ЗАСЕДАНИЕ № 2
КРИПТОГРАФИЯ ДЛЯ ГРАЖДАН И ГОСУДАРСТВА

УДК 519.671

КРИПТОЛОГИЯ И СТОХАСТИКА

Ю.С. ХАРИН
Научно-исследовательский институт прикладных проблем математики и информатики, Белорусский государственный университет, г. Минск, Республика Беларусь

Введение. В современных системах комплексной защиты информации важнейшим способом защиты информации является криптографический способ. Он позволяет с гарантированной стойкостью решить следующие четыре главные практические задачи: 1) конфиденциальность; 2) аутентификация источника сообщения; 3) проверка целостности; 4) невозможность отречения от авторства. Криптографический способ базируется на новой науке Криптологии [2], объединяющей Криптографию и Криптоанализ. Криптография – это отрасль математики, в которой разрабатываются модели, методы, алгоритмы и программные средства **математического преобразования информации** в целях сокрытия ее содержания, предотвращения видоизменения или несанкционированного использования; при этом преобразованное сообщение представляет собой хаотическую, чисто случайную последовательность символов. Криптоанализ – раздел математики, в котором разрабатываются модели, методы, алгоритмы и программные средства анализа криптосистемы или ее входных и выходных сигналов с целью извлечения конфиденциальных параметров, включая **открытый текст**. Стохастика – раздел математики, изучающий модели и методы исследования систем и процессов с учетом случайных элементов. Из этих определений видно, что Криптология и Стохастика тесно связаны: Стохастика представляет математический инструментарий для решения задач Криптологии, Криптология стимулирует Стохастику к разработке новых моделей для исследования сложных последовательностей, циркулирующих в криптосистемах. Следует заметить, что впервые на эту связь обратили внимание академик В.А. Котельников и К. Шеннон, разработав математические основы криптографии с помощью стохастических моделей исходного текста, ключа и зашифрованного сообщения. Настоящий доклад посвящен представлению и использованию новых стохастических моделей в криптологии.

1. Проблема «чистой случайности». Многие задачи криптологии (статистическое тестирование криптографических генераторов, статистический криптоанализ, разностный криптоанализ, линейный криптоанализ, криптоатаки по побочным каналам, стеганография) сводятся к задачам различения некоторой зарегистрированной последовательности символов $x_1, x_2, \ldots$ от «чисто случайной» последовательности и оценки величины этого различия.

Математической моделью последовательностей, порождаемых генераторами, а также последовательностей, возникающих в различных узлах СКЗИ, является дискретный временной ряд (ДВР). Дискретный временной ряд (ДВР) – это случайный процесс $x_t \in A$ на вероятностном пространстве $(\Omega, F, \mathbf{P})$ с дискретным временем $t \in \mathbb{T} = \{1, 2, \ldots\}$ и дискретным множеством состояний A мощности $|A| = N$, $2 \le N < +\infty$. Без потери общности полагаем пространство состояний (алфавит) $A = \{0, 1, \ldots, N-1\}$.

В криптологии в связи с Шенноновской теорией совершенных криптосистем большое внимание уделяется так называемому «чисто случайному» ДВР – равномерно распределенной случайной последовательности (РРСП).

РРСП – это последовательность дискретных случайных величин $x_1, x_2, \ldots \in A = \{0, 1, \ldots, N-1\}$, обладающая двумя свойствами [2]:

C_1) для любого числа $n \in \mathbb{T}$ и произвольных индексов $1 < t_1 < \ldots < t_n$ случайные элементы $x_{t_1}, \ldots, x_{t_n}$ независимы в совокупности;

C_2) для любого $t \in \square$ случайная величина x_t имеет равномерное на A распределение вероятностей: $\mathbf{P}\{x_t = i\} = N^{-1}$, $i \in A$.

В настоящее время известно более сотни методов и алгоритмов генерации последовательностей, по своим свойствам приближающихся к РРСП. Еще больше разработано методов статистического тестирования криптографических генераторов, заключающихся в проверке простой гипотезы $H_0 = \{\{x_t\}$ есть РРСП$\}$ против сложной альтернативы $H_1 = \bar{H}_0 = \{$нарушены свойства $C_1,\ C_2\}$.

Проведенный обзор существующих статистических тестов показывает:

1) многие из существующих тестов не ориентированы на проверку главного свойства C_1, а лишь частных случаев свойств C_1, C_2, т. е. частных случаев альтернативы $H_1 = \bar{H}_0$;

2) многие из известных тестов построены «эвристически» и не фиксируют семейство альтернатив H_1;

3) многие тесты не имеют оценок мощности;

4) при включении нескольких тестов в батарею не удается оптимизировать «составной» тест.

В связи с этим актуальна рассматриваемая далее проблема разработки адекватных стохастических моделей для описания отклонений H_1 от модели РРСП, построения статистических тестов для обнаружения и оценивания таких отклонений.

2. Модели ДВР на основе уклонений от s-мерной равномерности и их энтропийный анализ. Определим вложенное в H_1 семейство «альтернатив s-мерной неравномерности»:

$$H_{1(s)} = \{\{x_1, x_2, \ldots\} = \{X_1, X_2, \ldots\}\} \subset H_1,$$

где $X_1, X_2, \ldots \in A^s$ – независимые одинаково распределенные s-фрагменты (слова) над алфавитом A с некоторым s-мерным дискретным распределением вероятностей $\mathbf{P}_{i_1,\ldots,i_s} = \mathbf{P}\{x_1 = i_1, \ldots, x_s = i_s\}$, $i_1, \ldots, i_s \in A$, отличным от равномерного:

$$\Delta_s = \sum_{i_1,\ldots,i_s \in A} \left|\mathbf{P}_{i_1,\ldots,i_s} - N^{-s}\right| > 0, \quad \sum_{i_1,\ldots,i_s \in A} \mathbf{P}_{i_1,\ldots,i_s} \equiv 1.$$

Это семейство моделей ДВР обладает двумя свойствами: 1) при $s \to \infty$ семейство этих альтернатив имеет в пределе альтернативу $H_1 = \bar{H}_0$ общего вида; 2) чем меньше Δ_s, тем ближе альтернатива $H_{1(s)}$ к нулевой гипотезе H_0.

Обозначим: $\{x_1, x_2, \ldots, x_T\} = \{X_1, X_2, \ldots, X_M\}$ – наблюдаемая реализация выходной последовательности длиной $T = M \cdot s$, разбитая на M непересекающихся фрагментов длины s, $I\{B\}$ – индикатор события B,

$$\hat{\mathbf{P}}_{i_1,\ldots,i_s} = \frac{1}{M}\sum_{m=1}^{M} I\{X_m = (i_1, \ldots, i_s)\},\ i_1, \ldots, i_s \in A, - \qquad (1)$$

статистическая оценка для $\mathbf{P}_{i_1,\ldots,i_s}$.

Тест обобщенного отношения правдоподобия для проверки H_0, $H_{1(s)}$ на основе статистик (1) имеет вид:

$$\text{принимается} \begin{cases} H_0, \text{ если } \hat{H}_s - s\ln N > -\dfrac{1}{2M} G^{-1}_{N^s-1}(1-\varepsilon), \\ \qquad H_{1(s)} \text{ в противном случае,} \end{cases} \qquad (2)$$

$$\hat{H}_s = \sum_{i_1,\ldots,i_s \in A} \hat{\mathbf{P}}_{i_1,\ldots,i_s} \ln \hat{\mathbf{P}}_{i_1,\ldots,i_s} -$$

– статистическая оценка s-мерной энтропии Шеннона, $G_K^{-1}(\cdot)$ – обратная функция распределения хи-квадрат с K степенями свободы, $\varepsilon \in (0,1)$ – заданный уровень значимости теста.

Тест (1), (2) мы предлагаем использовать для визуализации процесса принятия решений в виде так называемого «энтропийного профиля (портрета)» – графика зависимости нормированного уклонения оценки s-мерной энтропии от ее математического ожидания при H_0 (см. рис. 1, 2, где штриховые линии обозначают границы области решений):

$$\alpha(s) = 2M\left(\hat{H}_s - s\ln N\right) / G_{N^s-1}^{-1}(1-\varepsilon), \quad s \in \{s_{min}, s_{min}+1, \ldots, s_{max}\}. \tag{3}$$

Отметим еще, что вместо энтропии Шеннона в (1) – (3) могут использоваться энтропийные функционалы Реньи и Тсаллиса [4].

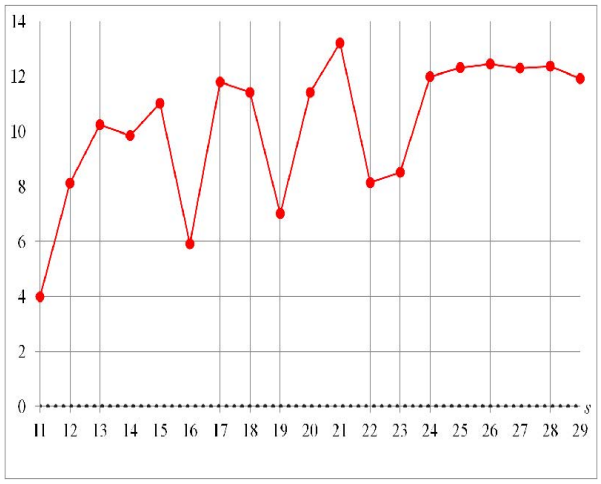

Рис. 1. Энтропийный профиль $\ln|\alpha(s)|$ нелинейного регистра сдвига порядка 24 ($N = 2, \varepsilon = 0.05, T = 2^{32}/s$)

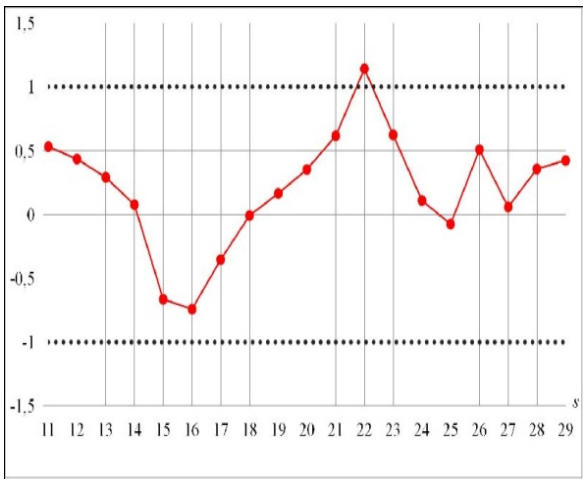

Рис. 1. Энтропийный профиль $\alpha(s)$ генератора BelT (СТБ 34.101.27-2011, $N = 2, \varepsilon = 0.05, T = 2^{29}/s$)

3. Универсальная модель ДВР на основе цепей Маркова высокого порядка. Учитывая, что универсальной моделью стохастической зависимости элементов выходной последовательности $\{x_t\}$ криптографического генератора является цепь Маркова достаточно высокого порядка s, определим вложенное в $H_1 = \bar{H}_0$ семейство альтернатив марковской зависимости: $H_1^{(s)} = \{\{x_t\}$ однородная цепь Маркова порядка s с матрицей переходов $\mathbf{P}\}$, где $\mathbf{P} = \left(p_{i_1,\ldots,i_{s+1}}\right)$, $i_1,\ldots,i_{s+1} \in A - (s+1)$-мерная матрица,

$$p_{i_1,\ldots,i_{s+1}} = \mathbf{P}\{x_{t+1} = i_{s+1} \mid x_t = i_s, \ldots, x_{t-s+1} = i_1\},$$

$$\Delta_s = \sum_{i_1,\ldots,i_s \in A} \left| p_{i_1,\ldots,i_{s+1}} - N^{-1} \right| > 0. \tag{4}$$

Тест обобщенного отношения правдоподобия для проверки гипотез H_0, $H_1^{(s)}$ основан на статистической оценке $\hat{h}_s$ условной энтропии $h_s = H\{x_t \mid x_{t-1}, \ldots, x_{t-s}\}$:

$$\text{принимается} \begin{cases} H_0, \text{ если } \hat{h}_s - \ln N > -G_f^{-1}(1-\varepsilon)/\left(2(T-s)\right), \ f = N^s(N-1), \\ H_1^{(s)} \text{ в противном случае.} \end{cases} \tag{5}$$

Аналогично (3) с помощью $\hat{h}_s$ строится энтропийный профиль для $\{x_1, \ldots, x_T\}$.

К сожалению, тесты (2), (5), анализирующие стохастические зависимости глубины s в выходной последовательности $\{x_t\}$, требуют экспоненциально растущей с ростом порядка s длины анализируемой последовательности $T = O\left(N^{s+1}\right)$. Для преодоления этой трудности целе-

сообразно использовать так называемые «малопараметрические модели цепей Маркова высокого порядка» [1, 2], т. е. модели цепей Маркова s-го порядка, для которых $\left(N^s \times N\right)$-матрица вероятностей переходов зависит от «малого» числа параметров $D \ll N^s\left(N-1\right)$; $\kappa = D/\left(N^s\left(N-1\right)\right) \ll 1$ – коэффициент сжатия, равный относительному числу параметров модели.

4. Подходы к построению малопараметрических цепей Маркова высокого порядка

Подход I. Этот подход состоит в «сжатии множества значений элементов матрицы» $\mathbf{P}$.

Пусть $Q = \left(q_{j_i,\ldots,j_r,j_{r+1}}\right)$ – некоторая $(r+1)$-мерная матрица, $1 \le r < s$,

$$\sum_{j_{r+1} \in A} q_{j_i,\ldots,j_r,j_{r+1}} \equiv 1, 0 \le q_{j_i,\ldots,j_r,j_{r+1}} \le 1;$$

$B(\cdot)\colon A^s \to A^r$ – некоторая дискретная функция. С помощью $B(\cdot)$ $(s+1)$-мерная матрица $\mathbf{P}$ «сжимается» в $(r+1)$-мерную матрицу Q преобразованием:

$$p_{i_1,\ldots,i_s,i_{s+1}} = q_{B(i_1,\ldots,i_s),i_{s+1}};\ \kappa_{\mathrm{I}} = N^{r-s} \le 1. \qquad (6)$$

Примеры малопараметрических ДВР в рамках подхода I: $\mathrm{MC}(s,r)$, $\mathrm{MCCO}(s,L)$, VLMC [5].

Подход II. Этот подход заключается в использовании порождающего уравнения для условного распределения вероятностей (4) будущего состояния $x_t \in A$ при условии предыстории $X_{t-s}^{t-1} = \left(x_{t-1},\ldots,x_{t-s}\right)' \in A^s$:

$$p_{i_1,\ldots,i_s,i_{s+1}} = q_{i_{s+1}}\left(\theta\left(i_1,\ldots,i_s;a\right)\right),\ i_1,\ldots,i_{s+1} \in A, \qquad (7)$$

где $\left\{q_j(\theta)\colon j \in A\right\}$ – некоторое вероятностное распределение на A, зависящее от параметра $\theta = \left(\theta_j\right) \in \Theta \subseteq R^L$; $\theta = \theta\left(i_1,\ldots,i_s;a\right)$ – некоторая функция, известная с точностью до вектора параметров $a = \left(a_k\right) \in R^m$. Коэффициент сжатия:

$$\kappa_{\mathrm{II}} = \frac{m}{N^s\left(N-1\right)} \le 1.$$

Примеры малопараметрических ДВР в рамках подхода II: модель Джекобса – Льюиса, MTD-модель, $\mathrm{DAR}(s)$, $\mathrm{BCNAR}(s)$, $\mathrm{BiCNAR}(s)$, $\mathrm{PCNAR}(s)$.

5. Малопараметрические модели ДВР на основе подхода I и их статистический анализ

Цепь Маркова $\mathrm{MC}(s,r)$ порядка s с r частичными связями. Эта модель определяется формулой (6) с $B\left(j_1\ldots,j_s\right) = \left(j_{m_1^0},\ldots,j_{m_r^0}\right)$ [1, 2]:

$$p_{J_1^{s+1}} = p_{j_1,\ldots,j_s,j_{s+1}} = q_{j_{m_1^0},\ldots,j_{m_r^0},j_{s+1}},\ J_1^{s+1} \in A^{s+1}, \qquad (8)$$

где $J_i^k = \left(j_i, j_{i+1},\ldots,j_k\right) \in A^{k-i+1}$ – последовательность $k-i+1$ индексов $\left(k \ge i\right)$; r – число связей; $M_r^0 = \left(m_1^0,\ldots,m_r^0\right)$ – вектор с r упорядоченными целыми компонентами $1 = m_1^0 < m_2^0 < \ldots < m_r^0 \le s$, называемый шаблоном связей; $Q = \left(q_{J_1^{r+1}}\right)_{J_1^{r+1} \in A^{r+1}}$ – $(r+1)$-мерная стохастическая матрица. Если $r = s$, то получаем полносвязную цепь Маркова порядка s.

Статистическую оценку $\hat{Q}$ удобно использовать для визуализации уклонения от гипотезы H_0 (для которой $q_{i_1,\ldots,i_{r+1}} = N^{-1}$). На рис. 3, 4 представлены результаты такой визуализации для генератора со случайной обратной связью и генератора BelT (СТБ 34.101.27-2011 в режиме гаммирования) соответственно; здесь красный цвет – оценка условной вероятности

перехода в «0» $\hat{q}_{K_1^r,0}$, зеленый – в «1» $\hat{q}_{K_1^r,1}$); здесь по оси абсцисс откладывается $K_1^r = B\left(J_1^s; \hat{M}_r\right) \in A^r$.

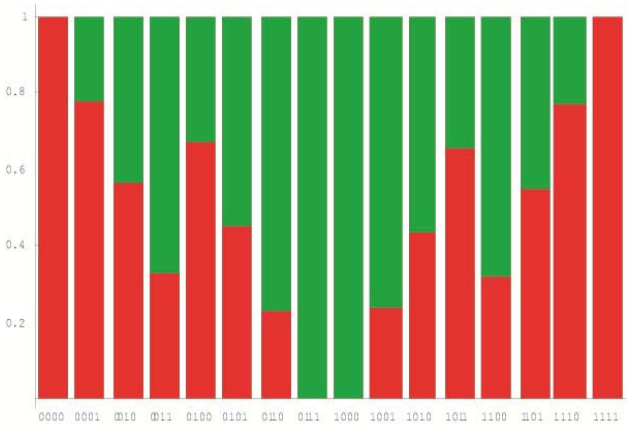

Рис. 3. Оценка $\hat{Q}$ $\left(s=64,\ r=4,\ T=10^5\right)$

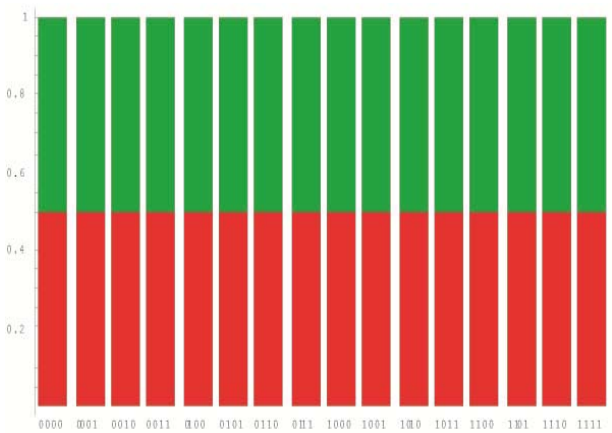

Рис. 4. Оценка $\hat{Q}$ $\left(s=32,\ r=4,\ T=8\cdot 10^6\right)$

С моделью $\mathrm{MC}(s,r)$ связана модель цепи Маркова переменного порядка [5].

6. Малопараметрические модели ДВР на основе подхода II и их статистический анализ

Модель Джекобса – Льюиса. Эта модель порождается стохастическим разностным уравнением [6]:

$$x_t = \mu_t x_{t-\eta_t} + (1-\mu_t)\xi_t, \tag{9}$$

где $t > s$, $\{\xi_t, \eta_t, \mu_t\}$ – независимые в совокупности случайные величины с вероятностными распределениями:

$$\mathbf{P}\{\mu_t = 1\} = 1 - \mathbf{P}\{\mu_t = 0\} = \rho;\ \mathbf{P}\{\xi_t = k\} = \pi_k\ ,\ k \in A\ ,\ \sum_{k\in A}\pi_k = 1\ ;$$

$$\mathbf{P}\{\eta_t = i\} = \lambda_i\ ,\ i \in \{1,2,\ldots,s\}\ ,\ \sum_{i=1}^{s}\lambda_i = 1,\ \lambda_s \neq 0\ ; \tag{10}$$

$$\mathbf{P}\{x_1 = k\} = \ldots = \mathbf{P}\{x_s = k\} = \pi_k\ ,\ k \in A\ .$$

Число параметров этой модели (9), (10) линейно (а не экспоненциально!) зависит от s, так что коэффициент сжатия

$$\kappa_{JL} = (N+s-1)/\left(N^s(N-1)\right).$$

Методы и алгоритмы статистического анализа модели Джекобса – Льюиса представлены в [2].

MTD-модель Рафтери. MTD (Mixture Transition Distribution) – модель [8] определяется следующим частным случаем уравнения (7):

$$p_{i_1,\ldots,i_s,i_{s+1}} = \sum_{j=1}^{s}\lambda_j q_{i_j,i_{s+1}}\ ,\ i_1,\ldots,i_{s+1} \in A\ ,$$

где $Q = \left(q_{i,k}\right)$ – некоторая стохастическая $(N\times N)$-матрица,

$$0 \le q_{i,k} \le 1\ ,\ \sum_{k\in A} q_{i,k} \equiv 1\ ,\ i,k \in A\ ,$$

$\lambda = (\lambda_1,\ldots,\lambda_s)'$ – некоторое дискретное распределение вероятностей, $\lambda_1 > 0$.

Обобщенная MTDg (generalized MTD)-модель определяется следующей параметризацией $(s+1)$-мерной матрицы $\mathbf{P}$:

$$p_{i_1,\ldots,i_s,i_{s+1}} = \sum_{j=1}^{s}\lambda_j q^{(j)}_{i_j,i_{s+1}}\ ,\quad i_1,\ldots,i_{s+1} \in A\ , \tag{11}$$

где $Q^{(j)} = \left(q_{i,k}^{(j)}\right)$ – некоторая стохастическая матрица для j-го лага.

Относительное число параметров MTDg-модели (11):

$$\kappa_{\text{MTDg}} = \left(s\left(N\left(N-1\right)/2+1\right)-1\right)/\left(N^s\left(N-1\right)\right).$$

Биномиальная условно нелинейная авторегрессионная модель BiCNAR(s). Эта модель порождается специальным (биномиальным) случаем порождающего уравнения (7):

$$p_{i_1,\ldots,i_s,i_{s+1}} = C_{N-1}^{i_{s+1}}\ \theta^{i_{s+1}}\left(1-\theta\right)^{N-1-i_{s+1}},\ i_{s+1} \in A = \left\{0,1,\ldots,N-1\right\},$$

$$\theta = \theta\left(\mathrm{I}_1^{\mathrm{s}}\right) = \mathrm{F}\left(a'\Psi\left(\mathrm{I}_1^{\mathrm{s}}\right)\right),\ \mathrm{I}_1^{\mathrm{s}} = \left(i_1,\ldots,i_s\right)' \in A^s,$$

где $\Psi\left(\mathrm{I}_1^{\mathrm{s}}\right) = \left(\psi_1\left(\mathrm{I}_1^{\mathrm{s}}\right),\ldots,\psi_m\left(\mathrm{I}_1^{\mathrm{s}}\right)\right)'$: $A^s \to R^m$ – вектор-столбец $m \le N^s$ линейно независимых функций, например, полиномов; $\mathrm{F}(\cdot)$: $R^1 \to [0,1]$ – некоторая функция распределения, например, логистическая, нормальная или Коши:

$$\Lambda(\zeta) = \frac{1}{1+e^{-\zeta}},\ \Phi(\zeta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\zeta} e^{-\frac{x^2}{2}}dx,\ C(\zeta) = \frac{1}{2} + \frac{\arctan(\zeta)}{\pi},\ \zeta \in R^1;$$

$a = \left(a_1,\ldots,a_m\right)'$ – вектор-столбец m неизвестных параметров модели.

Относительное число параметров модели: $\kappa = m\left(N^s\left(N-1\right)\right)^{-1} < 1$.

Методы и алгоритмы статистического анализа BiCNAR(s)-модели, ее частных случаев и обобщений представлены в [3, 7].

Заключение. В криптологии актуальна проблема построения и статистического анализа моделей дискретных временных рядов, адекватно описывающих отклонения от модели РРСП.

1. В статье представлены такие семейства моделей ДВР на основе уклонений от s-мерной равномерности и на основе цепей Маркова порядка s.

2. Для преодоления «проклятия размерности» представлены два подхода к построению малопараметрических моделей цепей Маркова высокого порядка.

3. Разработаны методы и алгоритмы статистического анализа (оценивание параметров, проверка гипотез) малопараметрических моделей, построенных на основе предложенных подходов.

4. Теоретические результаты иллюстрируются результатами компьютерных экспериментов по тестированию выходных последовательностей известных криптографических генераторов.

Список литературы

1. Харин, Ю. С. Цепи Маркова с s-частичными связями и их статистическое оценивание / Ю. С. Харин // Доклады НАН Беларуси. – 2004. Т. 48, № 1, с. 40 – 44.

2. Харин, Ю. С. Криптология / Ю. С. Харин [и др.]. – Минск: БГУ, 2014.

3. Харин, Ю. С. Биномиальные условно нелинейные авторегрессионные модели дискретных временных рядов и их вероятностные и статистические свойства / Ю. С. Харин, В. А. Волошко // Труды Института Математики НАН Беларуси. – 2019. – Т. 26, № 1. – С. 95–105.

4. Харин, Ю. С. Энтропийный анализ криптографических генераторов случайных и псевдослучайных последовательностей / Ю. С. Харин, В. Ю. Палуха // Веснік сувязі. – 2017. – Т. 146, № 1. – С. 46–49.

5. Buhlmann, P. Variable length Markov chains. / P. Buhlmann, A. J. Wyner // The Annals of Statistics. – 1999. – Vol. 27, No. 2. – P. 480–513.

6. Jacobs, P. A. Discrete time series generated by mixtures I: correlational and runs properties. / P. A. Jacobs, P. A. W. Lewis // Journal of the Royal Statistical Society. Ser. B. – 1978. – Vol. 40, No. 1. – P. 94–105.

7. Kharin, Yu. S. Statistical estimation of parameters for binary conditionally nonlinear autoregressive time series / Yu. S. Kharin, V. A. Voloshko, E. A. Medved // Mathematical Methods of Statistics. – 2019. – Vol. 26, No. 2. – P. 103–118.

8. Raftery, A. A model for high-order Markov chains / A. Raftery // Journal of the Royal Statistical Society. Ser. B. – 1985. – Vol. 47, No. 3. – P. 528–539.