Kriptoanaliza Plejferovog koda pomoću simuliranog kaljenja

Dragan Marković

Oktobar, 2015.

Sadržaj

1 Plejferov kod

Plejferov kod, ili Plejferov kvadrat je simetrična tehnika enkripcije i prva tehnika koja se zansvia na zameni digrama. Kod je osmisljen od strane Čarlsa Vetstona (eng. Charles Wheatstone), Engleskog naucnika iz Viktorijanskog doba, ali je popularizovan od strane Lorda Plejfera, po čemu je i dobio ime. Kod je osmišljen 1854. godine.

1.1 Opis ključa koda

Plejferov kod koristi kvadrat veličine $5 \cdot 5$ koji sadrži neku permutaciju slova Engleske abecede. Kako engleska abeceda ima 26 slova, a kvadrat je veličine 25 neka dva slova moramo tretirato kao isto. Ako su ta dva slova "I" (veliko i) i "J" onda su reči "Playfair" i "Playfajr" ekvivalentne, što ne predstavlja preveliki problem ljudskom oku koje čita poruku. Ovaj kvadrat je ključ koda i prijemnik i pošiljalac ga moraju posedovati kako bi slanje šifrovane poruke bilo uspešno.

1.2 Šifrovanje

Kako bismo šifrovali poruku, prvo je potrebno "razbiti" poruku na digrame, tj. grupe od po dva karaktera, tako da npr. "plejferova sifra" postaje "PL EJ FE RO VA SI FR AZ". Ako poruka ima neparan broj karaktera, možemo dodati još jedan karakter na kraju poruke po dogovoru. Šifrujemo svaki bigram posebno, tako što nalazimo pozicije ta dva karaktera u kvadratu (ključu) i poštujemo tri pravila:

1. Ako su karakteri u istom redu u kvadratu, zamenićemo ih sa njihovim prvim desnim susedom. Pri tom se red smatra cikličnim.

2. Ako su karakteri u istoj koloni, zamenićemo ih sa njihovim prvim donjim susedom, kolona se takođe smatra cikličnom.

3. Inače karakteri su suprotna temena pravougaonika, i njih zamenjujemo sa preostaval dva temena tog istog pravougaonika. Redosled je bitan, prvi karakter se zamenjuje sa temenom koje je u istom redu.

Neka je ključ:

P L A Y F
I R E X M
B C D G H
K N O Q S
T U V W Z

Bigram PF bi se šifrovao u LP (pravilo 1), bigram KT bi se šifrovao u TP (pravilo 2), a bigram MP u IF (pravilo 3).

1.3 Dešifrovanje

Dešifrovanje je analogno šifrovanju, samo se pravila primenjuju u suprotnom smeru.

2 Kriptoanaliza

Plejeferova šifra, kao većina starijih enkripcija, nije bezbedna i ključ se relativno lako generiše analizom frekvencije teksta. Za analize frenkvencije potrebna je dovoljno velika količina teksta i skoro je nemoguće primeniti ovu metodu na samo jednu ili dve šifrovane reči.

2.1 Kada je poznat dešifrovan tekst

Ako nam je poznat šifrovan i dešifrovan tekst naš problem se svodi na generisanje ključa koji ima jednaku frekvenciju bigrama u originalnom i šifrovanom tekstu. Algoritam koji bi rešio ovaj problem bi generisao sve moguće ključeve, sa velikim restrikcijama, koje se postižu minimizacijom razlike frekvencija u originalnom u šifrovanom tekstu.

2.2 Kada nije poznat dešifrovan tekst

Potrebno je znati na kom jeziku je šifrovan zadati tekst, što u praksi nije teško pretpostaviti. Koristi se jedan od algoritama za nalaženja globalnog maksimuma funkcije (npr. Simulirano kaljenje) i počevši od nekog nasumičnog izabranog ključa, u svakom koraku se zamene pozicije dva karaktera datog ključa sve dok se ne dobije ključ čiji dešifrovan tekst ima najpribližniju frekvenciju karaktera, bigrama, trigrama... zadatom jeziku.

2.3 Simulirano kaljenje

Simulirano kaljenje je proširenje klasičnog hill-climb algoritma. Hill-climb algoritam za neko stanje generiše sva susedna stanja i prebaci se u ono stanje koje ima najveći skor. Jasno se vidi da za proizvoljnu složenu funkciju hill-climb može da se "zaglavi" u lokalnom maksimumu i nikada da se ne približi globalnom maksimumu.

Simulirano kaljenje radi na sličan način, samo što dopušta sebi na pređe u lošije stanje. Verovatnoća da se pređe u lošije stanje je inverzno proporcijalna dužini izvrašavanja algoritama. Preciznije, simulirano kaljenje uvodi temperaturu sistema i u svakoj iteraciji ta temperaturu se smanjuje za određenju količinu. Algoritam se završava kada temperaturu postane 0, a verovatnoća da se izabere lošije stanje od prethodnog je $1 - \frac{1}{e^{t_i \ t_0}}$, gde je t_0 početna temperatura a t_i trenutna temperatura.

2.4 Skor funkcija

Kako bismo implementirali simulirano kaljenje, potrebna nam je matematički precizna funkcija stanja, koja opisuje da li je neko stanje lošije ili bolje od trenutnog i za koliko. Uzeto je procentualno pojavljivanje monograma, bigrama, trigrama i kvadgrama u dešifrovanom tekstu i delu "Rat i mir", koji se smatra dovoljno velikim uzorkom za analazu frekvencije karaktera.

Neka je K kljuc (permutacija slova "abcdefghijklmnopqrstuvwxyz") koji slika šifrovan tekst I u dešifrovan tekst O. Dalje, neka je $S = \{a_0, a_1, \cdots, a_r\}$ skup bigrama u tekstu O i broj pojavljivanja bigrama $a_0, a_1 \cdots$ u tekstu O, $b_{a_0}, b_{a_1}, \cdots, b_{a_r}$. Neka je broj pojavljivanja bigrama $a_0, a_1, \cdots, a_r$ u delu "Rat i mir" $c_{a_0}, c_{a_1}, \cdots, c_{a_r}$ i neka je ukupan broj bigrama u "Rat i mir"-u C. Onda je skor funkciju za dati ključ K:

$$f(K) = \sum_{i=0}^{i \leq r} b_{a_i} \cdot \log \frac{c_{a_i}}{C}$$

Cilj je pronaći ključ sa najvećom skor funkcijom. Za bigrami koji se nepojavljuju u "Rat i mir"-u, a pojavljuju se u O uzima se beskonačna mala vrednost. Analogni pristup se koristi za računanje skor funkcije za monograme, trigrame i kvadgrame.

Koriste se sve četiri skor funkcije (za monograme, bigrame, trigrame i kvadgrame) i na kraju simuliranog kaljenje uzima se tekst koji je najbolje

dešifrovan, tj. najčitljiv za čoveka.

3 Analiza efikasnosti i zaključak

Za testiranje su korišćenja tri ulazna teksta, koja se nalaze u folderu "input":

1. Izraz "Hello World" koji nije upešno dešifrovan jer nije dovoljno dugačak. Ovo je i očekivan rezultat.

2. Odsečak o Čarlsu Dikensu sa vikipedije, koji je potpuno dešifrovan analizom trigrama, i dovoljno dobro dešifrovan analizom monograma, bigrama i kvadragama da je čitljivo za čoveka.

3. Pesma Eminema, koja je potpuno dešifrovana analizom monograma i dešifrovana dovoljno dobro analizom bigrama, dok nije uspešno dešifrovana analizom trigrama i kvadgrama.

   Sa izuzetkom izraza "Hello World" svi tekstovi su uspešno dešifrovani kombinacijom nekih od pristupa analize monograma, bigrama trigrama i kvadgrama. Korisno je uraditi sve četiri analize i nema pravila kada je jedan pristup bolji od drugog.

   Dešifrovan tekst je dobijen uz relativno mali broj iteracija u simuliranom kaljenju. Plejferova šifra definitivno nije bezbedna u današnje vreme, osim ako se šifruje ekstremno mali tekst.