kddcup99.html. – Date of access: 20.04.2009.

4. Srilatha, C. et al. Feature deduction and ensemble design of intrusion detection systems / C. Srilatha, A. Ajith, Th. Johnson. – Computers & Security. – 2005. – №24. – P. 295–307.

5. Eskin, E. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data / E. Eskin // Data Mining for Security Applications; Eds.: D. Barbar, S. Jajodia. – Boston, Kluwer Academic Publishers, 2002.

6. Jahanbani, A. and Karimi, H. A new Approach for Detecting Intrusions Based on the PCA Neural Networks / A.Jahanbani, H.Karimi. – Journal of Basic and Applied Scientific Research. – 2012. – V. 2 (1). – P. 672–679.

7. Anyanwu, L.O. Scalable Intrusion Detection with Recurrent Neural Networks / L.O. Anyanwu, L. Keengwe, G.A. Arome. – International Journal of Multimedia and Ubiquitos Engineering. – 2011. – V. 6, №1.

8. Lee, W. A framework for constructing features and models for intrusion detection systems / W. Lee, S.J. Stolfo // ACM Trans. on Inform. and System Security. – 2000. – №3(4). – P. 227–261.

9. Frank, J. Artificial intelligence and intrusion detection: Current and future directions / J. Frank // The 17th National Computer Security Conference: proceedings, Baltimore, MD, 1999 / National Institute of Standards and Technology [Electronic resourse] – 1999. – Mode of access: http://svn.assembla.com/ svn/odinIDS/Egio/temp/ frank94artificialCiteSeer.pdf. – Date of access: 02.02.2010.

10. Pfahringer, B. Winning the KDD99 Classification Cup: Bagged Boosting / B. Pfahringer // SIGKDD Explorations. – 2000. – V. 1, №2. – P. 65–66.

11. Wang, G. A new approach to intrusion detection using Artificial Neural Networks and fuzzy clastering / G. Wang, J. Hao, J. Ma, L. Huang // Expert Systems with Applications. – 2010. – №2. – P. 6225–6232.

12. Жульков, Е.В. Построение модульных нейронных сетей для обнаружения классов сетевых атак: дис. … канд. техн. наук: 05.13.19 / Е.В. Жульков. – СПб., 2007. – 155 л.

13. Sabhnani, M. Application of Machine Learning Algorithms to KDD Intrusion detection dataset within Misuse detection context / M.Sabhnani, G. Serpen // The international conference on Machine Learning: Models, technologies and Applications: proceedings, 2003. – 2003. – P. 209–215.

14. Saravanakumar, S. Development and Implementation of Artificial Neural Networks for Intrusion Detection in Computer Network / S. Saravanakumar, Umamahchwari, D. Jayalakshmi, R. Sugumar // International Journal of Computer Science and Network Security. – 2010. – V. 10, №7. – P. 271–275.

15. Ghosh, A.K. Learning program behavior profiles for intrusion detection / A.K. Ghosh, A. Schwartzbard, M. Schatz // 1st USENIX Workshop on Intrusion Detection and Network Monitoring: proceedings, Santa Clara, CA, 1999, 9–12 April / USENIX. – USENIX, 1999. – P. 51–62.

16. Ali, A. Intelligent Adaptive Intrusion Detection Systems Using Neural Networks (Comparative study) / A. Ali, A. Saleh, T. Badawy // International Journal of Video & Image Processing and Network Security. – 2010. – V. 10, №1. – P. 1–12.

17. Kayacik, H.G. A Hierarchical SOM-Based Intrusion Detection System / H.G. Kayacik, A.N. Zincir-Heywood, M. Heywood // Engineering Applications of Artificial Intelligence. – 2006. – №9. – P. 439–451.

18. Novosad, T. Fast Intrusion Detection System based on Flexible Neural Tree / T. Novosad, J. Platos, V. Snasel, A. Ajith // Sixth International Conference on Information Assurance and Security (IAS): proceedings, USA. – 2010. – P. 142–147.

19. Muna, M.J. Design Network Intrusion Detection System usin hybrid Fuzzy-Neural Network / M.J. Muna, M. Mehrotra // International Journal of Computer Science and Security. – 2010. – V. 4 (3). – P. 258–294.

20. Liu, G. A hierarchical intrusion detection model based on the PCA neural networks / G. Liu, Z. Yi, S. Yang // Neurocomputing. – 2007. – P. 1561–1568.

Abstract

The approach to network intrusion detection and recognition with use of recirculation neural networks is presented. The technologies of anomaly detection and misuse detection are utilized and combined into one technology. Such technology shows high degree of computer system protection. Experimental results prove efficiency of technology.

Поступила в редакцию 18.05.2013 г.

О ВЫЯВЛЕНИИ ЗАВИСИМОСТИ БОЛЬШОЙ ГЛУБИНЫ В ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ НА ОСНОВЕ ЦЕПИ МАРКОВА УСЛОВНОГО ПОРЯДКА

УДК 519.2: 003.26

М.В. Мальцев,
НИИ ППМИ БГУ, г. Минск

Аннотация

В статье рассматривается цепь Маркова условного порядка, которая относится к классу так называемых «малопараметрических» моделей дискретных временных рядов. Данная модель использовалась для статистического анализа выходных последовательностей криптографических генераторов. Представлены результаты компьютерных экспериментов.

Введение

Методы теории вероятностей и математической статистики широко применяются в задачах защиты информации

для тестирования криптографических генераторов [1]. Генератор, используемый в реальных приложениях, должен порождать выходную последовательность, неотличимую от равномерно распределенной случайной последовательности (РРСП) [1], часто называемой также «чисто случайной» последовательностью, элементы которой независимы в совокупности и имеют равномерное распределение вероятностей. Известным примером отклонения от РРСП является используемый в протоколах SSL и WEP генератор RC4 [2], в котором несколько «смещено» распределение вероятностей второго байта: вероятность получить нуль в этом байте равна 1/128, а не 1/256. Для обнаружения отклонения от модели «чистой случайности» применяется статистическое тестирование, задача которого – установить, обладает ли выходная последовательность исследуемого генератора определенными свойствами РРСП. Статистические тесты объединяются в так называемые батареи тестов, примерами которых являются батарея NIST [3], разработанная Лабораторией Информационных технологий Национального Института стандартов и технологий США, батарея DIEHARD [4] разработанная Джорджем Марсальей, батарея Дональда Кнута [5].

Как уже отмечалось, элементы РРСП должны быть независимыми, однако вышеприведенные батареи тестов не позволяют определять зависимости большой глубины, поэтому для их выявления требуются дополнительные исследования. Математической моделью, в которой будущее состояние зависит от s прошлых, является цепь Маркова s-го порядка. Но число независимых параметров D полносвязной цепи Маркова с N состояниями возрастает экспоненциально с увеличением порядка: $D = (N–1)N^s$, поэтому использовать ее напрямую для определения зависимостей большой глубины в выходных последовательностях криптографических генераторов практически невозможно. Для решения этой проблемы разрабатываются так называемые «малопараметрические» модели цепи Маркова s-го порядка, которые представляют собой частные случаи полносвязной цепи Маркова, матрица вероятностей одношаговых переходов которых может быть описана значительно меньшим числом параметров, чем D. Примерами таких моделей являются цепь Маркова с частичными связями [6], модель Рафтери [7]. В настоящей статье рассматривается «малопараметрическая» модель, предложенная в [8, 9], – цепь Маркова условного порядка.

Цепь Маркова условного порядка

Приведем математическое описание цепи Маркова условного порядка. Примем обозначения: N – множество натуральных чисел; $A = \{0, 1,..., N – 1\}$ – пространство состояний мощности $N \in \mathrm{N}$, $2 \le N < \infty$; $J_n^m = (j_n, j_{n+1}, ..., j_m) \in A^{m-n+1}$, $m, n \in N$, $m \ge n$, – мультииндекс; $< J_n^m > = \sum_{k=n}^{m} N^{k-n} j_k \in \{0, 1, ..., N^{m-n+1} – 1\}$ – числовое представление мультииндекса J_n^m; $F_k^l G_n^m$ – конкатенация мультииндексов F_k^l и G_n^m; $\{x_t \in A : t \in \mathrm{N}\}$ – однородная цепь Маркова s-го порядка ($2 \le s < \infty$) с $(s + 1)$-мерной матрицей вероятностей одношаговых переходов $\mathrm{P} = (p_{J_1^{s+1}})$, $p_{J_1^{s+1}} = \mathrm{P}\{x_{t+s} = j_{s+1} | x_{t+s-1} = j_s, \ldots, x_t = j_1\}$, $J_1^{s+1} \in A^{s+1}$, $t \in \mathrm{N}$; $L \in \{1, \ldots, s–1\}$, $K = N^L – 1$ – натуральные числа; $Q^{(1)}, ..., Q^{(M)}$ – семейство M ($1 \le M \le K+1$) различных квадратных стохастических матриц порядка N: $Q^{(m)} = (q_{i,j}^{(m)})$, $0 \le q_{i,j}^{(m)} \le 1$, $\sum_{i \in A} q_{i,j}^{(m)} = 1$, $i, j \in A$, $1 \le m \le M$; $\nu_{l,y}(J_2^l) = \sum_{t=1}^{n-s} I\{x_{t+s-l-y+1} = j_1, X_{t+s-l+2}^{t+s} = J_2^l\}$, $l \ge 2$, $y \ge 0$, $l + y \le s + 1$, – частота состояния $J_1^l \in \mathrm{A}^l$ с пропуском в y символов между j_1 и J_2^l.

Цепь Маркова s-го порядка ($2 \le s < \infty$) $\{x_t \in A : t \in N\}$ называется цепью Маркова условного порядка [8, 9], если ее вероятности одношаговых переходов имеют вид:

$$p_{J_1^{s+1}} = \sum_{k=0}^{K} I\{< J_{s-L+1}^s >= k\} q_{j_{b_k}, j_{s+1}}^{(m_k)}, \quad J_1^{s+1} \in A^{s+1} \qquad (1)$$

где $1 \le m_k \le M \le K + 1$, $1 \le b_k \le s – L$, $0 \le k \le K$, $\min_{0 \le k \le K} b_k = 1$, при этом в последовательности $m_0, \ldots, m_K$ встречаются все элементы множества $\{1, 2, \ldots, M\}$. Последовательность L элементов J_{sL-+1}^s, определяющая условие в формуле (1), называется базовым фрагментом памяти (БФП), L – длина БФП; величина $s_k = s – b_k + 1$ называется условным порядком. Таким образом, состояние x_t процесса в момент времени t зависит не от всех s предыдущих состояний, как это было бы для полносвязной цепи Маркова порядка s, а от $L + 1$ состояний (j_{b_k}, J_{s-L+1}^s). Отметим, что при $L = s – 1$, $s_0 = \ldots = s_K = s$, получаем полносвязную цепь Маркова порядка s.

Если последовательность, порождаемая криптографическим генератором, представляет собой эргодическую цепь Маркова, то с течением времени текущее распределение вероятностей стремится к стационарному распределению вероятностей. Поэтому для исследования надежности таких генераторов важно определить условия, при которых стационарное распределение вероятностей является равномерным. Следующая теорема устанавливает такие условия для цепи Маркова условного порядка.

Теорема. Если цепь Маркова условного порядка (1) – эргодическая, то ее стационарное распределение вероятностей является равномерным тогда и только тогда, когда:

$$\begin{cases} Q^{(m_k)} = N^{-1} \cdot 1_{N \times N}, \text{ если } s_k \in \{L+1, \ldots, s-1\}, \\ Q^{(m_k)} - \textit{бистохастическая матрица, т.е.} \\ \\ \textit{т.е. } \sum_{i \in A} Q_{ij}^{(m_k)} = 1, \forall\, j \in A, \text{ если } s_k = s, \end{cases}$$

где $1_{N \times N}$ – (N×N)-матрица, все элементы которой равны 1, $k = 0,1, \ldots, K$.

В дальнейшем будем рассматривать случай, когда каждому значению БФП соответствует своя матрица вероятностей переходов, т.е. $m_k = k+1$, $k = 0, 1, \ldots, K$.

В [8] построены оценки максимального правдоподобия вероятностей переходов, условных порядков:

$$q_{j_0, j_{L+1}}^{(k+1)} = \begin{cases} \sum_{J_1^L \in A^L} I\{< J_1^L >= k\} \dfrac{\nu_{L+2, s_k - L - 1}(J_0^{L+1})}{\nu_{L+1, s_k - L - 1}(J_0^L)}, \text{ если } \nu_{L+1, s_k - L - 1}(J_0^L) > 0, \\ 1/N, \text{ если } \nu_{L+1, s_k - L - 1}(J_0^L) = 0, \end{cases} \qquad (2)$$

$$s_k = \arg\max_{L+1\le y\le s} \sum_{J_1^L\in A^L} I\{< J_1^L >= k\} \sum_{j_0,j_{L+1}\in A} \nu_{L+2,y-L-1}(J_0^{L+1}) \ln \hat{q}_{j_0,j_{L+1}}^{(k+1)},$$

$k = 0, \ldots, K$.

В [9] на основе асимптотических свойств оценок (2) построен статистический тест для проверки гипотез о значении матриц вероятностей одношаговых переходов $Q^{(k)}$, $k = 1, \ldots, K+1$: $H_0 = \{Q^{(1)} = Q_0^{(1)}, ..., Q^{(K+1)} = Q_0^{(K+1)}\}$, $H_1 = \overline{H}_0$, где $Q_0^{(1)}, ..., Q_0^{(K+1)}$ – некоторые фиксированные матрицы вероятностей переходов. Если $N = 2$, $A = \{0, 1\}$, все элементы матриц $Q_0^{(1)}, ..., Q_0^{(K+1)}$ равны 1/2, то тест позволяет обнаруживать отклонения от РРСП и имеет вид (порядок цепи Маркова, длина БФП и условные порядки фиксированы):

$$\text{принимается } \begin{cases} H_0(PPCП): \rho \le \Delta, \\ H_1: \rho > \Delta, \end{cases} \qquad (3)$$

где $\rho = \sum_{J_0^{L+1}\in A^{L+1}} \sum_{k=0}^{K} I\{< J_1^L >= k\} 2\nu_{L+1,s_k-L-1}(J_0^L)\left(\hat{q}_{j_0,j_{L+1}}^{(k+1)} - 1/2\right)^2$,

$\Delta = G_u^{-1}(1-\alpha)$ – квантиль уровня $1 - \alpha$ ($\alpha \in (0, 1)$) стандартного χ^2-распределения с $u = 2^{L+1}$ степенями свободы.

3 Результаты компьютерных экспериментов

Проведена серия компьютерных экспериментов на модельных и реальных данных. Во всех экспериментах уровень значимости $\alpha = 0.05$.

3.1 Модельные данные

При верной гипотезе H_0 генерировалось $U = 1000$ реализаций цепи Маркова условного порядка длительности $n = 100000$ с параметрами: $N = 2$, $A = \{0, 1\}$, $s = 16$, $L = 4$, $(b_0, \ldots, b_K) = (11, 1, 9, 3, 2, 7, 12, 5, 6, 10, 1, 4, 10, 1, 8, 9)$, все элементы матриц вероятностей одношаговых переходов равны $1/N = 1/2$. В ходе эксперимента вычислялась величина $\hat{\alpha} = \frac{1}{U}\sum_{u=1}^{U} I\{\rho_u > \Delta\}$, где ρ_u – значение статистики ρ, вычисленной по u-ой реализации, которая характеризует долю решений в пользу H_1. При вычислении ρ_u все значения параметров модели за исключением матриц вероятностей одношаговых переходов полагались известными. На рисунке 1 по горизонтальной оси откладывался номер реализации, по вертикальной – значение тестовой статистики ρ, точками отмечались величины ρ_u, сплошной линией

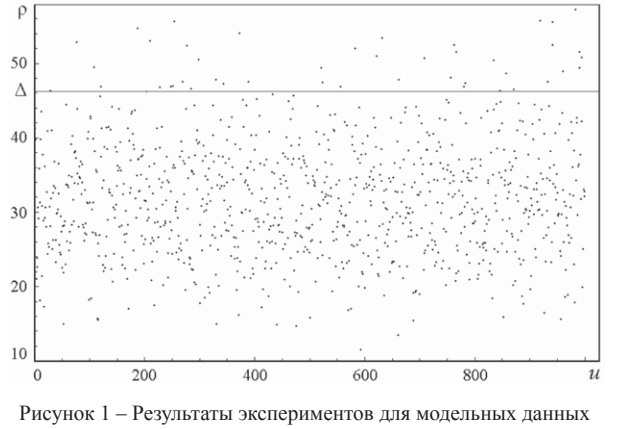

Рисунок 1 – Результаты экспериментов для модельных данных

изображен порог теста $\Delta = G_{32}^{-1}(0.95) \approx 46.2$. Значение $\hat{\alpha} = 0.045$, вычисленное в ходе эксперимента, согласуется с теоретическими результатами.

Аналогичный результат был получен в экспериментах, в которых параметры $(b_0, \ldots, b_K)$ полагались неизвестными и для вычисления тестовой статистики выбирались случайным образом из множества допустимых значений $(1 \le b_k \le s - L = 12)$.

3.2 Реальные данные

Исследовались выходные последовательности самосжимающего генератора [10] и регистра сдвига с переменной обратной связью [11]. Генерировалось по 100 реализаций каждого генератора длительности $n = 100\,000$ со случайно выбранным начальным заполнением регистров. При фиксированной длине БФП L изменялся порядок цепи Маркова от $L+1$ до 100. При этом первая половина последовательности длины $n_{0.5} = 50\,000$ использовалась для построения оценок условных порядков по формуле (2), вторая половина – для вычисления тестовой статистики ρ. Для каждой пары (s, L) вычислялась величина $\hat{\alpha}$, которая отмечалась на рисунках точкой (по горизонтальной оси откладывался порядок s). Сплошная линия на рисунках – уровень значимости α. Характеристический многочлен самосжимающего генератора $f(x) = x^{36} + x^{11} + 1$ (период выходной последовательности $T_1 = 2^{18}$), характеристические многочлены регистра сдвига с переменной обратной связью $f_1(x) = x^{23} + x^5 + 1$, $f_2(x) = x^{23} + x^9 + 1$, управляющий многочлен $g(x) = x^{18} + x^{16} + x^{11} + + x^4 + 1$ (период выходной последовательности $T_2 \ge 2^{18} - 1$). Все вышеперечисленные многочлены являются примитивными. Результаты представлены на рисунке 2 и рисунке 3.

Заключение

Таким образом, в статье представлены результаты обнаружения отклонения от модели «чисто случайной» последовательности выходных последовательностей самосжимающего генератора и регистра сдвига с переменной обратной связью. Для обнаружения отклонений использовалась цепь Маркова условного порядка – математическая модель, учитывающая зависимости высоких порядков при относительно небольшом числе независимых параметров. Результаты компьютерных экспериментов демонстрируют применимость теста (3) при определенных значениях s и L. Отметим нелинейный характер зависимости $\hat{\alpha}$ как от порядка s, так и от длины БФП L. К примеру, для самосжимающего генератора при $s = 3$ число решений в пользу H_1 больше, чем при $s = 4$, а при $L = 6$ число решений в пользу H_1 наибольшее при $L < 50$.

Литература:

1. Математические и компьютерные основы криптологии / Ю.С. Харин [и др.]. – Минск: Новое знание, 2003.

2. Sepehrdad, P. Discovery and Exploitation of New Biases in RC4 / P. Sepehrdad, S. Vaudenay, M. Vuagnoux // Lecture Notes in Computer Science. 2011. – V. 6544. – P. 74–91.

3. A statistiical test suite for random and pseudorandom number generators for cryptographic applicatiions: NIST Special Publication 800-22 Rev. 1a. National Institute of Standards and Technology, 2010. – 131 p.

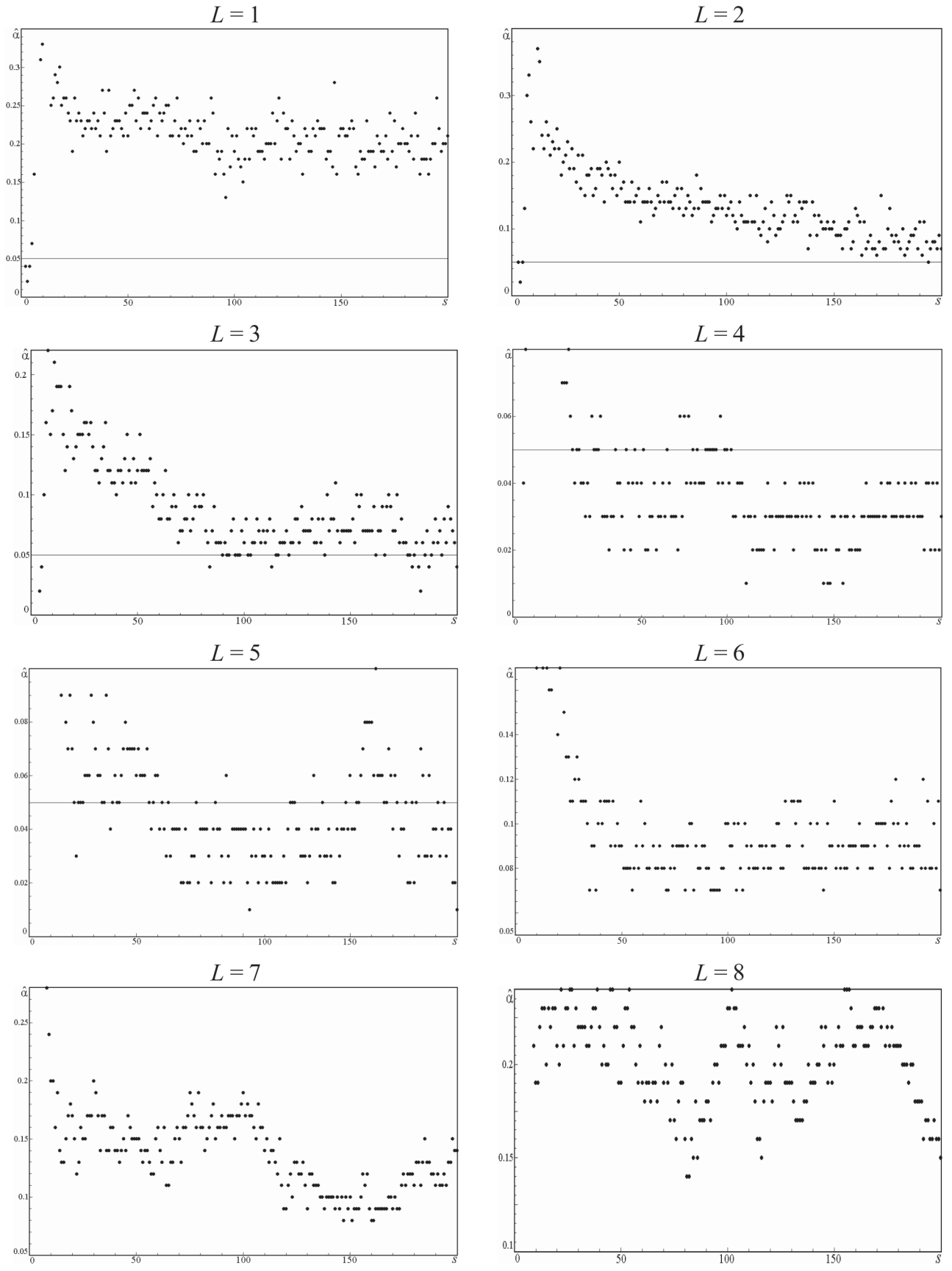

Рисунок 2 – Результаты экспериментов для самосжимающего генератора

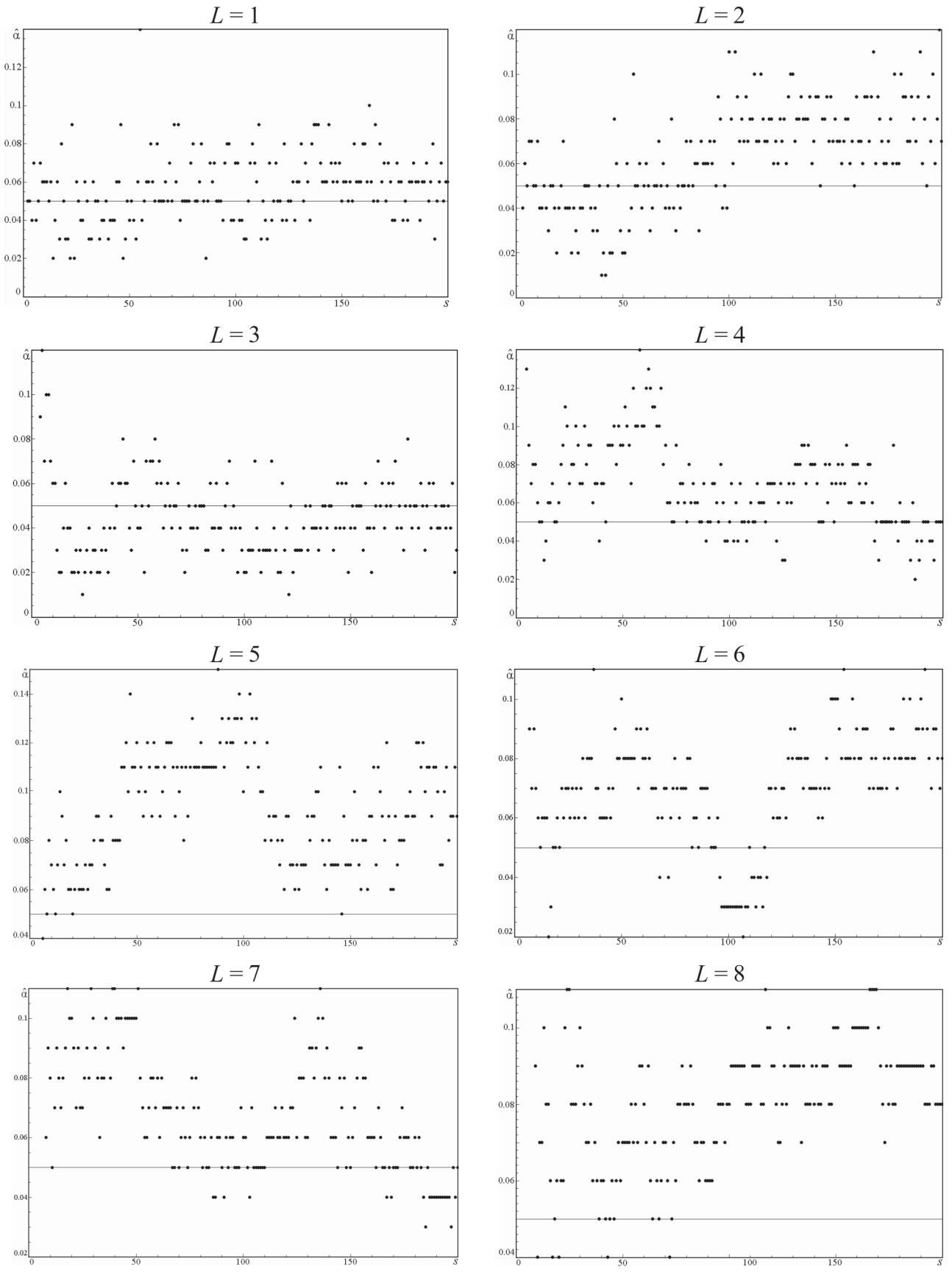

Рисунок 3 – Результаты экспериментов для регистра сдвига с переменной обратной связью

4. Marsaglia, G. The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness / G. Marsaglia [Electronic resource]. – Florida State University, 1995.

5. Искусство программирования: в 3 т. Т. 1: Полученные методы / Д.Э. Кнут. – М.: Вильямс, 2007.

6. Харин, Ю.С. Цепь Маркова с частичными связями ЦМ(s, r) и статистические выводы о ее параметрах / Ю.С. арин, А.И. Петлицкий // Дискретная математика. – 2007. – Т. 19, № 2. – С. 109–130.

7. Raftery, A.E. A model for high-order Markov chains / A.E. Raftery // J. Royal Statistical Society. – 1985. – V. B-47, № 3. – P. 528–539.

8. Харин, Ю.С. Алгоритмы статистического анализа цепей Маркова с условной глубиной памяти / Ю.С. Харин, М.В. Мальцев // Информатика. – 2011. – №1. – С. 34–43.

9. Харин, Ю.С. Статистическая проверка гипотез о параметрах цепи Маркова условного порядка / Ю.С. Харин, М.В. Мальцев // Весці НАН Беларуси, Серыя. фіз.-мат. навук. – 2012. – №3. – С. 5–12.

10. Meier, W. The self-shrinking generator / W. Meier, O Staffelbach // Advances in Cryptology – EUROCRYPT 94. 1995. Springer-Verlag LNSC 950. – P. 205–214.

11. Основы криптографии / А.П. Алферов [и др.]. М.: Гелиос АРВ, 2001.

Abstract

The paper deals with Markov chain of conditional order, which is a parsimonious model of discrete time series. The model was used for statistical analysis of cryptographic generators output sequences.

The results of computer experiments are presented.

Поступила в редакцию 18.05.2013 г.

ПОВТОРЯЕМОСТЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СОСТОЯНИЙ ПОЛНОЦИКЛОВЫХ АВТОНОМНЫХ АВТОМАТОВ

УДК 519.711.2,519.713.2.

А.Н. Гайдук,
НИИ ППМИ БГУ, г. Минск

Аннотация

В настоящей работе рассматривается полноцикловой генератор псевдослучайной последовательности. Попадание генератора псевдослучайной последовательности в ранее использованное состояние означает повторную выработку выходной последовательности, что является крайне нежелательным явлением. Анализ этого события для случая когда длины вырабатываемой псевдослучайной последовательности полноциклового генератора являются случайными величинами приведен в работе [1], случай когда длины вырабатываемой псевдослучайной последовательности являются произвольными фиксированными величинами рассмотрен в работе [2]. В настоящей работе для случая отрезков равной длины вырабатываемой псевдослучайной последовательности получена более простая модель, которая согласуется с результатами работы [1]-[2].

Введение

Будем рассматривать в качестве генератора псевдослучайной последовательности автономный автомат $U = (S, Z, \varphi, \psi)$, где $S = \{s_1, s_2, ..., s_N\}$ – конечное множество состояний, $Z = \{0,1\}$ – конечное множество выходных символов, $\varphi : S \to S$ – функция переходов, $\psi : S \to Z$ – функция выходов. Канонические уравнения функционирования автономного автомата U имеют вид:

$$s(t+1) = \varphi(s(t))$$
$$z(t) = \psi(s(t)), \quad t \geq 0,$$

$s(0)$ – начальное состояние.

Под состоянием генератора понимаем состояние описывающего его автономного автомата.

Постановка задачи

Пусть в процессе вычислений используется n отрезков последовательности псевдослучайных чисел, полученных с помощью полноциклового генератора. Будем отождествлять начальное состояние $s_i = (s_{i,1}, ..., s_{i,m})$ i-го отрезка последовательности псевдослучайных чисел с числом $X_i = \sum_{k=1}^{m} s_{i,k} 2^{m-k}$. Через L_i обозначим длину i-го отрезка.

Будем считать, что случайные величины $X_1, ..., X_n$ независимы в совокупности и при всех i:

$$P(X_i = k) = \frac{1}{2^m}, \; k = 0, ..., 2^m - 1.$$

Требуется оценить вероятность перекрытия n отрезков длин $L_i = \overline{1,n}$.

Повторяемость последовательностей состояний случайной длины (модель М1)

В работе [1] исследовался случай, когда длины отрезков $L_1, ..., L_n$ являются случайными величинами:

$$P(l_i = k) = p_k, \; k = 1, ..., N,$$

причем

$$P(1 \leq L_i \leq \frac{N}{2}) = 1.$$

В [1] доказана теорема, которая позволяет построить оценки снизу и сверху для вероятности неперекрытия отрезков $L_1, ..., L_n$.

Повторяемость последовательностей состояний произвольной длины (модель М2)

В работе [2] получены результаты для отрезков произвольных длины. Пусть отрезки имеют следующие длины: $L_1, ..., L_n$, где, вообще говоря, $L_i \neq L_j$ при $i \neq j$. Перекрытие отсутствует, если все пары отрезков не имеют перекрытий.

Теорема 1. Вероятность не перекрытия n отрезков длин $L_1, ..., L_n$, $\sum_{j=1}^{n} L_j = L$, при цикле длины N автомата равна

$$\frac{(n-1)! C_{N-L+n-1}^{n-1}}{N^{n-1}}. \qquad (1)$$