УДК 004.056.5
DOI: 10.12737/24905

М.Ю. Конышев, А.В. Козачок, О.М. Голембиовская, К.Е. Петров

СНИЖЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ МАШИННЫХ ЭКСПЕРИМЕНТОВ ПРИ ВЕРИФИКАЦИИ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ

Рассмотрена проблема получения набора выборок для оценки качества криптографических алгоритмов на основе использования статистических тестов. Описаны новые свойства двоичных цепей Маркова, учитывающие зависимости вероятностей двоичных векторов различной длины. Предложены аналитические выражения, позволяющие вычислить зависимости пределов диапазонов значений вероятностей многомерных двоичных случайных величин от вероятностей двоичных случайных величин меньшей размерности. Определены причины необходимости дополнительной процедуры отбраковки при симуляции реализаций двоичных марковских процессов. Рассмотрен метод направленного перебора значений вероятностей рядов распределений марковских двоичных последовательностей, позволяющий генерировать эргодические двоичные случайные последовательности, что позволяет полностью отказаться от процедуры отбраковки. Представлен реализующий указанный метод алгоритм, обладающий пониженной вычислительной сложностью по сравнению с известными алгоритмами организации вычислительного эксперимента по исследованию статистических свойств двоичных случайных последовательностей.

Ключевые слова: статистические тесты, цепи Маркова, двоичные последовательности, моделирование, вероятности двоичных векторов, дискретная случайная величина, вычислительная сложность, криптографические алгоритмы.

M.Yu. Konyshev, A.V. Kozachok, O.M. Golembiovskaya, K.E. Petrov

COMPUTATION COMPLEXITY DECREASE IN MACHINE EXPERIMENTS AT VERIFICATION OF CRYPTOGRAPHIC ALGORITHMS

The problem of obtaining a set of samples for the assessment of cryptographic algorithms quality on the basis of statistical tests use is considered. New properties of Markov binary chains taking into account dependences of probabilities of binary vectors with different length are described. The analytical expressions allowing the computation of dependences of range limits in values of probabilities of multidimensional binary random values upon probabilities of binary random values with smaller dimension are offered. The reasons for the necessity of an additional procedure of rejection at the simulation of the realization of Markov binary processes are defined. A method for the directed search of probability values of sets in the distribution of Markov binary sequences allowing the generation of ergodic binary random sequences that allows refusing completely the procedure of rejection is considered. An algorithm realizing a mentioned method possessing a lowered computational complexity in comparison with the well-known algorithms for the organization of a computational experiment on the investigation of statistical properties of binary random sequences is presented.

Key words: statistical tests, Markov chains, binary sequences (chains), simulation, probabilities of binary vectors, discrete random value, computational complexity, cryptographic algorithms.

Общие положения, термины и обозначения

В настоящее время одним из обязательных этапов верификации криптографических алгоритмов является их тестирование с использованием множеств двоичных последовательностей (ДСП), выступающих в роли шифруемых сообщений [1]. Суть указанного этапа заключается в проверке результатов шифрования указанных сообщений на основе статистических тестов на случайность, например тестов NIST [2; 3].

Современные подходы к решению задачи формирования множества ДСП основаны на методе статистических испытаний (Монте-Карло) [4]. При этом значительное распространение на практике получили 2 варианта метода. Первый вариант заключается в случайном выборе ДСП из заранее сформированного множества. Вто-

рой вариант основан на формировании ДСП с использованием математического аппарата сложных цепей Маркова (ЦМ) [5], позволяющего наиболее полно описывать статистические свойства ДСП, и метода обратной функции [6], обеспечивающего симуляцию ДСП за счет преобразования равномерного распределения в требуемое путем задания соответствующего отображения.

К настоящему времени наиболее изученными являются свойства равновероятных, или равномерно распределенных, согласно терминологии фундаментальной работы [7], ДСП. Это вызвано, во-первых, исключительной ролью равномерно распределенных ДСП в криптографии, а во-вторых – возможностью получения на их основе ДСП с требуемым рядом распределения двоичных векторов. При этом очевидно, что вариант формирования множества ДСП для верификации криптографических алгоритмов на основе ЦМ привлекательнее с точки зрения обеспечения возможностей по управлению качеством результатов вычислительного эксперимента. Рассмотрим особенности реализации указанного метода.

При использовании ЦМ отображение задается в виде матрицы переходных вероятностей (МПВ). В отличие от задач симуляции одномерных распределений особенностью симуляции ДСП на основе ЦМ является векторный характер получаемых распределений. Задача моделирования случайных векторов, элементы которых представляют собой различные случайные величины, рассмотрена в [8]. При этом моделирование требует указания совместного распределения нескольких случайных величин. В настоящей работе рассматривается другой случай, в котором все элементы векторов различной длины характеризуют одну двоичную случайную величину (ДСВ).

Иными словами, в зависимости от заданной связности двоичной ЦМ требуется определить ряд распределения двоичных комбинаций соответствующей длины. Затем на основе информации относительно значений вероятностей двоичных комбинаций несложно вычислить значения элементов МПВ цепи, требующихся для организации процесса симуляции ДСП.

Однако при проведении вычислительных экспериментов значения вероятностей двоичных векторов, составляющих в совокупности ряды распределений симулируемых ДСП, как правило, априорно неизвестны. Соответствующие численные значения требуется получать на основе некоторой исходной информации, характеризующей исследуемый случайный процесс при низкой степени его агрегирования.

Кроме того, в ряде случаев при моделировании двоичных векторов задание МПВ не приводит к требуемому результату. Иными словами, использование некоторой совокупности значений МПВ симулирует ДСП с рядом распределения, не соответствующим требуемому. Известные результаты теории марковских процессов не позволяют объяснить природу такого явления, но относят симулируемый процесс к так называемым неэргодическим.

Проблема наличия неэргодических марковских процессов усугубляется отсутствием условий эргодичности для класса двоичных марковских процессов, выполнение которых можно проверить до начала процесса симуляции, что значительно усложняет организацию вычислительных экспериментов при верификации криптографических алгоритмов. Фактически единственно возможным решением в этих условиях является организация эксперимента с обязательным включением в него дополнительной процедуры отбраковки ДСП, оказавшихся неэргодическими. Отбраковку несложно реализовать, например, на основе сравнения значений элементов МПВ, использованных при симуляции и вычисленных по ДСП, полученным в результате симуляции. Очевидно, такой подход избыточен с точки зрения затрачиваемых на реализацию вычислительных экспериментов временных и вычислительных ресурсов.

Следовательно, вопросы, связанные с формированием исходного набора ДСП, преобразуемых с использованием криптографических алгоритмов, недостаточно развиты с точки зрения организации направленного перебора значений рядов рас-

пределений двоичных векторов в ДСП. Значительный шаг в этом направлении сделан в работе [9], в которой предложен метод симуляции ДСП с заданными статистическими свойствами. Настоящая работа посвящена вопросам минимизации количества испытаний при исследовании качества криптографических алгоритмов на основе симуляции ДСП с заданными статистическими свойствами.

Формальная постановка задачи

Исходные данные:

1) $P(0)$ – вероятность события $0, P(0) \in [0,1]$;

2) v – масштаб распределения, т.е. длина двоичных векторов, для которых рассчитываются вероятности, $v \in Z, v \geq 2$;

3) L_i – количество интервалов значений вероятностей на масштабе i, $i \in Z, i \geq 2, L_v \in Z, L_v \geq 2$

4)n– номер варианта, $n \in Z, n \in [0, N)$.

Требуется разработать алгоритм расчёта значений вероятностей ряда распределения многомерных двоичных векторов, свободный от недостатков, связанных с необходимостью реализации процедуры отбраковки ДСП, обеспечивающий возможность направленного перебора статистических свойств симулируемых ДСП, варьирования точности их описания и обладающий пониженной вычислительной сложностью.

Свойства двоичных цепей Маркова, учитывающие зависимости вероятностей двоичных векторов различной длины

Рассмотрим множества двоичных последовательностей различной длины i с целью определения взаимосвязей вероятностей двоичных векторов различной длины. Соответствующая древовидная структура, узлами которой являются двоичные векторы, представлена на рис. 1. Каждый узел обозначен двухмерным индексом, где первый индекс i – длина вектора, а второй индекс j – десятичное значение двоичного вектора.

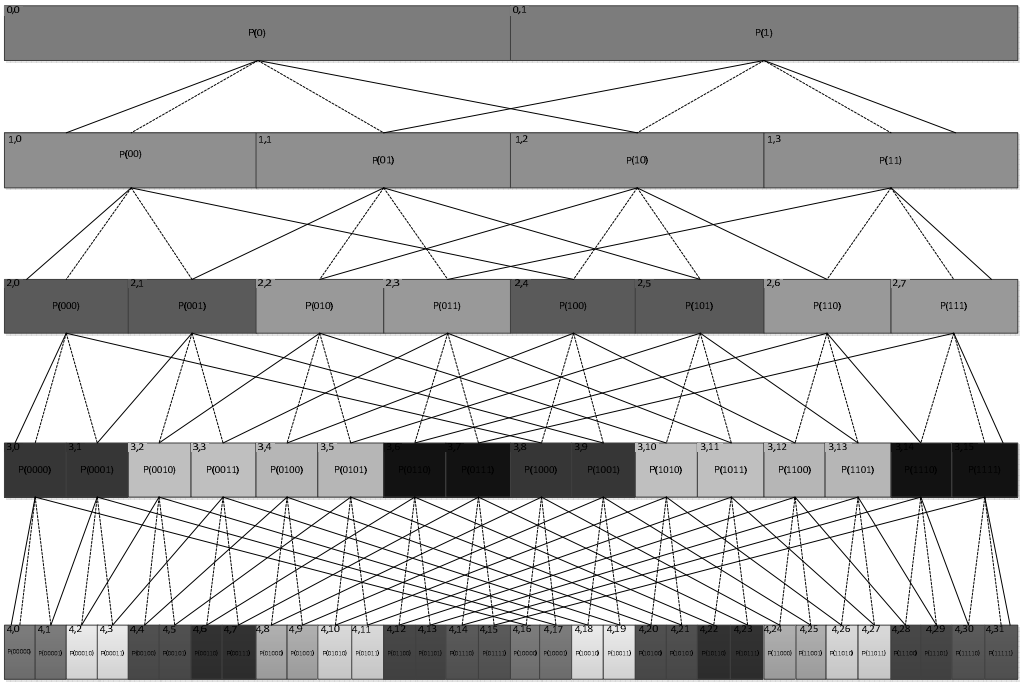

Рис. 1. Древовидная структура взаимосвязей вероятностей двоичных векторов различной длины

Очевидно, что значение вероятности двоичных векторов, исходя из определения полной группы событий, определяется выражениями

$$p(x_1, x_2, ..., x_v) = p(x_1, x_2, ..., x_v, x_{v+1}) + p(x_1, x_2, ..., x_v, \bar{x}_{v+1}),$$
$$p(x_1, x_2, ..., x_v) = p(x_0, x_1, x_2, ..., x_v) + p(\bar{x}_0, x_1, x_2, ..., x_v).$$

Тогда, учитывая номера позиций двоичных комбинаций в множествах комбинаций длины i, элемент модели на рис.1 с индексом (i, j) можно представить выражениями

$$p_{i-1,j} = p_{i,2j} + p_{i,2j+1}, \quad (1)$$
$$p_{i-1,j} = p_{i,j} + p_{i,j+2^i}. \quad (2)$$

Преобразование (1) и (2) с учетом введения номера четверки k позволило получить систему уравнений вида

$$\begin{cases} p_{i-1,k} = p_{i,2k} + p_{i,2k+1}; \\ p_{i-1,2k} = p_{i,2k} + p_{i,2k+}; \\ p_{i-1,2k+1} = p_{i,2k+1} + p_{i,2k+1+}; \\ p_{i-1,k+2^{i-1}} = p_{i,2k+2^i} + p_{i,2k+1+}. \end{cases} \quad (3)$$

Анализ значений индексов, определяющих значения двоичных векторов, показывает, что вероятности векторов, длина которых различается на единицу, образуют изолированные четверки. Таким образом, изменение значения вероятности любого двоичного вектора некоторой четверки оказывает влияние только на вероятности векторов этой четверки.

Решение системы (3) позволяет определить минимальные и максимальные значения вероятностей двоичных комбинаций длины i, исходя из значений свободных членов системы (3). Так, максимальные значения слагаемых уравнений системы (3) определяются выражениями

$$\max p_{i,2k} = \min[p_{i-1,k}; p_{i-1,2k}], \quad (4)$$
$$\max p_{i,2k+1} = \min[p_{i-1,k}; p_{i-1,2k+1}], \quad (5)$$
$$\max p_{i,2k+2} = \min[p_{i-1,2k}; p_{i-1,k+2^{i-1}}]$$
$$\max p_{i,2k+2} = \min[p_{i-1,2k+1}; p_{i-1,k+2^{i-1}}]$$

Минимальные значения слагаемых уравнений системы (3) определяются выражениями

$$\min p_{i,2k} = p_{i-1,k} - \max p_{i,2k+1} = p_{i-1,k} - \min[p_{i-1,k}; p_{i-1,2k+1}], \quad (6)$$
$$\min p_{i,2k+1} = p_{i-1,k} - \max p_{i,2k} = p_{i-1,k} - \min[p_{i-1,k}; p_{i-1,2k}],$$
$$\min p_{i,2k+2^i} = p_{i-1,2k} - \max p_{i,2k} = p_{i-1,2k} - \min[p_{i-1,k}; p_{i-1,2k}],$$
$$\min p_{i,2k+1+2^i} = p_{i-1,k+2^{i-1}} - \max p_{i,2k+2^i} = p_{i-1,k+2^{i-1}} - \min[p_{i-1,2k}; p_{i-1,k+2^{i-1}}],$$

Геометрически полученные зависимости удобно представить в виде трёх кругов, относительное расположение которых соответствует рис. 2. При этом углы секторов нижнего и верхнего кругов определяются парами значений свободных членов уравнений, а углы секторов среднего круга – четырьмя значениями слагаемых в правых частях уравнений, входящих в (3).

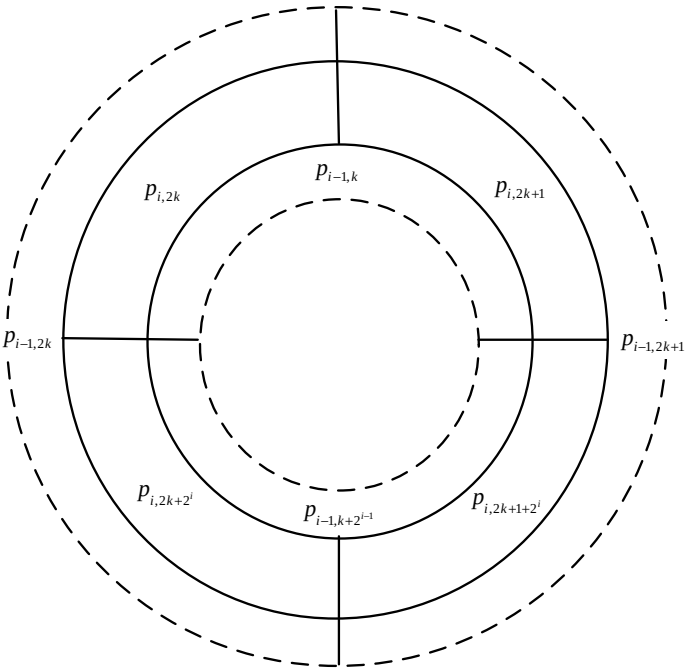

Рис. 2. Геометрическая интерпретация взаимосвязи значений вероятностей двоичных векторов, определяющих элементы четверок

Наличие информации относительно значений вероятностей векторов меньшей размерности позволяет судить о диапазонах изменений вероятностей векторов большей размерности.

Указанное обстоятельство позволило разработать новый метод направленного перебора рядов распределений в задачах моделирования ДСП, отличающийся от известного учетом ограничений на диапазоны значений вероятностей многомерных ДСВ, определяемых значениями вероятностей многомерных ДСВ меньшей размерности. Преимуществом метода является отсутствие необходимости осуществления операций, реализующих процесс отбраковки ДСП, для которых не выполняются требования к точности воспроизведения статистических свойств ДСП.

Алгоритм расчёта значений ряда распределения многомерных двоичных векторов

Разработанный метод реализуется алгоритмом, блок-схема которого представлена на рис. 3. Рассмотрим особенности указанного алгоритма. Результатом работы алгоритма является множество значений вероятностей двоичных векторов длины *v*. Номер варианта *n* определяет ряд распределения вероятностей масштаба *v* и выражается через десятичное представление многомерного числа, описывающего параметры ряда распределения. Количество разрядов указанного многомерного числа соответствует количеству четверок на всех масштабах и определяется выражением $N' = \sum_{i=2}^{v} 2^{i-2}$. Каждый разряд n_j задается в системе счисления по основанию $L_{f(j)}$, где $f(j) = 2 + [log_2(j+1)]$, $j = 0 \ldots N' - 1$, и определяет относительное положение значения вероятности ДСВ в интервале от минимального до максимального значения для каждой четверки. Десятичное значение варианта *n* определяется выражением

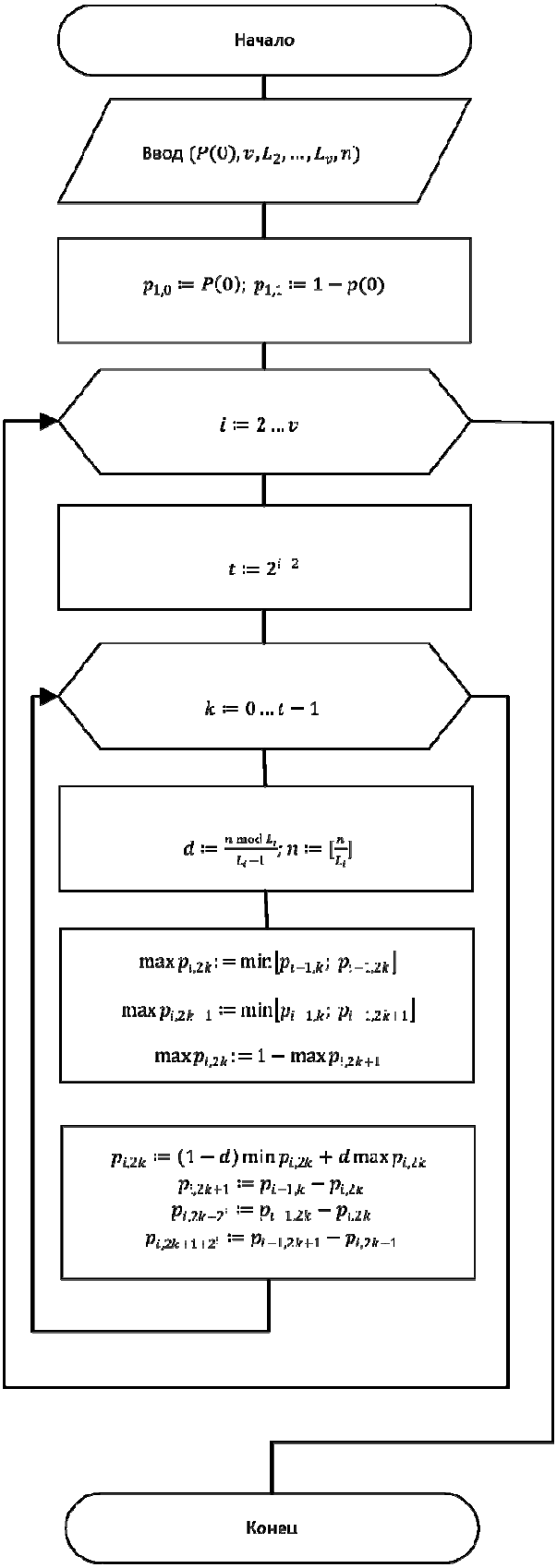

Рис. 3. Алгоритм расчёта значений вероятностей ряда распределения многомерных двоичных векторов на основании значения одномерной случайной величины и номера варианта

$$n = n_0 + \sum_{j=1}^{N^d-1} n_j \cdot \prod_{i=0}^{j-1} L_{f(i)},$$

а количество возможных вариантов – выражением

$$N = \prod_{i=2}^{v} (L_i)^{2^{i-2}} .$$

Рассмотрим функциональное назначение блоков алгоритма. Блок 2 определяет начальные значения вероятностей ДСВ на первом масштабе. В блоках 3–8 производится расчет вероятностей ДСВ на различных масштабах i= 2...v.

Количество различных комбинаций двоичных векторов длины i составляет величину 2^i. На каждом масштабе распределения все вероятности векторов $p_{i,m}$ длины i можно разбить на t непересекающихся групп по четыре элемента, m=0…2^i–1. Количество таких групп $t = \frac{2^i}{4} = 2^{i-2}$ рассчитывается в блоке 4.

Каждая четверка обрабатывается независимо друг от друга. Для обработки всех четверок вводится цикл (блоки 5–8) с параметром k=0…t-1, определяющим номер четверки.

В блоке 6 производится выделение младшего разряда из варианта n с последующим получением относительной позиции в интервале [0,1] для текущей обрабатываемой четверки в соответствии с выражением

$$d = \frac{n \bmod L_i}{L_i - 1}$$

и удаление младшего разряда за счет сдвига числа n вправо на один разряд:

$$n = \left[\frac{n}{L_i}\right].$$

Блок 7 предназначен для расчета минимального и максимального значений первого элемента обрабатываемой четверки в соответствии с выражениями (4), (5), (6).

В блоке 8 производится расчет значений финальных вероятностей для текущей обрабатываемой четверки в соответствии с выражениями

$$p_{i,2k} = (1-d)\cdot \min p_{i,2k} + d \cdot \max p_{i,2k}$$

,

$$p_{i,2k+1} = p_{i-1,k} - p_{i,2k},$$

$$p_{i,2k+2^i} = p_{i-1,2k} - p_{i,2k},$$

$$p_{i,2k+1+2^i} = p_{i-1,2k+1} - p_{i,2k+1}.$$

После обработки всех четверок текущего масштаба i получаем вероятности всех векторов длины i ($p_{i,m}, m = 0..(2^i - 1)$).

Результатом выполнения всех итераций цикла по i являются искомые (финальные) значения вероятностей векторов длины v.

Вычислительная сложность алгоритма

Каждый шаг цикла (блоки 6–8) требует 8 операций сложения/вычитания и 2 операций умножения (будем считать, что операция min требует одной операции вычитания, а формулы в блоке 6 в расчете вычислительной сложности не учитываются).

Каждый шаг цикла обрабатывает одну четверку. На масштабе v количество четверок есть величина 2^{i-2}. Количество четверок, обрабатываемых на всех масштабах до i=v, есть величина $\sum_{i=2}^{v} 2^{i-2} = 2^{v-1} - 1$. Следовательно, количество операций сложения/вычитания для представленного алгоритма есть величина $8*(2^{v-1}-1)$, т.е. сложность алгоритма $O(v)$=2^{v-1}.

Количество всех вариантов рассматриваемых рядов распределений есть величина $N = \prod_{i=2}^{v} (L_i)^{2^{i-2}}$. Пусть все значения L_i=L, тогда $N = \prod_{i=2}^{v} L^{2^{i-2}} = L^{2^{v-1}-1}$. Следовательно, вычислительная сложность алгоритма полного перебора рядов распределений масштаба k с количеством интервалов значений вероятностей в каждой четверке L определяется выражением

$$O(v, L) = 2^{v-1} \cdot L^{2^{v-1}-1}.$$

Выводы

Разработанный алгоритм обеспечивает возможность варьирования точности описания статистических свойств ДСП посредством наличия параметров, описывающих связность ЦМ (максимальный масштаб) и количество интервалов значений вероятностей в группах на каждом масштабе.

Проведенный сравнительный анализ вычислительной сложности известного алгоритма организации вычислительного эксперимента по исследованию статистических свойств ДСП и разработанного алгоритма, реализующего представленный в работе метод, позволил сделать вывод о том, что разработанный алгоритм обладает пониженной вычислительной сложностью при обеспечении выполнения заданных требований к точности воспроизведения статистических свойств симулируемых ДСП.

Для изучения вопросов по определению требований к генераторам ДСП с равномерным законом распределения, используемым в процедуре симуляции марковских ДСП, необходимо проведение дополнительных (желательно совместных с вьетнамскими коллегами) исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Фомичёв, В. М. Методы дискретной математики в криптологии/В.М. Фомичев. – М.: Диалог-МИФИ, 2010. – 424 с.
2. Gustafson. A computer package for measuring strength of encryption algorithms/Gustafson//Journal of Computers & Security. – 1994. - Vol. 13. - № 8. - P. 687-697.
3. Ritter, T. Randomness Tests and Related Topics/T. Ritter.-URL:http://www.ciphersbyritter.com/ RES/RANDTEST.HTM.
4. Бусленко, Н.П. Метод статистических испытаний (Монте-Карло) и его реализация на цифровых вычислительных машинах/Н.П.Бусленко, Ю.А.Шрейдер. - М.: ГИФМЛ, 1961. – 226 с.
5. Баруча-Рид, А. Т. Элементы теории марковских процессов и их приложения/А.Т. Баруча-Рид.-М.: Наука, 1969. – 512 с.
6. Ермаков, С.М. Статистическое моделирование. Ч. 1. Моделирование распределений: учеб. пособие/С.М. Ермаков. - СПб.: НИИМиМ им. В.И. Смирнова, 2006.-63 с.
7. Кейперс, Л. Равномерное распределение последовательностей/Л. Кейперс, Г. Нидеррейтер.– М.: Наука, 1985. – 408 с.
8. Советов, Б. Я. Моделирование систем / Б.Я. Советов, С.А. Яковлев.–М.: Юрайт, 2012. – 343 с.
9. Близнюк, В.И. Метод направленного перебора рядов распределений в задачах моделирования марковских двоичных последовательностей / В.И. Близнюк, М.Ю. Конышев, В.А. Иванов, С.В. Харченко //Промышленные АСУ и контроллеры. - 2015. – №5. - С. 40-45.

1. Fomichyov, V. M. *Methods of Discrete Mathematics in Cryptology* /V.M. Fomichyov. – M.: Dialogue-MEPI, 2010. – pp. 424.
2. Gustafson. A computer package for measuring strength of encryption algorithms/Gustafson//Journal of Computers & Security. – 1994. - Vol. 13. - № 8. - P. 687-697.
3. Ritter, T. Randomness Tests and Related Topics/T. Ritter.-URL:http://www.ciphersbyritter.com/ RES/RANDTEST.HTM.
4. Buslenko, N.P. *Method of Statistical Tests (Monte-Carlo) and Its Realization on Digital Computers/N.P.Buslenko,* Yu.A.Shreider. - M.: SPPML, 1961. – 226.
5. Barucha-Reed, A. T. *Elements of Theory of Markov Processes and Their Applications*/A.T. Barucha-Reed.-M.: Science, 1969. – pp. 512.
6. Yermakov, S.M. *Statistical Modeling. Part. 1. Simulation of Distributions:* manual/S.M. Yermakov. – S-Pb.: Smirnov RIM&M, 2006.-pp. 63.
7. Keipers, L. *Uniform Distribution of Sequences* /L. Keipers, G. Niderreiter.–M.: Science, 1985. – pp. 408.
8. Sovetov, B.Ya. *System Modeling* / B.Ya. Sovetov, S.A. Yakovlev.–M.: Yuright, 2012. – pp. 343.
9. Bliznyuk, V.I. Method of directed search of distribution sets in problems of simulation of Markov binary sequences / V.I. Bliznyuk, M.Yu. Konyshev, V.A. Ivanov, S.V. Kharchenko //*Industrial ASC and Controllers*. - 2015. – №5. - pp. 40-45.

Статья поступила в редколлегию 16.09.2016.
Рецензент: д.т.н., профессор Брянского государственного технического университета Аверченков В.И.

Сведения об авторах:

Конышев Михаил Юрьевич, к.т.н., доцент, сотрудник Академии ФСО России, e-mail: alex.totrin@gmail.com.
Козачок Александр Васильевич, к.т.н., сотрудник Академии ФСО России, e-mail: atotrin@gmail.com.

Голембиовская Оксана Михайловна, к.т.н., доцент, нач. отдела организации научно-исследовательской работы студентов, аспирантов и молодых ученых, e-mail: bryansk-tu@yandex.ru.
Петров Константин Евгеньевич, сотрудник Академии ФСО России, e-mail: pke.orel@bk.ru.

Konyshev Mikhail Yurievich, Can. Eng., Assistant Prof., worker of Academy of FSG of Russia, e-mail: alex.totrin@gmail.com.
Kozachok Alexander Vasilievich, Can. Eng., worker of Academy of FSG of Russia, e-mail: atotrin@gmail.com.

Golembiovskaya Oksana Mikhailovna, Can. Eng., Assistant Prof., Chief of Dep. for Students, Post Graduate Students and Young Scientists Research Work Organization, e-mail: bryansk-tu@yandex.ru.
Petrov Konstantin Yevgenievich, worker of Academy of FSG of Russia, e-mail: pke.orel@bk.ru.