ELECTRONICS AND ELECTRICAL ENGINEERING

ISSN 1392 – 1215 *2006. Nr. 2(66)*

ELEKTRONIKA IR ELEKTROTECHNIKA

TELECOMMUNICATION ENGINEERING

T 180

TELEKOMUNIKACIJŲ INŽINERIJA

Srautinių šifrų testavimas Euklido algoritmo raundų skaičiaus analizės testu

P. Nefas

Telekomunikacijų katedra, Kauno technologijos universitetas,
Studentų g. 50, LT–51368 Kaunas, Lietuva, tel. +370 37 223459, el. p. petrasnefas@gmail.com

Įvadas

Sisteminiu teoriniu požiūriu srautinių šifrų kriptografinis atsparumas nustatomas, įvertinus generuojamos sekos tinkamumą šifruoti[1]. Yra daug šifravimo sekos kokybės kriterijų, tačiau visus juos galima suskirstyti į dvi būdingas šifravimo sekos kokybės nustatymo metodikų grupes.

Pirmoji metodikų grupė susijusi su šifravimo sekos statistinių charakteristikų vertinimu ir tam tikro statistinio disbalanso, leidžiančio analitikui prognozuoti kito bito reikšmę su tikimybe, didesne nei paprastas spėjimas, paieška. Šiuo požiūriu reikalaujama, kad šifravimo seka turėtų tokias pat statistines charakteristikas kaip ir atsitiktinė seka.

Antroji metodikų grupė susijusi su bandymu, turint šifravimo sekos fragmentą, sukonstruoti generatorių, kuris atkartotų visą šifravimo seką. Šiuo požiūriu vertinamas šifravimo sekos kompleksiškumas ir bandoma atsakyti į klausimą, ar yra sunku atkurti seką iš turimo fragmento ir kokio ilgio fragmentą tam reikia turėti.

Bendruoju atveju šifro kriptografinis atsparumas vertinamas testuojant šifravimo seką įvairiais turimais statistiniais testais. Testavimo esmė – tai tikrinimas vadinamosios „nulinės hipotezės“, kuri tvirtina, kad analizuojamoji seka yra gauta Bernulio schema , t. y. vieneto pasirodymo kiekviename nepriklausomame bandyme tikimybė yra lygi ½.

Statistinį testą S_T dvejetainėms sekoms, kurių ilgis L ,galima traktuoti kaip funkciją [2]

$$S_T : V_L \to \{0,1\} = \{'Priimtl', 'Atmestl'\}, \qquad (1)$$

kuri dalija dvejetainių L ilgio sekų visumą V_L į neatsitiktinių sekų aibę $V_{L,0}$ ir atsitiktinių sekų $V_{L,1}$ aibę

$$V_{L,j} = \{s^L \in V_L : S_T(s^L) = j\}, j \in \{0,1\}, \qquad (2)$$

čia $s^L = (s_1, s_{2,}, .., s_L)$.

Tikimybė ρ, kad atsitiktinė L ilgio seka bus atmesta testo, yra $\rho = |V_{L,0}| \cdot 2^{-L}$. Dažniausiai parenkamos tokios testavimo sąlygos, kad ρ lygi 0,01...0,1. Kai L yra didelis, funkcijos S_T vertei nustatyti reikėtų didelių skaičiavimo išteklių, todėl testas realizuojamas efektyviai apskaičiuojama testine funkcija (statistika) f_{S_T}, kuri aibę V_L atvaizduoja realių skaičių aibe R^L.

Dažniausiai pasirenkama tokia f_{S_T} funkcija, kurios atsitiktinio dydžio $f_{ST}(R^L)$ pasiskirstymas būtų artimas kokiam nors gerai žinomam etaloniniam skirstiniui, paprastai – tai normalinis arba λ^2 skirstinys, kadangi yra sudarytos šių tikimybinių skirstinių lentelės, o tai palengvina įvertinti gautą statistiką. Yra sukurta daugybė testų [3],[4],[5], kurie skirtingais aspektais analizuoja sekos atsitiktinumo matą. Testų skaičius nėra galutinis, todėl jei tiriant sekas aptinkamas naujas statistinis silpnumas, konstruojamas naujas statistinis testas, kuris papildo žinomų statistinių testų rinkinį.

Euklido algoritmo raundų skaičiaus T_N statistikos analizė

Naujo statistinio testo pagrindas yra Euklido algoritmo raundų skaičiaus statistikos analizė. Euklido algoritmas skaičiuoja dviejų natūrinių skaičių u, v didžiausią bendrą vardiklį. Šis algoritmas yra rekurentinis. Raundų skaičius T_i priklauso nuo konkrečios natūrinių skaičių poros u, v vertės. Tarkime, natūriniai skaičiai u ir v yra nepriklausomi dydžiai ir tolydžiai pasiskirstę diapazone $1...N$. Ieškoma, kaip Euklido algoritmo raundų skaičius T_i priklauso nuo diapazono N. Yra žinoma [8], kad raundų skaičius yra diapazone

$$1 \le T_i \le 4{,}8 \lg N - 0{,}32, \qquad (4)$$

o raundų skaičiaus vidurkis [8]

$$\overline{T}_i = \frac{12 \ln 2}{\pi^2} * \ln N + Q(N), \qquad (5)$$

čia $Q(N)$ – korekcinė dedamoji.

Norint rasti tikimybinį raundų skaičiaus pasiskirstymą $P(T_i) = f(N)$, Euklido algoritmas pritaikytas visoms natūrinių skaičių u ir v poroms, čia $1 \le u, v \le N$. Bendras porų skaičius N^2. Buvo atliktas kompiuterinis modeliavimas, kai N=512, 1024, 2048, 4096, 8192.

Tarkim, skaičius K_{Ti}^N žymi raundų skaičiaus pasikartojimo dažnį, t. y. parodo, kelioms natūrinių skaičių $u, v \in [1...N]$ poroms, perrinkus visas įmanomas variacijas, Euklido algoritmo raundų skaičius yra lygus T_i, kur T_i, kaip išplaukia iš (1), turi tokias vertes:

$$T_i \in [1,..Int(4{,}8 \log_{10} N - 0{,}32)]. \qquad (6)$$

Apskaičiuota kiekvieno T_i pasikartojimo tikimybė

$$P(T_{Ni}) = \frac{K_{Ti}^N}{N^2}, \qquad (7)$$

empirinis vidurkis

$$\hat{T}_N = \frac{1}{N} \sum_1^N T_{Ni} = \sum_{Ti=1}^{Ti=\max} Ti * (P(T_{Ni})), \qquad (8)$$

empirinė dispersija

$$S_0^2 = \frac{1}{n} \sum_{i=1}^n (\hat{T}_N - T_{Ni})^2. \qquad (9)$$

Rezultatai pateikti 1 lentelėje. Grafike (1 pav.) pateiktas empirinis tikimybinis raundų skaičiaus pasiskirstymas.

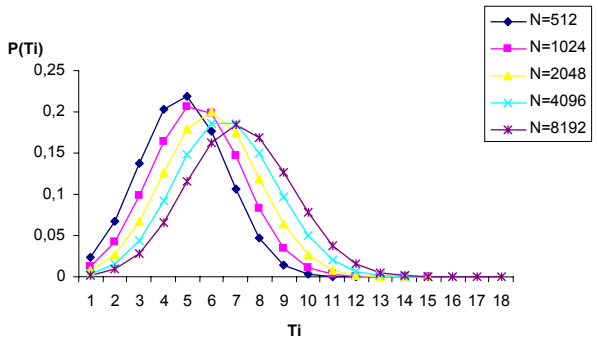
Modeliavimo rezultatai leidžia empirinį tikimybinį pasiskirstymą aproksimuoti normaliniu skirstiniu.

1 lentelė. Raundų skaičiaus analizės kompiuterinio modeliavimo rezultatai

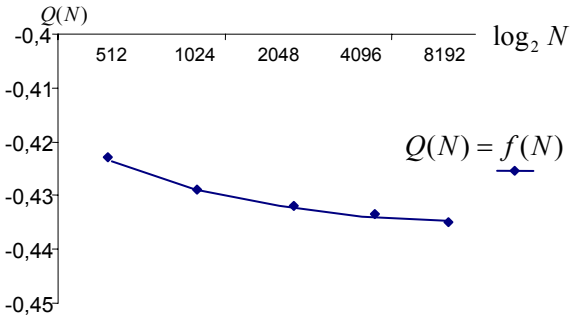
N	512	1024	2048	4096	8192
$\hat{T}_N$	4,83	5,41	5,99	6,58	7,16
S_0^2	3,07	3,44	3,82	4,18	4,54
Q(N)	-0,42	-0,42	-0,43	-0,43	-0,43

Normalinį skirstinį apibrėžia vidurkis ir dispersija. Galima išreikšti šių parametrų priklausomybę nuo diapazono N. Remiantis (5) ir (8) išraiškomis apskaičiuota korekcinės dedamosios $Q(N)$ priklausomybė nuo diapazono N:

$$Q(N) = \frac{1}{N} \sum_{N=1}^N T_N - \frac{12 \ln 2}{\pi^2} * \ln N. \qquad (11)$$



1 pav. Euklido algoritmo raundų skaičiaus empirinis tikimybinis pasiskirstymas



2 pav. Korekcinės dedamosios $Q(N)$ priklausomybė nuo diapazono N

Kaip matyti iš logaritminėje skalėje pateiktos empirinės $Q(N)$ priklausomybės nuo N grafiko (2 pav.), išraišką galima aproksimuoti kvadratine logaritmine lygtimi

$$Q(n) = a * \log_2^2 N + b * \log_2 N + c. \qquad (13)$$

Tada

$$Q(n) \approx 0{,}87 \cdot 10^{-3} \log_2^2 N - 0{,}02 \log_2 N - 0{,}29. \qquad (14)$$

Vadinasi,

$$\overline{T}_N = 0{,}62 \cdot 10^{-3} \log_2^2 N + 0{,}57 \log_2 N - 0{,}32. \qquad (15)$$

Paskaičiuojame empirinę dispersiją

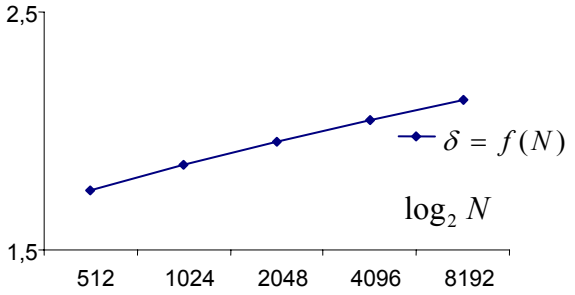
$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^n (\overline{T}_N - T_{Ni})^2}. \qquad (16)$$

Matyti (3 pav.), kad funkcija $\delta = f(N)$ artima tiesinei, todėl ją galima aproksimuoti tiesine logaritmine lygtimi

$$\delta \approx a * \log_2 N + b. \qquad (17)$$

Aproksimuojant tiesine lygtimi gaunama standartinio nuokrypio išraiška:

$$\delta \approx 0{,}095 \log_2 N + 0{,}896. \qquad (18)$$



3 pav. Dispersijos δ priklausomybė nuo diapazono N

Vadinasi, dviejų atsitiktinių natūrinių skaičių, tolydžiai pasiskirsčiusių diapazone [1, N], didžiausio bendro daliklio paieškos Euklido algoritmu raundų skaičiaus T_i tikimybinis tankis $p(T_i)$ yra aprašomas normaliniu skirstiniu, esant šiems parametrams:

$$p(T_i) = \phi(T_i, \overline{T}_N, \delta) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(T_i - \overline{T}_N)^2}{2\delta^2}}, \quad (19)$$

$$T_i \in [1, Int(4{,}8 \log_{10} N - 0{,}32)], \quad (20)$$

$$\overline{T}_N = 0{,}62 \cdot 10^{-3} \log_2^2 N + 0.57 \log_2 N - 0{,}32, \quad (21)$$

$$\delta = 0{,}095 \log_2 N + 0{,}896. \quad (22)$$

Srautinių šifrų testavimo Euklido algoritmo raundų testu metodika

Tai, kad, kaip anksčiau nustatyta, dviejų atsitiktinių natūrinių skaičių $a, b \in [1..N]$, kurių tikimybinis skirstinio tankis yra konstanta, t. y. $P(a) = P(b) = 1/N$, visame galimame verčių diapazone atitinkamo natūrinio skaičiaus T_i, kuris lygus Euklido algoritmo raundų skaičiui, reikalingo didžiausio bendro daliklio paieškai, tikimybinis skirstinio tankis $P(T_i)$ yra apibrėžtas normaliniu skirstiniu, leidžia sukonstruoti statistinį testą, kuris tikrintų šifravimo sekos atsitiktinumą. Tarkim, testuojamoji n ilgio dvejetaine seka yra $\varsigma = \varsigma_1 \varsigma_2 ... \varsigma_n$. Seka suskirstoma į $M = \frac{n}{N}$ viena kitos nedengiančių N ilgio atkarpų. Likę bitai atmetami. Kiekvienai *N* bitų atkarpai priskirtas natūrinis skaičius, kuris gaunamas konvertuojant *N*-bitį binarinį kodą į dešimtainės formos skaičių pridėjus 1 tam , kad nebūtų dalybos iš nulio. Turime natūrinių skaičių seką $a = a_1, a_2 ..., a_i, .., a_N$,čia $1 \le a_i \le 2^N$. Kiekvienai gretimai su cikliniu postūmiu skaičių porai $a_i : a_{i+1}$ (viso porų M) priskirtas tos poros Euklido algoritmo raundų skaičius T_i, kuris yra diapazone

$$T_i \in [1, Int(4{,}8 \log_{10} 2^N - 0.32)], \quad (23)$$

čia $T_i = 1$, jei skaičiai poroje tarpusavyje pirminiai ir $T_{i=\max} = Int(4.8 \lg 2^N - 0.32)$, jei raundų skaičius yra maksimaliai galimas.

Suformuojama statistika

$$\chi^2(obs) = \sum_{i=1}^{i=Int(4{,}8_{10} 2^M - 0{,}32)} \frac{(K_{Ti}^N - Mp(Ti))^2}{Mp(Ti)}, \quad (24)$$

čia K_{Ti}^N –raundų dažnumas, t. y. skaičius porų, kurių raundų skaičius T_i, o $p(Ti)$ – teorinis normalinis skirstinys

$$p(T_i) = \phi(T_i, \overline{T}_N, \delta) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(T_i - \overline{T}_N)^2}{2\delta^2}}. \quad (25)$$

Klaidos reikšmė:

$$P - value = igamc\left(\frac{\frac{Int(4{,}8 \lg 2^M - 0{,}32)}{2}}{\frac{\chi^2(obs)}{2}}\right). \quad (26)$$

Jei *P-value* daugiau nei 0,01, laikoma, kad generuojama seka šio testo požiūriu yra atsitiktinė.

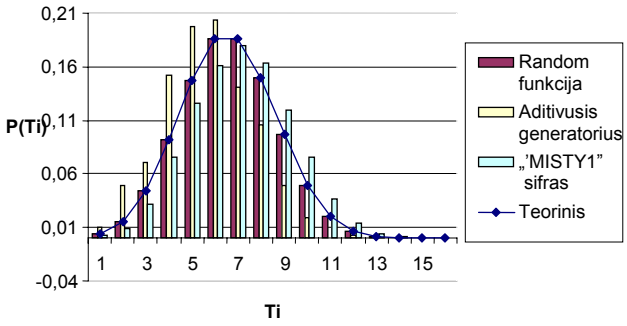
Funkcijos Random, adityviojo kongruencinio generatoriaus ir MISTY1 testavimas Euklido algoritmo testu

Naujuoju testu buvo patikrintas sekų, kurias generuoja PC atsitiktinė funkcija Random, kongruencinis generatorius $x_{n+1} = (430x_n + 374441) \bmod 1771875$ ir šifras MISTY1, kuris yra 3GPP šifro „Kasumi“ prototipas.

2.lentelė Testavimo rezultatai

	Random funkcija	Adityvusis generatorius	MISTY1 šifras
$\overline{T}_N$	6,58	5,63	7,04
δ	2,04	1,94	2,15
$\chi^2(obs)$	14,93	130532,9	265,1416
$P - value$	0,535	10^{-280}	$0{,}40 * 10^{-46}$

Šiuo testo požiūriu Random funkcijos generatoriaus generuojamos sekos yra atsitiktinės, o generatoriaus $x_{n+1} = (430x_n + 374441) \bmod 1771875$ ir MISTY1 šifro sekos yra neatsitiktinės.



4 pav. Testavimo rezultatai

Išvados

1. Atlikta analizė parodė, kad sisteminiu techniniu požiūriu srautinių šifrų kriptografinį atsparumą galima nustatyti testuojant šifravimo seką įvairiais statistiniais testais.
2. Pateiktas naujas, iki šiol mokslinėje techninėje literatūroje nepublikuotas statistinis testas, kurio pagrindas – Euklido algoritmo raundų skaičiaus statistikos analizė.
3. Įrodyta, kad raundų skaičiaus tikimybinis skirstinys yra normalinis. Nustatyta šio skirstinio vidurkio ir dispersijos priklausomybė nuo analizuojamų verčių diapazono N.
4. Nustatyta ir pateikta testavimo metodika.
5. Pateikti testavimo rezultatai. Nustatyta, kad sekos, kurias generuoja PC RAND funkcija šio testo požiūriu yra atsitiktinės, o sekos, generuojamos adityviojo generatoriaus, ir šifro MISTY1 generuojamos sekos šio testo požiūriu yra neatsitiktinės.

Literatūra

1. **Rueppel R.A.** Security Models and Notions for Stream Ciphers Cryptography and Coding 11, C. Mitchell, ed. Oxford: Clarendon Press, 1992, p. 213 – 230.
2. **Фомичев В.М.** Дискретная математика и криптология.– Москва:Диалог-МИФИ,2003.–400 с.
3. **NIST special Publication** A statistical test Suite for Validation of Random and pseudorandom Number Generators.
4. **Marsaglia G**. DIEHARD Statistical Tests. http://stat.fsu.edu/~geo/diehard.html
5. **Gustafson Helen, et. al**. Statistical test suite Crypt-SX //http://www.isrc.qut.edu.au/cryptx
6. **Menezes A., P. van Oorschot and S. Vinstone.** Handbook of Applied Cryptography.– CRC Press, 1996.– P. 175 – 184.
7. **Иванов М., Чугунков А.** Теория, применение и оценка качества генераторов псевдослучайных последовательностей. – Москва: Кудиц – образ, 2003 – 159p
8. **Knuth D.E.** The Art of Computer Programming, Vol. 2. Addison-Wesley, Reading, Mass., third edition, 1998.– P. 391–401.

Pateikta spaudai 2005 06

P. Nefas Stream Cipher Testing using the Test of Analysis of Number of Euclidian Algorithm Rounds // Electronics and Electrical Engineering. – Kaunas Technologija, 2006. – No. 2(66). – P. 92–95.

The new statistical test of gamma randomness of stream cipher is presented in this article. It is based on the round number analysis of the Euclidian algorithm. This work revealed that round number probabilistic distribution is normal distribution. Dependence of distribution average value and of variance on the range N of values analyzed was established. Methods of sequence testing as well as test results are presented. It was established that sequences, which are generated by the PC RAND function from the standpoint of this test are random, and sequences generated by the additive generator and by MISTYl cipher are from the standpoint of this test not rand. Ill. 4,. bibl. 8 (in Lithuanian; summaries in English, Russian and Lithuanian).

П. Нефас Тестирование поточных шифров посредством анализа числа раундов евклидового алгоритма . // Электроника и электротехника. – Каунас: Технология, 2006. - № 2(66). – С. 92–95.

Представлен новый, статистический тест. Этот тест основан на анализе статистики числа раундов евклидового алгоритма. Установлено, что вероятностное распределение числа раундов является нормальным. Установлена связь среднего значения и дисперсии этого распределения с диапазоном анализируемых величин. Представлена методика тестирования последовательностей. Представлены результаты тестирования. Установлено, что последовательности, которые генерирует функция RANDOM, являются случайными, а последовательности, которые генерируют аддитивный генератор и шифр MISTY1 с точки зрения данного теста, – неслучайными. Ил. 4, библ. 8 (на литовском языке; рефераты на английском, русском и литовском яз.).

P. Nefas Srautinių šifrų testavimas Euklido algoritmo raundų skaičiaus analizės testu // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2006. – Nr. 2(66). – P. 92–95.

Pateiktas naujas statistinis testas. Šio testo pagrindas – tai Euklido algoritmo raundų skaičiaus statistikos analizė. Įrodyta, kad raundų skaičiaus tikimybinis skirstinys yra normalinis. Nustatyta šio skirstinio vidurkio ir dispersijos priklausomybė nuo analizuojamų dydžių diapazono N. Pateikta sekų testavimo metodika. Pateikti testavimo rezultatai. Nustatyta, kad sekos, kurias generuoja PC RAND funkcija, šio testo požiūriu yra atsitiktinės, o sekos, generuojamos adityviuoju generatoriumi, ir šifro MISTY1 generuojamos sekos šio testo požiūriu yra neatsitiktinės Il. 4, bibl. 8 (lietuvių kalba; santraukos anglų, rusų ir lietuvių k.).