije za razmatranje u kriptografskim sistemima. Dosadašnja istraživanja govornog signala kao izvora slučajnosti odvijala su se u dva smera. Prvi je bio usmeren na metode za identifikaciju i autentifikaciju korisnika na osnovu prepoznatljivih akustičnih karakteristika pojedinca, kao što su osnovna frekvencija, tonalitet, tempo i drugo, dok drugi smer istraživanja pomera fokus sa identifikacije na generisanje kriptografskih ključeva, pri čemu govorni aparat postaje izvor entropije koji omogućava formiranje tajnih ključeva na osnovu prirodnih varijacija ljudskog govora. Međutim, istraživanja koja bi koristila govor kao izvor SCR u SKD sistemima nije zabeležen u dostupnoj literaturi.

Za evaluaciju predloženih sistema SKD baziranih na govoru korišćeni su uzorci iz javno dostupnih baza [56,57]. Ovaj skup su kreirala 2.618 različitih govornika, koji su snimili ukupno 105.829 govornih signala, pri čemu svaki uzorak predstavlja jednu od 35 različitih reči. Podaci su zapisani kao linearni 16-bitni jednokanalni PCM uzorkovani na 16kHz [57] pri čemu je svaki uzorak dužine jedne sekunde. Iz skupa od 35 mogućih reči slučajno je izabrana reč “house” koja pokriva različite foneme te predstavlja reprezentativan izbor. Da bi se izvršilo poređenje sa rezultatima dobijenih sa drugim biometrijskim izvorom, izabran je podskup od 100 signala te izgovorene reči.

Na slici 12 prikazan je slučajno izabran govorni signal iz prethodno opisanog skupa. Signal je prikazan u vremenskom domenu, a amplitudne vrednosti normalizovane su u opsegu od -1 do 1.



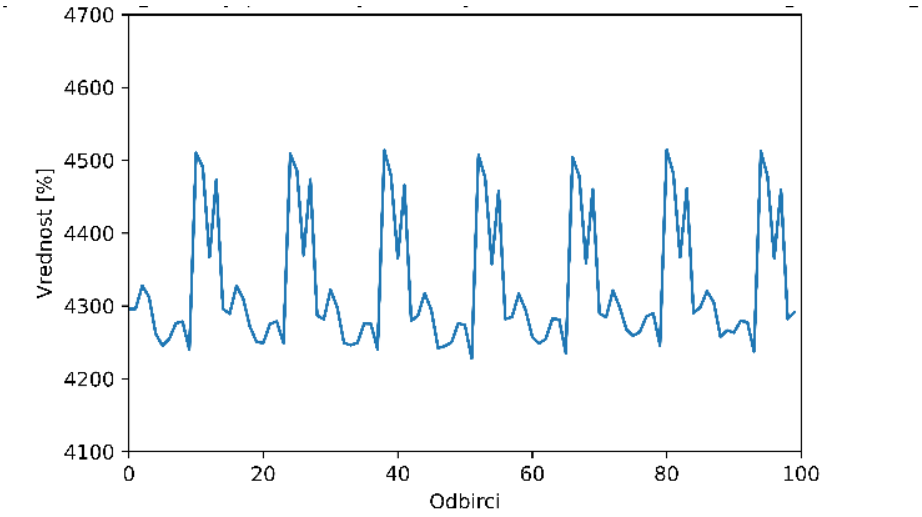
Slika 12. Talasni oblik slučajno izabranog govorni signala iz skupa podataka, prikazan u vremenskom domenu sa normalizovanom amplitudom.

Kod oba biometrijska izvora, skup za eksperimentalnu evaluaciju formiran je od tripleta signala koji modeluju komunikaciju između dva legitimna korisnika (Alise i Boba) i pasivnog napadača (Eve), pri čemu svaki triplet predstavlja jednu realizaciju posmatranog DMS-a. U okviru svakog tripleta, dva snimljena signala dodeljuju se Alisi i Bobu i predstavljaju legitimne sekvence, X i Y, dok se signal Eve, Z, nasumično bira iz skupa svih preostalih signala koji ne pripadaju posmatranom paru legitimnih korisnika. Na taj način simulira se napadač koji raspolaže sopstvenim biometrijskim uzorkom i nema pristup signalima legitimnih učesnika protokola. Kako je ukupan broj snimljenih signala 100, moguće je formirati 100*99/2=4950 različitih parova legitimnih korisnika. Svakom takvom paru pridružuje se nasumično izabrani signal Eve, čime se dobija skup od 4950 tripleta koji će biti korišćeni u eksperimentalnoj evaluaciji.

6 Eksperimentalna evaluacija predložene nove klase SKD sistema

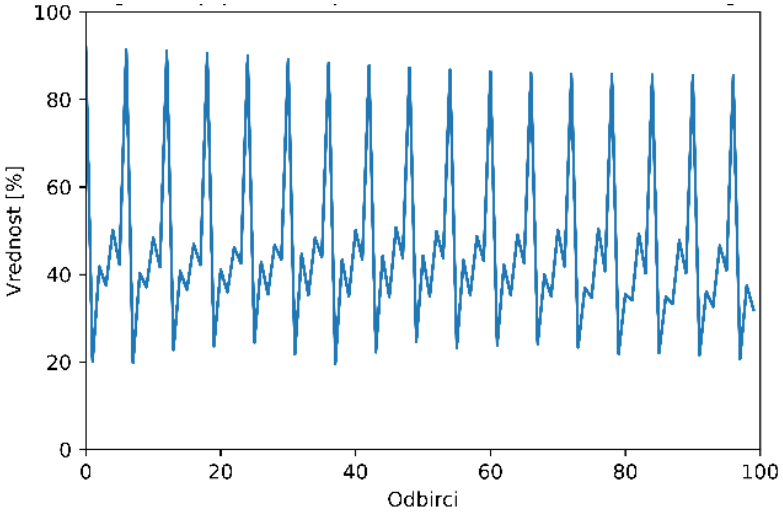
U prethodnom delu opisan je postupak izbora signala i način formiranja skupa podataka koji se koristi u daljoj analizi. Nakon procesa merenja, sekvence sirovog EEG signala dobijene merenjem sa više senzora, odnosno sekvence formirane na osnovu EEG metrika, prolaze kroz proces serijalizacije. Serijalizacija predstavlja transformaciju višedimenzionalnih vektora uzoraka, koji potiču iz paralelnih kanala posmatranja, u jedinstvenu jednodimenzionalnu sekvencu pogodnu za dalju numeričku obradu i primenu informacionih mera. U slučaju sirovog EEG signala, u svakom vremenskom trenutku registruju se vrednosti napona na više senzora, koje zajedno čine vektorski uzorak. Kod signala predstavljenog kroz EEG metrike, svaka metrika predstavlja komponentu takvog vektora. Proces serijalizacije obezbeđuje sukcesivno raspoređivanje svih komponenti vektora u jedinstvenu vremensku sekvencu, čime se dobija linearna reprezentacija originalnog višekanalnog signala.

U nastavku su ilustrovani rezultati procesa serijalizacije za oba razmatrana tipa EEG signala. Na slici 13 prikazan je serijalizovani EEG signal dobijen iz višekanalnih merenja sa senzora, dok je na slici 14 prikazan primer serijalizovanog EEG signala formiranog na osnovu EEG metrika. Ovi primeri jasno ilustruju transformaciju višedimenzionalnih vektorskih uzoraka u jedinstvenu jednodimenzionalnu sekvencu pogodnu za dalju analizu.



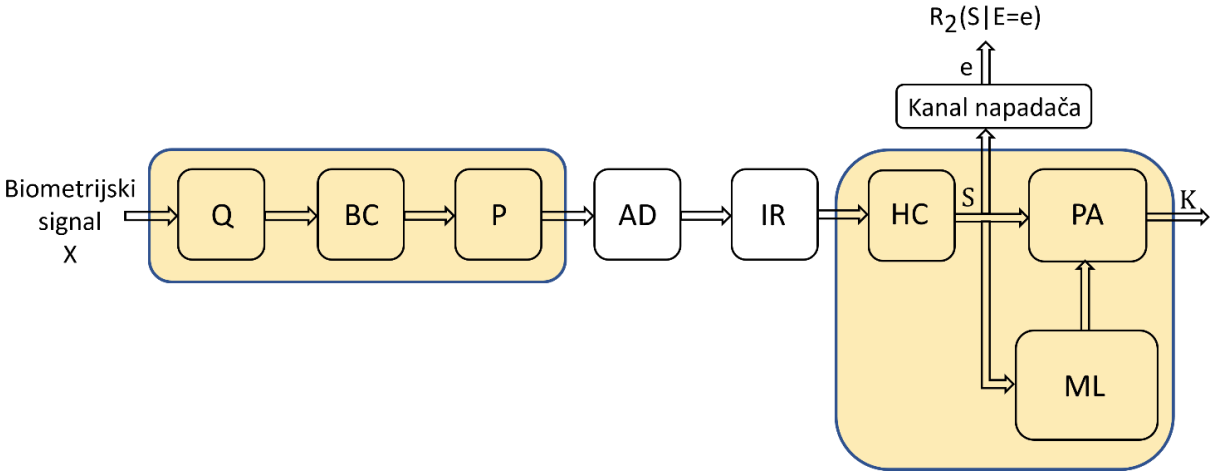
Slika 13. Serijalizovani EEG signal dobijen iz višekanalnih merenja sa senzora.

Ovakva reprezentacija omogućava da se dalja obrada i primena informacionih mera (poput entropije, blok entropije ili uslovne entropije) sprovode nad jednom sekvencom podataka, bez potrebe za zasebnim tretiranjem pojedinačnih kanala. Na taj način višedimenzionalni EEG podaci se prevode u format koji zadržava sve relevantne informacije o amplitudnim promenama i međukanalnim odnosima, koji su bitni za analizu algoritama u informacionom domenu.



Slika 14. Serijalizovani EEG signal formiran na osnovu EEG metrika.

Serijalizovana sekvenca se dalje koristi kao ulaz u blokove za analizu informacione složenosti, procenu entropijskih mera ili destilaciju kriptografskih ključeva. Na slici 15 je prikazana kompletna arhitektura sistema, koja pored prethodno dodatog ML bloka i bloka *Huffman*-ovog kodera obuhvata i blokove neophodne za početnu transformaciju ulaznog signala u odgovarajuću formu za dalju obradu.



Slika 15. Kompletna arhitektura nove klase SKD sistema. Blok Q označava kvantizacioni blok, blok BC predstavlja blok za pojačavanje korelacije, dok je sa P označen permutacioni blok.

Da bi se omogućila dodatna digitalna obrada signala i primena informaciono-teorijskih mera, neophodno je prethodno sprovesti kvantizaciju. Broj kvantizacionih nivoa, odnosno broj kvantizacionih bita, direktno određuje rezoluciju i tačnost digitalne reprezentacije signala dok u teoriji informacija oni utiču i na entropiju i korelacionu međuzavisnost signala.

U okviru bloka kvantizacije i pojačavanja korelacije, proces transformacije signala realizovan je kombinacijom stvarne kvantizacije i determinističkog mapiranja dodatnih bita. Prva komponenta odnosi se na kvantizaciju ulaznog signala, pri čemu se ograničen broj bita koristi za vernu diskretizaciju amplitudnih vrednosti i očuvanje statističkih karakteristika signala. Druga komponenta obuhvata deterministički generisane bitove, pri čemu se vodi računa o njihovoj međusobnoj uravnoteženosti. Na taj način obezbeđuje se približno jednaka zastupljenost determinističkih nula i jedinica, bez narušavanja osnovne informacije sadržane u kvantizovanom delu signala. Važno je

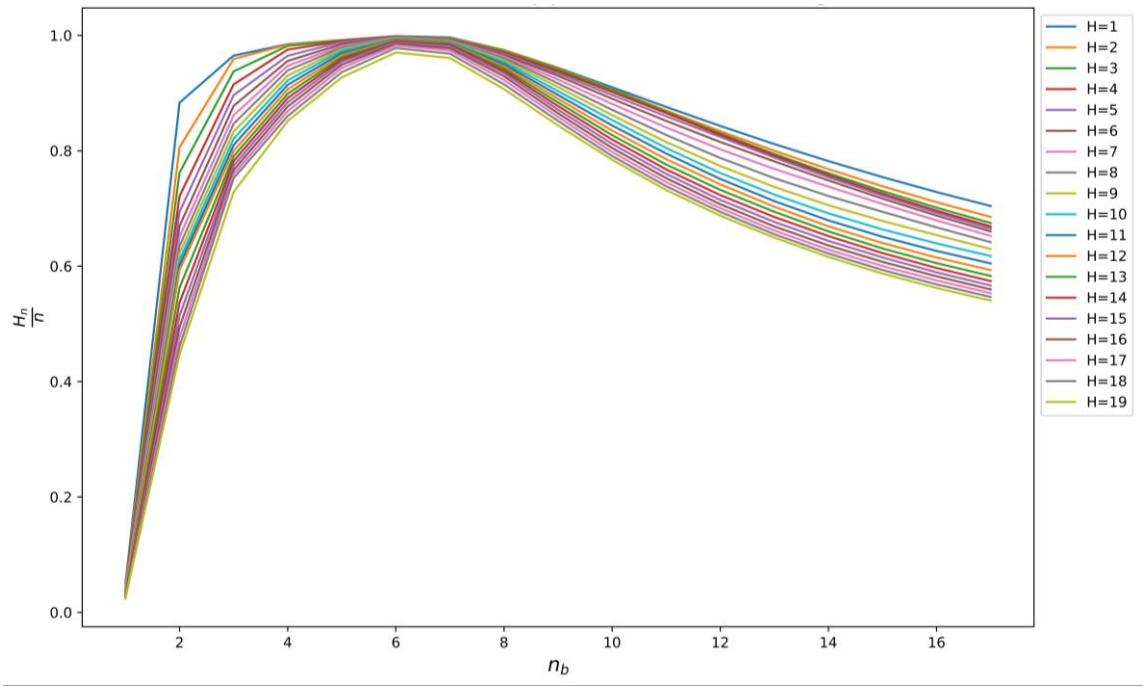
istaći da je ovaj postupak formiranja sekvence poznat potencijalnom napadaču, u skladu sa uobičajenim pretpostavkama u bezbednosnim modelima.

Dizajniran je posebni kvantizator koji uniformno kvantizuje sekvence na maksimalno 7 bita dok preostali biti kojim se opisuje odbirak predstavljaju deterministički deo i označavaju uvođenje dodatne korelacije.

Ukupan broj kvantizacionih bita kojim je određen proces transformacije ulaznog signala određuje se analizom *Shannon*-ove entropije koja se drugačije naziva i blok entropijom. Njena uopštena definicija data je sa (6.1). Za potrebe zavisnosti između uzastopnih simbola u binarnim sekvencama posmatra se uopšteni slučaj poznat kao n-blok entropija, H_n, koji se definiše kao

$$H_n = -\sum_{a \in \mathcal{A}^n} P(a) \log_2 P(a), \tag{6.1}$$

gde je $a = (x_1, x_2, \ldots, x_n)$ blok uzastopnih binarnih simbola definisani na alfabetu $\mathcal{A}^n$. Normalizovana blok entropija odnosi se na veličinu $\frac{H_n}{n}$, čija je asimptotska vrednost $\lim_{n\to\infty} \frac{H_n}{n}$ poznata kao brzina *Shannon*-ove ili blok entropije. U praksi se razmatra entropija konačne sekvence x dužine N. Ako se konačna sekvenca x smatra reprezentativnim izlazom nekog informacionog izvora, može se proceniti $P(a)$ na osnovu frekvencija obrazaca uočenih u x. Ako je x binarna sekvenca, frekvencije svih binarnih n-grama indukuju empirijsku raspodelu verovatnoća nad skupom svih mogućih binarnih nizova dužine n. Blok entropija reda n zatim se definiše na osnovu te raspodele, dok se normalizovana blok entropija dobija kao blok entropija podeljena sa n, čime se dobija entropija po bitu sekvence. Slika 16 prikazuje promenu normalizovane blok entropije analiziranog EEG izvora, kao funkciju od broja bitova po uzorku kvantizovanom uniformnim kvantizatorom. Ova funkcija je izračunata za vrednosti promene dužine bloka od 1 do 19 i za različit broj kvantizacionih bita po odbirku a čije su vrednosti uzimane u opsegu od 1 do 17.



Slika 16. Normalizovana blok entropija u zavisnosti od broja kvantizacionih bita.

Na prikazanom grafiku uočavaju se određeni trendovi. Sa povećanjem dužine bloka dolazi do do početnog porasta, a zatim smanjenja normalizovane blok entropije. Povećanje blok entropije u opsegu $n_b = [1,7]$, gde je n_b broj bitova po uzorku, odgovara boljem opisu informacionog sadržaja EEG izvora. Naknadno opadanje normalizovane blok entropije u opsegu $n_b = [8,16]$ može se tumačiti kao prekomerna kvantizacija, koja uvodi dodatnu redundantnost u primarni EEG izvor. Mnogi autori su

primetili [58], da prekomerna kvantizacija može povećati brzinu destilacije tajnih ključeva. Imajući u vidu navedeni fenomen, sistem je projektovan tako da za početak ispituje dve različite vrednosti kvantizacije: $n_b = 5$, što odgovara režimu grube kvantizacije, i $n_b = 10$ što predstavlja režim prekomerne kvantizacije, kako bi se ispitao njihov uticaj na ukupne performanse sistema.

U tabeli 3 prikazani su osnovni parametri EEG izvora koji podrazumevaju frekvenciju odabiranja, broj kanala (senzori ili metrike), dužinu sekvence u sekundama, broj kvantizovanih bita po odbirku i konačnu dužinu sekvence nakon kvantizacije i serijalizacije.

Tabela 3. Osnovni parametri EEG signala

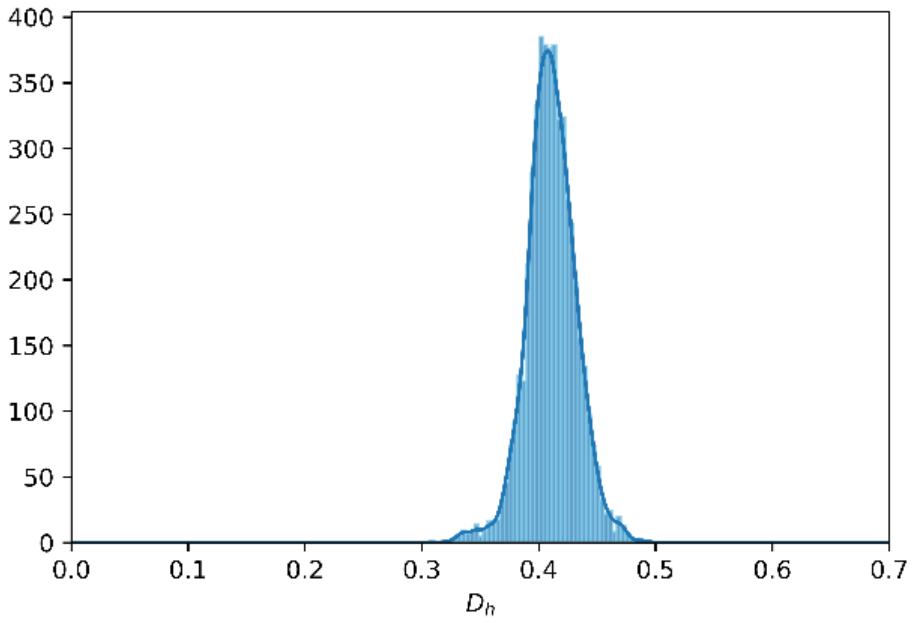
	Frekvencija odabiranja [Hz]	Broj Senzora/Metrike	Dužina [s]	n_b	Dužina [bit]
Sirovi EEG	128	14	2	5	17920
EEG metrike	2	6	300	5	18000
Sirovi EEG	128	14	2	10	35840
EEG metrike	2	6	300	10	36000

Na slikama 17 i 18 prikazani su histogrami normalizovanih Hamingovih rastojanja, D_h, svih parova ispitanika dobijenih na osnovu metrika EEG izvora za $n_b = 5$ i $n_b = 10$.

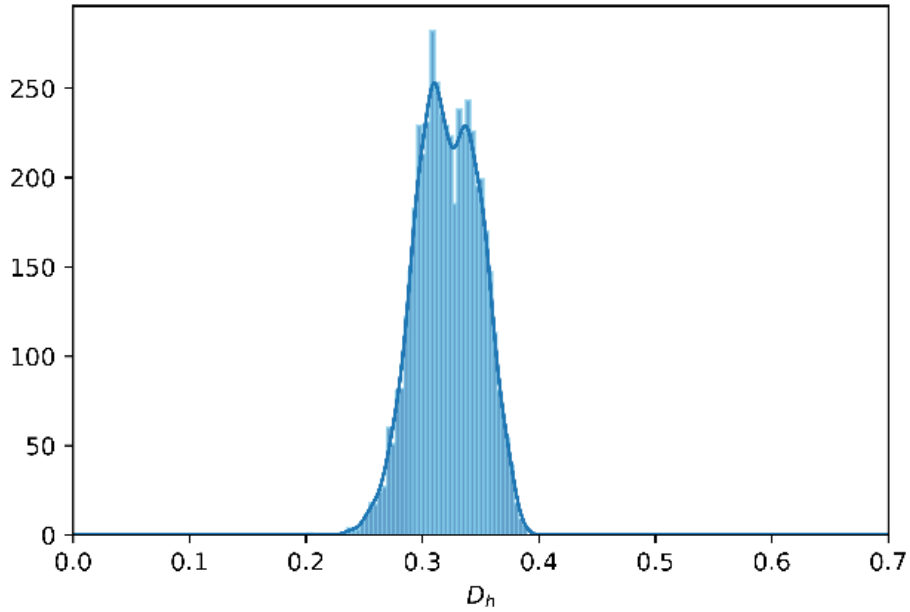
Normalizovana Hamingova rastojanja, prema definiciji, uzimaju vrednosti u intervalu od 0 do 1. U teoriji informacija, Hamingovo rastojanje veće od 0.5 ukazuje da su sekvence međusobno komplementarne, odnosno da su njihovi elementi suprotni (inverzni). Pošto je komplementarna sekvenca logički inverzna originalnoj, takva relacija ne označava veću različitost u informativnom smislu, već samo obrnutu reprezentaciju istog uzorka. Stoga se normalizovano Hamingovo rastojanje može smatrati periodičnom ili simetričnom merom u odnosu na tačku 0.5, pri čemu vrednost $D_h > 0.5$ ima ekvivalentnu interpretaciju kao $1 - D_h$. Ova interpretacija zasniva se na svojstvu

$$d(X, Y) + d(X, \bar{Y}) = n, \tag{6.2}$$

gde $d(X, Y)$ označava Hamingovo rastojanje između dve binarne sekvence dužine n, dok $\bar{Y}$ predstavlja komplement sekvence Y. Nakon normalizacije svojstvo postaje $D_h(X, Y) + D_h(X, \bar{Y}) = 1$, što objašnjava pomenutu ekvivalentnost vrednosti.



Slika 17. Histogram Hamingovih rastojanja između svih parova EEG sekvenci metrika za kvantizaciju $n_b = 5$. Srednja vrednost i standardna devijacija iznose, 0.4109 ± 0.0213.



Slika 18. Histogram Hamingovih rastojanja između svih parova EEG sekvenci metrika za kvantizaciju $n_b = 10$. Srednja vrednost i standardna devijacija iznose, 0.3218 ± 0.0269.

Treba se podsetiti da je u slučaju slučajnih i nezavisnih sekvenci, histogram njihovih međusobnih normalizovanih Hamingovih rastojanja usko centriran oko vrednosti 0.5. Da bi se uverilo u ovu tvrdnju, neka je S_i binarna slučajna promenljiva koja označava da li se dva binarna niza X i Y dužine n razlikuju na poziciji i. Ovih n slučajnih promenljivih su nezavisne, sa jednakom verovatnoćom za 0 i 1, tj., $P[S_i = 0] = P[S_i = 1] = \frac{1}{2}$. Iz linearnosti matematičkog očekivanja važi,

$$E[S_1 + S_2 + \cdots + S_n] = E[S_1] + E[S_2] + \cdots + E[S_n] = \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = \frac{n}{2}. \quad (6.3)$$

Posledično,

$$E[D_h(X,Y)] = \left(\frac{1}{n}\right) E[S_1 + S_2 + \cdots + S_n] = \left(\frac{1}{n}\right) \cdot \left(\frac{n}{2}\right) = \frac{1}{2}. \quad (6.4)$$

Poređenjem histograma na slikama 17 i 18, može se uočiti pomeranje prema manjim normalizovanim Hamingovim rastojanjima (tj. manjim razlikama između signala). Ovo ponovo pokazuje da prekomerna kvantizacija uvodi dodatnu korelaciju u ansambl realizacija primarnog izvora.

Prosečna stopa iskorišćenosti početnih sekvenci izražena preko stope destilovanih ključeva za sve tipove napadača koji su u radu [48] analizirani, iznosi 4,78% za 10-bitnu i 1,78% za 5-bitnu kvantizaciju, što predstavlja približno 2,7 puta bolji rezultat. Dodatno, procenat parova koji su usaglašeni dostiže 100% za 10-bitnu kvantizaciju. Prekomerna kvantizacija dakle, dovodi do izražajnije korelacije između legitimnih strana čime se povećava verovatnoća uspešnog usaglašavanja njihovih sekvenci nakon prolaska kroz faze SKD protokola. Ujedno, povećana je efikasnost u pogledu procenta iskorišćenosti početnih sekvenci, pri čemu se zadržava kriptografski kvalitet dobijenih ključeva. Dalja analiza i optimizacija faza SKD protokola biće izvršena na sekvencama kvantizovanim sa 10 bita.

6.1 Analiza performansi predloženog SKD sistema

U okviru analize performansi sprovedeno je testiranje AD algoritma, koji ima za cilj eliminaciju razlika između sekvenci legitimnih učesnika i povećanje razlike prema napadaču. Ovaj algoritam može samostalno da obezbedi usaglašavanje sekvenci. Međutim zbog velikog broja upita parnosti koji se šalju u okviru jedne iteracije, a koji su posledica podele na manje blokove, njegovo glavno ograničenje odnosi se na efikasnost. Naime, pri svakoj poslatoj blok parnosti odbacuje se jedan bit zarad očuvanja tajnosti, uz dodatno odbacivanje bita u slučaju kada je rezultat razmenjene parnosti neusaglašen. Stoga se u svakoj iteraciji eliminiše najmanje polovina njenih elemenata, što dovodi do značajnog smanjenja dužine finalne sekvence.

U radu [59] dat je izraz kojom se kvantifikuje verovatnoća greške između sekvenci A i B nakon i-te iteracije algoritma, u zavisnosti od početne verovatnoće greške:

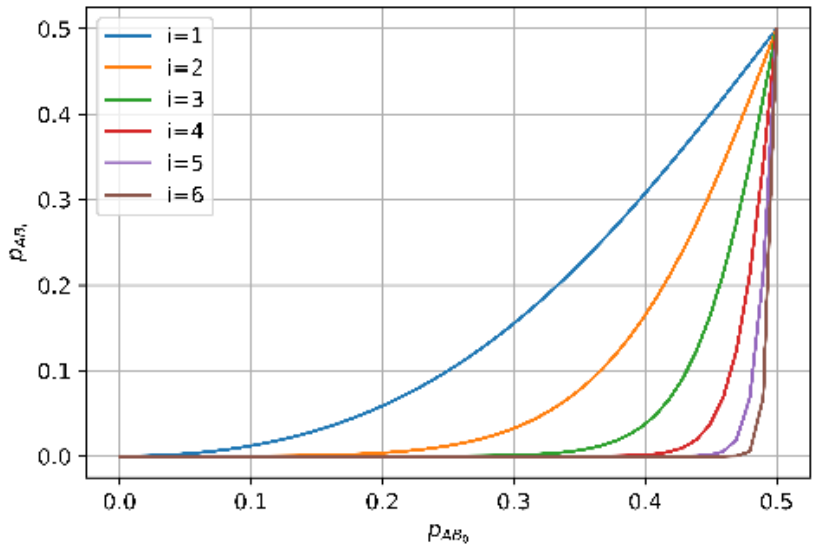
$$p_{AB_i} = \frac{(p_{AB_0})^{2^i}}{(p_{AB_0})^{2^i} + (1 - p_{AB_0})^{2^i}}, \tag{6.5}$$

dok se efikasnost zadržavanja podataka (*data remaining efficiency* [59]), tj. odnos između dužine preostalog niza i početnog niza može odrediti izrazom

$$eff = \frac{(p_{AB_0})^2 + (1 - p_{AB_0})^2}{2^i} \tag{6.6}$$

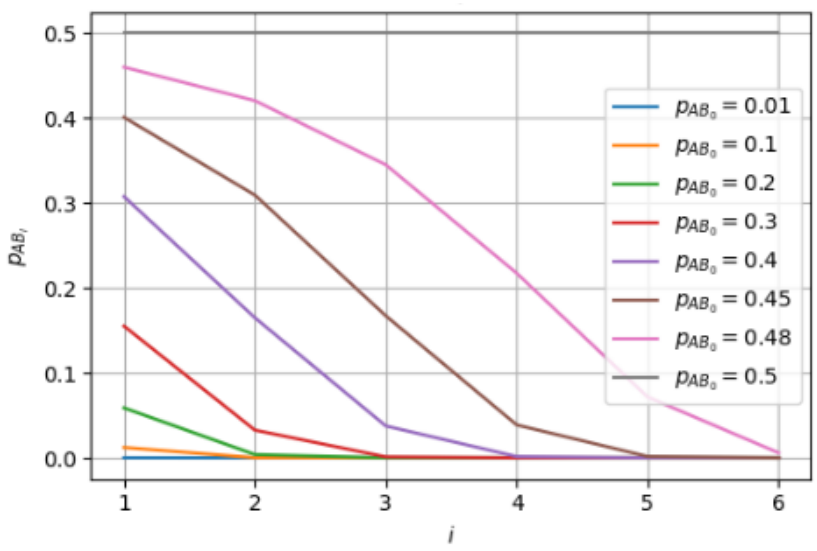
koji prikazuje koliko se podataka gubi sa svakom iteracijom algoritma.

Na slici 19 prikazan je grafik koji ilustruje kako se razlika između sekvenci menja tokom iteracija AD algoritma u zavisnosti od početne razlike između sekvenci.



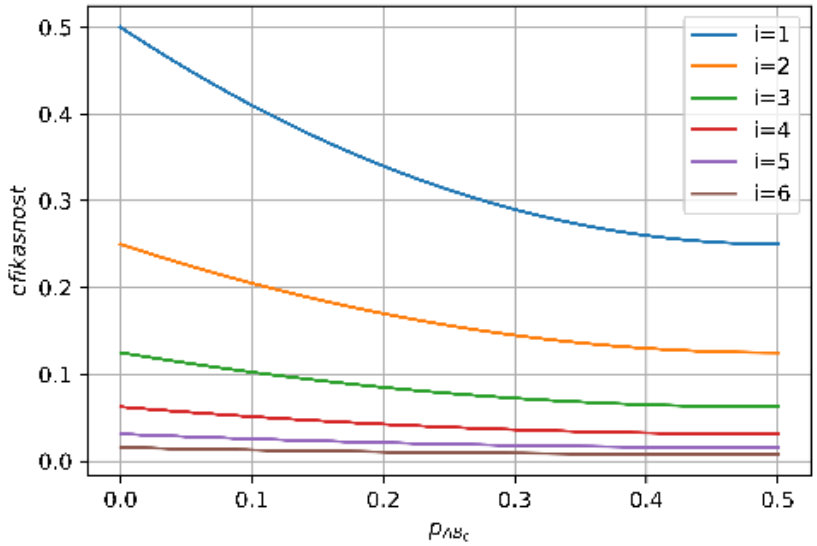
Slika 19. Zavisnost razlike između sekvenci od njihove početne razlike kroz iteracije AD algoritma.

Na slici 20 prikazan je grafik zavisnosti razlike između sekvenci tokom iteracija AD algoritma za različite početne razlike.



Slika 20. Zavisnost razlike između sekvenci od broja iteracija AD algoritma za različite početne razlike.

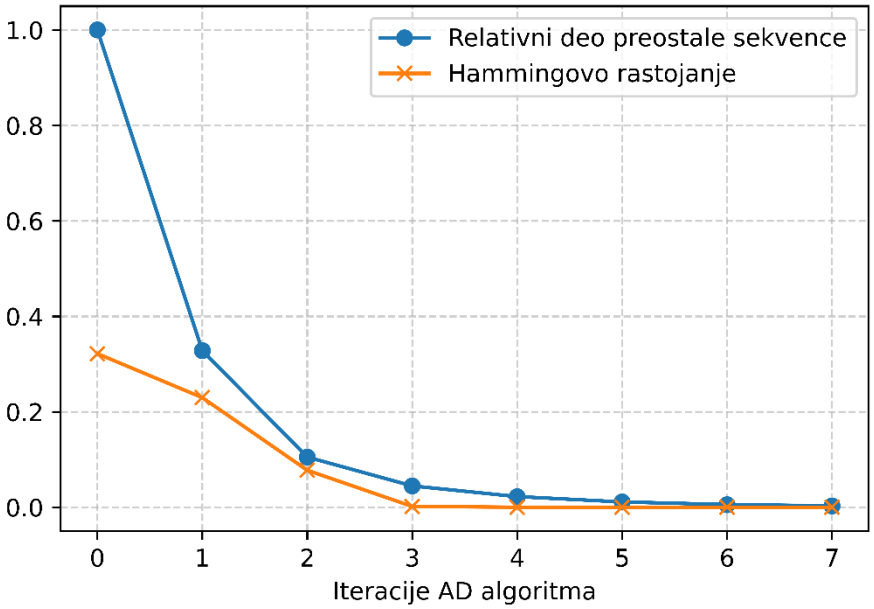
Na slici 21 prikazana je preostala efikasnost nakon iii iteracija AD algoritma u zavisnosti od početne razlike, za različit broj iteracija.



Slika 21. Preostala efikasnost u zavisnosti od početne razlike za različit broj iteracija AD algoritma.

Pri izboru broja iteracija neophodno je pronaći kompromis između efikasnosti, koja direktno utiče na brzinu generisanih kriptografskih ključeva, i grešaka koje se prenose u naredni IR blok sistema. Cilj je obezbediti što veću pouzdanost i što manji gubitak informacija, uz istovremeno očuvanje efikasnosti procesa generisanja zajedničkih ključeva.

Da bi se odredio optimalan broj iteracija AD algoritma, izvršeno je eksperimentalno testiranje kroz šest uzastopnih iteracija (za n_b=10), čiji su rezultati prikazani na slici 22.



Slika 22. Efikasnost AD algoritma kroz iteracije na skupu EEG signala metrika.

Relativni deo preostale sekvence dobijen je kao odnos dužine nakon izvršenih iteracija AD algoritma i njene inicijalne dužine. Na osnovu dobijenih rezultata utvrđeno je da se optimalne performanse postižu primenom dve iteracije AD algoritma. U ovom slučaju, razlika između sekvenci nakon primene algoritma dovoljno je mala da omogućava pouzdanu obradu u IR fazi rada sistema. Istovremeno, dužina nakon dve iteracije je redukovana ali značajno manje u poređenju sa dužinom nakon tri iteracije, što doprinosi očuvanju efikasnosti procesa generisanja kriptografskih ključeva.

Za EEG signale iz skupa za eksperimentalnu evaluaciju, sprovedena je analiza efikasnosti AD faze u dve iteracije kroz komparaciju propagacije Hamingovih rastojanja između sekvence Alise i Boba i između sekvenci Alise i Eve, Tabela 4.

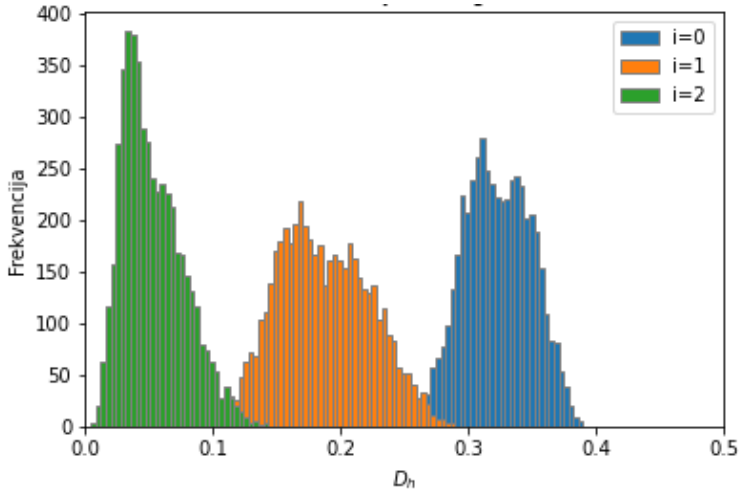
Tabela 4. Hamingova rastojanja između sekvence Alise i Boba i između sekvenci Alise i Eve kroz iteracije AD algoritma

	AB	AE
i=0	0.3218±0.0269	0.3200±0.0272
i=1	0.2296±0.0256	0.2290±0.0232
i=2	0.0774±0.0175	0.1826±0.0254

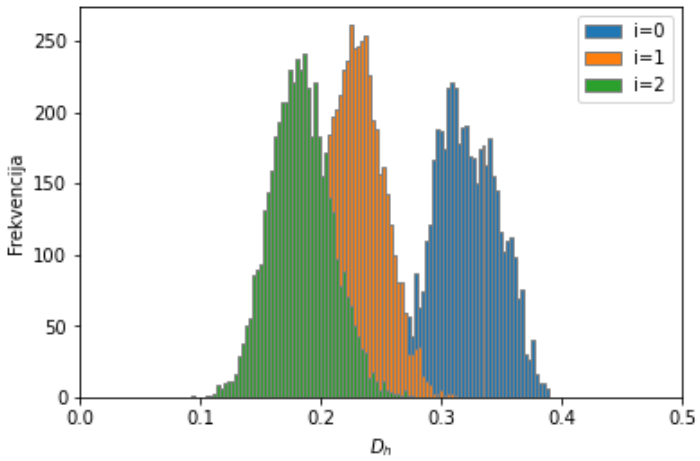
S obzirom na to da je svim učesnicima dostupan isti model izvora, očekivano je da se na početku eksperimenta Hamingova rastojanja između legitimnih sekvenci (Alise i Boba) i između sekvenci legitimnog korisnika i napadača (Alise i Eve) ne razlikuju značajno, odnosno da njihove statistike, srednja vrednost i standardna devijacija, budu približno jednake. Cilj primene AD algoritma jeste da otkloni sve eventualne prednosti koje potencijalni napadač može imati u odnosu na legitimne korisnike. Tokom razmene informacija putem javnog kanala, Alisa i Bob razmenjuju rezultate parnosti blokova, čime omogućavaju korekciju grešaka i usaglašavanje svojih sekvenci. Eva, s druge strane, iako ima pristup informacijama o parnosti, ne poseduje informaciju o stvarnim vrednostima bitova koji se zadržavaju, te je njena optimalna strategija ograničena na ponavljanje pokušaja procene, bez stvarne mogućnosti poboljšanja tačnosti svoje sekvence, što se može uočiti na histogramima. Nakon druge iteracije iz tabele se može videti da je rastojanje između sekvenci Alise i Boba manje od rastojanja između sekvenci Alise i Eve, što ukazuje na sticanje prednosti legitimnih korisnika. Sekvence Alise i Boba postaju visoko usaglašene, sa malim razlikama koje će u narednim blokovima

biti otklonjene, dok rastojanje između Alisine i Evine sekvence ostaje približno nepromenjeno. Time se potvrđuje da informacija kojom raspolaže napadač nije dovoljna da ugrozi bezbednost sistema, odnosno da AD algoritam uspešno povećava stepen zajedničke informacije između legitimnih korisnika, uz istovremeno ograničavanje mogućnosti treće strane da rekonstruiše ključeve.

Empirijska evaluacija efikasnosti AD algoritma nad SCR indukovanim EEG može se pratiti kroz evoluciju distribucije normalizovanih Hamingovih rastojanja tokom uzastopnih iteracija, slike 23 i 24. Početna iteracija (i=0), označena plavom bojom, predstavlja inicijalnu distribuciju normalizovanih Hamingovih rastojanja koja karakteriše sekvence primarnog SCR. Analizom srednjih vrednosti distribucija na kraju druge iteracije, označenih zelenom bojom na odgovarajućim dijagramima, zaključuje se da AD algoritam uspešno postiže željenu prednost i da inicijalna sekvenca napadača sa većom prednošću ne predstavlja nikakvo ograničenje u pogledu mogućnosti izdvajanje zajedničke informacije. Kvantitativno poređenje pokazuje da su Alisa i Bob uspostavili prednost nad Evom nakon primene dve iteracije ovog protokola. Uspostavljanjem prednosti omogućava se efikasno funkcionisanje narednih faza sekvencijalnog protokola.



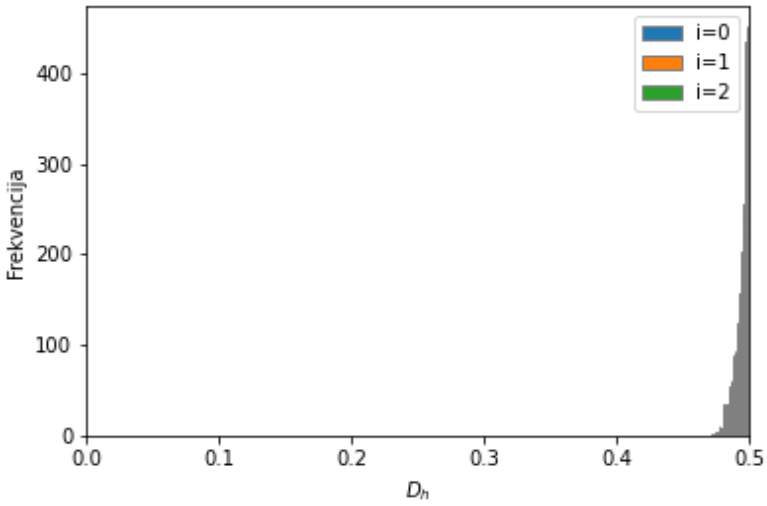
Slika 23. Histogram Hamingovih rastojanja između legitimnih sekvenci kroz iteracije AD algoritma (n_b=10).



Slika 24. Histogram Hamingovih rastojanja između legitimne sekvence i sekvence napadača kroz iteracije AD algoritma (n_b=10).

Izvršenom analizom u dve iteracije utvrđeno je da greške mogu biti grupisane, što povećava broj neophodnih iteracija u narednom koraku usaglašavanja. U cilju ujednačavanja rasporeda grešaka, kao i sprečavanja da napadač stekne dodatna znanja o sekvencama legitimnih korisnika, nad sekvencama legitimnih korisnika primenjena je unapred dogovorena permutacija koja nije poznata napadaču. U praktičnoj implementaciji, realizacija ovog koraka zahteva primenu kratkog zajedničkog ključa koji

služi kao seed za permutaciju. Na slici 25 grafički je predstavljen rezultat ostvarenih prednosti uz korišćenje početne permutacije. Ovim graficima je akcentovano početno dobijanje prednosti u odnosu na napadača kao i propagacija razlike između legitimnih sekvenci i približavanje usaglašenosti, odnosno vrednosti rastojanja $D_h = 0$.

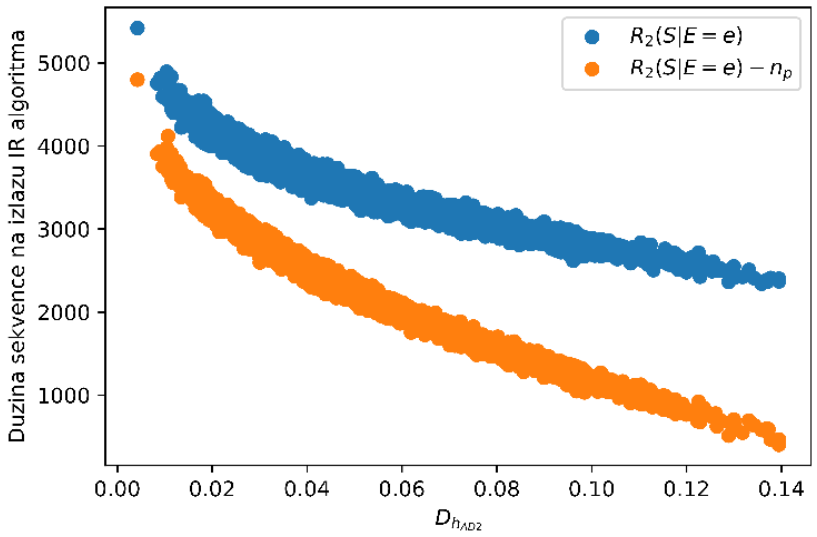


Slika 25. Histogram Hamingovih rastojanja između legitimnih sekvenci kroz iteracije AD algoritma, nakon primene početne permutacije (n_b=10).

Koristeći prethodno utvrđeni optimalan broj iteracija nastavlja se sa narednom IR fazom SKD sistema.

Prvo je testirano ponašanje kaskadnog algoritma. Algoritam se izvršava sve dok se legitimne sekvence ne usaglase, pri čemu je maksimalan broj iteracija ograničen na četiri. Efikasnost otklanjanja grešaka zavisi od razlike između sekvenci koje se usaglašavaju na ulazu u algoritam, kao i raspodele grešaka u okviru blokova na koje se sekvenca deli. Problem neuniformne raspodele grešaka otklonjen je mešanjem bita sekvence na početku svake iteracije kaskadnog algoritma. Izbor inicijalne dužine bloka kaskadnog algoritma predstavlja ključan parametar. U literaturi se preporučuju male dužine bloka, tipično reda veličine nekoliko bajtova, kako bi se povećala verovatnoća detekcije grešaka u ranim iteracijama kaskadnog algoritma [23,60,61]. Manja dužina bloka omogućava korekciju grešaka u manjem broju iteracija, ali istovremeno zahteva veći broj upita parnosti, što negativno utiče na efikasnost algoritma. Povećan broj upita parnosti utiče kako na vreme izvršavanja algoritma, tako i na konačnu dužinu sekvence, s obzirom da se svaki upit parnosti može ekvivalentno posmatrati kao kompromitovanje jednog bita sekvence. Uzimajući u obzir navedene faktore, izabrana je inicijalna dužina bloka od 10 bita kao kompromisno rešenje koje obezbeđuje ravnotežu između brzine konvergencije i efikasnosti algoritma.

Na slici 26 prikazana je maksimalna moguća dužina potencijalnih ključeva koja se može dobiti nakon AD i IR faze a koja je određena *Renyi*-jevom entopijom drugog reda. Sa $D_{h_{AD2}}$ označeno je Hamingovo rastojanje nakon druge iteracije AD algoritma, na ulazu u IR fazu. Prikazane su dve dužine, jedna, koja je direktno određena *Renyi*-jevom entropijom i druga, koja uzima u obzir informaciju o broju upita parnosti, n_p, oduzimajući je kako bi se simulirao broj bita koje je neophodno odbaciti u narednoj fazi algoritma kao potencijalno kompromitovane.



Slika 26. Uticaj razmenjenih upita parnosti na konačnu dužinu nakon kaskadnog algoritma.

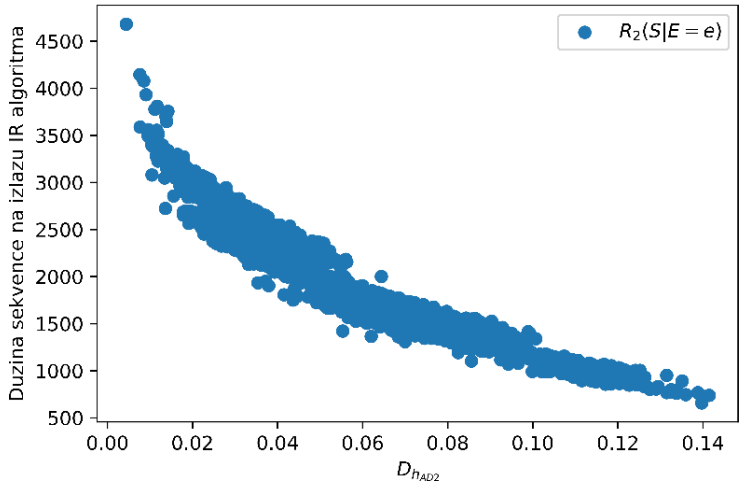
Sa grafika se uočavaju dve zakonitosti. Prvo, sa povećanjem razlika između sekvenci na ulazu kaskadnog algoritma opada vrednost *Renyi*-jeve entropije. Drugo, broj upita parnosti potrebnih za identifikaciju i ispravljanje greška raste sa povećanjem inicijalne razlike između sekvenci. Dakle, inicijalna razlika među sekvencama utiče na broj upita parnosti što direktno utiče na smanjenje konačne dužine tajnih ključeva koje je moguće izvući iz sekvenci.

Pored kaskadnog algoritma, u IR fazi primenjen je i *Winnow* algoritam. Osnovna odlika *Winnowog* algoritma jeste da se biti odbacuju tokom iteracija, za razliku od kaskade algoritma gde se greške samo identifikuju i ispravljaju, dok se odbacivanje vrši naknadno. U *Winnowom* algoritmu, za svaki upit parnosti ili slanje sindroma, eliminiše se onoliko bita koliko je tom prilikom potencijalno kompromitovano, što direktno utiče na dužinu sekvence na izlazu IR faze.

Algoritam je konfigurisan da se izvršava u maksimalno 10 iteracija, odnosno do trenutka dok se sekvence legitimnih učesnika ne usaglase. Izabrana veličina bloka iznosi 8 bita (2^3). Ovaj izbor motivisan je činjenicom da manji blokovi omogućavaju efikasniju detekciju grešaka. *Winnow* algoritam koristi Hamingov (8,4) kod koji može ispraviti jednu grešku po bloku, ili detektovati dve greške. Smanjenjem veličine bloka povećava se verovatnoća da blok sadrži najviše jednu grešku, što omogućava bržu konvergenciju ka usaglašenoj sekvenci i smanjuje ukupan broj potrebnih iteracija.

Kao i kod kaskadnog algoritma, raspodela grešaka u sekvenci ima značajan uticaj na efikasnost usaglašavanja. Zbog toga se na početku svake iteracije primenjuje permutacija koja omogućava uniformniju raspodelu grešaka, sprečavajući njihovo grupisanje u pojedinačnim blokovima i obezbeđujući konzistentniju performansu algoritma.

Na slici 27 prikazana je zavisnost konačne dužine sekvence na izlazu *Winnowog* algoritma u zavisnosti od Hamingovog rastojanja između sekvenci legitimnih korisnika na ulazu algoritma pri čemu treba uzeti u obzir da se potencijalno kompromitovani biti odbacuju u okviru samog algoritma.

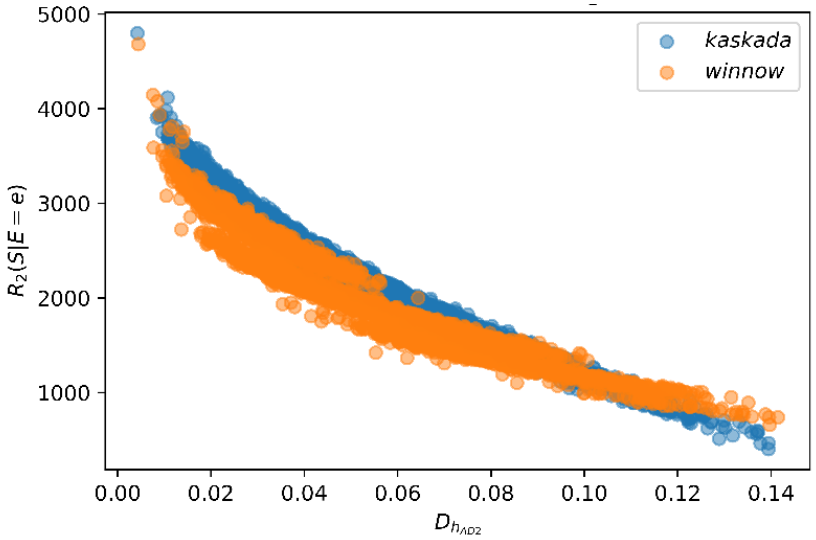


Slika 27. Uticaj razmenjenih upita parnosti na konačnu dužinu nakon *Winnowog* algoritma.

Sa priloženog grafika uočava se ista zakonitost kao i kod kaskadnog algoritma: sa povećanjem razlike između sekvenci na ulazu u IR algoritam, smanjuje se konačna dužina sekvence na izlazu, koja je predstavljena kroz Rényi-jevu entropiju drugog reda izračunatu u odnosu na sekvence napadača. Ova zakonitost je očekivana jer veća inicijalna razlika zahteva veći broj korekcija, što u slučaju *Winnowog* algoritma direktno vodi ka većem broju odbačenih bita.

U pogledu vremenske efikasnosti, *Winnow* algoritam pokazuje značajne prednosti u odnosu na kaskadni. Ova razlika je naročito izražena pri obradi dužih sekvenci, gde bi vremenska ušteda bila još izraženija. Ključni razlog za bolju vremensku efikasnost leži u tome što *Winnow* algoritam koristi linearne kodove (Hamingove kodove) koji omogućavaju direktnu korekciju grešaka putem sindroma, dok kaskadni algoritam zahteva binarnu pretragu unutar blokova sa nepoklapajućom parnošću. Dodatna prednost *Winnow* algoritma ogleda se u tome što se kompromitovani biti odmah odbacuju i ne prenose u sledeću iteraciju, čime se smanjuje veličina sekvence koja se obrađuje u narednim koracima i ubrzava proces konvergencije.

Međutim, treba napomenuti da *Winnow* algoritam, usled kontinualnog odbacivanja bita tokom iteracija, može rezultovati kraćom finalnom sekvencom u poređenju sa kaskadnim algoritmom pri istim inicijalnim uslovima. Ova karakteristika predstavlja kompromis između vremenske efikasnosti i očuvanja dužine sekvence, što treba uzeti u obzir pri izboru algoritma za konkretan sistem. Komparacija kvaliteta ova dva algoritma u okviru IR faze prikazana je na slici 28.



Slika 28. Uticaj razmenjenih upita parnosti na konačnu dužinu sekvence nakon IR algoritma.

Nakon faze usaglašavanja informacija, dobijena binarna sekvenca može sadržati određeni nivo statističke redundanse koji ne potiče iz algoritama SKD protokola, već iz procesa kvantizacije i binarizacije izvornog biometrijskog signala. Naime, inicijalna kvantizacija biometrijskog signala može proizvesti bitove sa neuniformnom raspodelom verovatnoća, pri čemu se javljaju lokalni šabloni koji smanjuju efektivnu entropiju sekvence. Iako se očekuje da se nakon IR faze sekvence legitimnih učesnika usklade, njihova statistička struktura ostaje nepromenjena, usled čega rezultujuće sekvence ne moraju imati maksimalnu entropiju.

U cilju eliminacije ove redundanse, nakon IR faze uvodi se blok koji koristi *Huffman*-ovo kodovanje da bi generisao optimalan prefiksni kod za zadatu raspodelu simbola, čime se minimalizuje prosečna dužina koda. Ovaj pristup eliminiše redundansu uvedenu tokom početne obrade signala i smanjuje količinu podataka koja se dalje obrađuje, postižući kompresiju blizu *Shannon*-ove donje granice. Ključna prednost je to što se kodna tabela generiše lokalno i simetrično na legitimnim stranama. Alisa i Bob rekonstruišu identičan koder iz iste zajedničke sekvence, pa nema potrebe za dodatnom razmenom pomoćnih podataka. S druge strane napadač, koji nema pristup istoj sekvenci, poseduje drugačiji koder i stoga dobija drugačiju transformaciju svojih podataka. Zbog ove asimetrije u lokalnom kodovanju, reprezentacija podataka kod legitimnih strana i napadača može se potpuno razlikovati. Njegova uloga se stoga svodi na predobradu sekvence, potencijalno smanjujući lokalne zavisnosti i menjajući distribuciju bita, čime se obezbeđuje pogodniji ulaz za PA fazu. Da bi sistem bio u potpunosti bezbedan potrebno je izmeriti sličnost među sekvencama i proceniti količinu bita koji se odbacuju da bi se garantovala informaciono-teoretska bezbednost konačnog ključa.

Broj bita po simbolu u *Huffman*-ovom kodovanju određuje se tako da obezbedi optimalnu ravnotežu između entropijske efikasnosti i složenosti kodera. Grupisanjem bitova u veće blokove povećava se razlika između verovatnoća simbola, što omogućava *Huffman*-ovom algoritmu efikasnost kodovanja čime se prosečna dužina koda približava teorijskoj granici određenoj *Shannon*-ovom entropijom. Na taj način, lokalna transformacija zadržava svu entropiju, ali ostvaruje bolju uniformnost raspodele i efikasnije korišćenje statističke strukture izvora, što predstavlja optimalan ulaz za fazu pojačanja privatnosti. Primena *Huffman*-ovog kodiranja omogućava efikasniju upotrebu preostale entropije u narednoj PA fazi, gde se finalna destilacija tajnih ključeva vrši primenom heš funkcija. Smanjenjem redundanse, ulazna sekvenca za PA fazu poseduje bolju uniformnost raspodele bitova, što direktno poboljšava kvalitet finalno destilovanih ključeva.

U nastavku je predstavljena eksperimentalna evaluacija SKD sistema zasnovanog na EEG SCR. Sproveden je niz eksperimenata koji ispituju potencijal EEG signala kao SCR u oviru SKD sistema. Testirana su tri različita sistema koja su zasnovana na osnovnom sistemu pri čemu je korišćen prethodno određen optimalan broj iteracija za BP AD algoritam. Sistemi se sastoje od sekvence blokova:

Sistem A: BP AD→*cascade*→*Universal Hash*

Sistem B: BP AD→*Winnow*→*Universal Hash*

Sistem C: BP AD→*Winnow*→*Huffman* kodovanje→*Universal Hash*

Obučavanje PIDNN mreže realizovano je primenom 10-tostruke unakrsne validacije na kompletnom skupu od 4950 formiranih tripleta. Prethodno je rečeno da je svaki triplet formiran tako da sadrži dve legitimne sekvence i jednu sekvencu napadača, nasumično izabranu iz skupa preostalih signala koji ne pripadaju posmatranom paru legitimnih korisnika. Za svaku iteraciju unakrsne validacije, predikcije donje granice izračunate su na odgovarajućim test skupovima. Ovakav pristup obezbeđuje da se, po završetku svih 10 rundi unakrsne validacije, dobiju predikcije donje granice *Renyi*-jeve entropije drugog reda prisluškivača za sve realizacije SKD procesa u posmatranom skupu podataka, uz očuvanje nezavisnosti između skupova za obučavanje i testiranje. Posmatrani skup podataka inicijalno je formiran od tripleta legitimnih korisnika i prisluškivača koji su prošli kroz faze SKD protokola, nakon čega su iz sekvenci legitimnih korisnika izdvojena obeležja korišćena kao ulaz modela, dok je odgovarajuća vrednost *Renyi*-jeve entropije drugog reda prisluškivača korišćena kao ciljna veličina za predikciju. Parametri algoritma za obučavanje izabrani su na osnovu preliminarnih

eksperimenata i preporuka iz literature za slične probleme. Broj epoha od 400 pokazao se dovoljnim za konvergenciju mreže na osnovu praćenja funkcije gubitka, pri čemu nisu uočeni znaci preobučavanja. Veličina *batch*-a od 32 izabrana je jer omogućava stabilnu konvergenciju dok je za optimizator izabran Adam optimizator sa stopom učenja 0.0005 kojim se postiže stabilan proces učenja. Izabrani parametri usvojeni su kao tipične vrednosti za ovaj tip regresionog problema sa neuralnim mrežama. Parametri PI algoritma (λ = 15, PICP = 0.95) postavljeni su u skladu sa standardnim postavkama koje obezbeđuju balans između širine intervala i pouzdanosti predikcije.

U okviru ovog nad sistemima A, B i C, korišćene su 4 različite PA strategije: strategija globalnog minimuma, optimalna strategija, strategija mašinskog učenja i hibridna strategija. Da bi se uporedile pojedinačne PA strategije, uvode se dva indikatora, dobitak i gubitak. Dobitak PA strategije SA u odnosu na PA strategiju SB definiše se na sledeći način.

$$G_{SB}^{SA} = \frac{|K_{SA}|}{|K_{SB}|}, \tag{6.7}$$

gde $|K_{SA}|$ i $|K_{SB}|$ označavaju ukupnu dužinu generisanih ključeva koristeći strategije SA i SB za iste ulazne sekvence S. Gubitak PA strategije SA se definiše kao

$$Loss_{SA} = \frac{|K_{Opt} - K_{SA}|}{K_{Opt}} \cdot 100\ [\%], \tag{6.8}$$

gde je K_{Opt} ukupna dužina generisanih ključeva koristeći optimalnu PA strategiju (12) za iste ulazne sekvence S.

Performanse sistema su merene sledećim indikatorima. PICP je dat sa (4.31), MPIW je definisan sa (4.32), R2 označava srednju vrednost, dok je R_{2min} minimalna vrednost ECRE2 preko cele populacije veličine L datog SCR

$$R2 = \frac{1}{L}\sum_{i=1}^{L} R_2(S_i|E = e_i), \tag{6.9}$$

$$R_{2min} = \min_i R_2(S_i|E = e_i). \tag{6.10}$$

Srednja vrednost donje granice L veličine ECRE2, može se dobiti iz PIDNN i označena je sa

$$R2_{ML} = \frac{1}{L}\sum_{i=1}^{L} L(F_i), \tag{6.11}$$

dok je srednja vrednost odgovarajuće donje granice L_{Hyb} iz hibridne PA strategije označena sa

$$R2_{Hyb} = \frac{1}{L}\sum_{i=1}^{L} L_{Hyb}(F_i). \tag{6.12}$$

Pokazatelj

$$G_{GLB\_R2min}^{Opt} = \frac{R2}{R_{2min}}, \tag{6.13}$$

predstavlja potencijalni dobitak u dužini generisanih ključeva pri primeni optimalne PA strategije u odnosu na strategiju zasnovanu na globalnoj donjoj granici R_{2min}. Slično,

$$G_{GLB\_R2min}^{ML} = \frac{R2_{ML}}{R_{2min}}, \tag{6.14}$$

predstavlja dobitak u dužini generisanih ključeva pri primeni PA strategije zasnovane na mašinskom učenju u poređenju sa standardnim postupkom zasnovanim na globalnoj donjoj granici R_{2min} za ECRE2, dok

$$G^{Hyb}_{GLB\_R2min} = \frac{R2_{Hyb}}{R_{2min}}, \tag{6.15}$$

predstavlja potencijalni dobitak u dužini generisanih ključeva pri primeni hibridne PA strategije u odnosu na strategiju zasnovanu na globalnoj donjoj granici R_{2min}. Konačno, pokazatelji

$$G^{Hyb}_{GLB\_R2\delta} = \frac{R2_{Hyb}}{R_{2\delta}}, \tag{6.16}$$

$$G^{Hyb}_{ML} = \frac{R2_{Hyb}}{R2_{ML}}, \tag{6.17}$$

predstavljaju potencijalni dobitak hibridne strategije u odnosu na globalnu strategiju i optimalnu PA strategiju zasnovanu na mašinskom učenju, respektivno.
Odgovarajući gubici

$$Loss_{GLB\_R2min} = \frac{|R2 - R_{2min}|}{R2} \cdot 100\ [\%], \tag{6.18}$$

$$Loss_{ML} = \frac{|R2 - R2_{ML}|}{R2} \cdot 100\ [\%], \tag{6.19}$$

$$Loss_{Hyb} = \frac{|R2 - R2_{Hyb}|}{R2} \cdot 100\ [\%]. \tag{6.20}$$

izražavaju procenat neiskorišćenosti datog SCR pri primeni globalne donje granice, PA strategije mašinskog učenja i hibridne strategije respektivno.

U Tabeli 5 prikazan je sveobuhvatan pregled svih indikatora performansi za sva tri analizirana sistema i oba SCR zasnovana na EEG signalima.

Tabela 5. Performanse sistema A, B i C analizirane sa oba tipa EEG signala

	A (cas-hash)		**B (win-hash)**		**C (win-huff-hash)**	
	Sirov	**Metrike**	**Sirov**	**Metrike**	**Sirov**	**Metrike**
$R > 0$ (%)	0.69 ± 0.51	0.08 ± 0.16	99.74 ± 0.48	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
PICP	0.9695 ± 0.0080	0.9822 ± 0.0114	0.9677 ± 0.0108	0.9883 ± 0.0058	0.9869 ± 0.0068	0.9927 ± 0.0022
MPIW	0.149 ± 0.024	0.085 ± 0.019	0.176 ± 0.041	0.095 ± 0.014	0.086 ± 0.036	0.070 ± 0.005
R2	1376.20 ± 20.27	1407.18 ± 11.12	819.71 ± 14.95	877.10 ± 9.69	1557.02 ± 33.26	1747.24 ± 17.71
R_{2min}	177.10 ± 3.44	845.01 ± 10.51	22.58 ± 0.05	370.62 ± 16.10	5.75 ± 0.02	631.10 ± 6.06
$R2_{ML}$	699.40 ± 88.19	975.24 ± 48.11	485.27 ± 39.16	665.64 ± 19.96	1543.30 ± 32.90	1734.75 ± 16.92
$G^{Opt}_{GLB\_R2min}$	7.77 ± 0.19	1.67 ± 0.02	36.31 ± 0.71	2.37 ± 0.08	270.55 ± 5.82	2.77 ± 0.04
$G^{ML}_{GLB\_R2min}$	3.96 ± 0.55	1.15 ± 0.06	21.49 ± 1.73	1.80 ± 0.08	268.17 ± 5.76	2.75 ± 0.04
$Loss_{GLB\_R2min}$ (%)	87.13 ± 0.33	39.95 ± 0.78	97.24 ± 0.05	57.75 ± 1.56	99.63 ± 0.01	63.88 ± 0.52
$Loss_{ML}$ (%)	49.17 ± 6.42	30.67 ± 3.80	40.77 ± 4.94	24.11 ± 2.09	0.88 ± 0.09	0.71 ± 0.06
k_{opt}	5.95 ± 4.73	0.08 ± 0.23	485.27 ± 39.16	665.64 ± 19.96	1543.30 ± 32.90	1734.75 ± 16.92

KR_{ML} (%)	0.0166 ± 0.0132	0.0002 ± 0.0006	1.35 ± 0.11	1.85 ± 0.06	4.30 ± 0.09	4.82 ± 0.05
KAR_{ML} (%)	0.69 ± 0.51	0.08 ± 0.16	99.74 ± 0.48	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
LR_{ML} (10^{-3})	6.38 ± 0.34	5.62 ± 0.02	4.82 ± 0.22	4.30 ± 0.10	0.74 ± 0.17	0.40 ± 0.03

Analiza dobijenih rezultata otkriva značajne razlike u efikasnosti razmatranih strategija. Globalna strategija pokazuje značajnu neiskorišćenost, što se manifestuje kroz parametar gubitka koji je sa izuzetno velikom vrednošću u slučaju sirovog EEG signala i iznosi preko 85% za sva tri sistema. Ovaj visoki procenat posledica je osetljivosti na autlajere koji snižavaju globalnu donju granicu i time podstiču neiskorišćenost sekvenci. Nasuprot tome ML strategija estimira donju granicu ECRE2 u posmatranom skupu podataka i time eliminiše problem autlajera, ujedno postižući veću iskorišćenost raspoloživog informacionog izvora. Delta parametar, izveden iz intervala pouzdanosti pokrivanja predviđanja (PICP), iznosi manje od 0.025 za sve analizirane sisteme. Prema Napomeni 2, ova veoma mala vrednost implicira da generisani ključevi poseduju gotovo maksimalnu entropiju sa Evine perspektive. Direktna posledica je izuzetno niska prosečna brzina curenja informacija, reda veličine 10−3 ili manja, za sve sisteme i tipove signala. Što se tiče komparativne analize samih sistema treći sistem postiže KR > 4% za oba tipa signala, čime je pokazana važnost uključivanja *Huffman*-ovog kodera koji se ponaša kao svojevrsna heš funkcija dodatno distancirajući Evinu sekvencu od legitimne što se ogleda kroz povećanje parametra R2 upoređujući vrednosti sistema B i C. Ako se pogleda uticaj IR algoritma može se zaključiti da *Winnow* algoritam u ovoj postavci prevazilazi kaskadni pristup zahvaljujući direktnom odbacivanju kompromitovanih bitova tokom iteracija.

Rezultati testiranja hibridne strategije dati su u Tabeli 6.

Tabela 6. Hibridna strategija, performanse sistema A, B i C analizirane sa oba tipa EEG signala

	A (cas-hash)		**B (win-hash)**		**C (win-huff-hash)**	
	Sirov	**Metrike**	**Sirov**	**Metrike**	**Sirov**	**Metrike**
$R > 0$ (%)	0.69 ± 0.51	0.08 ± 0.16	99.74 ± 0.48	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
PICP	0.9695 ± 0.0080	0.9822 ± 0.0114	0.9677 ± 0.0108	0.9883 ± 0.0058	0.9869 ± 0.0068	0.9927 ± 0.0022
MPIW	0.149 ± 0.024	0.085 ± 0.019	0.176 ± 0.041	0.095 ± 0.014	0.086 ± 0.036	0.070 ± 0.005
R2	1376.20 ± 20.27	1407.18 ± 11.12	819.71 ± 14.95	877.10 ± 9.69	1557.02 ± 33.26	1747.24 ± 17.71
R_{2min}	177.10 ± 3.44	845.01 ± 10.51	22.58 ± 0.05	370.62 ± 16.10	5.75 ± 0.02	631.10 ± 6.06
$R_{2\delta}$	469.18 ± 41.01	999.29 ± 51.71	59.25 ± 11.12	499.69 ± 19.96	37.89 ± 6.85	807.26 ± 25.46
$R2_{Hyb}$	727.03 ± 63.43	1031.60 ± 47.31	486.30 ± 39.10	668.29 ± 20.63	1543.37 ± 32.89	1735.04 ± 16.84
$G^{Hyb}_{GLB\_R2min}$	4.11 ± 0.41	1.22 ± 0.06	21.54 ± 1.73	1.81 ± 0.08	268.18 ± 5.76	2.75 ± 0.04
$G^{Hyb}_{GLB\_R2\delta}$	1.56 ± 0.18	1.03 ± 0.02	8.41 ± 1.18	1.34 ± 0.03	42.28 ± 8.72	2.15 ± 0.08
G^{Hyb}_{ML}	1.05 ± 0.06	1.06 ± 0.04	1.002 ± 0.001	1.004 ± 0.002	1.0000 ± 0.0000	1.0002 ± 0.001
$Loss_{Hyb}$ (%)	47.17 ± 4.63	26.68 ± 3.50	40.65 ± 4.93	23.81 ± 2.17	0.88 ± 0.09	0.71 ± 0.06
k_{opt}	5.95 ± 4.73	0.09 ± 0.23	486.30 ± 39.10	668.29 ± 20.63	1543.37 ± 32.89	1734.75 ± 16.92
KR_{hyb} (%)	0.0166 ± 0.0132	0.0003 ± 0.0006	1.36 ± 0.11	1.85 ± 0.06	4.30 ± 0.09	4.83 ± 0.05
KAR_{Hyb} (%)	0.69 ± 0.51	0.10 ± 0.16	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
LR_{Hyb} (10^{-3})	2.31 ± 0.93	0.93 ± 1.87	2.93 ± 0.04	3.24 ± 0.01	0.17 ± 0.09	0.47 ± 0.03
$R2_{Hyb} > R2$ (%)	2.83 ± 0.85	1.52 ± 1.09	3.03 ± 1.12	0.99 ± 0.57	1.39 ± 0.85	0.95 ± 0.43

Hibridna strategija ostvaruje neznatno bolje performanse u poređenju sa klasičnom ML strategijom, međutim razlike u kvalitetu i količini generisanih ključeva ostaju zanemarljive, dok se istovremeno povećava ukupna kompleksnost sistema. Relativni dobitak u performansama, izražen kroz parametar

dobitka, je reda veličine 10^{-2} za sistem A, 10^{-3} za sistem B i 10^{-4} za sistem C, što ukazuje na minimalan značaj poboljšanja. Sa stanovišta praktične implementacije, ECRE2 strategija se stoga izdvaja kao optimalniji izbor, jer obezbeđuje jednostavniju realizaciju uz gotovo identične performanse.

Istovremeno, obe razmatrane strategije ostvaruju značajno poboljšanje u odnosu na ranije korišćenu strategiju globalnog minimuma, uz očuvanje kvaliteta generisanih ključeva i minimalno curenje informacija ka napadaču. Posmatrano kroz ukupne performanse, sistem C ubedljivo nadmašuje sisteme A i B, pri čemu za oba tipa izvora postiže vrednosti KR > 4%, KAR = 100%, LR < 10^{-3} i *Loss* < 1%. Brzina generisanja ključeva je stoga ~800 b/s za 14-to kanalni sirovi EEG signal i ~200 b/s za signal okarakterisan metrikama EEG-a.

Tabela 7 i Tabela 8 sumiraju rezultate statističkih testova slučajnosti primenjenih na sve generisane sekvence ključeva. Testiranje slučajnosti je sprovedeno korišćenjem statističkog paketa testova razvijenog od strane Nacionalnog instituta za standarde i tehnologiju SAD (NIST) [62]. Svaki test ispituje specifičnu osobinu sekvence, a rezultat se izražava kroz p-vrednost, na osnovu koje se procenjuje da li sekvenca pokazuje ponašanje očekivano od slučajnog procesa. Test se smatra uspešnim ukoliko je dobijena p-vrednost veća od praga od 0.01.

Korišćeni su sledeći testovi:

- Test frekvencije (F - *frequency*) ispituje ravnotežu između broja nula i jedinica u celokupnoj sekvenci, čime se proverava osnovna uniformnost raspodele bitova.
- Test blok frekvencije (BF - *block frequency*) proširuje prethodni test analizom učestalosti bitova unutar manjih podsekvenci, čime se detektuju lokalna odstupanja od ravnomerne raspodele.
- Test ciklusa (R - *runs*) analizira raspodelu uzastopnih nizova identičnih bitova, odnosno učestalost promena između nula i jedinica.
- Test najdužeg ciklusa (LR - *longest run*) razmatra dužinu najdužeg neprekidnog niza identičnih bitova u sekvenci i poredi je sa očekivanim vrednostima za slučajne nizove.
- Spektralni test zasnovan na brzoj Furijeovoj transformaciji (FFT - *fast Fourier transformation*) koristi frekvencijsku analizu kako bi se identifikovale eventualne periodičnosti ili ponavljajući obrasci u sekvenci.
- Test serije (S - *serial*) ispituje učestalost svih mogućih preklapajućih obrazaca bita određene dužine, čime se procenjuje stepen zavisnosti između susednih bitova.
- Test aproksimativne entropije (AE - *approximate entropy*) meri složenost sekvence poređenjem frekvencija kraćih i dužih obrazaca bita, pri čemu veća entropija ukazuje na veću nepredvidivost.
- Test kumulativnih suma u smeru unapred (CSf - *cumulative sums forward*) prati odstupanje parcijalnih suma bitova od nulte vrednosti u toku sekvence, čime se ispituje globalna pristrasnost.
- Test kumulativnih suma u obrnutom smeru (CSr - *cumulative sums reverse*) predstavlja varijantu prethodnog testa, ali sa analizom sekvence u obrnutom redosledu, što omogućava dodatnu proveru simetrije odstupanja.

Tabela 7. Rezultati testa slučajnosti svih sekvenci ključeva generisanih pomoću ML strategije

		F	**BF**	**R**	**LR**	**FFT**	**S**	**AE**	**CSf**	**CSr**
A	Sirov	0.8165	0.4763	0.2544	0.4033	0.8535	0.1678	0.2183	0.4810	0.8636
	Metrike	0.3894	0.7751	0.4921	0.7234	0.7573	0.0846	0.8964	0.4783	0.5285
B	Sirov	0.8341	0.6606	0.4260	0.3764	0.8007	0.4116	0.4561	0.5415	0.6420
	Metrike	0.7106	0.2831	0.8854	0.7855	0.7748	0.3496	0.6214	0.8837	0.8560
C	Sirov	0.0983	0.3002	0.2403	0.6595	0.1543	0.4404	0.0790	0.0961	0.1360
	Metrike	0.3128	0.4011	0.7990	0.4171	0.7283	0.2780	0.5161	0.5815	0.5472

Tabela 8. Rezultati testa slučajnosti svih sekvenci ključeva generisanih pomoću PA hibridne strategije.

		F	BF	R	LR	FFT	S	AE	CSf	CSr
A	Sirov	0.1725	0.7343	0.3214	0.7653	0.2431	0.1347	0.5401	0.1972	0.1138
	Metrike	0.1874	0.2010	0.2874	0.1755	0.2746	0.4959	0.1305	0.3796	0.0949
B	Sirov	0.2140	0.3147	0.6757	0.3184	0.5932	0.2434	0.8971	0.1894	0.5623
	Metrike	0.2542	0.6322	0.0901	0.0993	0.2472	0.3020	0.4013	0.4091	0.3603
C	Sirov	0.4029	0.4452	0.2576	0.4368	0.6845	0.5021	0.3357	0.8104	0.6370
	Metrike	0.6111	0.2617	0.7254	0.7069	0.7147	0.3457	0.0576	0.6347	0.4807

Na osnovu prikazanih rezultata, može se zaključiti da sve generisane sekvence ključeva zadovoljavaju definisane kriterijume slučajnosti u svim sprovedenim testovima.

Analiza postojećih SKD pristupa pokazuje značajne razlike u ostvarivim brzinama generisanja tajnih ključeva u zavisnosti od korišćenog modela.

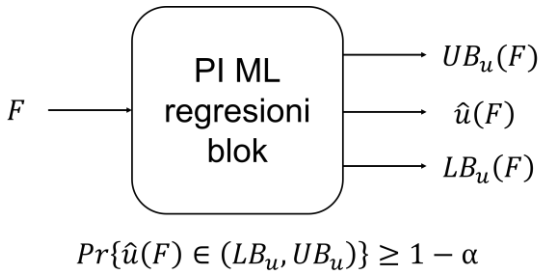
Sistemi zasnovani na kanalnom modelu koriste bežični komunikacioni kanal kao izvor zajedničke slučajnosti i mogu ostvariti znatno veće brzine, koje se, prema dostupnim pregledima, kreću od nekoliko stotina milibita u sekundi do približno 1.8 kb/s. Međutim, najviše prijavljene brzine u ovom modelu često su praćene povišenom stopom neusaglašenosti ključeva, kao i snažnim pretpostavkama o ograničenju slobode napadača [63,64], što dovodi u pitanje robusnost i praktičnu pouzdanost generisanih kriptografskih ključeva u realnim scenarijima.

U okviru modela izvora, gde se zajednička slučajnost posmatra kao inherentno svojstvo signala dostupnih legitimnim učesnicima, ostvarene brzine su uglavnom niske. Na primer, u mobilnim bežičnim mrežama gde se rastojanje između legitimnih čvorova koristi kao izvor zajedničke slučajnosti, postignute brzine generisanja ključeva kreću se u opsegu od 0.1 b/s do 0.6 b/s, u zavisnosti od brzine terminala i pozicije prisluškivača [49,50]. Nasuprot tome, prethodnim eksperimentom i dizajniranim SKD sistemom pokazan je potencijal modela izvora zasnovanih na EEG signalima, za koje su zabeležene znatno veće brzine.

U tom kontekstu, govorni signal se nameće kao perspektivan izvor zajedničke slučajnosti koji potencijalno omogućava prevazilaženje navedenih ograničenja. Kao prirodan, informaciono bogat i lako dostupan biometrijski signal, govor pruža osnovu za ostvarivanje znatno viših brzina generisanja tajnih ključeva u okviru izvornog modela. Time se omogućava povećanje bezbednosti kriptografskih primitiva zasnovanih na zajedničkom tajnom ključu, uz istovremeno smanjenje zavisnosti njihove bezbednosti od pretpostavki o računarskim mogućnostima napadača (veće dužine zajedničkih tajnih ključeva povećavaju donju granicu računarskih resursa napadača). Stoga se u nastavku analizira novi koncept SKD sistema sa većim brzinama generisanja ključeva zasnovanog na SCR nad govornim signalima učesnika protokola.

U daljoj analizi korišćen je prethodno definisan SKD sistem, uz minimalne modifikacije. PA blok je implementiran korišćenjem slučajnih binarnih matrica dimenzije n × r, koje imaju Toeplitz strukturu. Toeplitz heš funkcije su posebno pogodne za govorne signale, imajući u vidu da su tipične vrednosti za n reda 10^4.Za projektovanje PA bloka, neophodno je odrediti izlaznu dimenziju r univerzalne familije heš funkcija $G:\{0,1\}^n \rightarrow \{0,1\}^r$. Teorema 4 pokazuje da za ovu svrhu nije potrebna direktna predikcija $D_H(S_i, e_i)$, već samo njegova donja granica, koja važi sa verovatnoćom najmanje 1−δ.

ML blok prikazan na slici 29 daje tri različita izlaza. Pored stvarne vrednosti $u = D_H(S, e)$, on daje i vrednosti intervalnih granica, tako da se stvarna vrednost nalazi unutar tog intervala sa visokom verovatnoćom većom od $1 - \alpha$. U skladu sa Teoremom 4, tada se koristi LB_u tog intervala kao procenu *Spoiling Knowledge* donje granice za ECRE2. Ako važi $P\{\hat{u} \leq LB_u\} = P\{\hat{u} \geq UB_u\}$, sledi da je $\delta = \frac{\alpha}{2}$.



Slika 29. PIDNN blok sa predikcionim intervalima za procenu Hamingovog rastojanja i indirektnu procenu ECRE2. Adaptirano iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Stepen kompresije PA bloka je određen dimenzijom r Toeplitz heš matrice i dobija se na osnovu PIDNN izlaza, prema izrazima

$$R_{2Spoil}(LB_u, \delta) = nh\left(\frac{LB_u}{n}\right) - \log_2 \sqrt{2n} \tag{6.21}$$

$$k(F) = R_{2Spoil}(LB_u(F), \delta) - t. \tag{6.22}$$

Teorema 5 i Napomena 5 garantuju da su tajni ključevi tako generisani apsolutno tajni i imaju maksimalnu neizvesnost.

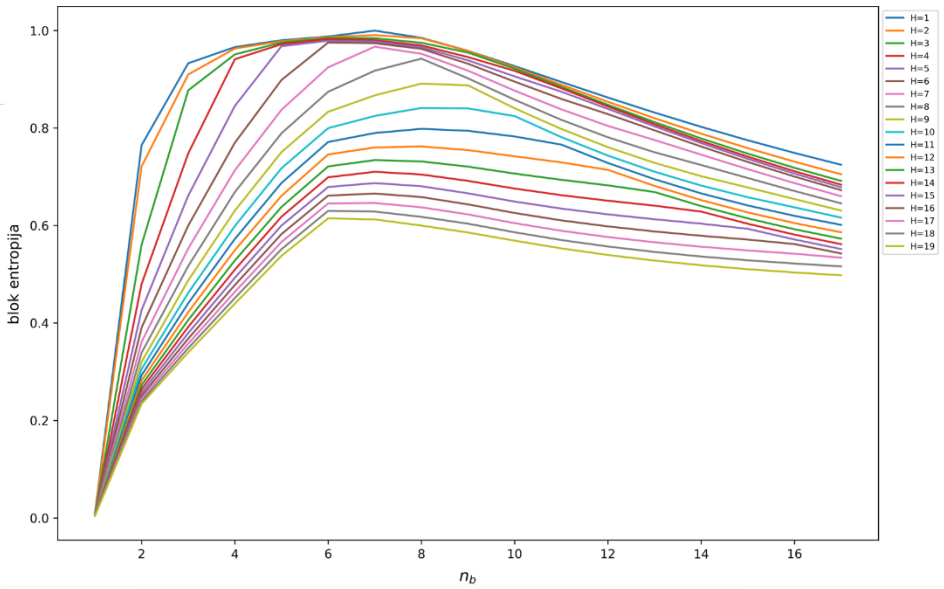
U daljnjoj analizi, razmatraju se dve varijante predloženog SKD sistema: bez bloka *Huffman*-ovog kodovanje (sistem A) i sa blokom *Huffman*-ovog kodovanja (sistem B). U sistemu B, *Huffman*-ovo kodovanjem se vrši nad izvornim alfabetom veličine 2^8 [43]. Preciznije, sistemi se sastoje od niza blokova:

Sistem A: Govor → Permutacija → BP AD → *Winnow* → *Toeplitz Hash* → Tajni ključevi

Sistem B: Govor → Permutacija → BP AD → *Winnow* → *Huffman* kodovanje → Toeplitz Hash → Tajni ključevi

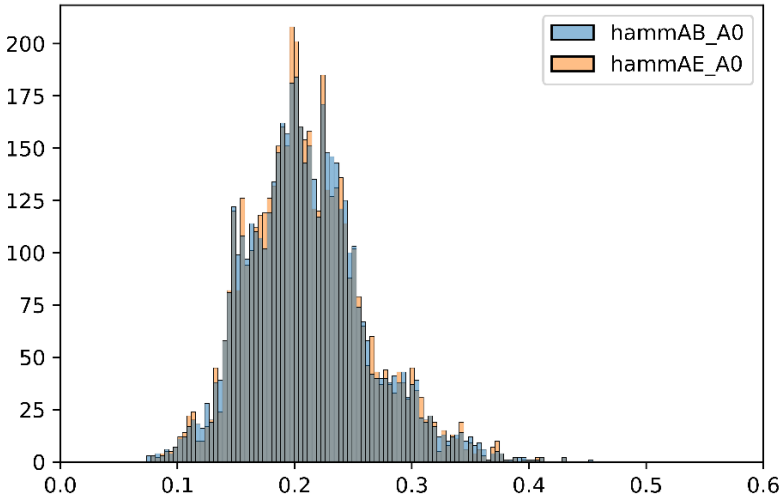
U odeljku 5 opisan je način formiranja skupa od 4950 tripleta govornog signala koji su podležu daljoj obradi. Tokom dizajna SKD sistema zasnovanih na govornom modelu izvora, predprocesiranje i analogno digitalna konverzija značajno određuju korelaciona svojstva rezultujućeg izvora SCR (XYZ, P_{XYZ}) i otuda, maksimalnu brzinu tajnih ključeva. Pri odabiru pogodne vrednosti broja bitova po uzorku, dominantan kriterijum treba da bude brzina generisanih tajnih ključeva.

Sledeći predlog 5.6. iz [65], preporučuje se izbor najjednostavnije moguće procedure kvantovanja ako se ne nameće ograničenje na brzinu prenosa na javnom kanalu. Pošto se u ovom radu ne razmatraju ograničenja brzine prenosa podataka preko javnog kanala, izabrana su skalarna uniformna kvantovanja. Brzina izvlačenja tajnih ključeva može rasti sa prekomernim kvantovanjem, stoga je odlučeno da se dizajnira sistem koji radi sa $n_b = 16$, tj. u režimu prekomernog kvantovanja. Kako su uzorci kodirani kao linearni 16-bitni jednokanalni PCM (*Pulse Code Modulation*) i pri čemu svaka reč traje 1s, sa ovim izborom kvantizacije dobija se dužina početnih sekvenci 16.000×1×16 = 256.000 bita. Sa grafika na slici 30, očigledno je da će povećanjem broja bitova po uzorku, brzina blok entropije najpre porasti a zatim opadati (kao i kod EEG-a). Već je pomenuto da brzina izvlačenja tajnih ključeva može rasti sa prekomernim kvantovanjem. Stoga je odlučeno da se u nastavku razmatra sistem koji radi sa $n_b = 16$, tj. u režimu prekomernog kvantovanja.



Slika 30. Uticaj broja kvantizacionih bita na blok entropiju kod govornih signala. Preuzeto iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

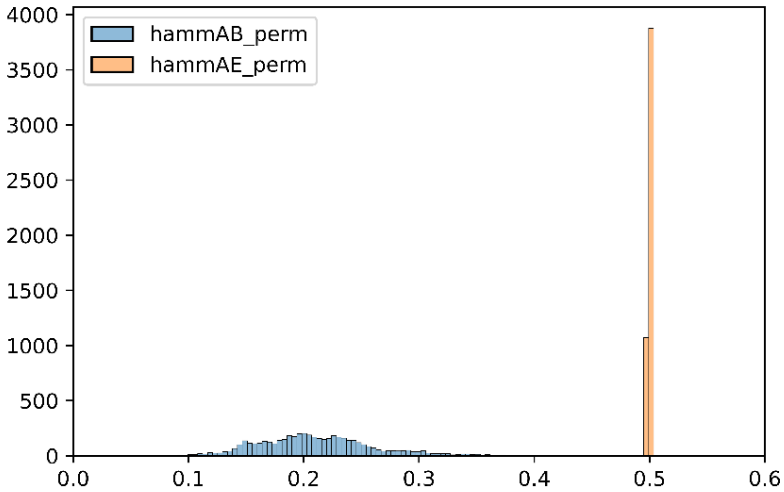
Normalizovana Hamingova rastojanja svih parova legitimnih učesnika protokola, Alise i Boba, kao i rastojanja, Alise i Eve, prikazana su na slici 31. Treba podsetiti da je Eva izabrana iz skupa preostalih učesnika na slučajan način. U pogledu evaluacije bezbednosnih aspekata ovog protokola, ovaj izbor Eve odgovara napadu iznutra, što se može smatrati najgorim slučajem sa stanovišta legitimnih učesnika. Stoga su izvršena analiza i bezbednosni parametri pouzdani pokazatelji praktične bezbednosti celog sistema.



Slika 31. Histogram normalizovanih Hamingovih rastojanja između Alise i Boba i Alise i Eve na početku. Preuzeto iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Rezultujuća brzina tajnih ključeva će biti mnogo veća kada je Evina sekvenca manje korelisana sa sekvencama Alise i Boba. Stoga, ako Alisa i Bob primene istu slučajno izabranu permutaciju na svoje sekvence, kao i kod EEG signala, njihovo normalizovano Hamingovo rastojanje će ostati nepromenjeno, dok će matematičko očekivanje rastojanja Evine sekvence do ovih permutovanih sekvenci biti 0.5.

Drugim rečima, ako legitimni učesnici protokola primene istu slučajnu permutaciju, o kojoj Eve nema informacija, doći će do potpune dekorelacije između njene sekvence sa sekvencama Alise i Boba. Ova situacija je prikazana na slici 32.



Slika 32. Histogram normalizovanih Hamingovih rastojanja između Alise i Boba i Alise i Eve nakon primene permutacije. Preuzeto iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Prednost dekorelacije se posmatra u svim fazama SKD-a. Evolucija normalizovanih Hamingovih rastojanja u prvoj i drugoj iteraciji primenjenog BP AD-a predstavljena je kroz Tabelu 9 u kojoj su prikazane srednje vrednosti i standardne devijacije rastojanja na početku (AD0), nakon početne permutacije (perm), nakon prve i druge iteracije AD algoritma (AD1 i AD2) kao i nakon IR algoritma (IR).

Tabela 9. Statistički parametri Hamingovih rastojanja parova sekvenci Alise i Boba i Alise i Eve u različitim fazama SKD-a.

	AD0	**perm**	**AD1**	**AD2**	**IR**
AB	0.211 ± 0.051	0.211 ± 0.051	0.074 ± 0.044	0.010 ± 0.017	0.000 ± 0.000
AE	0.211 ± 0.050	0.499 ± 0.001	0.500 ± 0.002	0.500 ± 0.003	0.499 ± 0.003

Vidi se direktni uticaj permutacije na statističke parametre, dok vrednosti normalizovanih Hamingovih rastojanja prema Evi ostaju 0.5, rastojanja sekvenci između Alise i Boba brzo opadaju. Očekivana vrednost rastojanja nakon druge iteracije BP AD algoritma je oko 0.01, što omogućava veoma efikasnu primenu *Winnow* protokola u IR fazi [59]. Nakon primene *Winnow* protokola, svi parovi legitimnih sekvenci usaglašeni (normalizovano Hamingovo rastojanje je jednako 0), dok je Eve ostala na rastojanju od 0.5.

Za generisanje permutacije, potreban je početni kratak tajni ključ prethodno razmenjen između Alise i Boba. Pošto osnovnom SKD protokolu prethodi autentifikacija javnog kanala, to jest, autentifikacija legitimnih korisnika, ovaj početni tajni ključ može biti lokalno generisan na osnovu kriptografskih ključeva korišćenih u fazi autentifikacije [66,67].

Druga mogućnost je sinteza autonomnog SKD sistema koji ne zavisi od faze autentifikacije korisnika. Prvo se destiluje kratak tajni ključ preko originalnih (nepermutovanih) signala sa nižom brzinom ključa, a zatim sledi glavna faza destilacije sa permutovanim signalima.

Obeležja IT 4, IT 5 i IT 6 zahtevaju razmenu informacija između Alise i Boba preko javnog kanala, što predstavlja dodatni komunikacioni trošak i potencijalnu bezbednosnu ranjivost. Stoga je odlučeno da se ova tri obeležja zamene novim informaciono-teorijskim obeležjima koja se mogu lokalno izračunati. Konkretno, uvedena su tri dodatna obeležja zasnovana na normalizovanoj blok entropiji izračunatoj na usaglašenoj sekvenci pre faze pojačanja privatnosti, za veličine blokova 2, 14 i 20, Tabela 10. Prethodno je ovo obeležje razmatrano za veličinu bloka 8, dok prošireni skup omogućava sveobuhvatniju karakterizaciju entropijskih svojstava na različitim skalama. Time informaciono-teorijska obeležja formiraju skup od 14 obeležja.

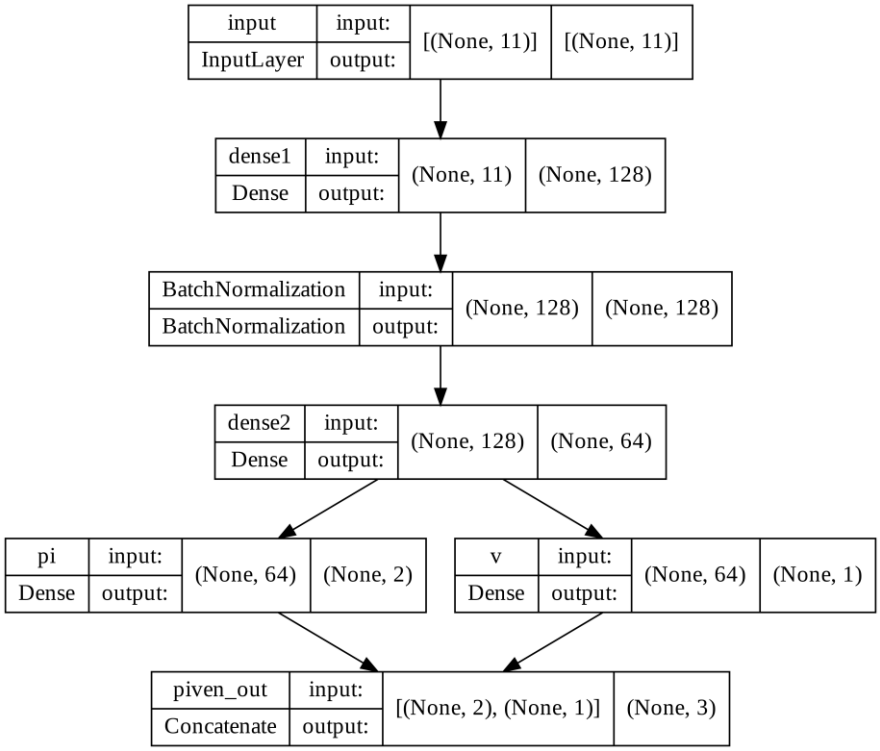
Tabela 10. Dodatna informaciono-teorijska obeležja

IT 12 - IT 14	Normalizovana blok entropija usaglašene sekvence pre faze pojačanja privatnosti (veličine bloka = 2,14,20)

Pored navedenog skupa informaciono-teorijskih obeležja, u eksperimentu će biti uključena i sva stilometrijska obeležja definisana u Tabeli 2, koja će zajedno sa odabranim informaciono-teorijskim obeležjima činiti ulazni skup karakteristika za dalju analizu.

Rangiranje karakteristika iz kombinovanog skupa ovih karakteristika izvršeno je računanjem njihovih SHAP (*SHapley Additive exPlanations*) vrednosti [68,69]. *Shapley* vrednosti imaju svoje poreklo u kooperativnoj teoriji igara. Zbog svoje teorijske osnove i praktične upotrebljivosti, Shapley vrednosti postaju dominantan metod korišćen u mašinskom učenju za evaluaciju i rangiranje važnosti karakteristika.

Arhitektura korišćenog PIDNN-a je prikazana na slici 33 prema notaciji usvojenoj u Keras API [68]. Prvi sloj je *dense* sloj sa 128 neurona. Nakon primene *batch* normalizacije, sledi drugi *dense* sloj sa 64 neurona. Konačno treći *dense* sloj ima ukupno 3 neurona koja predstavljaju izlaze mreže. Izlazi su organizovani u dva bloka i to dva izlaza u prvom bloku, koji predstavljaju gornju i donju granicu vrednosti koja se predviđa, i jednim izlazom u drugom bloku, koji predstavlja vrednost koja se predviđa.



Slika 33. Arhitektura predloženog PIDNN modela. Adaptirano iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

U fazi inženjeringa karakteristika (eng. *feature engineering*), ulaz u neuronsku mrežu je dimenzije 36, što je ukupna dimenzija kombinovanih informaciono-teorijskih i stilometrijskih karakteristika (14 + 22 = 36). Nakon izbora 11 najinformativnijih karakteristika, finalni PIDNN je prethodno opisane arhitekture.

Računanje SHAP vrednosti izvršeno je unutar desetostruke unakrsne validacije, što znači da su prikazane vrednosti jednake prosečnim vrednostima svih 10 podskupova. Slike 34 i 35 prikazuju prvih 11 rangiranih karakteristika za sisteme A i B, respektivno.

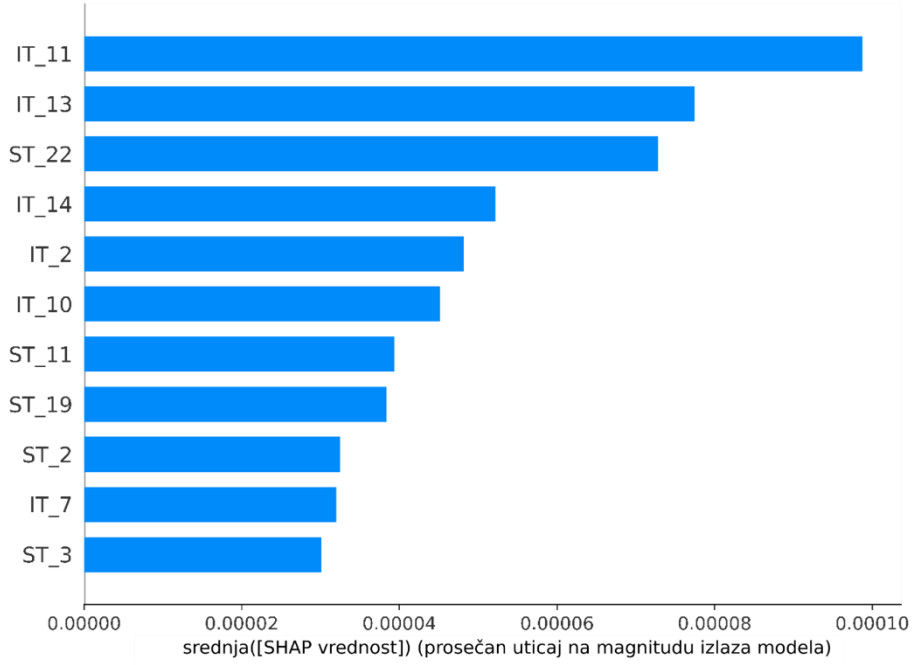
Primećuje se da je sinergija informaciono-teorijskih i stilometrijskih karakteristika za oba sistema rezultovala skupom od 11 najinformativnijih karakteristika, koje ne zahtevaju dodatnu razmenu informacija preko javnog kanala. Naime, nema karakteristika IT 4, IT 5 i IT 6 u tim skupovima. Ovo svojstvo je važno ne samo sa aspekta bezbednosti, već se takođe omogućava nezavisna lokalna sinteza PIDNN blokova na strani legitimnih korisnika.

Interesantno je istaći relativnu važnost informaciono-teorijskih i stilometrijskih karakteristika. U tu svrhu, definišu se sledeći indikatori:

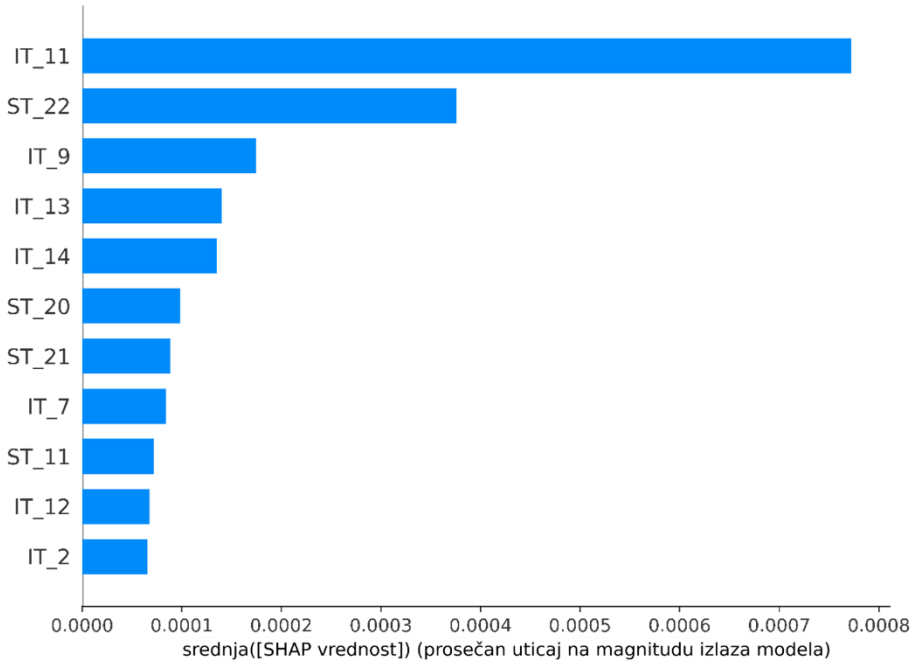
$$Sig_{IT} = \frac{\sum_{i=1}^{11} Shap(IT_i)}{\sum_{i=1}^{11}[Shap(IT_i)+Shap(ST_i)]} \times 100 \ [\%], \tag{6.23}$$

$$Sig_{ST} = \frac{\sum_{i=1}^{11} Shap(ST_i)}{\sum_{i=1}^{11}[Shap(IT_i)+Shap(ST_i)]} \times 100 \ [\%]. \tag{6.24}$$

Oni daju relativan doprinos informaciono-teorijskih i stilometrijskih karakteristika u procentima, respektivno. Za Sistem A, dobijaju se vrednosti $Sig_{IT} = 62.4\%$, $Sig_{ST} = 37.6\%$, dok se za sistem B dobijaju vrednosti $Sig_{IT} = 69.4\%$, $Sig_{ST} = 30.6\%$.



Slika 34. Značaj obeležja na ulazu u PIDNN za SKD sistem bez HC bloka (sistem A), rangirani po svojim SHAP vrednostima. Adaptirano iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.



Slika 35. Značaj obeležja na ulazu u PIDNN za SKD sistem sa HC blokom (sistem B), rangirani po svojim SHAP vrednostima. Adaptirano iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Veći uticaj stilometrijskih karakteristika u Sistemu A može se objasniti činjenicom da ove karakteristike bolje opisuju nemodelovane serijalne korelacije SCR-a, koje su više prisutne u ovom sistemu. Uticaj stilometrijskih karakteristika opada u sistemu B jer se ove nemodelovane serijalne korelacije SCR-a uklanjaju *Huffman*-ovim kodovanjem.

Metodološki tok eksperimenta prikazan je Algoritmom 6. Kao diskretni memorijski izvor u eksperimentu primenjen je govorni signal i to na oba opisana sistema a dobijene vrednosti prikazane su u Tabeli 11.

Tabela 11. Vrednosti parametara za eksperiment sa govornim signalima i sistemima A i B

	A (Win-Hash)	**B (Win-Huff-Hash)**
$R_{2Spoil} > 0$ [%]	100 ± 0.00	99.96 ± 0.09
PICP	0.9994 ± 0.0009	0.9996 ± 0.0008
MPIW	0.047 ± 0.009	0.048 ± 0.048
$R2$	28,944.50 ± 351.67	28,859.68 ± 352.83
R_{2min}	445.08 ± 34.24	323.97 ± 27.57
R_{2Spoil}	28,899.21 ± 7443.42	28,811.07 ± 352.74
$G^{Opt}_{GLB\_R2min}$	65.3366 ± 4.0328	89.5971 ± 6.0219
$G^{ML}_{GLB\_R2min}$	65.2343± 4.0265	89.4362 ± 6.0119
$Loss_{GLB\_R2min}$ [%]	98.46 ± 0.11	98.88 ± 0.09
$Loss_{ML}$ [%]	0.16 ± 0.01	0.17 ± 0.01
KR [%]	11.29 ± 0.14	11.25 ± 0.14
KAR [%]	100 ± 0.00	99.96 ± 0.09
LR [10^{-3}]	0.0286 ± 0.0027	0.0286 ± 0.0030

Radi provere slučajnosti dobijenih sekvenci iz sistema A i B, primenjeni su NIST testovi, a rezultati su prikazani u Tabeli 12.

Tabela 12. Rezultati NIST testova

	F	**BF**	**R**	**LR**	**FFT**	**S**	**AE**	**CSf**	**CSr**
A	0.0661	0.8932	0.9852	0.2289	0.4825	0.2203	0.4928	0.0104	0.0108
B	0.1268	0.5856	0.8277	0.5529	0.7190	0.2801	0.4120	0.1366	0.0835

Na osnovu rezultata prikazanih u Tabelama, mogu se izvesti sledeći zaključci:

Sistemi A i B demonstriraju vrlo slične performanse u pogledu glavnih indikatora: brzine ključa (eng. *key rate* - KR), stopa prihvatanja ključa (eng. *key acceptance rate* - KAR), gubitak mašinskog učenja ($Loss_{ML}$) i brzina curenja (eng. *leakage rate* - LR). Ova empirijska zapažanja ukazuju da dominantan uticaj na performanse sistema potiče od ulazne dekorelacione permutacije, dok *Huffman*-ovo kodovanje izvora ima sekundarnu ulogu. Drugim rečima, ključni faktor efikasnosti sistema nije sama kompresija izvora putem *Huffman*-ovog kodiranja, već proces permutacije koji eliminiše statističke međuzavisnosti između ulaznih podataka. Ova permutacija smanjuje statističku zavisnost među simbolima govornog signala, čime se povećava entropija izvora i omogućava efikasnije generisanje tajnih ključeva. *Huffman*-ovo kodovanje, iako korisno za smanjenje redundanse, ne doprinosi značajno sigurnosnim performansama sistema u ovom kontekstu.

Ostvarena vrednost KR = 11% predstavlja najvišu vrednost brzine ključeva dobijenu u okviru SKD sistema zasnovanog na biometrijskim SCR. Uzimajući u obzir da SCR zasnovan na govornom signalu ima brzinu od 256 kb/s (16 bitova po uzorku pri frekvenciji uzorkovanja od 16 kHz), apsolutna brzina generisanja potpuno tajnih ključeva iznosi približno 28 kb/s. Ova vrednost daleko prevazilazi zahteve tipičnih kriptografskih primena i omogućava primenu sistema u scenarijima visokog protoka podataka.

Rezultujući LR je reda veličine 3×10−5 bita po jednom bitu generisanog tajnog ključa. To znači da Eva dobija zanemarljivih 0.85 bita na svakih 28 kb generisanog ključa. Ova količina „procurelih“ informacija raspoređena je po celoj sekvenci koju Eva prisluškuje, što praktično onemogućava rekonstrukciju bilo kog bita destilovanih tajnih ključeva.

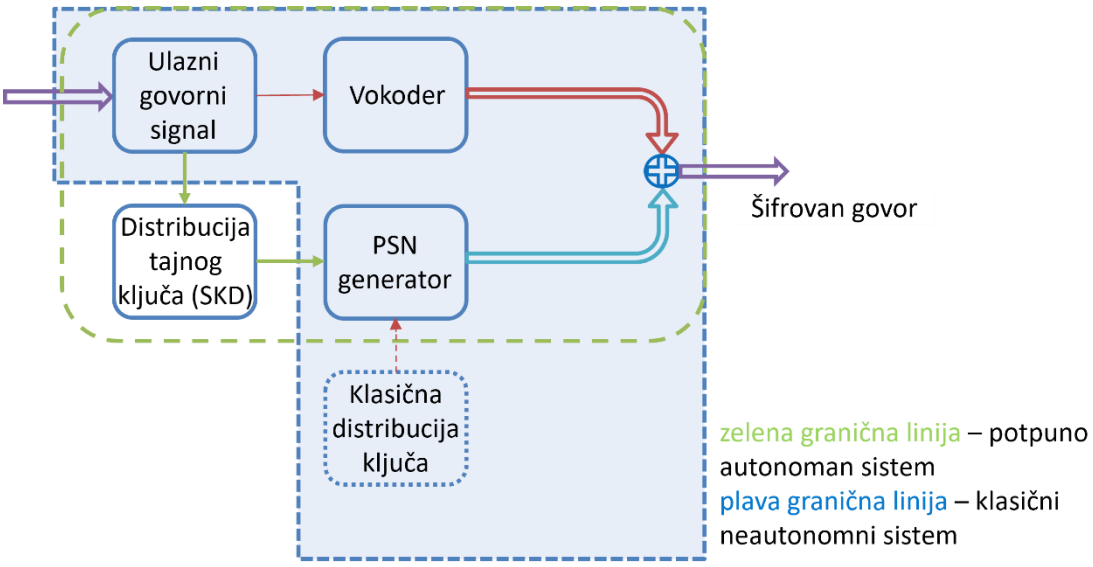
U dostupnoj literaturi ne postoje radovi koji eksplicitno navode LR za ovakve sisteme, osim radova [34,48]. Oni su prijavili LR vrednosti u rasponu do 4×10−3 za SCR zasnovane na EEG signalima, što je više od jednog reda veličine lošiji rezultat. Pošto sistemi iz [34,48] ne sadrže inicijalnu dekorelacionu permutaciju, opravdano je zaključiti da njen značajan uticaj na performanse SKD sistema dolazi do izražaja ne samo kroz povećanje KR, već i kroz značajno smanjenje LR.

Vrednost $Loss_{ML} = 0.16 \pm 0.01\%$ pokazuje da predloženi sistem maksimalno koristi slučajnost datog SCR-a. Ovo je ujedno potvrda da je koncept adaptivnog PA zasnovan na PIDNN i donjoj granici ECRE2 praktično optimalan. U idealnom slučaju, vrednost gubitka bila bi $Loss_{ML} = 0$.

Visoke vrednosti veličina $G_{GLB\_R2min}^{Opt}$ i $G_{GLB\_R2min}^{ML}$ (u rasponu od 65 do 410) ukazuju na superiornost dizajna PA bloka zasnovanog na proceni donje granice ECRE2, u poređenju sa klasičnim pristupima zasnovanim na globalnom minimumu.

NIST test potvrđuje slučajnost destilovanih tajnih ključeva. Uz zanemarljiv LR, predloženi sistem se može koristiti za generisanje i distribuciju kriptoloških ključeva u sistemima koji obezbeđuju apsolutnu tajnost u *Shannon*-ovom smislu.

Predloženi SKD sistem omogućava dizajn potpuno autonomnih sistema za zaštitu govora, videti sliku 36.



Slika 36. Prikaz potpuno autonomnog sistema za zaštitu govora na predajnoj strani. Plava granična linija označava klasičan sistem, dok zelena granična linija označava potpuno autonoman sistem. Adaptirano iz rada [36], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Autonomija predloženog sistema se odnosi na nezavisnost od dodatnog sistema za generisanje i distribuciju tajnih ključeva. Ona važi u slučaju ispunjenosti opštih uslova za funkcionisanje SKD sistema sa javnom diskusijom, što po pravilu podrazumeva raspolaganje autentifikovanim javnim kanalom odgovarajućeg propusnog opsega.

7 Ka praktičnoj realizaciji apsolutno tajnog autonomnog sistema zaštite govora na niskim brzinama prenosa

Kao što je poznato iz teorije informacione bezbednosti [1], Vernamova šifra [71] ili One-Time-Pad (OTP) [72] zadovoljava uslov apsolutne tajnosti, koji podrazumeva da brzina tajnih ključeva mora biti jednaka brzini poruka što predstavlja praktičnu prepreku u mnogim realnim sistemima. Potreba za generisanjem tajnih ključeva smanjuje se proporcionalno sa brzinom poruka, što opravdava upotrebu vokodera sa niskom brzinom prenosa u praktičnim implementacijama [73-75].

U kontekstu izbora izvora zajedničke slučajnosti, većina postojećih rešenja oslanja se na bežični komunikacioni kanal [76-78], što je u suprotnosti sa zahtevom za autonomijom sistema. S druge strane, prethodno analizirani pristupi zasnovani na biometrijskim signalima iako obećavajući, nisu demonstrirali sposobnost garantovanja brzina generisanja tajnih ključeva koje odgovaraju zahtevima standardnih vokodera sa niskom brzinom prenosa (1.2 kb/s ili 2.4 kb/s).

Većina postojećih radova fokusirana je na protokole generisanja ključeva u specifičnim okruženjima, dok se veoma malo njih fokusira na integraciju tih protokola sa OTP-om. Jednostavan način podrazumeva kaskadno ili paralelno generisanje ključeva i OTP, ali se u literaturi pojavljuju i složenije konstrukcije. Na primer, u radu [79] dokazano je da upotreba neusklađenog ključa za OTP prevazilazi klasični OTP sa identičnim ključem.

Nadovezujući se na prethodno razmatrane SKD metode i principe, u ovom poglavlju biće predstavljen pristup koji obezbeđuje nezavisnost izvora zajedničke slučajnosti od telekomunikacionog kanala, uz istovremeno zadovoljenje zahteva za brzinom generisanja ključeva kompatibilnom sa standardnim vokoderima niske brzine. Što se tiče modela napada na ovu klasu sistema, analiza se ograničava na model pasivnog kriptoanalitičara i pretpostavku postojanja autentifikovanog javnog kanala. Problem neautentifikovanog javnog kanala može se rešiti primenom postojećih klasičnih metoda autentifikacije, pri čemu je jedina posledica skraćenje finalnog tajnog ključa, bez narušavanja njegove informaciono-teorijske bezbednosti [27].

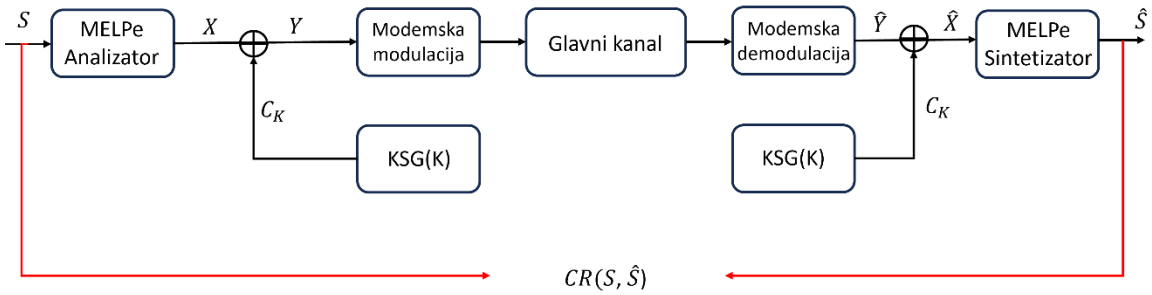
7.1 Arhitektura sistema

Generički sistem za sigurnu govornu komunikaciju sa niskom brzinom prenosa prikazan je na slici 37. Prema klasifikaciji prikazanoj u [80], sistem pripada kategoriji sigurne govorne komunikacije korišćenjem modem-baziranih kriptografskih tehnika. Ključni aspekt informaciono-teorijskog pristupa verifikaciji nivoa bezbednosti sistema je količina informacija koju prisluškivač može dobiti o porukama na osnovu posmatranja šifrata. Dobro je poznato da, ako je C_K generisan korišćenjem bilo kog kriptografskog algoritma sa konačnim tajnim ključem K, njegova ekvivokacija sa stanovišta napadača, $H(K|Y)$, brzo konvergira ka nuli nakon dovoljno dugog posmatranja šifrata [1]. U praksi, ovo znači da takav sistem nije informaciono-teorijski siguran, i njegova bezbednost zavisi od računske moći protivnika. Jednom kada količina posmatranog šifrata zadovolji uslov $H(K|Y) = 0$, ključ K ima jedinstveno rešenje, što znači da je, sa stanovišta kriptoanalize, sistem kompromitovan. Međutim, ako je C_K potpuno slučajna sekvenca nezavisna od poruka X, može se pokazati da je međusobna informacija $I(Y;X) = 0$. Ovo implicira da napadač ne može da povrati poruke bez obzira na njihove računske resurse, čineći sistem potpuno sigurnim.

Za analizu izabran je standardni vokoder sa niskom brzinom prenosa MELPe (Enhanced Mixed Excitation Linear Prediction) sa brzinom od 1.2kb/s [74]. MELP vokoder je parametarski

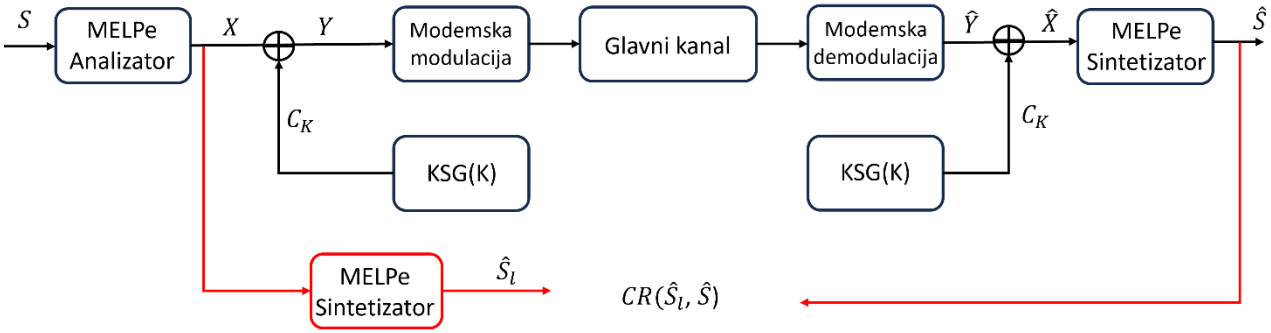
vokoder sa linearnom predikcijom, razvijen za izuzetno niske brzine prenosa (≤2.4kb/s), prvenstveno korišćen za robusnu govornu komunikaciju u degradiranim kanalima. Za razliku od klasičnog LPC vokodera, koji koristi čisto pobuđivanje impulsima ili šumom**,** MELP uvodi mešovitu pobudu, čime se značajno poboljšava kvalitet govora pri niskim brzinama prenosa. Iako je u ovom radu korišćen MELPe vokoder, zaključci se mogu proširiti i na druge standardne vokodere sa sličnim performansama i parametarskom strukturom, pri čemu treba voditi računa o poziciji i ulozi lokalnog generatora šuma u sintezi izabranog vokodera [81].

Siguran prenos govornog signala vrši se na sledeći način. Ulazni govorni signal S je odabran na 8 kHz, diskretizovan sa 16 bita po uzorku i podeljen u okvire koji traju 67.5 ms (540 uzoraka). U MELPe analizatoru, 81 bit se generiše za svaki ulazni okvir, od čega 80 bitova kodira 10 LSP (*Linear Spectral Pairs*) parametara LP (*Linear Prediction*) modela produkcije govora, dok je 1 bit sinhronizacioni bit. Ova sekvenca bita se šifruje sabiranjem po modulu 2 sa binarnom pseudo-slučajnom sekvencom C_K generisanom generatorom niza ključeva KSG(K) (eng. *keystream generator*) sa tajnim ključem K, koji mora biti razmenjen između legitimnih strana pre početka komunikacije. Šifrat Y se zatim prenosi preko glavnog kanala nakon odgovarajuće modulacije, a na prijemnoj strani vrši se demodulacija i dešifrovanje. Dešifrovanje se obavlja sabiranjem po modulu 2 sa sinhronizovano generisanom binarnom sekvencom C_K na prijemnoj strani. Kao rezultat, dobijaju se identični LSP parametri, koji proizvode rekonstruisani govorni signal $\hat{S}$ u MELPe bloku sintetizatora. Sa $CR(S,\hat{S})$, označen je prvi kandidat za izvor zajedničke slučajnosti, formiran od ulaznog govornog signala S na strani predajnika i govornog signala $\hat{S}$ sintetisanog na strani prijemnika; videti sliku 37. Kriptografski deo sistema se sastoji od KSG(K) generatora binarne pseudo-slučajne sekvence C_K, koja je sabrana po modulu 2 sa binarnom sekvencom koja dolazi iz analizatora. $CR(S,\hat{S})$ je označen kao izvor zajedničke slučajnosti, formiran od ulaznog govornog signala S na strani predajnika i govornog signala $\hat{S}$ sintetisanog na strani prijemnika.



Slika 37. Generička šema sigurnih komunikacija sa niskom brzinom prenosa korišćenjem MELPe govornog kodera. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Slika 38 prikazuje proširenje generičke šeme sa slike 37 sa lokalnom MELPe sintezom na strani predajnika. $CR(\hat{S}_l,\hat{S})$ je označen kao drugi kandidat za izvor zajedničke slučajnosti, formiran od lokalno sintetisanog govornog signala $\hat{S}_l$ na strani predajnika i govornog signala $\hat{S}$ sintetisanog na strani prijemnika. Ako nema grešaka prenosa, filteri sinteze na prijemnoj i predajnoj strani su jednaki, a razlike u sintetisanim signalima potiču od različitih lokalnih izvora slučajnosti, koji se koriste za formiranje složene ekscitacije MELPe vokoder sintetizatora.



Slika 38. Proširenje generičke šeme sa lokalnom MELPe sintezom na strani predajnika. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Kao što je istaknuto u literaturi [3], tokom izvođenja SKD protokola, u AD i IR fazi, određene informacije se razmenjuju između Alise i Boba putem javnog kanala, te su samim tim dostupne i Evi. U kontekstu AD faze, kada se primenjuje protokol bitskog pariteta (BP) [59], količina informacija o Alisinoj sekvenci koja se otkriva Evi može se kvantifikovati. Konkretno, broj paritetnih bitova dostupnih Evi određen je sledećom lemom.

Lema 1. Neka su početni stringovi X i Y, koje Alisa i Bob poseduju na početku SKD protokola, binarne iid slučajne sekvence dužine N na Hamingovom rastojanju ε, $\varepsilon \in [0,0.5]$. Tada je očekivani broj bitova pariteta koje Alisa razmenjuje sa Bobom preko javnog kanala dat sa

$$N_{ADparity} = \left\lfloor \frac{N}{2} \right\rfloor + \sum_{i=1}^{s-1} \left\lfloor \frac{N}{2^{i+1}} \cdot \frac{\varepsilon^{2^i} + (1-\varepsilon)^{2^i}}{\prod_{j=0}^{i-1} \left(\varepsilon^{2^j} + (1-\varepsilon)^{2^j} \right)} \right\rfloor, \tag{7.1}$$

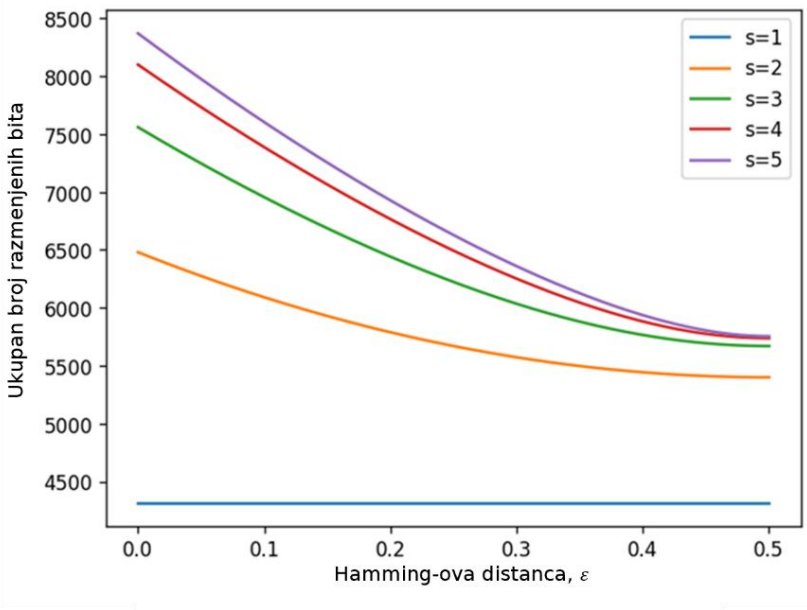
gde je s broj iteracija BP algoritma.

Dokaz. Dokaz direktno sledi iz činjenice da je ukupna količina bitova pariteta razmenjenih u BP algoritmu jednaka zbiru razmena za svaku iteraciju. S druge strane, za svaku iteraciju, ova vrednost je jednaka polovini dužine sekvenci na početku iteracije. Ova dužina je takođe jednaka dužini sekvenci na kraju prethodne iteracije. U [21], Teorema 2.2.3, str. 17, dat je izraz za stopu kompresije BP algoritma u svakoj iteraciji. Na osnovu ove formule, dobija se dužina sekvenci posle i iteracija, $\frac{N}{2^{i+1}} \cdot \frac{\varepsilon^{2^i} + (1-\varepsilon)^{2^i}}{\prod_{j=0}^{i-1} \left(\varepsilon^{2^j} + (1-\varepsilon)^{2^j} \right)}$. Sumiranjem ovih vrednosti preko svih iteracija, dobija se tvrdnja (7.1). Treba napomenuti da se operator najmanjeg celog broja $\lfloor \cdot \rfloor$ primenjuje zbog same prirode provere pariteta 2-bitnih blokova. Ovime se završava dokaz leme. □

Broj bitova od značaja za prisluškivača može biti neznatno manji od $N_{ADparity}$ pošto neke jednačine pariteta mogu biti linearno zavisne. Međutim, pošto je $\left\lfloor \frac{N}{2} \right\rfloor$ bitova pariteta u prvoj iteraciji BP algoritma međusobno linearno nezavisno, uvek važi:

$$N_{ADparity} \geq \left\lfloor \frac{N}{2} \right\rfloor. \tag{7.2}$$

Ovo znači da se neizvesnost govornog signala na ulazu u sistem barem prepolovi, posmatrano sa strane prisluškivača. Na slici 39 prikazan je primer zavisnosti $N_{ADparity}$ kao funkcije od ε za $N = 8640$ i $s = 1,2,3,4,5$.



Slika 39. Primer zavisnosti broja bitova pariteta razmenjenih preko javnog kanala u AD fazi, kao funkcije početnog Hamingovog rastojanja Alisinih i Bobovih sekvenci. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Napomena 7. Pošto je naš cilj apsolutno tajan sistem prenosa govora, može se zaključiti da ova dva kandidata za izvore zajedničke slučajnosti ne mogu biti korišćeni tokom zaštićene komunikacije. Naime sprovođenje SKD protokola po javnom kanalu bi otkrilo napadaču parcijalne informacije o signalima S i $\hat{S}$ u slučaju izvora $CR(S,\hat{S})$, ili o signalima $\hat{S}_l, \hat{S}$ u slučaju izvora $CR(\hat{S}_l,\hat{S})$, što po definiciji narušava kriterijum apsolutne tajnosti. Međutim, ovi izvori se mogu koristiti u režimu otvorenog rada, koji po pravilu prethodi zaštićenom. Naime, u tipičnoj upotrebi takvih sistema, legitimne strane uspostavljaju komunikacioni link kroz uobičajenu otvorenu komunikaciju. Nakon provere kvaliteta veze i međusobnog pristanka strana, vrši se prelazak na šifrovanu komunikaciju. Praksa pokazuje da ovaj deo otvorene komunikacije traje 2 do 10 sekundi. Pošto je komunikacija otvorena, curenje informacija o signalima $S, \hat{S}$ ili $\hat{S}_l$, kao različitim kodnim slikama otvorenog govora S ka prisluškivaču, ne igra nikakvu ulogu. Stoga se ovi, izvori zajedničke slučajnosti mogu koristiti za bezbednu destilaciju tajnih ključeva.

Prateći prethodnu logiku, dobar izvor zajedničke slučajnosti bi mogao biti $CR(\hat{S}_{lA},\hat{S}_{lB})$. Signali $\hat{S}_{lA}$ i $\hat{S}_{lB}$ su dobijeni korišćenjem lokalnih LP sinteza sa istim LSP parametrima o kojima Eva nema informacija. Razmatraju se dva FIFO (*First In First Out*) bafera identičnog tajnog slučajnog sadržaja na Alisinoj i Bobovoj strani, videti sliku 40. Ukoliko se ovaj sadržaj interpretira kao skup slučajno odabranih LSP parametara o kojima Eva nema informacija, sinhronizovanim čitanjem obe strane mogu sintetisati potrebne signale $\hat{S}_{lA}$ i $\hat{S}_{lB}$. U tom slučaju, Alisa i Bob mogu koristiti SKD protokol preko izvora $CR(\hat{S}_{lA},\hat{S}_{lB})$ za destilaciju tajnih ključeva bez da Eva primi i jedan bit informacija sa javnog kanala o ulaznom govornom signalu S. Naime,

$$I(S,\hat{S}_{lA}) = 0,\ I(S,\hat{S}_{lB}) = 0, \tag{7.3}$$

imajući u vidu način na koji su signali $\hat{S}_{lA}$ i $\hat{S}_{lB}$ generisani. FIFO bafer se kontinuirano dopunjava upravo destilovanim tajnim ključevima, dok se niz tajnog ključa C_K sinhronizovano čita na prijemnoj i predajnoj strani. Sabiranjem po modulu 2 sa izlaznom sekvencom MELPe analizatora, X,

$$Y = X \oplus C_K, \tag{7.4}$$

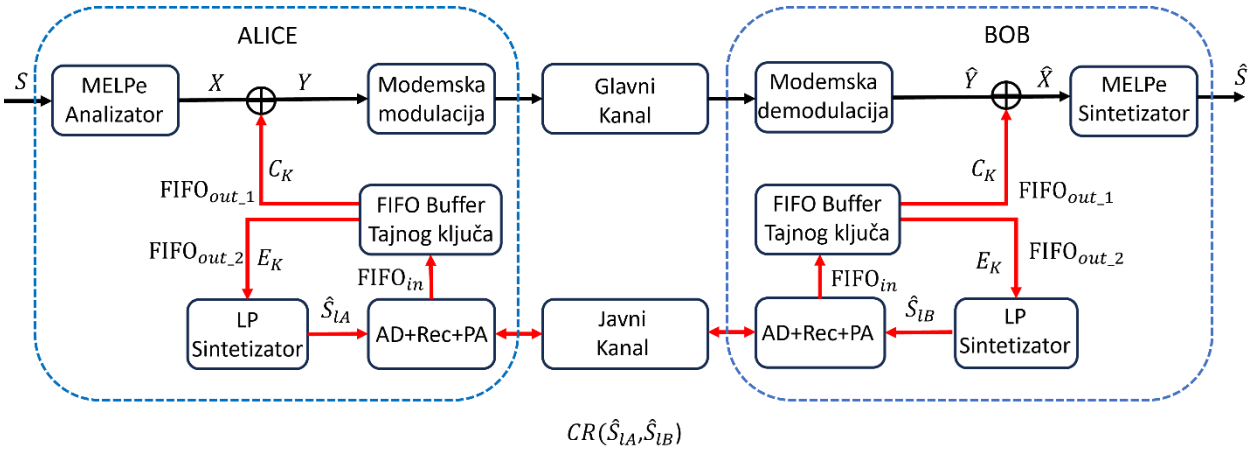
Formira se apsolutno tajna Vernamova šifra. Neophodan uslov za održavanje sistema u kontinuiranoj apsolutnoj tajnosti je da brzina punjenja FIFO memorija na predajnoj i prijemnoj strani ne sme biti manja od brzine čitanja, tj.,

$$R_K \geq R_C + R_E. \tag{7.5}$$

U (7.5), R_K je brzina destilacije tajnog ključa, R_C je brzina potrošnje tajnog ključa Vernamove šifre, a R_E je brzina potrošnje LP sintetizatora. Treba napomenuti da R_C mora biti jednaka brzini izlazne sekvence MELPe vokodera u glavnom kanalu

$$R_C = R_{MELPe}. \tag{7.6}$$

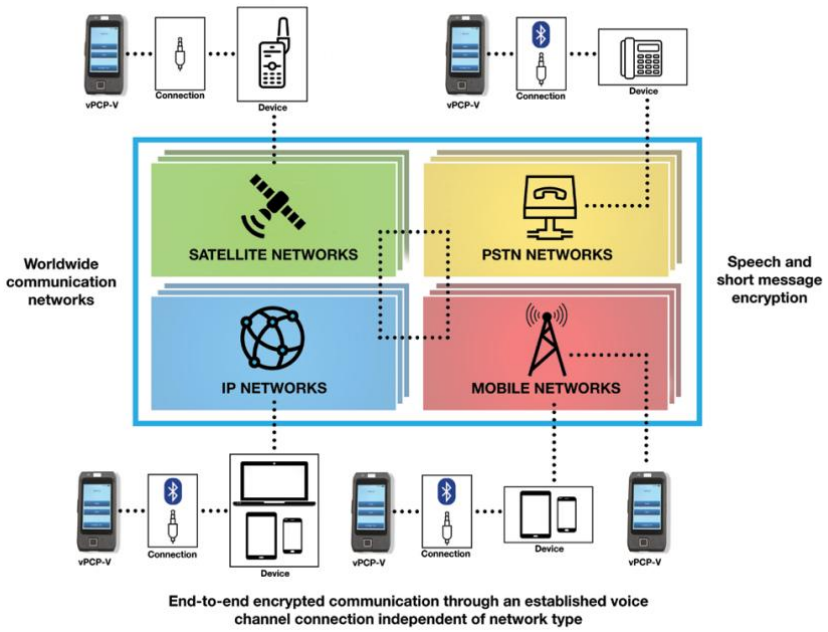
Slika 40 prikazuje generičku šemu ovog APS-VCS (Autonomous Perfectly Secure Low-Bit-Rate Voice Communication System) koncepta. Pseudo-slučajan niz C_K je jednokratni tajni ključ Vernamove šifre, koji se čita iz FIFO bafera tajnih ključeva sinhronizovano na Alisinoj i Bobovoj strani. Sadržaj FIFO bafera se puni destilisanim tajnim ključevima primljenim primenom SKD protokola na izvor zajedničke slučajnosti $CR(\hat{S}_{lA},\hat{S}_{lB})$. Signali $\hat{S}_{lA}$ i $\hat{S}_{lB}$ su dobijeni lokalnim LP sintezama preko istih LSP parametara sinhronizovano očitanih iz FIFO bafera.



Slika 40. Generička šema predloženog APS-VCS. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

7.2 Identifikacija izvora zajedničke slučajnosti

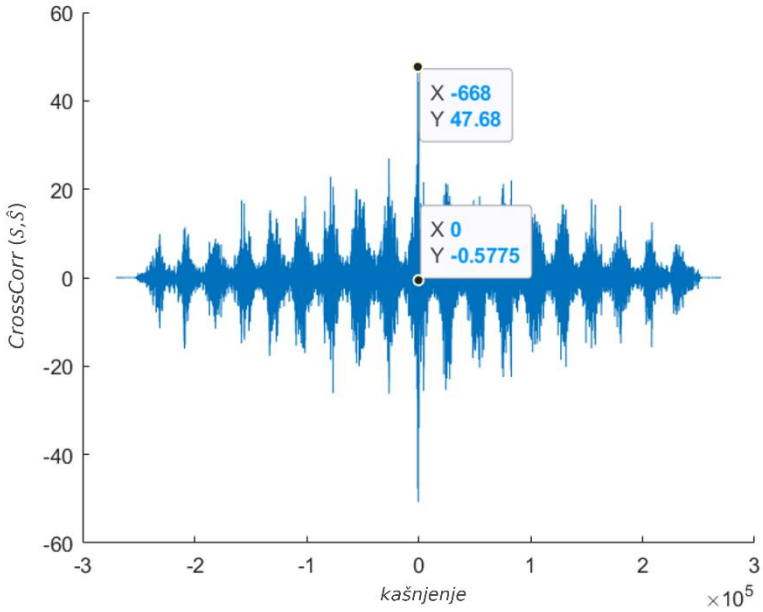
Prethodna analiza pokazuje da su u fazi otvorenog govora dostupna dva izvora zajedničke slučajnosti, $CR(S,\hat{S})$ i $CR(\hat{S}_l,\hat{S})$. Kako bi se procenilo koji od ovih izvora je pogodniji, sprovedena je eksperimentalna evaluacija u realnim uslovima komunikacije sa vPCP-V sistemom (*Voice over Internet Protocol* – VoIP, javnom, fiksnom, mobilnom ili satelitskom), videti sliku 41 i Vlatacom generatorom pravih slučajnih brojeva (vTRNG) [83,84].



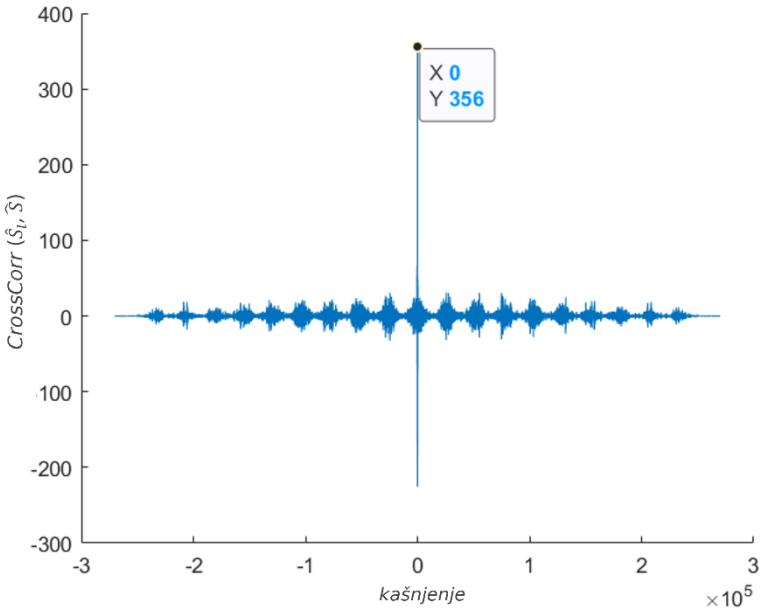
Slika 41. Vlatacom-ova personalna kripto platforma za šifrovanje glasa dizajnirana za upotrebu u bilo kom dostupnom komunikacionom sistemu. Preuzeto iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Za eksperimentalnu evaluaciju, formiran je test skup koji se sastoji od 24 govorna signala sa govornicima koji čitaju unapred izabran tekst. U 14 slučajeva, tekst je bio jedinstven, dok su u preostalih 10 slučajeva govornici snimili ponovljeni tekst. Snimljeni signali su trajanja između 32 i 59 sekundi, uzorkovani su na frekvenciji od 8 kHz i diskretizovani sa 16 bitova po uzorku.

Slika 42 prikazuje funkciju unakrsne korelacije između originalnog ulaznog govornog signala S i sintetisanog primljenog signala $\hat{S}$ za uzorak broj 1 iz test skupa. Slika 43 prikazuje odgovarajuću funkciju unakrsne korelacije između lokalno sintetisanog govornog signala $\hat{S}_l$ i sintetisanog primljenog signala $\hat{S}$ za isti govorni uzorak. Iz prikazanih primera je jasno da je korelacija izvora zajedničke slučajnosti $CR(\hat{S}_l, \hat{S})$ za red veličine veća od izvora $CR(S, \hat{S})$.

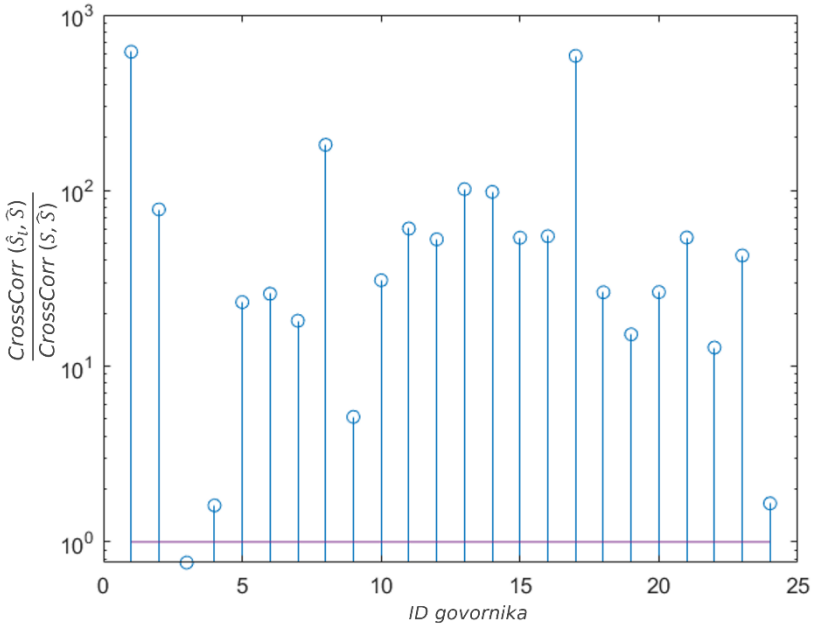


Slika 42. Unakrsna korelacija između originalnog ulaznog govornog signala S i sintetisanog primljenog signala $\hat{S}$ za govornika broj 1 iz test skupa. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.



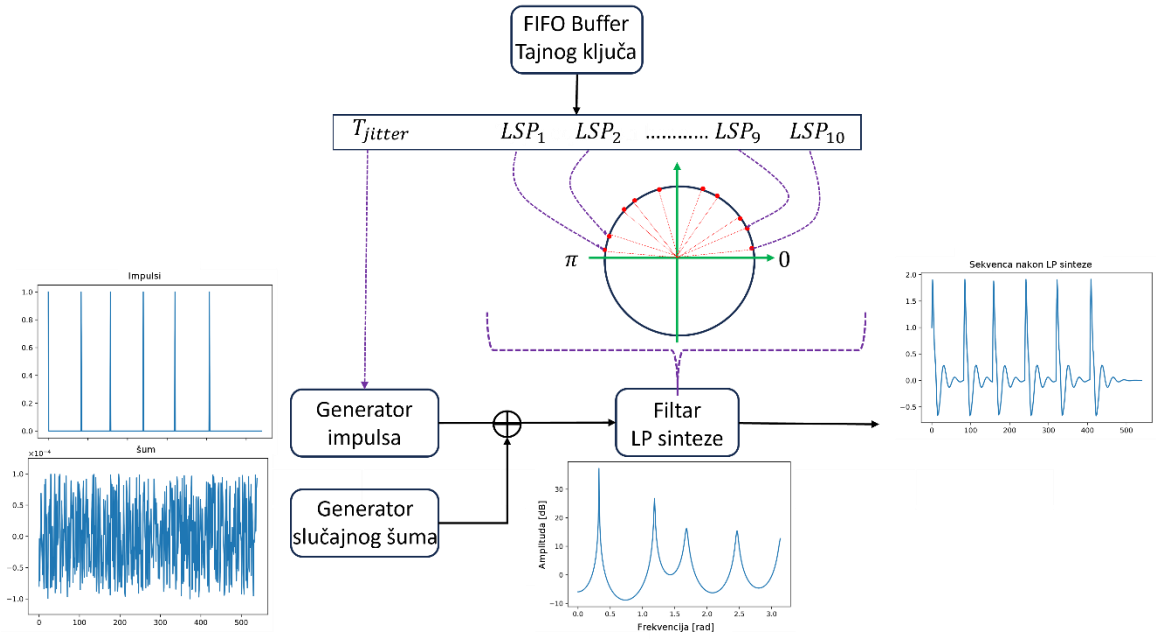
Slika 43. Unakrsna korelacija između lokalno sintetisanog govornog signala $\hat{S}_l$ na strani predajnika i sintetisanog signala na strani prijemnika $\hat{S}$ za govornika broj 1 iz test skupa. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Ova činjenica je potvrđena na celom test uzorku. Slika 44 prikazuje logaritamski odnos $\frac{CrossCorr(\hat{S}_l,\hat{S})}{CrossCorr(S,\hat{S})}$ za svih 24 test uzorka govornih signala. Samo u slučaju test uzorka broj 3 je ovaj odnos manji od 1. Ovo ukazuje da je unakrsna korelacija $CrossCorr(\hat{S}_l,\hat{S})$ gotovo uvek značajno veća od unakrsne korelacije $CrossCorr(S,\hat{S})$. Stoga je odlučeno da se koristimi izvor $CR(\hat{S}_l,\hat{S})$ za izvršavanje SKD protokola u otvorenoj fazi komunikacije APS-VCS sistema.

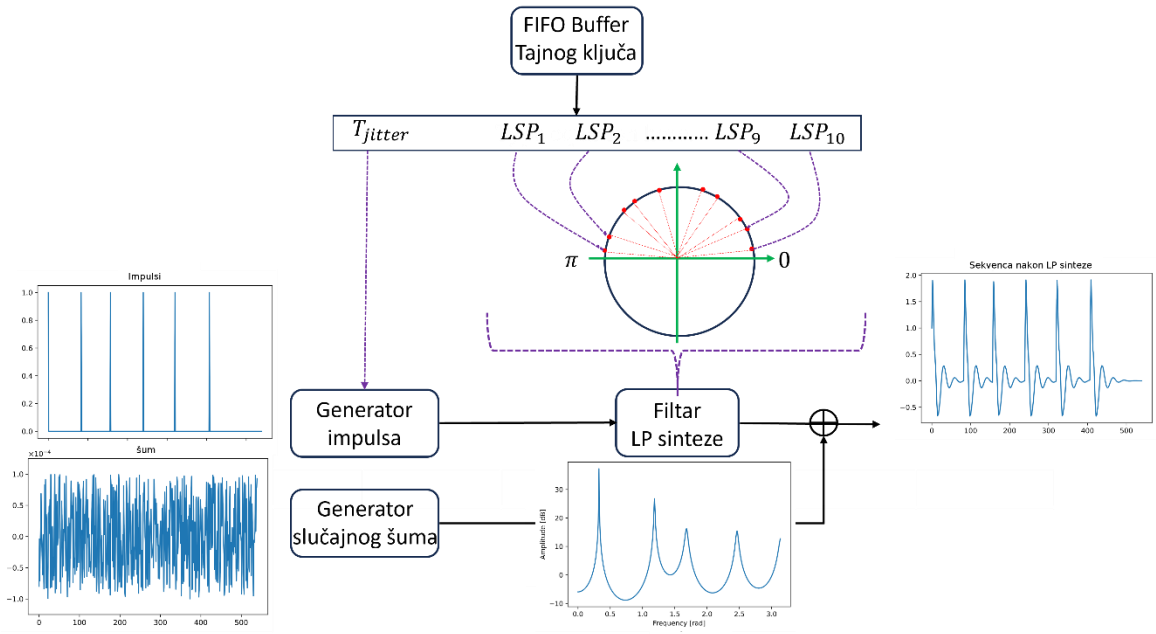


Slika 44. Logaritamski odnos $\frac{CrossCorr(\hat{S}_l,\hat{S})}{CrossCorr(S,\hat{S})}$ za svih 24 test uzorka govornih signala. Samo u slučaju test uzorka broj 3 je ovaj odnos manji od 1, što znači da je unakrsna korelacija $CrossCorr(\hat{S}_l,\hat{S})$ gotovo uvek veća od unakrsne korelacije $CrossCorr(S,\hat{S})$. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

U fazi otvorene komunikacije, analizator i sintetizator ugrađenog MELPe vokodera se koriste za formiranje izvora $CR(\hat{S}_l, \hat{S})$ i destilaciju tajnih ključeva koji se koriste za početno punjenje FIFO memorija. Za potrebe zaštićene komunikacije, potrebno je formirati izvor $CR(\hat{S}_{lA}, \hat{S}_{lB})$ zasnovan na LP sintetizatoru, koji može biti mnogo jednostavniji od MELPe sintetizatora. Naime, kompleksnost MELPe sintetizatora potiče od složenog procesa formiranja signala ekscitacije kako bi se ispunili zahtevni kriterijumi razumljivosti i prirodnosti sintetisanog govora. Ovaj zahtev nema značaj u formiranju sintetisanih signala $\hat{S}_{lA}$ i $\hat{S}_{lB}$. Stoga je ekscitacija značajno pojednostavljena i sastoji se od periodičnog niza jediničnih impulsa, sa moguće kontrolisanim jitter-om i dodavanjem lokalno generisanog čisto slučajnog šuma. Za ovu svrhu, u eksperimentalnoj evaluaciji sistema smo koristili vTRNG zasnovan na izvoru entropije prirodnog procesa sa ugrađenim sistemom za proveru slučajnosti. Slika 45 i slika 46 prikazuju generičku šemu generatora lokalno sintetisanih signala zasnovanih na slučajnim LSP parametrima, periodičnom impulsnom ulazu i aditivnom šumu na ulazu, odnosno izlazu LP sintetizatorskog filtera.



Slika 45. Generator lokalno sintetisanih signala $\hat{S}_{lA}$ zasnovan na slučajnim LSP parametrima, periodičnom impulsnom ulazu δ i aditivnom šumu ξ na ulazu u LP sintetizatorskog filtra. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.



Slika 46. Generator lokalno sintetisanih signala $\hat{S}_{lA}$ zasnovan na slučajnim LSP parametrima, periodičnom impulsnom ulazu δ i aditivnom šumu ξ na izlazu LP sintetizatorskog filtra. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

LP sintetski filter je dat prenosnom funkcijom:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}}, \tag{7.7}$$

gde je p red LP filtera

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}. \tag{7.8}$$

Ako je LP filter minimalne faze, tj. ako su sve njegove nule unutar jediničnog kruga u Z ravni, LP sintetski filter H(z) je stabilan.

Koeficijenti LP filtera $\{a_i\}$ se dobijaju na osnovu učitanog skupa LSP parametara iz FIFO memorije

$$\left\{\varphi_1, \theta_1, \varphi_2, \theta_2, \dots, \varphi_{\frac{p}{2}}, \theta_{\frac{p}{2}}\right\} \tag{7.9}$$

uz ograničenje

$$0 < \varphi_1 < \theta_1 < \varphi_2, < \theta_2, \dots, \varphi_{\frac{p}{2}} < \theta_{\frac{p}{2}} < \pi. \tag{7.10}$$

Kao što je poznato [85], LP filter A(z) može se dekomponovati u obliku

$$A(z) = \frac{1}{2}[P(z) + Q(z)] \tag{7.11}$$

gde su P(z) i Q(z) takozvani parni i neparni polinomi definisani LS frekvencijama (7.10)

$$P(z) = (1 + z^{-1}) \prod_{i=1}^{\frac{p}{2}} \left(1 - e^{-j\varphi_i} z^{-1}\right) \left(1 - e^{j\varphi_i} z^{-1}\right), \tag{7.12}$$

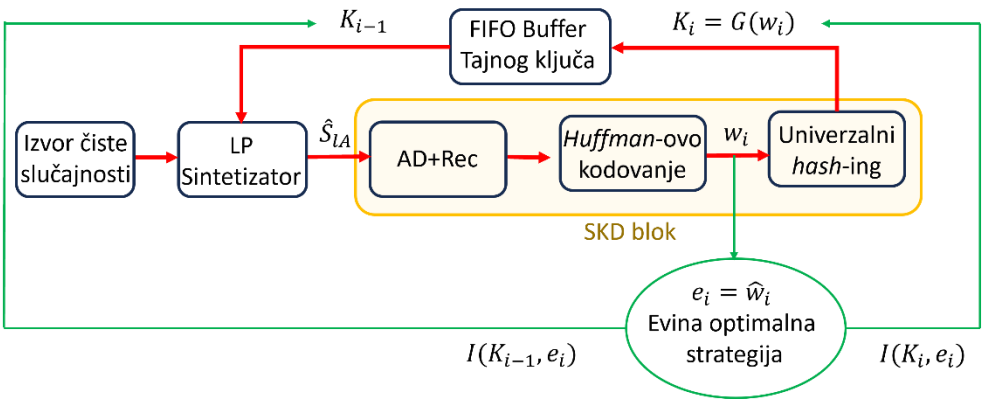
$$Q(z) = (1 - z^{-1}) \prod_{i=1}^{\frac{p}{2}} \left(1 - e^{-j\theta_i} z^{-1}\right) \left(1 - e^{j\theta_i} z^{-1}\right). \tag{7.13}$$

Ako se znaju LSP parametri iz (7.9), zamenom (7.12) i (7.13) u (7.11), a zatim u (7.7), dobija se LP sintetizatorski filtar $H(z)$. Ovo omogućava sintezu signala $\hat{S}_{lA}$ ai $\hat{S}_{lB}$, koji se koriste u SKD protokolu u zaštićenoj fazi rada sistema.

Napomena 8. Ako i samo ako je uslov (7.10) striktno zadovoljen, tj. ako se sve nule polinoma $P(z)$ i $Q(z)$ smenjuju u opsegu $(0, \pi)$, može se pokazati da je LP filter $A(z)$ generisan na ovaj način minimalne faze [86,87]. Stoga, ako se slučajno izabere p LSP-ova iz opsega $(0, \pi)$ koji se zatim sortiraju u rastućem redosledu, nakon čega se sprovede gornja procedura za formiranje polinoma $P(z)$, $Q(z)$ i $A(z)$, rezultujući LP sintetski filter $H(z)$ će biti stabilan.

7.3 Informaciono-teorijska analiza izvora zajedničke slučajnosti zasnovane sa slučajnom izboru LSP parametara

Ako predloženi sistem obezbeđuje zanemarljivo curenje destilovanih tajnih ključeva ka Evi (videti slike 40 i 47), tada istovremeno postoji zanemarljivo curenje tajnih ključeva Vernamove šifre u glavnom kanalu, tj. sistem je u stanju da održi apsolutnu tajnost. SKD protokol na slici 47 se sastoji od BP algoritma za AD fazu, *Winnow* algoritma za IR fazu i optimalnog *Huffman*-ovog kodera. PA faza je zasnovana na univerzalnim heš funkcijama.



Slika 47. Tokovi informacija od značaja za analizu destilovanih tajnih ključeva u sistemu. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Neizvesnost signala $\hat{S}_{lA}$ i $\hat{S}_{lB}$ na izlazu LP sintetizatora potiče od slučajnosti primenjenih LSP koeficijenata, kao i aditivnog čisto slučajnog lokalno generisanog šuma, bez obzira da li se nalazi na ulazu ili izlazu sintetizatora. Zbog zatvorene petlje koja uključuje LP sintetizator, SKD blok, FIFO bafer tajnih ključeva i LP sintetizator, ako bi lokalni izvor čiste slučajnosti imao zanemarljiv doprinos entropiji izlaza LP sintetizatora, efektivna brzina ključa generisanih tajnih ključeva za Vernamovu šifru bi, tokom vremena, težila ka nuli. Formalno, važi:

$$\hat{S}_{lA} = h_{LSP} * \delta + \xi, \tag{7.14}$$

gde je h_{LSP} impulsni odziv LP sintetizatorskog filtra $H(z)$, δ je periodični impulsni ulaz, ξ je aditivni čisti slučajni šum, a $*$ je konvolucioni operator. Na osnovu klasičnih rezultata [88,89], pošto

$$\lim_{z\to\infty} H(z) = 1, \tag{7.15}$$

sledi da sintetizatorski filtar $H(z)$ čuva ulaznu entropiju, tj. da važi

$$H\left(\hat{S}_{lA}\right) = H(h_{LSP}, \delta) + H(\xi), \tag{7.16}$$

bez obzira da li postoji aditivni šum na ulazu ili izlazu filtra $H(z)$. Dalje,

$$H\left(\hat{S}_{lA}\right) = H(h_{LSP}, \delta) + H(\xi) = H(LSP) + H(\xi) \tag{7.17}$$

pošto postoji 1-1 korespondencija između LSP i $\{a_i\}$ parametara filtera $H(z)$ [86].

Napomena 9. Uzimajući u obzir (7.17), jasno je da $H(\xi)$ mora biti značajno dominantna u odnosu na H(LSP) da bi $\hat{S}_{lA}$ održao dovoljan nivo "inovacione" entropije neophodne za destilaciju apsolutno tajnih ključeva za Vernamovu šifru u glavnom kanalu.

Ukupna entropija sintetisanih signala (7.17) može se izraziti kao funkcija odnosa signal-šum (eng. *signal-to-noise ratio* - SNR).

Lema 2. Neka je SNR dat u dB. Tada je entropija šuma $H(\xi)$ po jednom uzorku signala jednaka

$$H(\xi) = \log(2 \cdot A_\xi) + \left[\log\left(A_\xi \cdot 2^{n+1}\right)\right] \tag{7.18}$$

$$A_\xi = \left(3 \cdot \|h_{LSP}\|^2 \cdot 10^{-\frac{SNR}{10}}\right)^{\frac{1}{2}} \tag{7.19}$$

gde je n broj bitova korišćenih za kodovanje uzoraka signala, ξ je šum u intervalu $\left[-A_\xi, A_\xi\right]$, h_{LSP} je impulsni odziv LP sintezatorskog filtera, a $\|\cdot\|$ ima značenje operatora Euklidske norme.

Dokaz. Na osnovu Teoreme 8.3.1 [2], entropija kontinuirane Riemann integrisane slučajne promenljive X, funkcije gustine verovatnoće $f(x)$, kvantizovane sa n bitova je

$$H(X) = -\int f(x) \cdot \log x \, \mathrm{d}x + n. \tag{7.20}$$

Dalje se dokazuje (7.18). Neka je šum ξ uniformno raspoređen u intervalu $\left[-A_\xi, A_\xi\right]$. Tada je prvi član u (7.20) jednak

$$-\int_{-A_\xi}^{A_\xi} \frac{1}{2A_\xi} \cdot \log \frac{1}{2A_\xi} \mathrm{d}x = \log 2A_\xi \tag{7.21}$$

dok je drugi član jednak broju bitova koji kodiraju signal šuma u opsegu $\left[-A_\xi, A_\xi\right]$. Pošto je on jednak broju zauzetih nivoa kvantizacije, dobijamo

$$\left\lfloor \log \frac{2A_\xi}{2^{-n}} \right\rfloor = \left\lfloor \log(A_\xi \cdot 2^{n+1}) \right\rfloor. \tag{7.22}$$

Time je dokazana ispravnost tvrdnje (7.18). Da bi se dokazala tvrdnju (7.19), dovoljno je direktno slediti definiciju SNR [dB], naime

$$SNR = 10 \cdot \log_{10} \frac{E_{h_{SLP}}}{E_\xi} = 10 \cdot \log_{10} \frac{3 \cdot \|h_{LSP}\|^2}{{A_\xi}^2}, \tag{7.23}$$

pošto

$$E_\xi = Var(\xi) = \frac{{A_\xi}^2}{3}. \tag{7.24}$$

Iz (7.23), rešavanjem za A_ξ, dobija se (7.19), čime se završava dokaz. □
Notacija log se uvek odnosi na $\log_2$, osim ako nije eksplicitno navedeno drugačije.

Napomena 10. Na osnovu Leme 2, sledi da se odgovarajućim izborom SNR može kontrolisati veličina inovacione entropije $H(\xi)$ i njena dominacija u odnosu na $H(LSP)$. Treba napomenuti da je $H(LSP)$ fiksna veličina jednaka broju bitova korišćenih za kodovanje LSP parametara. U slučaju MELPe vokodera, njegova odgovarajuća brzina R_{MELPe} (7.6) je jednaka 1.2 kb/s.

U sistemima sa primenom SKD protokola preko klasičnih SCR, osnovni kriterijum kvaliteta sistema je količina informacija $I(K_i, e_i)$ koju Eva može dobiti o generisanim tajnim ključevima nakon komunikacije preko javnog kanala. Formalno,

$$I(K_i, e_i), \;\; e_i = \widehat{w}_i, K_i = G(w_i) \tag{7.25}$$

gde je w_i slučajan n-bitni string sa uniformnom raspodelom nad $\{0,1\}^n$ na izlazu optimalnog *Huffman*-ovog kodera [43], e_i je vrednost optimalne Evine procene w_i, dok je $K_i = G(w_i)$ destilovani tajni ključ. G je izabrana slučajno iz univerzalne klase heš funkcija iz $\{0,1\}^n$ u $\{0,1\}^{|K_i|}$ [29]. Prema dobro poznatim rezultatima iz [16], konkretno Napomena 10, Evine informacije o K_i za specifične e_i i G se eksponencijalno smanjuju po višku kompresije $c - |K_i|$

$$I(K_i; G, e_i) = H(K_i) - H(K_i^*|G, e_i) \leq \frac{2^{-(c-|K_i|)}}{\ln 2}, \tag{7.26}$$

gde je c donja granica Evine uslovne *Renyi*-jeve entropije drugog reda za w_i, odnosno važi.,

$$R(w_i|e_i) \geq c. \tag{7.27}$$

Međutim, pošto u APS-VCS sistemu, $\hat{S}_{lA}$ i $\hat{S}_{lB}$ zavise od prethodno destilovanih tajnih ključeva, potrebno je ispitati da li su Evine informacije $I(K_{i-1}, e_i)$ o K_{i-1} takođe zanemarljivo male. Prema (7.26), važi

$$I(K_i^*; G, e_i, K_{i-1}) = H(K_i^*) - H(K_i^*|G, e_i, K_{i-1}) \leq \frac{2^{-(R^*(w_i|e_i, K_{i-1}) - |K_i^*|)}}{\ln 2} \leq \frac{2^{-(c^* - |K_i^*|)}}{\ln 2}, \tag{7.28}$$

gde je c^* donja granica Evine uslovne *Renyi*-jeve entropije drugog reda za w_i, odnosno,

$$R^*(w_i|e_i, K_{i-1}) \geq c^*. \tag{7.29}$$

Treba napomenuti da se u (7.28) sa K_i^* označavaju destilovani tajni ključevi kada Eva poseduje neke informacije o K_{i-1}. Pošto će ova činjenica uticati na njenu optimalnu strategiju, i samim tim na dužinu destilovanih ključeva, u opštem slučaju $|K_i^*| \neq |K_i|$.
Stoga, optimalna PA strategija mora da se zasniva na Evinoj uslovnoj *Renyi*-jevoj entropiji, koja je

$$\min\{R(w_i|e_i), R^*(w_i|e_i, K_{i-1})\}, \tag{7.30}$$

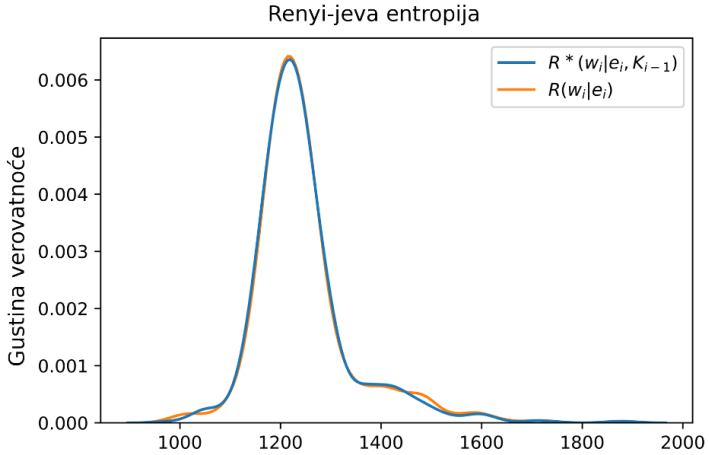
ili u smislu njihovih minimalnih vrednosti

$$\min\{c, c^*\}. \tag{7.31}$$

Napomena 11. Ako bi uslovne *Renyi*-jeve entropije $R(w_i|e_i)$ i $R^*(w_i|e_i, K_{i-1})$ bile identične ili neznatno različite ($c \approx c^*$), to bi značilo da Evine informacije o ključu K_{i-1} ne utiču na njene informacije o destilovanom ključu K_i, i da, prema (7.28) i (7.29), ove informacije opadaju eksponencijalno u višku kompresije $c^* - |K_i^*| \approx c - |K_i|$.

Da li je ovo tačno ili ne za APS-VCS, testirano je empirijski, procenjujući ove dve distribucije u eksperimentu sa 1000 lokalno sintetisanih signala $\hat{S}_{lA}$ i $\hat{S}_{lB}$ dužine 540 uzoraka kodiranih sa 16 bitova. SKD protokol se sastoji od BP algoritma za AD fazu, *Winnow* algoritma za IR fazu praćenog optimalnim *Huffman*-ovim koderom i univerzalnim heširanjem, videti sliku 47. Izbor i optimizacija algoritama, kao i definisanje ključnih parametara SKD protokola, izvršeni su u prethodnim poglavljima ove disertacije, gde je težište bilo na postizanju maksimalne brzine generisanja ključa uz minimalno curenje informacija. U skladu sa tim prethodno utvrđenim postavkama, AD algoritam se primenjuje kroz dve iteracije, čime se nivo greške svodi na meru koju *Winnow* algoritam može efikasno da ispravi. Za potrebe IR faze usvojen je *Winnow* algoritam sa dužinom bloka od 8 bita, za koji je ranije u radu pokazano da pruža optimalne rezultate u pogledu minimizacije informacija dostupnih prisluškivaču i brzine generisanja tajnih ključeva. Konačno, u okviru univerzalne klase heš funkcija, koristi se binarna matrica sa Toeplitz strukturom, odabrana zbog njene niske računske složenosti.

Slika 48 prikazuje distribucije uslovnih *Renyi*-jevih entropija $R(w_i|e_i)$ i $R^*(w_i|e_i, K_{i-1})$ i to, plavom linijom je označena funkcija gustine verovatnoće uslovne *Renyi*-jeve entropije $R^*(w_i|e_i, K_{i-1})$ kada su slučajni LSP parametri Alisinih, Bobovih i Evinih sintetizatora isti, tj., $LSP_A = LSP_B = LSP_E$ dok je narandžastom linijom označena funkcija gustine verovatnoće uslovne *Renyi*-jeve entropije $R(w_i|e_i)$ kada je $LSP_A = LSP_B \neq LSP_E$. U Tabeli 13 prikazane su njihove srednje vrednosti i varijanse. Može se zaključiti da su distribucije gotovo identične i da ekstremno male razlike potiču od inherentnih svojstava slučajnih eksperimenata na konačnim uzorcima.



Slika 48. Funkcija gustine verovatnoće uslovne *Renyi*-jeve entropije u slučaju kada Alisa, Bob i Eva imaju isti *Huffman*-ov koder. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Tabela 13. Srednje vrednosti i standardne devijacije odgovarajućih *Renyi*-jevih entropija sa slike 13.

Renyi entropija	Srednja vrednost ± Standardna devijacija
$R^*(w_i\|e_i, K_{i-1})$	1244.01 ± 97.84
$R(w_i\|e_i)$	1246.66 ± 101.49

Napomena 12. Predstavljena teorijska analiza i eksperimentalna verifikacija omogućavaju zaključak da je predloženi APS-VCS sistem otporan na napade na sadržaj FIFO memorija. Ovo svojstvo logički sledi iz svojstava PA zasnovane na univerzalnim heš funkcijama, kao i dovoljne količine inovacione entropije koja osvežava informacioni sadržaj sintetisanih signala $\hat{S}_{lA}$ i $\hat{S}_{lB}$. Stoga je odgovor na pitanje da li postoji strategija Eve koja joj pruža prednost u odnosu na SKD sisteme zasnovane na klasičnim SCR negativan.

7.4 PA strategija zasnovana na *Huffman-Renyi* rastojanju

Kao što je prethodno objašnjeno, kako bi se efikasno iskoristio određeni SCR, potrebno je adaptivno odrediti stepen kompresije PA bloka. Na ovaj način se brzina generisanih tajnih ključeva prilagođava informacijama o ključevima koje su dostupne Evi. Složeni sistemi mašinskog učenja razvijeni za ove svrhe mogu se zameniti jednostavnijim ali još uvek vrlo efikasnim procedurama za određene klase SCR. Slika 49 prikazuje histogram razlike

$$D_{HR} = |w_i| - R(w_i|e_i) \tag{7.32}$$

između dužine sekvence $|w_i|$ na izlazu *Huffman*-ovog kodera i uslovne *Renyi*-jeve entropije $R(w_i|e_i)$ te iste sekvence posmatrane od strane Eve za sintetisane signale sa SNR = 39.9 dB. Veličina D_{HR} nazva se *Huffman-Renyi* rastojanje. Uočava se da je srednja vrednost veoma blizu 0, preciznije 1.84 bita, i da postoji zanemarljiv broj uzoraka van opsega od $3\sigma = 9.02$ bita. Na osnovu (7.32), može se izvesti jednostavan estimator

$$\hat{R}(w_i|e_i) = |w_i| - \hat{D}_{HR}. \tag{7.33}$$

Iz (7.33) vidi se da sa kvalitetnom procenom za $\hat{D}_{HR}$ i poznavanjem $|w_i|$, takođe može se dobiti kvalitetna procena za $R(w_i|e_i)$. Ako se postavi da je $\hat{D}_{HR}$ jednak srednjoj vrednosti $\overline{D}_{HR}$, tada sa visokom verovatnoćom

$$\hat{R}(w_i|e_i) \geq |w_i| - \overline{D}_{HR} - 3 \cdot \hat{\sigma}_{D_{HR}}. \tag{7.34}$$

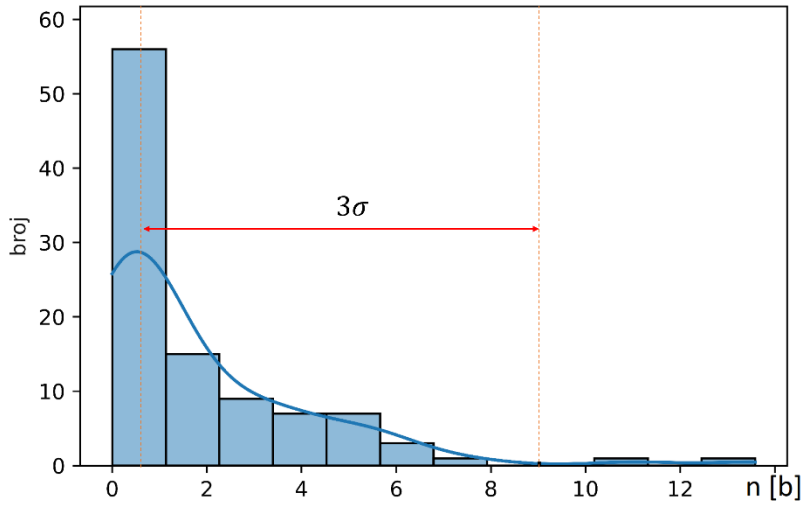
Imajući u vidu da je stepen kompresije PA bloka jednak $\hat{R}(w_i|e_i) - |K_i|$, dolazi se do tri moguće PA strategije:

$$|K_i| = |w_i| - \overline{D}_{HR}, \qquad \text{“mean” srednja} \tag{7.35}$$

$$|K_i| = |w_i| - \overline{D}_{HR} - 3 \cdot \hat{\sigma}_{D_{HR}} \qquad \text{“}3\sigma\text{”} \tag{7.36}$$

$$|K_i| = |w_i| - \overline{D}_{HR} - 3 \cdot \hat{\sigma}_{D_{HR}} - s \qquad \text{“}3\sigma + s\text{”} \tag{7.37}$$

Strategije su poređane prema rastućem stepenu kompresije. Strategija (7.37) omogućava izbor bezbednosne margine s prema unapred definisanoj vrednosti brzine curenja informacija ka Evi.

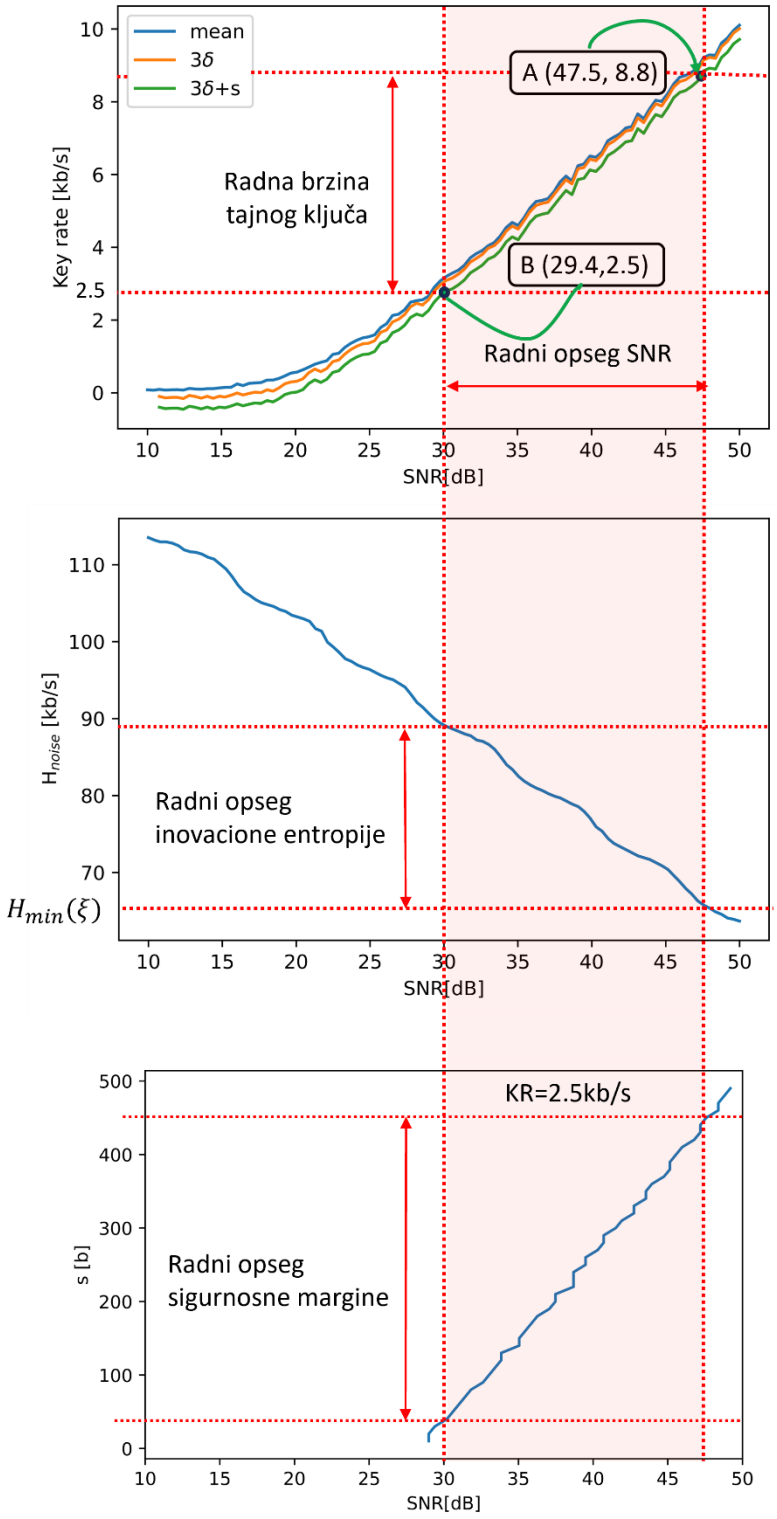


Slika 49. Empirijska distribucija *Huffman-Renyi* rastojanja D_{HR} ($\overline{D}_{HR} = 1.84, \hat{\sigma}_{D_{HR}} = 2.39$) za sintetisane signale sa parametrom SNR = 39.9 dB. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

7.5 Eksperimentalna evaluacija predloženog sistema

U prvom koraku sinteze APS-VCS sistema, potrebno je izabrati operativne vrednosti za glavne sistemske parametre, kao što su brzina generisanja tajnog ključa (eng. *secret key rate*), SNR sintetisanih signala, brzina inovacione entropije i sigurnosna margina. Slika 50 prikazuje međuzavisnosti operativnih opsega ovih veličina, dobijene na realnom APS-VCS sistemu, usrednjavanjem 100 vrednosti za svaku vrednost SNR u opsegu od 10 do 50 dB. Destilacija tajnih ključeva je izvršena korišćenjem tri različite PA strategije: srednja (plava linija), 3σ (narandžasta liniija), i $3\sigma + s$, $s = 20$ (zelena linija). Ako se PA strategija $3\sigma + s$ uzme kao referentna, i KR je najmanje 2.4 kb/s, dobija se za SNR [dB] radni opseg [29.4, 47.5], inovativna entropija [kb/s] [65, 88], KR = 2.5 kb/s, i sigurnosna margina [b] [70, 460]. Redosled izbora je sledeći (videti sliku 50):

- Bira se željeni KR. Uz napomenu da mora zadovoljiti ograničenja (7.5) i (7.6).
- Sigurnosna margina s se bira u skladu sa zahtevima ukupne bezbednosti sistema. Uzimajući u obzir (7.26) i (7.28), sa povećanjem s, stepen kompresije u PA bloku se povećava, i samim tim, informacije koje Eva može dobiti o generisanim ključevima opadaju eksponencijalno.
- Izbor sigurnosne margine s jedinstveno određuje SNR.
- Dobijena vrednost za SNR jedinstveno određuje inovacionu entropiju.

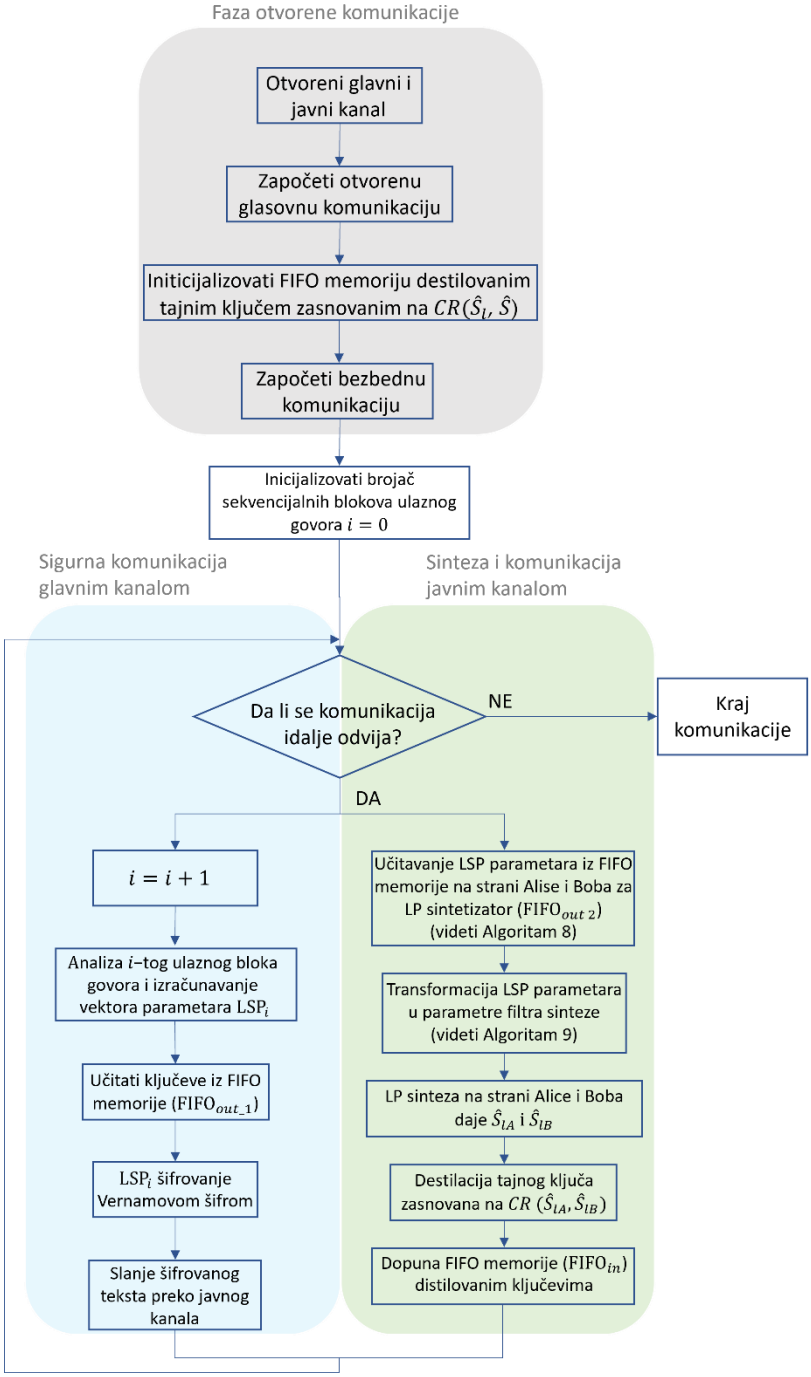


Slika 50. Međuzavisnosti operativnih opsega glavnih veličina: brzina generisanja tajnog ključa, SNR sintetisanih signala, brzina inovacione entropije i sigurnosna margina u sintezi APS-VCS sistema. Zavisnost sigurnosne margine od SNR je prikazana za brzinu tajnog ključa od 2.5 kb/s. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Napomena 13. Pošto za svaki unapred fiksiran KR bezbednosna margina može biti u širokom opsegu, dizajner sistema ima veliku slobodu da lako izabere parametre sintetizatora koji će istovremeno zadovoljiti zahteve za brzinom generisanja tajnih ključeva, njihovom maksimalnom entropijom, neponovljivošću i zanemarljivim curenjem informacija ka Evi. Svi ovi elementi

potvrđuju osnovne zahteve koje moraju ispuniti tajni ključevi Vernamove šifre kako bi se održala njena apsolutna tajnost.

Slika 51 prikazuje funkcionalni opis rada APS-VCS sistema. Podsistem za generisanje tajnih ključeva počinje da radi nakon završetka faze otvorene komunikacije. Neophodan uslov za funkcionisanje ovog podbloka je početno uspešno punjenje FIFO memorija na strani Alise i Boba sa destilisanim tajnim ključevima od najmanje 160 bitova. Prvih 80 bitova će se koristiti kao tajni ključ za šifrovanje LSP prvog bloka ulaznog govornog signala, dok će sledećih 80 bitova biti korišćeno za LP sintetizator u SKD bloku.



Slika 51. Funkcija APS-VCS sistema u prenosnom režimu. Funkcija sistema u režimu prijema je fundamentalno iste strukture. Adaptirano iz rada [82], licencirano pod Creative Commons Attribution (CC BY 4.0) licencom.

Algoritam 8 pruža detaljno objašnjenje procesa transformacije binarne sekvence u LSP parametre, dok Algoritam 9 nudi detaljno objašnjenje transformacije LSP parametara u parametre LP sintetskog filtera $H(z)$. Tabela 14 prikazuje rezultat eksperimentalne evaluacije destilisanih tajnih ključeva zasnovanih na izvoru $CR(\hat{S}_l,\ \hat{S}$ tokom otvorene faze komunikacije. Pošto je prosečna vrednost KR=12.88 kb/s, samo 1 sekunda otvorene komunikacije u proseku puni FIFO memorije sa celih 12.88 kb, što daleko prevazilazi potrebnih 160 bitova. Treba napomenuti da brzina prihvatanja ključa (KAR) ne mora imati maksimalnu vrednost od 100%, što je važan indikator efikasnosti SKD protokola u uobičajenim primenama [76]. Izmerena vrednost razmenjenih bitova na javnom kanalu (prosek - 6175 bitova, maksimum - 7063 bita) po bloku od 8640 bitova pokazuje značajno curenje informacija o ulaznom govoru, što nije sigurnosna pretnja jer nije šifrovan u ovoj fazi. Brzina curenja tajnih ključeva korišćenih za punjenje početnog sadržaja FIFO memorija je samo 0.0012 b/b. Ova vrednost može biti smanjena na zahtev dizajnera uvođenjem dodatne sigurnosne margine, koja može biti reda veličine nekoliko stotina bitova (videti sliku 50).

Algoritam 8. Transformacija binarne sekvence (FIFO_{out_2}) u LSP parametre

Ulaz: Binarna sekvenca
Output: LSP parametri

1: Čitanje slučajne sekvence E_k iz FIFO, $|E_k| = L$
2: Podela E_k na p podsekvenci, odnosno, $E_k = [E_{k1}, E_{k2}, \dots E_{kp}]$, $|E_{ki}| = \left\lfloor \frac{L}{p} \right\rfloor, i = 1, \dots p$
3: Transformacija svakog E_{ki} u LSP_i skaliranjem decimalne vrednosti E_{ki} faktorom $\frac{\pi}{2^{\left\lfloor \frac{L}{p} \right\rfloor}-1}$
4: Sortiranje dobijenih LSP parametara prema (7.10)

Algoritam 9. Transformacija LSP parametara u LP parametre filtra sinteze H(z)

Ulaz: LSP parametri
Izlaz: H(z) parametri

1: Izračunati P(z) prema (7.12)
2: Izračunati Q(z) prema (7.13)
3: Izračunati LP filtar prema (7.11)
4: Izračunati LP filtar sinteze prema (7.7)

Tabela 14. Rezultati SKD protokola izvršenog preko 24 govornika u prvoj otvorenoj fazi komunikacije, izmereni na vPCP-V sistemu.

Mere performansi	Srednja vrednost
KR [kb/s]	12.88
KR [%]	9.98
KAR [%]	51.93
Srednja vrednost broja bita parnosti [po bloku]	6175
Maksimalni broj bita parnosti [po bloku]	7063
LR [b/b]	0.0012

Nakon faze otvorene komunikacije i uspešnog punjenja FIFO memorija početnim sadržajem, sistem prelazi na sigurnu komunikaciju. Istovremeno se izvršavaju sinteza i komunikacija preko javnog kanala. Destilacija tajnih ključeva zasnovana na izvoru $CR(\hat{S}_{lA},\hat{S}_{lB})$ sada se odvija. U Tabeli

15 prikazani su rezultati ovog SKD protokola izvršenog preko 1000 blokova i izmereni na vPCP-V sistemu. Pošto je grafički prikaz ovih rezultata dat na slici 50, Tabela 15 prikazuje numeričke rezultate samo za pet karakterističnih SNR, u opsegu od 10 do 50 dB. Primećuje se da u već pomenutom radnom opsegu [29.4, 47.5], KAR ne pada ispod 100%, dok je LR reda veličine 0.0011-0.0042 b/b. Predstavljeni rezultati pokazuju da SKD preko izvora $CR(\hat{S}_{lA},\hat{S}_{lB})$ sa velikim sigurnosnim marginama obezbeđuje stabilno osvežavanje FIFO memorija sa novogenerisanim tajnim ključevima, koji se zatim koriste kao tajni ključevi Vernamove šifre u glavnom kanalu.

Tabela 15. Rezultati SKD protokola izvršenog preko 1000 blokova u sigurnoj fazi komunikacije, izmereni na vPCP-V sistemu.

Veštački sintetizator baziran na LSP-u					
SNR [dB]	50	39.90	29.80	19.70	10
KR [kb/s]	9.96 ± 1.54	6.44 ± 1.35	3.04 ± 1.06	0.53 ± 0.47	0.08 ± 0.08
KR [%]	7.78	5.03	2.38	0.42	0.05
KAR [%]	100	100	100	75	7
LR [b/b]	0.0011	0.0015	0.0042	0.0399	0.0191

Kao što je primećeno na slici 50, sa povećanjem SNR sintetizatora, KR se povećava bez obzira na primenjenu PA strategiju. Ovo ponašanje se može lako objasniti na osnovu modela (7.14) sintetisanih signala, prema kojem se međusobna korelacija između $\hat{S}_{lA}$ i $\hat{S}_{lB}$ povećava sa povećanjem SNR; odnosno, sa smanjenjem uticaja lokalnog šuma ξ u odnosu na deterministčku komponentu $h_{LSP} * \delta$. Smanjenje inovacione entropije sa povećanjem SNR sledi isti mehanizam: kako se SNR povećava, udeo šuma ξ u ukupnom sintetisanom signalu se smanjuje, i samim tim, odgovarajuća inovaciona entropija opada.

Povećanje sigurnosne margine s sa rastućim SNR je direktno povezano sa njenom definicijom kao razlike $c - |K_i|$, gde c predstavlja donju granicu Evine uslovne Rényi entropije drugog reda, a $|K_i|$ je stvarna dužina destilovanog tajnog ključa; videti (7.26). Pošto c direktno određuje maksimalni KR određenog sistema (viši c vodi do višeg maksimalnog destilovanog KR), jasno je da će varijacija sigurnosne margine slediti istu zavisnost sa promenama u SNR, kao što je takođe primećeno na slici 50. Stoga se može zaključiti da eksperimentalna evaluacija potvrđuje teorijski očekivane rezultate, kako u pogledu brzine ključa tako i u promeni inovacione entropije i sigurnosne margine u odnosu na SNR modela (7.14).

Predstavljeni redosled izbora ključnih sistemskih parametara KR, SNR i s može se takođe interpretirati na sledeći način. Radna tačka B na slici 50 se dobija na preseku zavisnosti KR od SNR i željene vrednosti KR. Ova tačka određuje donju granicu za SNR. Tačka A na slici 50 se dobija određivanjem maksimalno dozvoljenog SNR na osnovu minimalno dozvoljene inovacione entropije. Dozvoljeni interval varijacije SNR direktno diktira opseg mogućih vrednosti za KR, inovacionu entropiju i sigurnosnu marginu s.

Napomena 14. Pošto se SKD preko izvora $CR(\hat{S}_{lA}, \hat{S}_{lB})$ izvršava nezavisno od rada sistema na glavnom kanalu, njegovi parametri, kao što su brzina odabiranja i rezolucija sintetisanih signala, mogu biti gotovo proizvoljno različiti, omogućavajući brzine generisanja tajnih ključeva u mnogo širem opsegu vrednosti. Jedini ograničavajući faktor brzine destilacije tajnih ključeva je komunikacioni kapacitet javnog kanala. Stoga, predloženi APS-VCS sistem može pouzdano raditi na drugim standardnim brzinama vokodera (2.4 kb/s, 4.8 kb/s) sa odgovarajućom propusnom širinom javnog kanala.

Cena plaćena za apsolutnu tajnost APS-VCS sistema je uspostavljanje i održavanje javnog kanala tokom kripto zaštićenog razgovora. Međutim, glavni i javni kanali rade u asinhronom režimu,

što značajno pojednostavljuje praktičnu implementaciju. Jedini uslov koji mora biti ispunjen je održavanje konstantne brzine čitanja FIFO memorije od 1.2 kb/s za Vernamovu šifru.

Tabela 16 prikazuje BER za četiri tipična komunikaciona kanala, sa i bez upotrebe koda za korekciju grešaka (eng. *Error-Correcting Code* - ECC). Za ECC se koristi Golay(12,24), koji je specifično dizajniran za zaštitu 15% najosetljivijih bitova binarne reprezentacije LSP parametara koji podležu šifrovanju. Važno je napomenuti da prva dva tipa GSM kanala daju izuzetno dobre rezultate, uzimajući u obzir uticaj bloka ulazne kompresije u GSM uređajima. Ovi eksperimentalni rezultati potvrđuju suštinsku funkcionalnost koju sistem za zaštitu glasa visokog nivoa bezbednosti mora da ispuni [90]. Rezultati pokazuju da je BER prikazan u Tabeli 16 nezavisan od SKD sistema, pod uslovom da su parametri sintetizatora izabrani tako da garantuju brzinu generisanja ključa veću od 2.4 kb/s.

Tabela 16. BER za različite audio kanale sa i bez ECC.

Audio kanal	BER bez FEC	BER sa FEC
GSM 3G	$1.20 \cdot 10{-}3$	$1.02 \cdot 10{-}3$
GSM VoLTE	$1.50 \cdot 10{-}3$	$1.20 \cdot 10{-}3$
WhatsApp	$5.50 \cdot 10{-}3$	$4.70 \cdot 10{-}3$
Google meet	$7.90 \cdot 10{-}3$	$3.40 \cdot 10{-}3$

Predloženi APS-VCS sistem nudi značajne prednosti u aplikacijama iz realnog sveta gde tradicionalne kriptografske metode ne uspevaju da obezbede i autonomiju i apsolutnu tajnost. U nastavku su ključni scenariji gde APS-VCS prevazilazi postojeća rešenja:

1. Vojne i vladine komunikacije. APS-VCS eliminiše potrebu za eksternom pouzdanom infrastrukturom za distribuciju ključeva, čineći ga idealnim za vojne i vladine operacije gde su potrebni visoka bezbednost i operativna autonomija. Za razliku od sistema zasnovanih na QKD, koji zahtevaju specijalizovanu optičku infrastrukturu, APS-VCS radi preko postojećih digitalnih i mobilnih mreža, pružajući potpuno sigurnu glasovnu komunikaciju u realnom vremenu čak i u udaljenim ili neprijateljskim okruženjima. Rat u Ukrajini može poslužiti kao svež i relevantan primer potencijalne primene i značaja APS-VCS. Kombinacija Starlink-a kao otpornog i široko dostupnog javnog kanala i APS-VCS kao sigurnog komunikacionog sistema omogućava vojnim jedinicama da održe komandnu koordinaciju čak i u najtežim okolnostima bez straha od prisluškivanja ili dešifrovanja od strane protivnika.
2. Bezbednost u tajnim misijama. Tradicionalni sigurni komunikacioni uređaji čuvaju unapred distribuirane tajne ključeve, pa ako neprijatelj zarobe uređaj, ceo sistem enkripcije bi mogao biti kompromitovan. U obaveštajnim, anti-pobunjeničkim ili tajnim operacijama, APS-VCS obezbeđuje da nijedan osetljiv materijal tajnog ključa nije uskladišten ili nošen od strane terenskog operativca. Ako je operativac zarobljen ili je prebegao, nikakve informacije o tajnom ključu ne mogu biti ekstrahovane da bi se kompromitovale tekuće operacije.
3. Prilagodljivost bez potrebe za prethodnom. Za razliku od tradicionalnih sigurnosnih sistema koji zahtevaju prethodnu distribuciju ključeva (što može biti logistički izazovno i rizično), APS-VCS omogućava korisnicima da dinamički uspostave sigurne komunikacije. Ovo ga čini idealnim za brzo promenljive parametre misije gde nove komunikacione čvorove možda treba integrisati bez fizičke razmene ključeva
4. Industrijska i korporativna bezbednost. Preduzeća koja se bave osetljivom intelektualnom svojinom ili poslovnim tajnama često se oslanjaju na šifrovane komunikacione kanale koji zavise od konvencionalne PKI (*Public Key Infrastructure*) infrastrukture. APS-VCS eliminiše potrebu za upravljanjem ključevima preko eksternih strana, sprečavajući

potencijalne unutrašnje pretnje i sigurnosne prekršaje povezane sa centralizovanim skladištenjem ključeva za enkripciju.

5. Taktičke i hitne službe. Timovi za hitnu intervenciju zahtevaju sigurne sisteme za glasovnu komunikaciju koji funkcionišu nezavisno od centralizovane infrastrukture, posebno u katastrofalnim scenarijima gde konvencionalne mreže mogu biti kompromitovane. APS-VCS pruža pouzdan, autonoman sistem enkripcije koji obezbeđuje potpunu tajnost komunikacija između prvih koji reaguju, policije i timova za upravljanje u kriznim situacijama.

Ukazujući na ove praktične primene, APS-VCS demonstrira jasne prednosti u odnosu na postojeće kriptografske metode, posebno u scenarijima gde su nezavisnost od infrastrukture, apsolutna tajnost i bezbedna govorna komunikacija u realnom vremenu ključni zahtevi.

8 Zaključak i budući rad

8.1 Ključni rezultati i doprinosi disertacije

U okviru ove disertacije razvijen je informaciono-teorijski utemeljen pristup analizi i sintezi SKD protokola zasnovanih na SCR poteklih od entropijski bogatih biometrijskih izvora kao što su EEG i govor. Razvijena metodologija je praktično primenjena u dizajnu autonomnog sistemima apsolutno tajne govorne komunikacije zasnovanog na MELPe vokoderu. Glavni naučni doprinosi i praktični značaj ostvarenih rezultata mogu se sažeti na sledeći način:

1. Razvijena je teorijsko-empirijska metodologija analize i projektovanja PA bloka u SKD sistemima, koja omogućava kvantitativnu procenu iskorišćenja zajedničke slučajnosti, curenja informacija ka prisluškivaču i stvarne bezbednosne margine generisanih tajnih ključeva (Teoreme 1-5, poglavlje 4). Osnovu ovog pristupa predstavlja lokalna pouzdana estimacija ECRE2 na strani legitimnih korisnika protokola.
2. Pokazano je da govorni signal može predstavljati efikasan izvor zajedničke slučajnosti za SKD sisteme, čime je identifikovan novi biometrijski izvor pogodan za destilovanje tajnih ključeva na visokim brzinama, uz zadržavanje informaciono-teorijskih bezbednosnih garancija (Lema 1-2, poglavlje 7).
3. Predložen je adaptivni mehanizam za maksimizaciju iskorišćenja zajedničke slučajnosti, zasnovan na slabljenju pozicije prisluškivača i minimizaciji neiskorišćene slučajnosti. Ovaj mehanizam je realizovan adaptivnim određivanjem maksimalne dužine tajnog ključa za dati dozvoljeni nivo curenja informacija, posredstvom estimiranih vrednosti za ECRE2.
4. Razvijen je blok pojačanja privatnosti zasnovan na metodama mašinskog učenja (PIDNN), koji funkcioniše isključivo na osnovu lokalnih informacija dostupnih učesnicima protokola i ne zahteva dodatnu razmenu podataka preko javnog kanala.
5. Uspostavljen je pouzdan metod merenja stepena curenja informacija, uz dokaz optimalnosti strategije prisluškivača i dokazivo ograničenje njegove informacije o generisanim tajnim ključevima sa unapred zadatim nivoom pouzdanosti.
6. Analizom klasičnih PA strategija pokazano je da pojedini široko primenjivani pristupi (npr. LHL) mogu imati značajno curenje informacija, uprkos prolasku standardnih testova slučajnosti, dok predložene ML i hibridne strategije ostvaruju superiorne performanse u pogledu efikasnosti i bezbednosne margine.
7. Na osnovu predloženih metoda projektovan je autonomni APS-VCS sistem apsolutno tajne govorne komunikacije, zasnovan na MELPe vokoderu brzine 1,2 kb/s i Vernamovoj šifri, u kome se tajni ključevi generišu i distribuiraju korišćenjem dva izvora zajedničke slučajnosti i SKD protokola uz dodatni autentifikovani javni kanal.
8. Eksperimentalnim ispitivanjem pokazano je da sistem omogućava postizanje željene brzine generisanja ključa od 1,2 kb/s uz maksimalnu bezbednosnu marginu od približno 460 bita i zanemarljivo curenje informacija, kao i robusno funkcionisanje preko različitih komunikacionih kanala, uključujući GSM 3G i VoLTE.
9. Praktična analiza ukazuje da APS-VCS predstavlja posebno pogodno rešenje za primene koje zahtevaju autonomiju sistema i apsolutnu tajnost komunikacije, kao što su vojne i državne komunikacije, prikrivene i krizne operacije, kao i sistemi bezbedne govorne komunikacije u uslovima ograničene ili kompromitovane kritične infrastrukture.

8.2 Ograničenja predloženog rešenja

Iako predloženi metodološki pristup i sistemi ostvaruju visok nivo informaciono-teorijske bezbednosti i praktične primenljivosti, određena ograničenja su neizbežna i proizlaze kako iz fundamentalnih teorijskih pretpostavki, tako i iz praktičnih aspekata implementacije:

1. Predloženi SKD sa javnom diskusijom zahteva postojanje dodatnog pouzdanog i autentifikovanog javnog kanala. U radu je razmatran scenario pasivnog napadača, zbog čega se pretpostavlja postojanje autentifikovanog javnog kanala koji omogućava proveru identiteta učesnika komunikacije i sprečava umetanje, izmenu ili lažno predstavljanje poruka od strane napadača tokom javne diskusije. Iako se ovaj zahtev može smatrati prihvatljivim u većini realnih komunikacionih scenarija, on predstavlja infrastrukturno ograničenje u okruženjima gde autentifikaciju javnog kanala nije moguće obezbediti ili gde su komunikacioni resursi strogo ograničeni. *Maurer* i *Wolf* [27] pokazuju da je informaciono-teorijski bezbedna destilacija tajnog ključa moguća i preko neautentifikovanog javnog kanala, kombinacijom particionisanja parcijalno tajnog ključa, autentifikacije sa parcijalno tajnim ključem i interaktivnih *challenge-response* tehnika. Cena neautentifikovanog kanala se ogleda u dvema strožim zahtevima u odnosu na pasivni (autentifikovani) slučaj: *Renyi*-jeva entropija parcijalno tajnog ključa mora biti veća od dve trećine njegove dužine, a dužina destilovanog ključa je proporcionalno smanjena. Drugim rečima, nedostatak autentifikovanosti kanala se nadoknađuje kraćim finalnim ključem, bez narušavanja informaciono-teorijske bezbednosti.
2. Performanse predloženih SKD sistema direktno zavise od kvaliteta i statističkih svojstava SCR. U slučajevima kada SCR ne obezbeđuje dovoljnu količinu inovativne entropije ili kada dolazi do povećane korelacije između sekvenci legitimnih učesnika i prisluškivača, maksimalna brzina destilacije tajnih ključeva može biti značajno smanjena. Ovo ograničenje je posebno izraženo kod prirodnih i biometrijskih izvora, čija statistička svojstva mogu varirati u vremenu i zavisiti od konteksta upotrebe.
3. Primena adaptivnih mehanizama zasnovanih na metodama mašinskog učenja uvodi dodatnu složenost u projektovanje i verifikaciju sistema. Iako su predložene ML-strategije projektovane tako da ne narušavaju informaciono-teorijske bezbednosne garancije, njihova obuka zahteva pažljiv izbor skupova podataka i parametara, kao i dodatne računske resurse u fazi dizajna sistema.
4. Integracija SKD sistema u vokoderske sisteme malih brzina prenosa nameće stroge zahteve u pogledu sinhronizacije, kašnjenja i stabilnosti rada u realnom vremenu. U nepovoljnim uslovima kanala, povećana greška koji unosi kanal, može uticati na efikasnost usaglašavanja informacija i, posredno, na brzinu destilacije ključeva, iako ne narušava osnovna bezbednosna svojstva sistema.
5. Predloženi pristup podrazumeva kompromis između postizanja savršene tajnosti i složenosti sistema. U poređenju sa klasičnim računarski bezbednim šemama, informaciono-teorijski bezbedni sistemi zahtevaju veće količine slučajnosti, dodatne protokole i precizno projektovanje svih faza obrade, što može ograničiti njihovu primenu u izuzetno resursno ograničenim uređajima.

Navedena ograničenja ne umanjuju validnost i značaj predloženih rešenja, već ukazuju na realne granice njihove primene i predstavljaju polaznu osnovu za dalja istraživanja i unapređenja, od kojih su neka razmatrana u narednom odeljku.

8.3 Mogući pravci za dalja istraživanja

Na osnovu identifikovanih ograničenja predloženog rešenja i ostvarenih rezultata disertacije, mogu se definisati sledeći pravci daljih istraživanja:

1. Unapređenje i pojednostavljenje SKD protokola sa javnom diskusijom. U cilju ublažavanja zahteva za dodatnim autentifikovanim javnim kanalom, u vezi sa ograničenjem 1, buduća istraživanja mogu biti usmerena ka razvoju SKD protokola sa smanjenim obimom javne razmene, kao i ka efikasnom multipleksiranju javnog i glavnog komunikacionog kanala. Poseban interes predstavlja analiza minimalnih uslova autentifikacije potrebnih za očuvanje informaciono-teorijske bezbednosti.
2. Napredno modelovanje i adaptivno upravljanje SCR. U skladu sa ograničenjem 2, dalji rad može biti usmeren ka razvoju adaptivnih modela SCR-a koji u realnom vremenu procenjuju entropijska i korelaciona svojstva izvora. Posebnu pažnju treba posvetiti hibridnim SCR-ovima, kod kojih se kombinovanjem prirodnih i determinističkih komponenti omogućava kontrola brzine destilacije i stabilnosti sistema u promenljivim uslovima.
3. Jedinstvene ML arhitekture za celokupni SKD lanac obrade. Kao odgovor na ograničenje 3, perspektivan pravac istraživanja predstavlja razvoj integrisanih arhitektura zasnovanih na mašinskom učenju koje bi mogle zameniti kompletan lanac AD-IR-HC-PA jedinstvenim adaptivnim modelom. Poseban izazov u ovom pravcu jeste očuvanje dokazivih informaciono-teorijskih bezbednosnih garancija u prisustvu naučenih strategija.
4. Zajednički dizajn vokodera i SKD sistema za rad u realnom vremenu. U vezi sa ograničenjem 4, dalja istraživanja mogu se fokusirati na ko-dizajn vokoderskih i kriptografskih komponenti sistema, sa ciljem smanjenja kašnjenja, povećanja robusnosti i boljeg prilagođavanja promenljivim uslovima kanala. Među nekim od najinteresantnijih mogućnosti izdvajaju se multipleksiranje glavnog i javnog kanala, kao i usavršavanje modela sintetizatora istraživanjem alternativnih metoda ekscitacije (npr. sinteza zasnovana na neuronskim vokoderima). Ovo uključuje razmatranje adaptivnih vokodera i naprednih mehanizama korekcije grešaka koji su specifično optimizovani za SKD sisteme.
5. Hibridne SKD-QKD arhitekture za dugoročnu otpornost sistema. Kao odgovor na fundamentalni kompromis između apsolutne tajnosti i složenosti sistema (ograničenje 5), poseban pravac budućih istraživanja predstavlja integracija klasičnih SKD sistema sa kvantnom distribucijom ključeva. U takvim hibridnim arhitekturama, QKD bi mogao biti korišćen kao dodatni sloj za periodično osvežavanje ključeva, dok bi SKD obezbeđivao autonomiju i praktičnu primenljivost u okruženjima bez kvantne infrastrukture.

Literatura

[1] Shannon, C.th E. (1949). Communication theory of secrecy systems. *Bell System Technical Journal*, *28*(4), 656–715.

[2] Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. Wiley-Interscience.

[3] Maurer, U. M. (1993). Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, *39*(3), 733–742.

[4] Renner, R. (2005). *Security of quantum key distribution* [Doctoral dissertation, ETH Zurich]. arXiv.

[5] Katz, J., & Lindell, Y. (2015). *Introduction to modern cryptography* (2nd ed.). CRC Press.

[6] Gisin, N., Ribordy, G., Tittel, W., & Zbinden, H. (2002). Quantum cryptography. *Reviews of Modern Physics*, *74*(1), 145–195.

[7] Bennett, C. H., & Brassard, G. (1984). Quantum cryptography: Public key distribution and coin tossing. *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (pp. 175–179). IEEE.

[8] Kish, S., Pieprzyk, J., & Camtepe, S. (2026). Trends in quantum key distribution (QKD). In J. Jang-Jaccard, P. Caroff, E. Blezinger, V. Mulder, A. Mermoud, & V. Lenders (Eds.), *Quantum technologies* (pp. 119–131). Springer.

[9] Wyner, A. D. (1975). The wire-tap channel. *Bell System Technical Journal*, *54*(8), 1355–1387.

[10] Ahlswede, R., & Csiszár, I. (1993). Common randomness in information theory and cryptography—Part I: Secret sharing. *IEEE Transactions on Information Theory*, *39*(4), 1121–1132.

[11] Csiszár, I., & Narayan, P. (2004). Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory*, *50*(12), 3047–3061.

[12] Jost, D., Maurer, U. M., & Ribeiro, J. L. (2018). Information-theoretic secret-key agreement. In *Theory of cryptography conference* (TCC 2018). IACR ePrint.

[13] Jana, S., Premnath, S. N., Clark, M., Kasera, S. K., Patwari, N., & Krishnamurthy, S. V. (2009). On the effectiveness of secret key extraction from wireless signal strength in real environments. *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking* (MobiCom '09) (pp. 321–332). ACM.

[14] Bloch, M., & Barros, J. (2011). *Physical-layer security: From information theory to security engineering*. Cambridge University Press.

[15] Ahlswede, R., & Csiszár, I. (1998). Common randomness in information theory and cryptography. II. CR capacity. *IEEE Transactions on Information Theory*, *44*(1), 225–240.

[16] Bennett, C. H., Brassard, G., Crépeau, C., & Maurer, U. M. (1995). Generalized privacy amplification. *IEEE Transactions on Information Theory*, *41*(6), 1915–1923.

[17] Ricci, S., Dobias, P., Malina, L., Hajny, J., & Jedlicka, P. (2024). Hybrid keys in practice: Combining classical, quantum and post-quantum cryptography. *IEEE Access*, *12*, 23206–23219.

[18] Staat, P., Dörpinghaus, M., Sheikholeslami, A., Paar, C., Fettweis, G., & Goeckel, D. (2024). *Key exchange in the quantum era: Evaluating a hybrid system of public-key cryptography and physical-layer security*. arXiv.

[19] Katz, J., & Lindell, Y. (2020). *Introduction to modern cryptography* (3rd ed.). CRC Press.

[20] Gander, M., & Maurer, U. (1994). On the secret-key rate of binary random variables. *IEEE International Symposium on Information Theory* (pp. 351). IEEE.

[21] Liu, S. L. (2002). *Information-theoretic secret key agreement* [Doctoral dissertation, Technische Universiteit Eindhoven]. TU/e Repository.

[22] Bennett, C. H., Bessette, F., Brassard, G., Salvail, L., & Smolin, J. (1992). Experimental quantum cryptography. *Journal of Cryptology*, *5*(1), 3–28.
[23] Brassard, G., & Salvail, L. (1994). Secret-key reconciliation by public discussion. In T. Helleseth (Ed.), *Advances in cryptology—EUROCRYPT '93* (pp. 410–423). Springer.
[24] Maab, H. U., Hussain, I., & Alvi, Z. (2025). Analysis of information reconciliation algorithms with randomness extractors in quantum key distribution post-processing. *Quantum Information Processing*, *24*, 287.
[25] Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, *29*(2), 147–160.
[26] Bennett, C. H., Brassard, G., & Robert, J.-M. (1988). Privacy amplification by public discussion. *SIAM Journal on Computing*, *17*(2), 210–229.
[27] Maurer, U. M., & Wolf, S. (2003). Secret-key agreement over unauthenticated public channels—Part III: Privacy amplification. *IEEE Transactions on Information Theory*, *49*(4), 839–885.
[28] Yakovlev, V., Korzhik, V. I., Morales-Luna, G. B., & Bakaev, M. (2010). *Key distribution protocols based on extractors under the condition of noisy channels in the presence of an active adversary*. arXiv.
[29] Carter, J. L., & Wegman, M. N. (1979). Universal classes of hash functions. *Journal of Computer and System Sciences*, *18*(2), 143–154.
[30] Impagliazzo, R., Levin, L. A., & Luby, M. (1989). Pseudo-random generation from one-way functions. In D. S. Johnson (Ed.), *Proceedings of the 21st Annual ACM Symposium on Theory of Computing* (pp. 12–24). ACM.
[31] Impagliazzo, R., & Zuckerman, D. (1989). How to recycle random bits. *30th Annual Symposium on Foundations of Computer Science* (pp. 248–253). IEEE.
[32] Krawczyk, H. (1994). LFSR-based hashing and authentication. In *Annual International Cryptology Conference* (pp. 129–139). Springer.
[33] Cachin, C., & Maurer, U. M. (1997). Linking information reconciliation and privacy amplification. *Journal of Cryptology*, *10*(2), 97–110.
[34] Radomirović, J., Milosavljević, M., Kovačević, B., & Jovanović, M. (2022). Privacy amplification strategies in sequential secret key distillation protocols based on machine learning. *Symmetry*, *14*(10), 2028.
[35] MacWilliams, F. J., & Sloane, N. J. A. (1977). *The theory of error-correcting codes* (1st ed.). North-Holland.
[36] Radomirović, J., Milosavljević, M., Banjac, Z., & Jovanović, M. (2023). Secret key distillation with speech input and deep neural network-controlled privacy amplification. *Mathematics*, *11*(6), 1524.
[37] Keren, G., Cummins, N., & Schuller, B. (2018). Calibrated prediction intervals for neural network regressors. *IEEE Access*, *6*, 54033–54041.
[38] Kivaranovic, D., Johnson, K., & Leeb, H. (2020). Adaptive, distribution-free prediction intervals for deep networks. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics* (Vol. 108, pp. 4346–4356). PMLR.
[39] Simhayev, E., Katz, G., & Rokach, L. (2021). PIVEN: A deep neural network for prediction intervals with specific value prediction. arXiv.
[40] Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, *22*(3), 337–346..
[41] Elisim. (2022). *PIVEN* [Computer software]. GitHub. https://github.com/elisim/piven
[42] Adamović, S., Miškovic, V., Maček, N., Milosavljević, M., Šarac, M., Saračević, M., & Gnjatović, M. (2020). An efficient novel approach for iris recognition based on stylometric features and machine learning techniques. *Future Generation Computer Systems*, *107*, 144–157.

[43] Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, *40*(9), 1098–1101.
[44] Xu, W., Revadigar, G., Luo, C., Bergmann, N., & Hu, W. (2016). Walkie-talkie: Motion-assisted automatic key generation for secure on-body device communication. In *Proceedings of the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks*. ACM.
[45] Xu, W., Javali, C., Revadigar, G., Luo, C., Bergmann, N., & Hu, W. (2017). Gait-key: A gait-based shared secret key generation protocol for wearable devices. *ACM Transactions on Sensor Networks*, *13*(1).
[46] Guglielmi, A. V., Muraro, A., Cisotto, G., & Laurenti, N. (2021). Information theoretic key agreement protocol based on ECG signals. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM)* (pp. 1–6). IEEE.
[47] Milosavljević, M., Adamović, S., & Jevremović, A. (2019). Secret keys generation from mouse and eye tracking signals. In *Proceedings of the 6th International Conference on Electrical, Electronic and Computing Engineering—IcETRAN 2019*. IEEE.
[48] Galis, M., Milosavljević, M., Jevremović, A., Banjac, Z., Makarov, A., & Radomirović, J. (2021). Secret-key agreement by asynchronous EEG over authenticated public channels. *Entropy*, *23*(10), 1327.
[49] Gungor, O., Chen, F., & Koksal, C. E. (2011). Secret key generation from mobility. In *2011 IEEE GLOBECOM Workshop*.
[50] Gungor, O. (2014). *Information theory enabled secure wireless communication, key generation and authentication* [Doctoral dissertation, The Ohio State University]. OhioLINK Electronic Theses and Dissertations Center.
[51] Nguyen, T. D., Pham, D. T., & Dang, T. K. (2017). On the study of EEG-based cryptographic key generation. *Procedia Computer Science*, 936–943.
[52] Monrose, F., Reiter, M. K., Li, Q., & Wetzel, S. (2001). Cryptographic key generation from voice. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 202–213). IEEE.
[53] Bajwa, G., & Dantu, R. (2016). NeuroKey: Towards a new paradigm of cancelable biometrics-based key generation using electroencephalograms. *Computers & Security*, *62*, 95–113.
[54] Fant, G. (1971). *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations*. De Gruyter Mouton.
[55] Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals*. Prentice Hall.
[56] *Speech commands v0.02* [Data set]. Kaggle. https://www.kaggle.com/datasets/mok0na/speech-commands-v002
[57] Warden, P. (2018). *Speech commands: A dataset for limited-vocabulary speech recognition*. arXiv.
[58] Ye, C., Mathur, S., Reznik, A., Shah, Y., Trappe, W., & Mandayam, N. B. (2010). Information-theoretically secret key generation for fading wireless channels. *IEEE Transactions on Information Forensics and Security*, *5*, 240–254.
[59] Wang, Q., Wang, X., Lv, Q., Ye, X., Luo, Y., & You, L. (2015). Analysis of the information theoretically secret key agreement by public discussion. *Security and Communication Networks*, *8*, 2507–2523.
[60] Yan, H., Ren, S., Chen, Y., & Yang, J. (2017). Information reconciliation protocol in quantum key distribution system. In *2017 4th International Conference on Systems and Informatics (ICSAI)* (pp. 1364–1369).
[61] Elkouss, D., Martinez-Mateo, J., & Martin, V. (2011). Information reconciliation for quantum key distribution. *Quantum Information & Computation*, *11*(3–4), 226–238.
[62] National Institute of Standards and Technology. (2010). *A statistical test suite for random and pseudorandom number generators for cryptographic applications* (NIST Special Publication 800-22 Rev. 1a). U.S. Department of Commerce.
[63] Pierrot, A. J., Chou, R. A., & Bloch, M. R. (2013). *The effect of eavesdropper's statistics in experimental wireless secret-key generation*. arXiv.

[64] Mitev, M., Pham, T. M., Chorti, A., Barreto, A. N., & Fettweis, G. (2022). *Physical layer security-From theory to practice*. arXiv.
[65] Watanabe, S., & Oohama, Y. (2011). Secret key agreement from vector Gaussian sources by rate limited public communication. *IEEE Transactions on Information Forensics and Security*, *6*, 541–550.
[66] Schneier, B. (1996). *Applied cryptography: Protocols, algorithms, and source code in C* (2nd ed.). John Wiley & Sons.
[67] Stallings, W. (2003). *Cryptography and network security: Principles and practices* (3rd ed.). Prentice Hall.
[68] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates.
[69] *SHAP (SHapley Additive exPlanations)* [Software documentation]. https://shap.readthedocs.io/en/latest/.
[70] Keras. *Model plotting utilities* [Software documentation]. https://keras.io/api/utils/model_plotting_utils/#plot_model-function.
[71] Vernam, G. S. (1919). *Secret signaling system* (U.S. Patent No. 1,310,719). U.S. Patent and Trademark Office.
[72] Lugrin, T. (2023). One-time pad. In V. Mulder, A. Mermoud, V. Lenders, & B. Tellenbach (Eds.), *Trends in data protection and encryption technologies* (pp. 3–6). Springer Nature.
[73] McCree, A. V. (2008). Low-bit-rate speech coding. In J. Benesty, M. M. Sondhi, & Y. A. Huang (Eds.), *Springer handbook of speech processing* (pp. 331–350). Springer.
[74] MELPe TSVCIS. *MELPe TSVCIS*. https://melpe.com/melpe-tsvcis/
[75] NATO Standardization Office. *NATO standards*. https://nso.nato.int/nso/nsdd/main/standards
[76] Zhang, J., Duong, T. Q., Marshall, A., & Woods, R. (2016). Key generation from wireless channels: A review. *IEEE Access*, *4*, 614–626.
[77] Li, G., Sun, C., Zhang, J., Jorswieck, E., Xiao, B., & Hu, A. (2019). Physical layer key generation in 5G and beyond wireless communications: Challenges and opportunities. *Entropy*, *21*(5), 497.
[78] Zhang, J., Duong, T. Q., Woods, R., & Marshall, A. (2017). Securing wireless communications of the Internet of Things from the physical layer: An overview. *Entropy*, *19*(8), 420.
[79] Li, G., Zhang, Z., Zhang, J., & Hu, A. (2021). Encrypting wireless communications on the fly using one-time pad and key generation. *IEEE Internet of Things Journal*, *8*(1), 357–369.
[80] Pekerti, A. A., Sasongko, A., & Indrayanto, A. (2024). Secure end-to-end voice communication: A comprehensive review of steganography, modem-based cryptography, and chaotic cryptography techniques. *IEEE Access*, *12*, 75146–75168.
[81] McCree, A. V., & Barnwell, T. P. (1995). A mixed excitation LPC vocoder model for low bit rate speech coding. *IEEE Transactions on Speech and Audio Processing*, *3*(4), 242–250.
[82] Radomirović, J., Milosavljević, M., Čubrilović, S., Kuzmanović, Z., Perić, M., Banjac, Z. & Perić, D. (2025). A Class of Perfectly Secret Autonomous Low-Bit-Rate Communication Systems. *Symmetry*, *17* (3), 365.
[83] Vlatacom Institute. *Encryption & authentication*. https://www.vlatacominstitute.com/encryption-authentication.
[84] Perić, M., Milićević, P., Banjac, Z., Orlić, V., & Milićević, S. (2013). High speed random number generator for section key generation in encryption devices. In *Proceedings of the 21st Telecommunications Forum (TELFOR)* (pp. 117–120). IEEE.
[85] Itakura, F. (1975). Line spectrum representation of linear predictive coefficients of speech signals. *Journal of the Acoustical Society of America*, *57*, S35.
[86] Kabal, P., & Ramachandran, R. P. (1986). The computation of line spectral frequencies using Chebyshev polynomials. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *34*, 1419–1426.

[87] Soong, F. K., & Juang, B.-H. (1984). Line spectrum pair (LSP) and speech data compression. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 37–40). IEEE.

[88] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

[89] Aaron, M. R., McDonald, R. A., & Protonotarios, E. N. (1967). Entropy power loss in linear sampled data filters. *Proceedings of the IEEE, 55*, 1093–1094.

[90] Ntantogian, C., Veroni, E., Karopoulos, G., & Xenakis, C. (2019). A survey of voice and communication protection solutions against wiretapping. *Computers & Electrical Engineering*, *77*, 163–178.

[91] Martinez-Mateo, Elkouss, D., J., & Martin, V. (2011). Blind reconciliation. *Quantum Information & Computation*, *12*(9–10), 791–812.

[92] Reis, A. (2019). *Quantum Key Distribution Post Processing - A Study on the Information Reconciliation Cascade Protocol* [Master's Thesis, Faculdade de Engenharia, Universidade do Porto].

Biografija

Jelica Radomirović je rođena 3.7.1997. godine u Beogradu. Završila je osnovnu školu „Rade Končar“ u Zemunu kao đak generacije. Nakon osnovne škole upisuje Matematičku gimnaziju u Beogradu koju završava sa odličnim uspehom uz osvajanje nagrade na državnom takmičenju iz fizike. Elektrotehnički fakultet u Beogradu upisuje 2016. godine. Bila je student na katedri za Signale i sisteme, a diplomirala je u septembru 2020. godine sa prosečnom ocenom 9,13. Diplomski rad pod nazivom „Prepoznavanje karaktera znakovnog jezika knn metodom“ pod mentorstvom prof. dr Aleksandre Krstić odbranila je u septembru 2020. godine sa ocenom 10.

Diplomske akademske – master studije na Elektrotehničkom fakultetu u Beogradu, na modulu za Signale i sisteme upisala je u oktobru 2020. godine a završila u septembru 2021. godine sa prosečnom ocenom 9,83. Master rad pod nazivom „Fuzija slika termalne kamere i kamere u vidljivom delu spektra“ pod mentorstvom prof. dr Veljka Papića odbranila je sa ocenom 10.

Doktorske akademske studije upisala je u oktobru 2021. godine na Elektrotehničkom fakultetu u Beogradu na modulu Upravljanje sistemima i obrada signala. Oblast njenog istraživanja obuhvata obradu signala, teoriju informacija, kriptografiju i neuralne mreže. Rezultat ovih istraživanja je osam objavljenih radova u međunarodnim naučnim časopisima (po jedan rad kategorija M21a+ i M21, dva u kategoriji M21a i četiri rada kategorije M22) i dva rada na međunarodnoj konferenciji.

Od avgusta 2021. zaposlena je u Vlatacom institutu visokih tehnologija kao istraživačko razvojni inženjer. Na toj poziciji dominantno je radila na razvoju algoritama za generisanje tajnih ključeva baziranih na izvorima zajedničke slučajnosti. Osim toga, u okviru oblasti mašinskog učenja radila je i na modelovanju visokopreciznih prediktora iz više različitih problemskih domena.

Изјава о ауторству

Име и презиме аутора Јелица Радомировић

Број индекса 5028/2021

Изјављујем

да је докторска дисертација под насловом

Једна нова класа система за дестилацију апсолутно тајних криптографских кључева заснованих на заједничкој случајности

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 24.04.2026.

Радомировић Јелица

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Јелица Радомировић

Број индекса 5028/2021

Студијски програм Електротехника и рачунарство
(модул Управљање системима и обрада сигнала)

Наслов рада Једна нова класа система за дестилацију апсолутно тајних криптографских кључева заснованих на заједничкој случајности

Ментор др Бранко Ковачевић, професор емеритус

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 24.04.2026.

Радомировић Јелица

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Једна нова класа система за дестилацију апсолутно тајних криптографских кључева заснованих на заједничкој случајности

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 24.04.2026.

Радомировић Јелица

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.