Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Универзитет у Београду

Математички факултет

Катедра за рачунарство и информатику

мастер рад

Анализа излазног низа генератора RC4

студент: Јована Радуловић

ментор: др Миодраг Живковић

Београд

2015.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Садржај

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

1.Увод

Са развојем првих писама настаје и потреба да се садржај писаних порука заштити, поготово порука које се користе у војним и дипломатским круговима. То доводи до развоја криптографије и појаве првих једноставних шифара попут спартанске шифре "скитале" и Цезарове шифре.

Међутим, појава рачунара и развој комуникационих технологија довели су до експанзије ове области, нарочито у последњих тридесетак година. У том периоду уводе се први стандарди и развија се велики број нових алгоритама за шифровање.

Један од алгоритама насталих у периоду вртоглавог развоја криптографије је и алгоритам RC4. Алгоритам RC4 генерише низ кључа, на основу кога се врши шифровање порука. Овај алгоритам је 1987. године дизајнирао Рон Ривест (Ron Rivest), за потребе компаније RSA Data Security, Inc. Шифра је све до 1994. године била држана у тајности, када је неко анонимно послао њен изворни код на листу Cypherpunks. Од тог тренутка, алгоритам RC4 постао је доступан широј јавности.

Данас је RC4 једна од најпопуларнијих проточних шифара. Алгоритам на коме се ова шифра заснива уграђен је у велики број комерцијалних производа (Lotus Notes, Oracle Secure SQL, Microsoft Windows...) и користи се за шифровање интернет саобраћаја као саставни део мрежних протокола (SSL, TLS, WPA, WEP...). Његова популарност се заснива на једноставном, брзом и ефикасном алгоритму. Због своје једноставности, алгоритам RC4 је од датума објављивања па све до данас био мета многобројних напада, али и поред бројних напада алгоритам није озбиљно угрожен, тако да је његова употреба безбедна уз извесна ограничења.

Предмет овог рада је анализа статистичких особина алгоритма RC4, када се алгоритам користи као генератор псеудослучајних бројева. Овај рад поред тога садржи и одељак у коме су приказана статистичка тестирања експериментално добијених резултата. Циљ овог дела рада је евентуално проналажење тестова чији се излаз из алгоритма RC4 статистички битно разликује од равномерно расподељеног случајног низа.

Наставак рада подељен је на следећа поглавља:

- *Основни појмови.* У овом поглављу су уведени основни појмови.

- *Алгоритам RC4.* Ово поглавље описује алгоритам RC4, његове фазе и у њему је укратко изложен преглед основних својстава која се користе у даљем тексту.

- *Особине излазног низа генератора RC4.* У овом поглављу приказују се особине излазног низа генератора алгоритма RC4.

- *Резултати симулације.* У овом поглављу је изложена програмска реализација и изнете су статистичке анализе и обрађени резултати.

- *Закључак.* Ово поглавље поред закључка садржи и предлоге за будућа истраживања.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

2. Основни појмови

Ово поглавље садржи два одељка. У првом одељку, *Криптографски алгоритам*, уведени су основни криптографски појмови. Други одељак, *Проточне шифре*, садржи дефиницију и неке основне карактеристике проточних шифара.

2.1. Криптографски алгоритам

Криптографски алгоритам је функција која пресликава један низ битова у други, при чему није нужно да низови буду једнаке дужине. Улазни низ битова, тј. низ битова који је аргумент функције, зове се **отворени текст**. **Шифровање** је процес примене криптографског алгоритма, а резултат шифровања назива се **шифратом**. Ако је отворени текст означен са **P**, шифрат са **C**, а функција шифровања са **E**, процес шифровања описује следећа једнакост

$$C = E(P).$$

Поступак инверзан поступку шифровања, назива се **дешифровањем**. Ако је функција дешифровања означена са **D**, тада важе једнакости:

$$P = D(C)$$

и

$$P = D\big(E(P)\big).$$

У пракси, криптографски алгоритми садрже још један параметар – низ битова који се зове **кључ**. Пре употребе алгоритама који су засновани на кључу, безбедност криптографског алгоритма се заснивала на тајности алгоритма. Увођењем кључа, алгоритми нису више морали бити тајни, већ се њихова сигурност почела заснивати на тајности кључа. Такође, уколико је тајни алгоритам који користи кључ разбијен, довољно је променити кључ да би се алгоритам могао и даље користити. Захваљујући увођењу кључа, јавност се укључила у откривање недостатака постојећих алгоритама и у проналажење нових, безбеднијих алгоритама.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Криптографски алгоритми који користе кључеве могу се поделити на симетричне и асиметричне алгоритме. Асиметрични алгоритми су они алгоритми код којих се при шифровању и дешифровању користе различити кључеви – јавни кључеви се користе за шифровање, а тајни за дешифровање, док симетрични алгоритми користе исти тајни кључ и за шифровање и дешифровање. У овом раду од значаја су само симетрични алгоритми, тако да се на даље сматра да је исти кључ коришћен и за шифровање и за дешифровање.

Ако је кључ означен са $\boldsymbol{k}$, процесе шифровања и дешифровања описују следеће једнакости:

$$C = E_k(P)$$

и

$$P = D_k(C).$$

Такође, важи и следећа једнакост

$$P = D_k\big(E_k(P)\big).$$

2.2. Проточне шифре

Проточне шифре отворени текст трансформишу симбол по сибол, тј. бит по бит. Ако су са $m_1, m_2, m_3\ldots$ означени битови поруке коју треба шифровати, са $k_1, k_2, k_3\ldots$ означени битови низа кључа, а битови шифрата са $c_1,c_2,c_3\ldots$ тада је поступак шифровања описан следећом једнакошћу

$$m_i \oplus k_i = c_i \ , \qquad i \geq 1.$$

Слична једнакост важи за процес дешифровања

$$c_i \oplus k_i = m_i \ , \qquad i \geq 1.$$

Главне предности оваквог начина шифровања су што се операција $\oplus$ (XOR) једноставно и брзо извршава, као и то што се шифровање и дешифровање обављају на исти начин што олакшава имплементацију алгоритма.

Сигурност овакве шифре у потпуности зависи од квалитета генератора који генерише низ кључа. Уколико генератор низа кључа даје бесконачан низ случајних битова, тада се оваква шифра своди на случајну шифру (one-time-pad) и има савршену сигурност. Међутим, у пракси сигурност проточне шифре је мања од сигурности случајне шифре, јер генератор низа кључа генерише низ битова који је у ствари детерминистички, иако наизглед делује случајно (тзв. псеудослучајни низ). Што је излаз генератора сличнији случајном низу, нападач има тежи проблем пред собом.

Како је излаз генератора који генерише низ кључа псеудослучајни низ, треба нагласити да је добра пракса повремено променити кључ, како би се обезбедила већа сигурност система.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

3.Алгоритам RC4

Ово поглавље подељено је у четири одељка. Први одељак овог поглавља садржи опис основних елемената алгоритма. У другом одељку су уведене и објашњене ознаке које се користе у даљем раду. У трећем одељку детаљно је изложена фаза KSA алгоритма и дат је преглед основних својстава KSA која се користе у даљем тексту, док је у четвртом одељку детаљно описана фаза PRGA. Сва тврђења у овом поглављу написана су на основу трећег и четвртог поглавља литературе [1].

3.1. Основно о алгоритму

RC4 је проточна шифра, са тајним кључем чија је величина најчешће између 40 и 148 битова, а низ кључа је независан од отвореног текста (тзв. OFB режим, Output Feedback).

Основни елеменат ове шифре је низ $\boldsymbol{S} = (S[0], \dots, S[N-1])$, дужине $N = 2^n$, а елементи за низ S су пермутација скупа $\{0,1, \dots, 2^n - 1\}$. Сабирање индекса за S врши се по модулу N. Поред низа S, користи се и низ $\boldsymbol{K}$ и индекси $\boldsymbol{i}$ и $\boldsymbol{j}$. Низ K се добија од тајног кључа $\boldsymbol{k}$ поновљеног довољан број пута да би се попунио низ $(K[0], \dots, K[N-1])$.

Обично се узима да је $n = 8$ тј. $N = 256$.

Алгоритам RC4 састоји се из две фазе. Прва фаза је фаза **KSA** (Key-scheduling algorithm) у којој се врши иницијализовање пермутације S у зависности од кључа. За њом следи фаза **PRGA** (Pseudo-random generation algorithm) у којој се из пермутације S генерише псеудослучајни низ бајтова који ће се користити приликом шифровања отвореног текста.

3.2. Фаза KSA

Фаза KSA је прва фаза приликом извршења алгоритма RC4. У овој фази врши се иницијализовање низа S у зависности од кључа. У даљем тексту детерминистички индекс, псеудослучајни индекс и пермутација током фазе KSA редом су означени са *i, j* и *S.*

Следећи псеудокод детаљно описује *KSA*:

```
Ulaz:  Niz K[0... N - 1] dobijen periodičnim ponavljanjem tajnog ključa

Izlaz:  Permutacija S

//Inicijalizacija identičke permutacije

for i = 0 to N - 1 do

        S[i] = i

end

//Inicijalizacija indeksa j

 j = 0

// Mešanje elemenata niza S

for i = 0 to N - 1 do

        j = j + S[i] + K[i]  (mod N)

        swap(S[i], S[j])

end
```

Током фазе **KSA** врши се иницијализација низа S, тако што се сваком елементу додели вредност позиције на којој се налази у низу тј. од низа S се направи идентичка пермутација. У следећем кораку се врши иницијализација индекса ***j*** на 0, а потом се

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

врши премештање елемената низа S у N итерација. У свакој итерацији одређује се вредност елемента на позицији ***j*** која зависи од тренутног стања низа S и од вредности у низу K (тј. кључа), а потом се врши замена елемената који се налазе на позицијама ***i*** и ***j***.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

3.2.1. Основна својства фазе KSA

Рус (Roos) [6] и Вагнер (Wagner) [7] су независно један од другог, 1995. године открили прве слабости у фази KSA, на основу којих су издвојили неколико класа слабих кључева. У овом одељку дат је преглед особина фазе KSA до којих су Рус и Вагнер дошли у свом раду, а које су значајне за анализу фазе PRGA.

Ознака f_y уведена је ради краћег и једноставнијег записа и дефинише се на следећи начин

$$f_y \stackrel{\text{def}}{=} \frac{y(y+1)}{2} + \sum_{x=0}^{y} K[x].$$

Лема 3.1.

Највероватнија вредност елемента на позицији ***y*** у пермутацији S након KSA, за првих неколико вредности ***y*** дата је једнакошћу $S_N[y] = f_y$.

■

Рус је закључио, да када је y довољно мало, постоји велика вероватноћа да ниједан од два елемента која учествују у замени, није учествовао у претходним заменама. На основу тога се са великом сигурношћу може претпоставити да је $S_y[y] = y$ пре замене у *(y+1)*-ој рунди. Одатле се даље може закључити да је највероватнија вредност елемента на позицији $S[1]$ једнака $K[0] + K[1] + 0 + 1$, а да је највероватнија вредност елемента на позицији $S[2]$ једнака $K[0] + K[1] + K[2] + 0 + 1 + 2$. Како је вероватноћа да j не узима одређену вредност током фазе KSA једнака $\left(\frac{N-1}{N}\right)^N$, закључак до кога је дошао Рус, може се формализовати у виду следећег тврђења.

Последица 3.2.

По завршетку фазе KSA највероватније вредности другог и трећег елемента у пермутацији S дате су следећим изразима:

1. $P(S_N[1] = K[0] + K[1] + 1) \approx \left(\frac{N-1}{N}\right)^N$
2. $P(S_N[2] = K[0] + K[1] + K[2] + 3) \approx \left(\frac{N-1}{N}\right)^N$.

■

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

3.3. Фаза PRGA

Током фазе PRGA из пермутације S генерише се псеудослучајни низ бајтова који се касније користи приликом шифровања отвореног текста. У току фазе PRGA детерминистички индекс, псеудослучајни индекс и пермутација редом су означени са i^G, j^G и S^G, док је са t је означен индекс на основу кога се из пермутације S^G добија излазни бајт z, за који важи да је $z = S^G[t]$. Улога експонента G у ознакама детерминистичког индекса i, псеудослучајног индекса j и пермутације S у оквиру фазе PRGA је да разликује наведене елементе од истих елемената који се користе у оквиру фазе KSA.

Следећи псеудокод детаљно описује *PRGA*:

Ulaz: *Permutacija S[0 … N - 1] koja zavisi od tajnog ključa*

l dužina pseudoslučajnog niza

Izlaz: *Pseudoslučajan niz bajtova z*

// Inicijalizacija

$i^G = j^G = 0$

// Generisanje pseudoslučajnog niza bajtova dužine l

for *r = 0* ***to*** *l-1* ***do***

$i^G = i^G + 1 \pmod N$

$j^G = j^G + S^G[i^G] \pmod N$

$swap(S^G[i^G], S^G[j^G])$

$t = S^G[i^G] + S^G[j^G] \pmod N$

$z[r]=S^G[t]$

end

PRGA фаза започиње иницијализацијом индекса i^G и j^G на 0. Након тога приступа се генерисању задатог броја бајтова (означимо тај број са l). Сваки бајт генерише се у четири корака, а цео процес се понавља док се не изгенерише свих l бајтова. Индекс i^G се увек мења на исти начин, док индекс j^G зависи од вредности која се у пермутацији S^G налази на позицији i^G. Након израчунавања индекса i^G и j^G врши се замена елемената који се налазе на позицији i^G и позицији j^G у низу S^G. Последњи корак у генерисању једног бајта је рачунање вредности t која представља позицију елемента у пермутацији S^G који је излаз генератора. Излазни бајт је онај бајт који се налази на позицији t у пермутацији S^G.

На основу претходно изнетог, RC4 се може разматрати и као коначни аутомат, чије је унутрашње стање уређена тројка (S,i,j).

У даљем тексту уз променљиву је често у индеску наведен и параметар r који означава да је у питању вредност одређене променљиве у r-тој рунди. На пример, S_0^G означава стање пермутације S пре прве рунде PRGA, док i_r^G означава вредност параметра i у r-тој рунди PRGA.

3.3.1. Финијеви циклуси

Убрзо након објављивања алгоритма RC4 Хал Фини (Hal Finney) [10] је открио следеће својство:

Теорема 3.3.

Ако је $j_r^G = i_r^G + 1$ и $S_r^G[j_r^G] = 1$ у r-тој рунди фазе PRGA, тада важе следећа тврђења:

1. наведена релација важи у свим наредним рундама
2. формира се циклус дужине $N(N-1)$ у простору стања
3. излазни бајтови $z_{r+(N-1)}, z_{r+2(N-1)}, \dots, z_{r+N(N-1)}$ су добијени цикличком ротацијом пермутације S_r^G.

Доказ

1. Нека је након r-те рунде фазе PRGA $j_r^G = i_r^G + 1$ и $S_r^G[j_r^G] = 1$. Тада је у наредној рунди $i_{r+1}^G = i_r^G + 1$ и $j_{r+1}^G = j_r^G + S_r^G[i_{r+1}^G] = j_r^G + S_r^G[i_r^G + 1] = j_r^G + S_r^G[j_r^G] = i_r^G + 1 + 1 = i_r^G + 2$.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Након замене елемената на позицијама i и j, добија се да је $S_{r+1}^G[i_r^G + 2] = 1$ и $S_{r+1}^G[i_r^G + 1] = S_r^G[i_r^G + 2]$. Дакле, однос између пермутације S^G и индекса i и j, који важи у r-тој рунди фазе PRGA, важи и у r+1-ој рунди фазе PRGA. Тиме је тврђење доказано математичком индукцијом.

2. У свим рундама фазе PRGA, врши се замена елемената који су на позицијама $S^G[y + 1]$ и $S^G[y]$ и вредност 1 се циклички помера за једно место удесно у свакој рунди.

Након N рунди, вредност 1 се враћа на своју почетну позицију, а остале вредности се циклички померају за једно место улево. Након *N-1* понављања *N* рунди, све вредности се налазе на својим почетним позицијама и индекси i и j имају почетне вредности. Одатле је јасно да се у простору стања формира циклус дужине *N(N-1)*.

3. За фиксно i^G, вредност на позицији $S^G[i^G]$ се понавља након сваких *N-1* корака, док вредност $S^G[\mathrm{j}^G]$ увек остаје 1. Због тога се индекс $S^G[i^G] + S^G[j^G]$ понавља на сваких *N-1* корака. Како се све вредности померају циклички за једно место улево, тако *N* излазних бајтова међу којима су свака два одвојена *N-1* рундом, представља циклички померену пермутацију S.

■

Уколико у неком стању важи да је $j_r^G = i_r^G + 1$ и $S_r^G[j_r^G] = 1$ тада се то стање назива **Финијевим стањем**. На основу теореме **3.3**, јасно је да се из Финијевог стања не може прећи у стање које није Финијево, нити се из стања које није Финијево може прећи у стање које јесте Финијево. Циклус у простору стања који се формира на овај начин назива се **Финијевим циклусом**.

Захваљујући иницијализацији $i = j = 0$ пре прве рунде PRGA, избегнута је појава Финијевих циклуса приликом рада RC4. Самим тим избегнути су и озбиљни недостаци који би се појавили у псеудослучајном низу бајтова који генерише алгоритам RC4. Међутим, ово својство отвара питање да ли се могу појавити неки други кратки циклуси у оквиру фазе PRGA. За сада постојање таквих циклуса није познато.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4. Особине излазног низа генератора RC4

У овом поглављу приказују се особине излазног низа генератора алгоритма RC4. Изложена својства су организована у два одељка. У првом одељку су наведена својства која се односе на први бајт излазног низа, док се својства чији је преглед дат у другом одељку односе на све бајтове излазног низа, почев од првог бајта. Сва тврђења која су наведена у овом поглављу написана су на основу четвртог поглавља у литератури [1].

4.1. Особине које се односе на први бајт излазног низа

У одељку 4.1.1. изложена су својства која се односе на условљеност првог излазног бајта пермутацијом S, док су у одељку 4.1.2. изложена својства која се односе на условљеност првог излазног бајта кључем.

4.1.1. Условљеност првог излазног бајта пермутацијом S

Вредност индекса t_1 на основу кога се добија први излазни бајт z_1 нема равномерну расподелу вероватноће. Вероватноћа да је $t_1 = 2$ је готово двоструко већа, него да индекс t_1 узима било коју другу вредност. Формално речено:

Теорема 4.1.

Ако се претпостави да је пермутација S_0^G, добијена после фазе KSA, случајно одабрана из скупа свих могућих пермутација Z_N са равномерном расподелом вероватноћа, тада је расподела вероватноће индекса t_1, на основу кога се одређује први излазни бајт

$$P(t_1 = x) = \begin{cases} \frac{1}{N} & \text{за непарно } x; \\ \frac{1}{N} - \frac{2}{N(N-1)} & \text{за парно } x \neq 2; \\ \frac{2}{N} - \frac{1}{N(N-1)} & \text{за } x = 2. \end{cases}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Доказ

У првој рунди PRGA $i_1{}^G$ постаје 1, а $j_1{}^G$ добија вредност $S_0{}^G[1]$. Након замене елемената на позицијама *i* и *j* добија се:

$$S_1^G[i_1^G] = S_0^G[j_1^G] = S_0^G\left[S_0^G[1]\right]$$

и

$$S_1^G[j_1^G] = S_0^G[i_1^G] = S_0^G[1].$$

Ако је $S_1^G[j_1^G] = 1$ тада је и $S_0^G[1] = 1$ јер је $S_1^G[j_1^G] = S_0^G[1]$. Из $S_0^G[1] = 1$ и својства PRGA да је $j_1^G = S_0^G[1]$ следи да је $j_1^G = 1$, а на основу полазне чињенице да је $S_1^G[j_1^G] = 1$ и резултата из претходног корака добија се да је $S_1^G[1] = 1$. Најзад је и $S_1^G[i_1^G] = 1$ јер је $i_1^G = 1$.

Како је

$$P(S_1^G[i_1^G] = u, S_1^G[j_1^G] = v) = \begin{cases} \frac{1}{N}, & \text{ако је } u = 1 \text{ и } v = 1; \\ 0, & \text{ако је } u \neq 1 \text{ и } v = 1; \\ 0, & \text{ако је } u \neq 1 \text{ и } v = u; \\ \frac{1}{N(N-1)}, & \text{иначе.} \end{cases}$$

расподела вероватноће за индекс $t_1 = S_1^G[i_1^G] + S_1^G[j_1^G]$ може се израчунати на следећи начин:

- за непарно x:

$$P(t_1 = x) = \sum_{\substack{v=0 \\ v \neq 1}}^{N-1} P(S_1^G[j_1^G] = v, S_1^G[i_1^G] = N - v + x) = (N-1) \cdot \frac{1}{N(N-1)} = \frac{1}{N}.$$

- за парно $x \neq 2$:

$$P(t_1 = x) = \sum_{\substack{v=0 \\ v \neq 1, \frac{x}{2}, \frac{N+x}{2}}}^{N-1} P(S_1^G[j_1^G] = v, S_1^G[i_1^G] = N - v + x) = (N-3) \cdot \frac{1}{N(N-1)} =$$

$$= \frac{1}{N} - \frac{2}{N(N-1)}.$$

- за $x = 2$:

$$P(t_1 = x) = P(S_1^G[j_1^G] = 1, S_1^G[i_1^G] = 1) + \sum_{\substack{v=0 \\ v \neq 1, \frac{N+2}{2}}}^{N-1} P(S_1^G[j_1^G] = v, S_1^G[i_1^G] = N - v + 2) =$$

$$\frac{1}{N} + (N-2) \cdot \frac{1}{N(N-1)} = \frac{2N-3}{N(N-1)} = \frac{2}{N} - \frac{1}{N(N-1)}.$$

■

4.1.2. Условљеност првог излазног бајта кључем

Први излазни бајт условљен је са прва три бајта кључа. Ову условљеност први је открио Рус [6].

Ради краћег и једноставнијег записа уведимо ознаку φ_N , дефинисану на следећи начин

$$\varphi_N \stackrel{\text{def}}{=} \left(\frac{N-1}{N}\right)^N \left(1 - \frac{1}{N} - \frac{1}{N^2}\right) + \frac{1}{N^2}.$$

За управо уведену ознаку φ_N, важи да је $\varphi_N \approx e^{-1}$.

Лема 4.2.

$P(S_1^G[2] = K[0] + K[1] + K[2] + 3) \approx \varphi_N.$

Доказ

Нека је $f(K) = K[0] + K[1] + K[2] + 3.$
Из последице **3.2.** важи да је $P\left(S_0^G[2] = f(K)\right) \approx (\frac{N-1}{N})^N.$

Доказ се може раздвојити на два случаја:

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

1. Ако је после KSA $S_0^G[2] = f(K)$ и ако у премештању у првој рунди PRGA не учествује индекс два. Вероватноћа овог догађаја је

$$P\left(S_0^G[2] = f(K)\right) \cdot P(S_0^G[1] \neq 2) = \left(\frac{N-1}{N}\right)^N \left(1 - \frac{1}{N}\right).$$

2. Ако је после фазе KSA $S_0^G[2] \neq f(K)$ и ако $f(K)$ након прве рунде PRGA долази на позицију са индексом два. Вероватноћа овог догађаја је

$$P\left(S_0^G[2] \neq f(K)\right) \cdot P(S_0^G[1] = f(K), S_0^G[1] = 2) =$$

$$P\left(S_0^G[2] \neq f(K)\right) \cdot P(S_0^G[1] = 2) \cdot P(f(K) = 2) \approx \left(1 - \left(\frac{N-1}{N}\right)^N\right) \cdot \frac{1}{N^2}.$$

Сабирањем добијених вероватноћа добија се да је

$$P\left(S_0^G[2] = f(K)\right) \approx \left(\frac{N-1}{N}\right)^N \left(1 - \frac{1}{N} - \frac{1}{N^2}\right) + \frac{1}{N^2} = \varphi_N.$$

■

Теорема 4.3.

За произвољан тајни кључ, веза између бајтова тајног кључа и првог излазног бајта дата је изразом

$$P(z_1 = K[0] + K[1] + K[2] + 3) \approx \frac{1}{N}(1 + \varphi_N).$$

Доказ

Нека је: $f(K) = K[0] + K[1] + K[2] + 3$.

Онда је:

$$P\left(z_1 = f(K)\right) = P\left(S_1^G[t_1] = f(K)\right) = \sum_{x=0}^{N-1} P(t_1 = x) \cdot P(S_1^G[t_1] = f(K) | t_1 = x) =$$

$$\sum_{x=0}^{N-1} P(t_1 = x) \cdot P\left(S_1^G[x] = f(K)\right) =$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

$$P(t_1 = 2) \cdot P\left(S_1^G[2] = f(K)\right) + \sum_{\substack{x=0 \\ \text{парно } x \neq 2}}^{N-1} P(t_1 = x) \cdot P\left(S_1^G[x] = f(K)\right) +$$

$$+ \sum_{\substack{x=0 \\ \text{непарно}}}^{N-1} P(t_1 = x) \cdot P\left(S_1^G[x] = f(K)\right)$$

На основу теореме **4.1:**

$$P\left(z_1 = f(K)\right) = \left(\frac{2}{N} - \frac{1}{N(N-1)}\right) \cdot P\left(S_1^G[2] = f(K)\right) +$$

$$+ \sum_{\substack{x=0 \\ \text{парно } x \neq 2}}^{N-1} \left(\frac{1}{N} - \frac{2}{N(N-1)}\right) \cdot P\left(S_1^G[x] = f(K)\right) + \sum_{\substack{x=0 \\ \text{непарно}}}^{N-1} \frac{1}{N} \cdot P\left(S_1^G[x] = f(K)\right) = I,$$

где је ознака *I* уведена због једноставности.

На основу чињенице да је $\frac{1}{N(N-1)} \ll \frac{1}{N}$ следи да је:

$$I \approx \frac{2}{N} \cdot P\left(S_1^G[2] = f(K)\right) + \sum_{\substack{x=0 \\ \text{парно } x \neq 2}}^{N-1} \frac{1}{N} \cdot P\left(S_1^G[x] = f(K)\right) + \sum_{\substack{x=0 \\ \text{непарно}}}^{N-1} \frac{1}{N} \cdot P\left(S_1^G[x] = f(K)\right) =$$

$$\frac{1}{N} \cdot P\left(S_1^G[2] = f(K)\right) + \frac{1}{N} \cdot P\left(S_1^G[2] = f(K)\right) + \sum_{\substack{x=0 \\ \text{парно } x \neq 2}}^{N-1} \frac{1}{N} \cdot P\left(S_1^G[x] = f(K)\right) +$$

$$+ \sum_{\substack{x=0 \\ \text{парно } x \neq 2}}^{N-1} \frac{1}{N} \cdot P\left(S_1^G[x] = f(K)\right) = \frac{1}{N} \cdot P\left(S_1^G[2] = f(K)\right) + \frac{1}{N} \cdot \sum_{x=0}^{N-1} P\left(S_1^G[x] = f(K)\right)$$

Последњи израз на основу леме **4.2.** има вредност

$$\frac{1}{N} \cdot \varphi_N + \frac{1}{N} = \frac{1}{N}(1 + \varphi_N).$$

■

Експерименталним путем Рус [6] је дошао до закључка, да уколико додатно важи да је $K[0] + K[1] \equiv 0 \ (mod\ N)$, условљеност првог излазног бајта кључем расте. Формално, ово својство је представљено наредном теоремом.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Теорема 4.4.

Уколико је $K[0]+K[1]\equiv 0\ (mod\ N)$, тада је веза између кључа и индекса t_1 на основу кога се одређује први излазни бајт дата вероватносним изразом

$$P\big(t_1=2\big|K[0]+K[1]\equiv 0\ (mod\ N)\big)>\left(\frac{N-1}{N}\right)^N.$$

Доказ

$P(t_1=2|K[0]+K[1]\equiv 0\ )=P(S_1^G[i_1^G]+S_1^G[j_1^G]=2|K[0]+K[1]\equiv 0)=$

$P(S_1^G[i_1^G]=1,S_1^G[j_1^G]=1\mid K[0]+K[1]\equiv 0)+$

$$+\sum_{\substack{x=0\\ x\neq 1,\frac{N+1}{2}}}^{N-1} P(S_1^G[j_1^G]=x,S_1^G[i_1^G]=N-x+2\mid K[0]+K[1]\equiv 0)>$$

$>\ P(S_1^G[i_1^G]=1,S_1^G[j_1^G]=1\mid K[0]+K[1]\equiv 0)$

Како је $(S_1^G[i_1^G]=1\ \wedge S_1^G[j_1^G]=1)\Leftrightarrow S_0^G[1]=1$ последњи израз своди се на

$P(S_0^G[1]=1|K[0]+K[1]\equiv 0)=P(S_0^G[1]=K[0]+K[1]+1|K[0]+K[1]\equiv 0)\approx$

$\approx P(S_0^G[1]=K[0]+K[1]+1)\approx\left(\frac{N-1}{N}\right)^N$ (на основу последице **3.2, 1.**)

■

Теорема 4.5.

Уколико је $K[0]+K[1]\equiv 0\ (mod\ N)$, тада је веза између кључа и првог излазног бајта дата вероватносним изразом

$$P\big(z_1=K[2]+3\big|K[0]+K[1]\equiv 0\ (mod\ N)\big)>\left(\frac{N-1}{N}\right)^N\varphi_N.$$

Доказ

$P(z_1=K[2]+3\mid K[0]+K[1]\equiv 0)=P(S_1^G[t_1]=K[2]+3\mid K[0]+K[1]\equiv 0)=$

$$\sum_{x=0}^{N-1} P(t_1 = x \mid K[0] + K[1] \equiv 0) \cdot P(S_1^G[t_1] = K[2] + 3 \mid K[0] + K[1] \equiv 0, t_1 = x) =$$

$$\sum_{x=0}^{N-1} P(t_1 = x \mid K[0] + K[1] \equiv 0) \cdot P(S_1^G[x] = K[2] + 3 \mid K[0] + K[1] \equiv 0) >$$

$$> P(t_1 = 2 \mid K[0] + K[1] \equiv 0) \cdot P(S_1^G[2] = K[2] + 3 \mid K[0] + K[1] \equiv 0) =$$

$$P(t_1 = 2 \mid K[0] + K[1] \equiv 0) \cdot P(S_1^G[2] = K[0] + K[1] + K[2] + 3 \mid K[0] + K[1] \equiv 0) \approx$$

$$\approx P(t_1 = 2 \mid K[0] + K[1] \equiv 0) \cdot P(S_1^G[2] = K[0] + K[1] + K[2] + 3) \approx \left(\frac{N-1}{N}\right)^N \varphi_N$$

(последња једнакост добија се на основу теореме **4.4.** и леме **4.2.**)

■

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

4.2. Особине које се односе на све бајтове излазног низа

У одељку 4.2.1. изложене су Џенкинсонова теорема и њена последица, док се одељак 4.2.2. односи на низ вредност у у оквиру фазе PRGA.

4.2.1. Џенкинсова теорема

Џенкинсова теорема је једна од најважнијих слабости PRGA.

Теорема 4.6. (Џенкинсова)

У r-тој рунди PRGA, r ≥ 1, важи

$$P(S_r^G[j_r^G] = i_r^G - z_r) = P(S_r^G[i_r^G] = j_r^G - z_r) \approx \frac{2}{N}.$$

Доказ

Доказаћемо да је $P(S_r^G[j_r^G] = i_r^G - z_r) \approx \frac{2}{N}$, док се доказ да је $P(S_r^G[i_r^G] = j_r^G - z_r) \approx \frac{2}{N}$ изводи аналогно.

Доказ се може поделити на два случаја:

1. ако је $i_r^G = S_r^G[i_r^G] + S_r^G[j_r^G]$

$$z_r \overset{\text{def}}{=} S_r^G\left[S_r^G[i_r^G] + S_r^G[j_r^G]\right] = S_r^G[i_r^G] = i_r^G - S_r^G[j_r^G]$$

Догађај $i_r^G = S_r^G[i_r^G] + S_r^G[j_r^G]$ јавља се са вероватноћом $\frac{1}{N}$, док се догађај $S_r^G[j_r^G] = i_r^G - z_r$ јавља са вероватноћом 1.

Одатле следи да је вероватноћа овог случаја $\frac{1}{N} \cdot 1 = \frac{1}{N}$.

2. ако је $i_r^G \neq S_r^G[i_r^G] + S_r^G[j_r^G]$

Догађај $i_r^G \neq S_r^G[i_r^G] + S_r^G[j_r^G]$ јавља се са вероватноћом $\frac{N-1}{N}$, док се догађај $S_r^G[j_r^G] = i_r^G - z_r$ јавља са вероватноћом $\frac{1}{N}$.

Одатле следи да је вероватноћа овог случаја $\frac{N-1}{N} \cdot \frac{1}{N} = \frac{N-1}{N^2}$.
Сабирањем вероватноћа израчунатих у 1. и 2. случају добија се да је

$$P(S_r^G[j_r^G] = i_r^G - z_r) = \frac{2}{N} - \frac{1}{N^2} \approx \frac{2}{N}.$$

■

Последица 4.7.

$$P(z_r = r - S_{r-1}^G[r]) \approx \frac{2}{N}, \qquad \text{за} \quad r \geq 1.$$

Доказ

Како је у r-тој рунди $i_r^G = r$ на основу теореме **4.6.** важи да је

$$P(z_r = r - S_r^G[j_r^G]) \approx \frac{2}{N}, \qquad \text{за } r \geq 1.$$

На основу премештања у r-тој рунди, $S_r^G[j_r^G]$ добија вредност $S_{r-1}^G[r]$. Одатле је

$$P(z_r = r - S_{r-1}^G[r]) \approx \frac{2}{N}, \qquad \text{за } r \geq 1.$$

■

4.2.2. Низ вредности индекса j у фази PRGA

Ако се посматра један корак током фазе PRGA, тада је:

$$i'^G = i^G + 1$$

и

$$j'^G = j^G + S^G[i^G + 1],$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

где су i'^G и j'^G вредности у наредној рунди PRGA.

Ако су u и v дефинисани на следећи начин

$$u \stackrel{\text{def}}{=} S^G[i'^G] = S^G[i^G + 1]$$

и

$$v \stackrel{\text{def}}{=} S^G[j'^G] = S^G[j^G + u],$$

тада је после замене $S^G[j^G + u] = u$ и $S^G[i^G + 1] = v$, а излазни бајт $z = S^G[u + v]$.

Нека је $\psi(i^G, j^G, z)$ број пермутација од укупних $N!$ пермутација скупа $Z_N = \{0, \dots, N-1\}$, таквих да је за дате вредности i^G и j^G, излаз након једног корака PRGA баш z. Тада важи следеће својство:

Теорема 4.8.

Нека је N парно. Тада важи:

1. Ако је i^G парно, тада је

$$\psi(i^G, j^G, z) = \begin{cases} (N-1)! - (N-2)!, & \text{ако је } z = j^G; \\ (N-1)! + 2(N-3)!, & \text{иначе.} \end{cases}$$

2. Ако је i^G непарно и $j^G = \frac{i^G+1}{2}, \frac{i^G+N+1}{2}$, тада је

$$\psi(i^G, j^G, z) = \begin{cases} 2(N-1)! - (N-2)!, & \text{ако је } z = j^G; \\ (N-1)! - 2(N-2)!, & \text{ако је } z = j^G + \frac{N}{2}; \\ (N-1)! - (N-2)! + 2(N-3)!, & \text{иначе.} \end{cases}$$

3. Ако је i^G непарно и $j^G \neq \frac{i^G+1}{2}, \frac{i^G+N+1}{2}$, тада је

$$\psi(i^G, j^G, z) = \begin{cases} (N-1)! - (N-2)! + 2(N-3)!, & \text{ако је } z = h; \\ (N-1)! + 4(N-3)!, & \text{иначе.} \end{cases},$$

где је $h = j^G, i^G + 1 - j^G, \frac{i^G+1}{2}, \frac{i^G+N+1}{2}$.

■

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Доказ наведеног својства се може пронаћи у литератури [5].

Последица 4.9.

Под претпоставком да је свака од пермутација из скупа Z_N подједнако могућа пред сваки корак PRGA, важе следећа тврђења:

1. $P(j^G = z \mid i^G \text{ је непарно}) \geq \frac{1}{N} + \frac{1}{N^2}$
2. $P(j^G = z \mid i^G \text{ је парно}) \leq \frac{1}{N} - \frac{1}{N^2}$
3. $P(j^G = z \mid 2z = i^G + 1) = \frac{2}{N} + \frac{1}{N(N-1)}$.

Доказ

1. $P(j^G = z \mid i^G \text{ је непарно}) = \sum_m P(j^G = m)P(z = m \mid j^G = m,\ i^G \text{ је непарно})$

На основу претходне теореме (први случај у 2. и 3.) закључује се да претходни израз има вредност

$$\frac{1}{N} \cdot \left(2 \cdot \frac{2(N-1)! - (N-2)!}{N!} + (N-2) \cdot \frac{(N-1)! - (N-2)! + 2(N-3)!}{N!}\right) =$$

$$= \frac{1}{N} + \frac{2}{N^2} - \frac{1}{N(N-1)} + \frac{2}{N^2(N-1)} \geq \frac{1}{N} + \frac{1}{N^2}.$$

2. Из 1. дела ове теореме важи да је $P(j^G = z \mid i^G \text{ је непарно}) \geq \frac{1}{N} + \frac{1}{N^2}$.
Вероватноћа да је $P(j^G = z \mid i^G \text{ је парно}) = 1 - P(j^G = z \mid i^G \text{ је непарно})$.
Одатле следи да је $P(j^G = z \mid i^G \text{ је парно}) \leq \frac{1}{N} - \frac{1}{N^2}$.

3. $P(j^G = z \mid z = \frac{i^G+1}{2}) = \frac{P(j^G=z \mid z=\frac{i^G+1}{2})}{P\left(z=\frac{i^G+1}{2}\right)}$

На основу претходне теореме (првог случаја у 2.) добија се да је

$$P(j^G = z \mid z = \frac{i^G + 1}{2}) = \frac{1}{N} \cdot \frac{2(N-1)! - (N-2)!}{N!}$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Такође на основу претходне теореме (другог случаја у 2. и првог случаја у 1.) важи да је

$$P\left(z=\frac{i^G+1}{2}\right)=\frac{1}{N}\left(\frac{2(N-1)!-(N-2)!}{N!}+\frac{(N-1)!-2(N-2)!}{N!}+(N-2)\cdot\frac{(N-1)!-(N-2)!+3(N-3)!}{N!}\right)$$

Одатле је $P(j^G=z \mid z=\frac{i^G+1}{2})=\frac{2}{N}-\frac{1}{N(N-1)}$.

Доказ да је $P(j^G=z \mid z=\frac{i^G+N+1}{2})=\frac{2}{N}-\frac{1}{N(N-1)}$ се изводи аналогно.

Ако је дато $2z=i^G+1\ (mod\ N)$ догађаји $z=\frac{i^G+1}{2}$ и $z=\frac{i^G+N+1}{2}$ су једнако вероватни, чиме је доказ завршен.

■

Ако се разматрају два узастопна корака током фазе PRGA долази се до закључка да индекс j_{r+2}^G зависи од индекса j_r^G.

Лема 4.10.

За свако $r \geq 0$ следеће једнакости су еквивалентне:

1. $S_r^G[i_{r+1}^G]=\ i_r^G-j_r^G+2$
2. $j_{r+1}^G=i_r^G+2$
3. $j_{r+2}^G=2j_{r+1}^G-j_r^G$.

Доказ

Доказ да су прве две једнакости еквивалентне изводи се у два корака.

Најпре се доказује да $1.\ \Rightarrow\ 2,$ а потом и да $2.\ \Rightarrow\ 1.$
На основу својстава фазе PRGA закључује се да је $j_{r+1}^G=j_r^G+S_r^G[i_{r+1}^G]$, а на основу поставке теореме важи да је $S_r^G[i_{r+1}^G]=i_r^G-j_r^G+2$.

Одатле се заменом $S_r^G[i_{r+1}^G]$ у првој једнакости добија да је

$$j_{r+1}^G=j_r^G+i_r^G-j_r^G+2=i_r^G+2.$$

Аналогно се изводи доказ у другом смеру.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Доказ да су друга и трећа једнакост еквивалентне такође се изводи у два корака.

Најпре се доказује да $2. \Rightarrow 3.$ а потом и да $3. \Rightarrow 2.$
На основу својстава фазе PRGA је $j_{r+2}^G = j_{r+1}^G + S_{r+1}^G[i_{r+2}^G] = j_{r+1}^G + S_{r+1}^G[i_r^G + 2]$, а на основу поставке теореме је $j_{r+1}^G = i_r^G + 2$.

Одатле се заменом $i_r^G + 2$ у првој једнакости добија

$$j_{r+2}^G = j_{r+1}^G + S_{r+1}^G[j_{r+1}^G] = j_{r+1}^G + S_r^G[i_{r+1}^G] = j_{r+1}^G + (j_{r+1}^G - j_r^G) = 2j_{r+1}^G - j_r^G.$$

Аналогно се изводи доказ у другом смеру.

■

На основу претходно наведене леме изводи се следећа теорема.

Теорема 4.11.

$P(j_{r+2}^G = 2i_r^G + 4 - j_r^G) = \frac{2}{N}, \quad r \geq 0.$

Доказ

Догађај $j_{r+2}^G = 2i_r^G + 4 - j_r^G$ може да се одигра на два начина:

1. ако је $S_r^G[i_{r+1}^G] = i_r^G - j_r^G + 2$

Тада је на основу претходне леме

$$j_{r+2}^G = 2j_{r+1}^G - j_r^G = 2(i_r^G + 2) - j_r^G = 2i_r^G + 4 - j_r^G.$$

Вероватноћа овог дела је $P(S_r^G[i_{r+1}^G] = i_r^G - j_r^G + 2) = \frac{1}{N}$.

2. ако је $S_r^G[i_{r+1}^G] \neq i_r^G - j_r^G + 2$, али је због случајне расподеле $j_{r+2}^G = 2i_r^G + 4 - j_r^G$

Из претходне леме $S_r^G[i_{r+1}^G] \neq i_r^G - j_r^G + 2 \Rightarrow j_{r+2}^G \neq 2j_{r+1}^G - j_r^G$. На основу тога j_{r+2}^G може узети било коју од преосталих *N-1* вредности.

Одатле се добија да је вероватноћа овог дела

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

$$P(S_r^G[i_{r+1}^G] \neq i_r^G - j_r^G + 2) \cdot \frac{1}{N-1} = \left(1 - \frac{1}{N}\right) \cdot \frac{1}{N-1} = \frac{1}{N}.$$

Сабирањем добијених вероватноћа добија се да је

$$P(j_{r+2}^G = 2i_r^G + 4 - j_r^G) = \frac{2}{N}, \qquad r \geq 0.$$

■

Теорема 4.12.

$$P(j_{r+2}^G = i_{r+2}^G + j_{r+1}^G - j_r^G) = \frac{2}{N}, \qquad r \geq 0.$$

Доказ

Догађај $(j_{r+2}^G = i_{r+2}^G + j_{r+1}^G - j_r^G)$ може се одиграти на два начина:

1. ако је $j_{r+1}^G = i_r^G + 2$, тада је на основу леме **4.10.**

$$j_{r+2}^G = 2j_{r+1}^G - j_r^G = j_{r+1}^G + (i_r^G + 2) - j_r^G.$$

Вероватноћа овог случаја је $P(j_{r+1}^G = i_r^G + 2) = \frac{1}{N}$.

2. ако је $j_{r+1}^G \neq i_r^G + 2$, али је због случајне расподеле $j_{r+2}^G = i_{r+2}^G + j_{r+1}^G - j_r^G$

Вероватноћа овог случаја је $P(j_{r+2}^G = i_{r+2}^G + j_{r+1}^G - j_r^G \,|\, j_{r+1}^G \neq i_r^G + 2) = \left(1 - \frac{1}{N}\right) \cdot \frac{1}{N-1}$.

Сабирањем претходно добијених вероватноћа добија се да је

$$P(j_{r+2}^G = i_{r+2}^G + j_{r+1}^G - j_r^G) = \frac{2}{N}, \qquad r \geq 0.$$

■

Последица 4.13.

$$P(j_r^G = i_{r+2}^G - S_{r+2}^G[j_{r+2}^G]) = \frac{2}{N}, \qquad r \geq 0.$$

■

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Теорема 4.14.

$$P(z_{r+2} = j_r^G) \approx \frac{1}{N}\left(1 + \frac{2}{N}\right), \qquad r \geq 0.$$

Доказ

Догађај $(z_{r+2} = j_r^G)$ може се одиграти на два начина:

1. ако је $z_{r+2} = i_{r+2}^G - S_{r+2}^G[j_{r+2}^G]$ и $j_r^G = i_{r+2}^G - S_{r+2}^G[j_{r+2}^G]$

Тада је на основу теореме **4.6.** и последице **4.13.** вероватноћа овог случаја:

$$\frac{2}{N} \cdot \frac{2}{N} = \frac{4}{N^2}.$$

2. ако је $z_{r+2} \neq i_{r+2}^G - S_{r+2}^G[j_{r+2}^G]$, али је због случајне расподеле $z_{r+2} = j_r^G$.

Тада је вероватноћа овог случаја $\left(1 - \frac{2}{N}\right) \cdot \frac{1}{N}$.

Сабирањем добијених вероватноћа добија се да је

$$P(z_{r+2} = j_r^G) = \frac{4}{N^2} + \left(1 - \frac{2}{N}\right) \cdot \frac{1}{N} = \frac{1}{N} + \frac{2}{N^2} = \frac{1}{N}\left(1 + \frac{2}{N}\right).$$

■

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

5. Резултати симулације

У овом поглављу приказује се програм који је коришћен за експерименталну проверу тврђења приказаних у претходном поглављу, као и анализа добијених резултата. Већина тврђења изложених у раду дају приближне вероватноће. Циљ овог поглавља је спровођење статистичких тестова којима би се проверило да ли постоје статистички значајна одступања од приближних вредности параметара из тврђења у претходном поглављу.

5.1. Програмска реализација

Програм коришћен за практичну проверу резултата изведених теоријским путем састоји се из три логичке целине.

Прву целину представљају функције које симулирају рад основних компоненти алгоритма RC4 – фазу KSA и фазу PRGA. Другу целину чине функције које имају улогу у генерисању и проширивању случајних кључева намењених за симулацију рада алгоритма RC4. Последњу целину чине функције чија је улога генерисање резултата, као и оне функције које помажу при каснијој анализи добијених резултата.

Програм се покреће са унапред задатим бројем $m \in \{10000,20000,40000,80000\}$, где m представља број кључева који се генеришу. У свакој од функција на основу којих се добијају експериментални резултати, генерише се задати број кључева, са сваким од тих кључева се покрене алгоритам RC4 и потом се проверава колики број генерисаних кључева задовољава својство које се испитује. Овај процес представља један експеримент.

Сходно томе, један експеримент (у ознаци X) са m кључева можемо посматрати као низ независних догађаја $I_1, \ldots I_m$, при чему сваки од догађаја узима или вредност 1 у случају успеха или вредност 0 у случају неуспеха. Сматра се да је догађај I_i окончан успехом, ако је након покретања алгоритма RC4 установљено да је за кључ k_i задовољено својство које се испитује. Из претходно изложеног, јасно је да експеримент X представља број успешно реализованих догађаја међу догађајима $I_1, \ldots I_m$.

Ако I_i узима вредност 1 са неком вероватноћом p, онда I_i узима вредност 0 са вероватноћом q, где је $q = 1 - p$.

Дакле, вероватноћа догађаја I_i је

$$I_i : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

На основу тога очекивање догађаја I_i је

$$\mathrm{E}(I_i) = p,$$

док је његова дисперзија

$$D(I_i) = E(I_i^2) - E(I_i)^2 = p - p^2 = p(1-p).$$

Очекивање збира једнако је збиру очекивања

$$E(X) = E\left(\sum_1^m I_i\right) = E(I_1) + \cdots + E(I_m) = mp,$$

а из чињенице да су догађаји $I_1, \ldots I_m$ независни, следи да је

$$D(X) = D\left(\sum_1^m I_i\right) = D(I_1) + \cdots + D(I_m) = mp(1-p).$$

Експеримент X има биномну расподелу $B(m,p)$, која се може апроксимирати нормалном расподелом $N(\mu,\sigma^2)$, где је $\mu = E(X)$ и $\sigma = \sqrt{D(X)}$.

Расподела $N(\mu,\sigma^2)$ се своди на нормалну $N(0,1)$ расподелу трансформацијом у z-вредност, где је

$$z = \frac{x-\mu}{\sigma}.$$

Нека је за фиксирано m, експеримент X извршен $n = 10000$ пута и нека су реализације експеримента редом $x_1, \ldots, x_n$. Резултати реализације експеримента, $x_1, \ldots, x_n$, називају се тестом. Тест има биномну расподелу $B(m,p)$, која се може апроксимирати нормалном расподелом $N(\mu,\sigma^2)$, где је

$$\mu = \frac{x_1 + \cdots + x_n}{n}$$

и

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2},$$

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

а n и x_i представљају величину узорка и i-ти елемент у узорку.

У овом случају се на основу добијених резултата, спровођењем Пирсоновог χ^2 теста, проверава хипотеза H_0 да се резултати добијени експерименталним путем имају биномну $B(m,p)$ расподелу.

Нека је

$$\chi_n^2 = \sum_{k=1}^{l} \frac{(M_k - np_k)^2}{np_k},$$

где је $n \geq 50$ и $np_k \geq 5$. Број n представља величину узорка, l број категорија, M_k број узорака у k-тој категорији, а p_k вероватноћу у k-тој категорији. Хипотеза H_0 се одбацује ако важи да је

$$\sum_{k=1}^{l} \frac{(M_k - np_k)^2}{np_k} \geq \chi_{l-1,\alpha}^2 ,$$

где се вредност $\chi_{l-1,\alpha}^2$ добија очитавањем из одговарајуће статистичке таблице.

У случају да се врши само једно извођење експеримента, приликом анализе добијених резултата користи се чињеница да око 95% резултата за z-вредност треба да припада интервалу [-2,2].

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

5.2. Анализа добијених резултата

У одељку 5.2.1. приказани су резултати тестирања теореме **4.3**, док су у одељку 5.2.2. изложени резултати тестирања теореме **4.1**, теореме **4.6.** (Ценкинсова теорема) и теореме **4.14.**

5.2.1. Провера условљености првог излазног бајта кључем

Провера условљности првог излазног бајта кључем спроведена је тестирањем теореме **4.3.** Тестирање је спроведено са $m \in \{10000, 20000, 40000, 80000\}$ кључева. За сваки од ових тестова спроводи се $n = 10000$ експеримената.

У овом случају, провера добијених резултата, врши се спровођењем Пирсоновог χ^2 теста. Резултати експеримената су подељени у категорије, тако да у свакој категорији буде приближно једнак број елемената биномне расподеле $B(m, p)$. Подела исхода биномних расподела $B(10000, p)$ и $B(20000, p)$ извршена је на 7 категорија, док је подела исхода биномних расподела $B(40000, p)$ и $B(80000, p)$ извршена на 8 категорија. Вероватноћа p у биномној расподели добијена је на основу теореме **4.3.**

Хипотеза H_0 је тестирана са прагом значајности $\alpha = 0.05$, за коју се гранична вредност очитава из наредне таблице.

степени слободе	**α=0.05**
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07

Добијени резултати приказани су наредним графиконима и табелама. Најпре су изложени резултати за $m = 10000$ кључева, а потом и за остале вредности $m \in \{20000,40000,80000\}$.

У наредној табели приказане су експерименталне и очекиване вредности вероватноће p за све спроведене тестове.

	10000	**20000**	**40000**	**80000**
Тест 1	0.005216	0.005217	0.005217	0.005219
Тест 2	0.005208	0.005213	0.005221	0.005218
Очекивана вредност	0.005345			

На слици 1 представљени су хистограми за два различита случајна теста са $m = 10000$ кључева. Приказани хистограми представљају расподелу вредности експеримента X у оквиру једног теста.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Вероватноћа да $\sum_{i=1}^{10000} x_i$ припада одређеном интервалу, број очекиваних погодака по интервалу, као и резултати Пирсоновог χ^2 теста спроведеног са 10000 кључева приказани су наредном табелом.

Категорија	**Интервал I**	$P(z \in I)$	np_k	M_k	$(M_k - np_k)^2$	$\frac{(M_k - np_k)^2}{np_k}$
1	≤ 44	0.144	1440	1430	100	0.069444
2	[45, 47]	0.115	1150	1209	3481	3.026957
3	[48, 50]	0.150	1500	1543	1849	1.232667
4	[51, 53]	0.165	1650	1626	576	0.349091
5	[54, 56]	0.153	1530	1462	4624	3.022222
6	[57, 60]	0.150	1500	1498	4	0.002667
7	≥ 61	0.123	1230	1232	4	0.003252
						7.706299

Резултати поновљеног тестирања са 10000 кључева добијају се поступком приказаним у претходној табели, а вредност χ_n^2 за овај тест износи **6.194857.**

На слици 2 представљени су хистограми за два различита случајна теста са $m = 20000$ кључева. Приказани хистограми представљају расподелу вредности експеримента X у оквиру једног теста.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Резултати првог теста са 20000 кључева приказани су наредном табелом.

Категорија	**Интервал I**	$P(z \in I)$	np_k	M_k	$(M_k - np_k)^2$	$\frac{(M_k - np_k)^2}{np_k}$
1	≤ 93	0.144	1440	1385	3025	2.100694
2	[94, 98]	0.139	1390	1462	5184	3.729496
3	[99, 102]	0.145	1450	1517	4489	3.095862
4	[103, 106]	0.156	1560	1552	64	0.041026
5	[107, 110]	0.143	1430	1410	400	0.279720
6	[111, 115]	0.136	1360	1339	441	0.324265
7	≥ 116	0.137	1370	1335	1225	0.894161
						10.46522

Вредност χ_n^2 за други тест са 20000 кључева је **12.54346.**

На слици 3 представљени су хистограми за два различита случајна теста са $m = 40000$ кључева. Приказани хистограми представљају расподелу вредности експеримента X у оквиру једног теста.

Резултати првог теста са 40000 кључева приказани су наредном табелом.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Категорија	Интервал I	$P(z \in I)$	np_k	M_k	$(M_k - np_k)^2$	$\frac{(M_k - np_k)^2}{np_k}$
1	≤ 192	0,131	1310	1294	256	0.19542
2	[193, 198]	0.109	1090	1121	961	0.881651
3	[199, 203]	0.120	1200	1241	1681	1.400833
4	[204, 208]	0.136	1360	1317	1849	1.359559
5	[209, 213]	0.136	1360	1367	49	0.036029
6	[214, 218]	0.121	1210	1227	289	0.238843
7	[219, 225]	0.126	1260	1236	576	0.457143
8	≥ 226	0.121	1210	1197	169	0.139669
						4.709148

Вредност χ_n^2 за други тест са 40000 кључева је **8.334824.**

На слици 4 представљени су хистограми за два различита случајна теста са $m = 80000$ кључева. Приказани хистограми представљају расподелу вредности експеримента X у оквиру једног теста.

Резултати првог теста спроведеног са 80000 кључева приказани су у наредној табели.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Категорија	**Интервал I**	$P(z \in I)$	np_k	M_k	$(M_k - np_k)^2$	$\frac{(M_k - np_k)^2}{np_k}$
1	≤ 393	0,119	1190	1167	529	0.444538
2	[394, 402]	0.112	1120	1133	169	0.150893
3	[403, 409]	0.116	1160	1127	1089	0.938793
4	[410, 415]	0.114	1140	1177	1369	1.200877
5	[416, 421]	0.116	1160	1173	169	0.14569
6	[422, 428]	0.128	1280	1333	2809	2.194531
7	[429, 437]	0.132	1320	1303	289	0.218939
8	≥ 438	0.163	1630	1587	1849	1.134356
						6.428617

Вредност χ_n^2 за други тест са 80000 кључева је **6.646295.**

На основу добијених резултата закључује се да тестирани узорци пролазе Пирсонов χ^2 тест.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

5.2.2. Тестирање теорема 4.1, 4.6. и 4.14.

Тестирање осталих својстава такође се врши са по 10000, 20000, 40000 односно 80000 кључева, али се због специфичности тврђења која се проверавају, спроводи само по један експеримент. Из резултата експеримента се добија вероватноћа на основу које се рачуна z-вредност, а добијени резултати су приказани табеларно.

На овај начин тестирана су следећа својстава:

- Зависност првог излазног бајта од пермутације S - теорема **4.1.**
- Ценкинсонова теорема - теорема **4.6.**
- Зависност излазног бајта у (r+2)-ој рунди од индекса j у r-тој рунди - теорема **4.14.**

Тестирања свих својстава су ограничена на генерисање првих 128 вредности, док је првих 10 приказано у табелама. Уколико су узорци у складу са предвиђеном расподелом вероватноћа, очекивано је да око 5% z-вредности буде ван интервала [-2,2].

Резултати су приказани у наредној табели.

Зависност првог излазног бајта од пермутације S				
	n=10000	n=20000	n=40000	n=80000
$z \notin (-2, 2)$	6	7	5	6
$z \in [-2, 2]$	122	121	123	122
Ценкинсонова теорема				
	n=10000	n=20000	n=40000	n=80000
$z \notin (-2, 2)$	4	3	5	7
$z \in [-2, 2]$	124	125	123	121
Зависност излазног бајта у (r+2)-ој рунди од индекса j у r-тој рунди				
	n=10000	n=20000	n=40000	n=80000
$z \notin (-2, 2)$	3	6	4	5
$z \in [-2, 2]$	125	122	124	123

На основу резултата приказаних у претходној табели, може се закључити да заиста око 95% z-вредности припада интервалу [-2,2].

У наредној табели приказани су резултати добијени тестирањем теореме **4.1.**

x	n=10000	z-вредност	n=20000	z-вреднсот	n=40000	z-вредност	n=80000	z-вредност
0	32	-1.09	**58**	**-2.22**	**118**	**-2.98**	**253**	**-3.25**
1	34	-0.81	69	-1.03	173	1.34	291	-1.22
2	70	-0.91	149	-0.56	278	-1.93	**571**	**-2.12**
3	36	-0.49	71	-0.81	151	-0.42	328	0.88
4	28	-1.73	75	-0.29	164	0.72	303	-0.40
5	28	-1.77	72	-0.69	141	-1.22	313	0.03
6	32	-1.09	79	0.17	131	-1.93	276	-1.94
7	36	-0.49	73	-0.58	149	-0.58	278	-1.96
8	29	-1.57	84	0.74	140	-1.21	316	0.34
9	39	-0.01	68	-1.15	173	1.34	291	-1.22

У наредној табели приказани су резултати тестирања **Ценкинсонове теореме**.

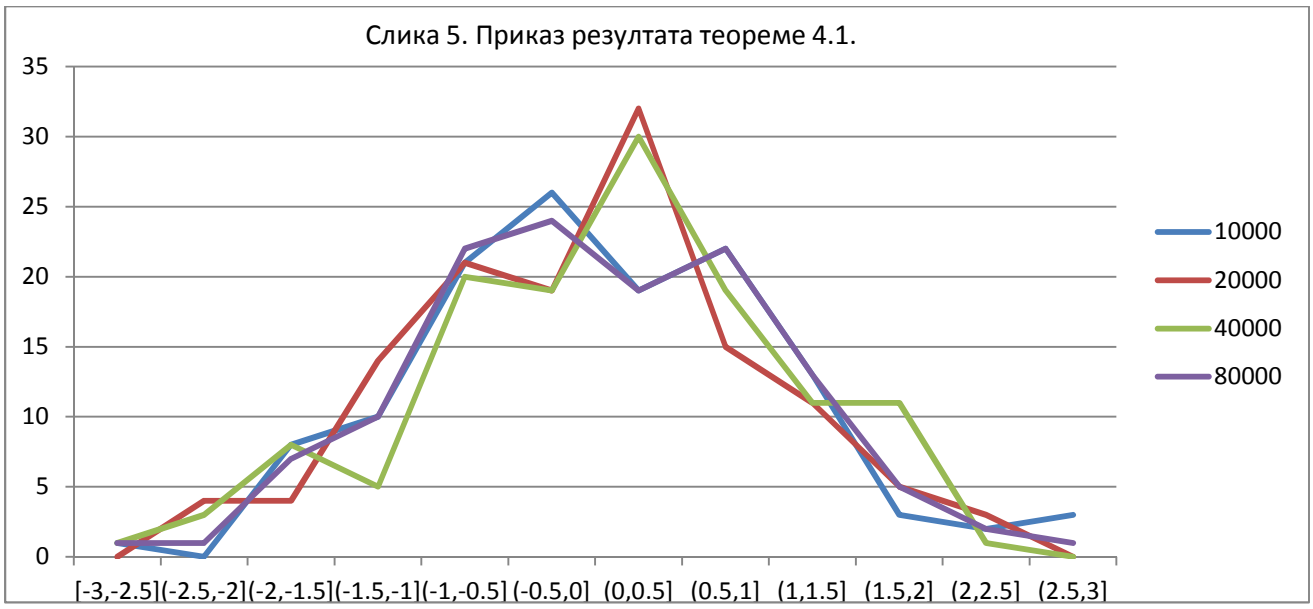
бајт	n=10000	z-вредност	n=20000	z-вредност	n=40000	z-вредност	n=80000	z-вредност
1	78	-0.01	164	0.62	295	-0.99	632	0.28
2	77	-0.13	169	1.02	294	-1.05	577	-1.93
3	**104**	**2.94**	171	1.18	345	1.85	636	0.44
4	75	-0.35	152	-0.34	336	1.33	629	0.16
5	88	1.12	158	0.14	323	0.60	604	-0.84
6	73	-0.58	172	1.26	285	-1.56	601	-0.96
7	76	-0.24	142	-1.14	323	0.60	638	0.52
8	71	-0.81	150	-0.50	329	0.94	638	0.52
9	73	-0.58	149	-0.58	308	-0.26	**676**	**2.05**
10	71	-0.81	157	0.06	291	-1.22	607	-0.72

У наредној табели приказани су резултати провере теореме **4.14.**

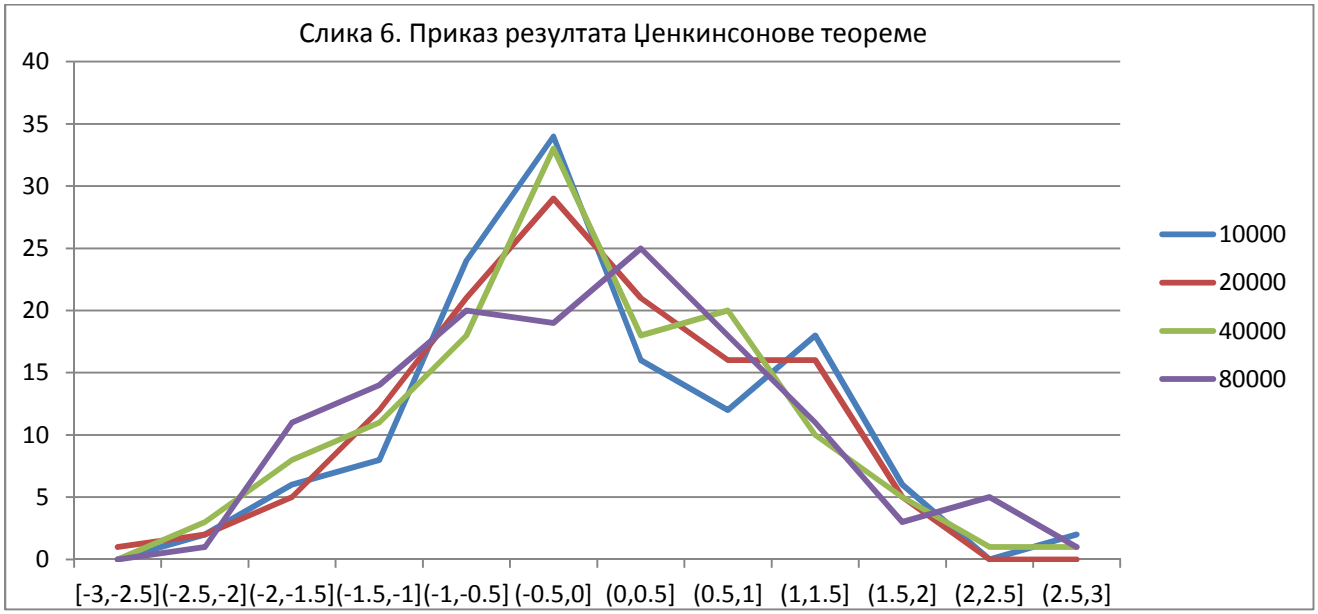
бајт	n=10000	z-вредност	n=20000	z-вредност	n=40000	z-вредност	n=80000	z-вредност
1	30	-1.50	91	1.54	163	0.44	313	-0.11
2	43	0.58	78	0.06	143	-1.16	329	0.79
3	32	-1.18	85	0.86	165	0.60	329	0.79
4	31	-1.34	72	-0.62	150	-0.60	309	-0.33
5	39	-0.06	83	0.63	144	-1.08	317	0.11
6	28	-1.82	65	-1.41	154	-0.28	288	-1.52
7	38	-0.22	83	0.63	153	-0.36	334	1.08
8	43	0.58	67	-1.19	162	0.36	296	-1.07
9	46	1.06	72	-0.62	168	0.84	314	-0.05
10	40	0.10	81	0.40	**189**	**2.52**	287	-1.58

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Из практичних разлога у претходним табелама приказани су само резултати за првих 10 бајтова. На основу комплетних резултата добијених спровођењем тестова направљени су следећи хистограми. Хистограми су добијени тако што су z-вредности које су у интервалу од [-3,3] груписане у мање интервале.



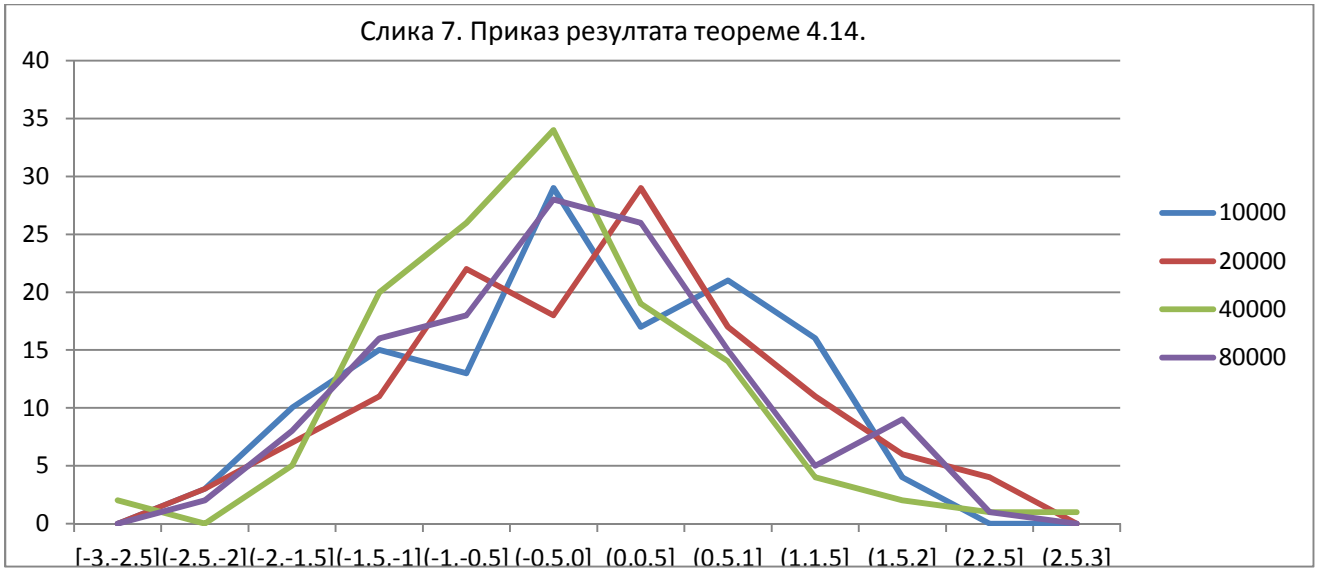
На слици 5 приказан је хистограм z-вредности добијених тестирањем теореме **4.1.** са $m \in \{10000, 20000, 40000, 80000\}$ кључева.



Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

На слици 6 приказан је хистограм z-вредности добијених тестирањем **Ценкинсонове теореме** са $m \in \{10000, 20000, 40000, 80000\}$ кључева.

На наредној слици (слика 7) приказан је хистограм z-вредности добијених тестирањем теореме **4.14.** са $m \in \{10000, 20000, 40000, 80000\}$ кључева.



На основу добијених вредности и приказаних хистограма, може се закључити да се најмањи број z-вредности налази на ободима интервала [-3,3], а да тај број расте приближавањем средини овог интервала.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

5.2.3. Одступање експерименталне вероватноће p' од вероватноће p дате теоремом 4.3.

У тестовима у одељку 5.2.1, p је биномна вероватноћа из теореме 4.3. и статистички тестови не показују значајно одступање узорака од ње. Због тога се спроводе други, једноставнији тестови са већим узорцима чији је циљ да се провери да ли постоји статистички значајно одступање процене биномне вероватноће из узорка од биномне вероватноће p из теореме 4.3.

За довољно велик број кључева (у ознаци m), вероватноћа p добијена на основу теореме 4.3. статистички ће се разликовати од експериментално добијене вероватноће (у ознаци p'). Може се сматрати да се вероватноће p и p' статистички значајно разликују ако је

$$|p - p'| > 4\sigma,$$

при чему се користи чињеница да се вредности стандардне девијације из тврђења и њена процена из узорка практично не разликују.

Из чињенице да је $\sigma = \sqrt{\frac{p(1-p)}{m}}$, претходна неједнакост своди се на

$$\Delta p > 4\sqrt{\frac{p \cdot (1-p)}{m}},$$

где је $\Delta p = p - p'$.

Како је p мало, израз $1 - p$ може се апроксимирати јединицом, а претходна неједнакост своди се на

$$m > \frac{16 \cdot p}{\Delta p^2}.$$

Заменом вредности p и p' у претходној неједнакости добија се да се процена величине узорка m при којој се може очекивати да вредност p' из узорка статистички значајно одступа од вредности p из теореме 4.3. Заменом конкретних вредности у претходној неједнакости, добија се да се за $m > 10^7$ вероватноће p и p' статистички значајно разликују. Ради провере наведених својстава извршено је по 10 тестова са $m \in \{10^4, 10^5, 10^6, 10^7, 10^8\}$ кључева.

У наредној табели приказани су резултати $z = \frac{|p-p'|}{\sigma}$ за различите величине узорка m, где је вероватноћа p на основу теореме 4.3. приближно једнака $p \approx 0.005345$.

	10^4		10^5		10^6		10^7		10^8	
	p'	z	p'	z	p'	z	p'	z	p'	z
1	0.005900	0.76	0.004990	1.54	0.005215	1.78	0.005264	3.51	0.005221	17.05
2	0.005300	0.06	0.005060	1.23	0.005219	1.73	0.005234	4.82	0.005210	18.54
3	0.005200	0.20	0.005230	0.50	0.005284	0.84	0.005218	5.49	0.005221	17.02
4	0.004200	1.57	0.005090	1.11	0.005133	2.91	0.005204	6.10	0.005222	16.83
5	0.004000	1.84	0.005220	0.54	0.005215	1.78	0.005200	6.28	0.005216	17.70
6	0.005000	0.47	0.005460	0.50	0.005239	1.45	0.005196	6.45	0.005209	18.64
7	0.005400	0.08	0.004720	2.71	0.005334	0.15	0.005232	4.90	0.005218	17.46
8	0.005700	0.49	0.005200	0.63	0.005226	1.63	0.005186	6.91	0.005211	18.44
9	0.006000	0.90	0.005130	0.93	0.005223	1.67	0.005239	4.59	0.005236	14.98
10	0.005000	0.47	0.005250	0.41	0.005103	3.32	0.005184	6.97	0.005216	17.67

На основу резултата приказаних у табели јасно је да се за $m = 10^7$ и $m = 10^8$ кључева вероватноће p и p' статистички значајно разликују. Слично понашање очекује се и за остала тврђења наведена у овом раду.

На основу тестова спроведених приликом израде овог рада, тестови у одељцима 5.2.1. и 5.2.2. нису открили одступања од приближних вредности параметара у тврђењима изложеним у четвртом поглављу, док је у одељку 5.2.3. откривено статистички значајно одступање.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

6.Закључак

У овом раду уведени су основни криптографски појмови, дат детаљан опис алгоритма RC4 као и преглед особина излазног низа који генерише овај алгоритам. На основу изнетих особина, спроведени су тестови и урађена је статистичка анализа добијених резултата која је изложена у 5. поглављу.

Опис алгоритма RC4, особине излазног низа генератора RC4, као и поједини делови кода који су коришћени у програму, написани су коришћењем литературе [1].

Као помоћно средство за добијање и обраду добијених резултата, коришћен је програм писан у програмском језику C који се налази у прилогу овог рада. Кључеви за тестирање својстава наведених у 4. поглављу, генеришу се у два корака. У првом кораку се простим линеарним генератором псеудослучајних бројева изгенерише низ бројева, а потом се у другом кораку добијени низ шифрује алгоритмом RC4. На тај начин добија се случајан кључ. Тестирање одређеног својства, спроведи се по поступку укратко описаном у првом одељку 5. поглавља.

У овом раду извршена је практична провера централних тврђења из сваког одељка у 4. поглављу, осим условљености првог излазног бајта кључем, уколико додатно важи да је $K[0] + K[1] \equiv 0 \ (mod\ N)$. Описани начин за добијање кључева користи се и за тестирање овог својства, али се приликом генерисања кључева одбацују сви кључеви који не задовољавају услов $K[0] + K[1] \equiv 0 \ (mod\ N)$. На тај начин је очекивано да генерисање резултата за 10.000 експеримената са само 10.000 кључева траје преко 80 сати.

На основу статистичких тестова који су спроведени у одељку 5.2.3, може се закључити да је пронађен тест који даје статистички значајно одступање од приближних вредности параметра из тврђења 4.3.

Даља истраживања би се могла усмерити на унапређивање алгоритма за генерисање кључева који задовољавају услов $K[0] + K[1] \equiv 0 \ (mod\ N)$.

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs

7.Литература

[1] G. Paul, S. Maitra, RC4 Stream Cipher and its Variants, CRC Press, 2012

[2] М. Живковић, Криптографија , Математички факултет, Београд, 2012

[3] B. Schneier, Примењена криптографија, Микро књига, Београд, 2007

[4] С. Синг, Књига о шифрама, Плато, Београд, 2003

[5] R. Basu, S. Ganguly, S. Maitra, G. Paul, A Complete Characterization of the Evolution of RC4 Pseudo Random Generation Algorithm, Journal of Mathematical Cryptology, 2008

[6] A. Roos, A Class of Weak Keys in the RC4 Stream Cipher, sci.crypt, 1995

[7] D. Wagner, My RC4 weak keys, sci.crypt, 1995

[8] П. Младеновић, Вероватноћа и статистика, Математички факултет, Београд, 2008

[9] stattrek.com/online-calculator/binomial.aspx

[10] H. Finney, An RC4 cycle that can't happen, sci.crypt, 1994

Virtual Library of Faculty of Mathematics - University of Belgrade
elibrary.matf.bg.ac.rs