СИСТЕМИ ТА МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ

УДК 621.327

В.В. Баранник, доктор технических наук, профессор,
С.А. Сидченко, кандидат технических наук,
В.В. Ларин

МЕТОДИКА СТАТИСТИЧЕСКОГО ТЕСТИРОВАНИЯ ДЕШИФРИРУЕМО-СТОЙКОГО ПРЕДСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ

Предложена методика тестирования для оценки статистических характеристик дешифрируемо-стойкого представления изображений, и на их основе проведено тестирование последовательностей кодов-номеров. Из результатов статистического тестирования видно, что в последовательностях между элементами дешифрируемо-стойкого представления изображений отсутствует зависимость.

Ключевые слова: *дешифрируемо-стойкое представление, статистическое тестирование.*

Запропоновано методику тестування для оцінки статистичних характеристик дешифровано-стійкого представлення зображень, і на їх основі проведено тестування послідовностей кодів-номерів. З результатів статистичного тестування видно, що в послідовностях між елементами дешифровано-стійкого представлення зображень відсутня залежність.

Ключові слова: *дешифровано-стійке представлення, статистичне тестування.*

Testing methodology for an estimation of the statistical characteristics of a decoded-proof representation of images is offered and on their basis the testing of sequences of codes-numbers is carried out. From the statistical testing results it is visible that in sequences between the elements of images decoded-proof representation there is no any dependence.

Keywords: *decoded-proof representation, statistical testing.*

Развитие мультимедийных приложений и их внедрение в различные сферы деятельности человека с использованием современных информационно-телекоммуникационных систем требует решения ***актуальной научно-прикладной задачи***, заключающейся в сокращении времени на обработку и доставку постоянно повышающегося объема видеоинформации с заданным уровнем конфиденциальности.

Одним из способов оценки качества криптографических алгоритмов является оценка их статистических характеристик.

Каждый раз, когда встает вопрос о том, насколько случайна тестируемая последовательность, следует иметь в виду, что спектр критериев оценки последовательностей чрезвычайно широк. За последнее время разработано большое количество статистических тестов для анализа того, насколько последовательность способна демонстрировать случайное поведение.

Для определения меры случайности бинарных последовательностей, порожденных аппаратными или программными генераторами случайных чисел, разработано большое количество тестов. Наибольшую популярность получил пакет из 15 статистических тестов, разработанный Лабораторией информационных технологий (Information Technology Laboratory), являющейся главной исследовательской организацией Национального института стандартов и технологий (NIST) [1].

Однако основной задачей предложенного в работах [2, 3] метода построения дешифрируемо-стойкого преобразования (ДШСП) изображений является не формирование псевдослучайных последовательностей, а сокрытие семантического смысла изображения с учетом как статистических, так и структурных особенностей источника информации. Кроме того, в процессе предложенного подхода для построения ДШСП изображений происходит интеграция нескольких исходных битовых последовательностей в одну последовательность переменной длины. Это приводит к несоответствию формируемых кодограмм дешифрируемостойкого представления требованиям относительно входных последовательностей для некоторых тестов пакета NIST, предназначенных для криптографических алгоритмов стандартной серии. Поэтому для статистического тестирования последовательностей ДШСП изображений предлагается использовать только часть тестов комплекса NIST. При этом статистическое тестирование предлагается провести без учета незначимых элементов в ДШСП (незначимых нулевых бит в начале каждой битовой последовательности кодов-номеров).

Целью исследований статьи является разработка подхода к оценке статистических характеристик дешифрируемо-стойкого представления изображений и проведение тестирования.

1. Частотный побитовый тест. Цель теста – проверка равновероятности появления 0 и 1 в исследуемой последовательности.

Пусть $\varepsilon = \varepsilon_1\varepsilon_2...\varepsilon_n$ – тестируемая двоичная последовательность длиной n бит. Количество 1 в тестируемой битовой последовательности ε определяется выражением $n_1 = \sum_{i=1}^{n}\varepsilon_i$. Количество нулей в тестируемой битовой последовательности определяется выражением $n_0 = n - n_1$.

Исходя из целей теста оценивается близость вероятности появления единиц к 1/2, то есть количество 0 и 1 в последовательности должно быть приблизительно одинаковым:

$$P(n_1) = \frac{n_1}{n} \approx P(n_0) = \frac{n_0}{n} \approx \frac{1}{2},$$

где $P(n_1)$ и $P(n_0)$ – соответственно вероятности появления 1 и 0 в тестируемой последовательности.

Диаграммы зависимости вероятности появления 1 и 0 в тестируемых последовательностях ($P(n_1)$ и $P(n_0)$ соответственно) для разной степени насыщенности изображений приведены на рис. 1.

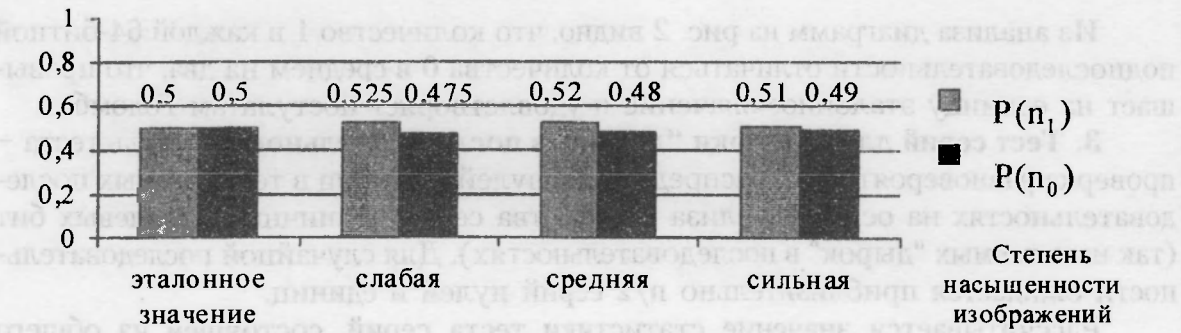


Рис. 1. Диаграмма значений величин $P(n_1)$ и $P(n_0)$

Из анализа диаграмм на рис. 1 видно, что количество 1 в последовательностях больше количества 0 в среднем на 2–5 %, а вероятность появления единиц отклоняется от $\frac{1}{2}$ всего на 1–2,5 %.

2. Частотный тест в подпоследовательностях. Цель теста – проверка равновероятности появления 0 и 1 в подпоследовательностях.

Двоичная последовательность $\varepsilon = \varepsilon_1\varepsilon_2...\varepsilon_n$ длиной n бит разбивается на N=[n/M] непересекаемых M-битных последовательностей. Лишние биты отбрасываются. После этого определяется доля (вероятность появления) единиц η_i в каждой подпоследовательности с помощью выражения:

$$\eta_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)M+j}}{M}, \quad 1 \le i < M.$$

При тестировании двоичной формы представления ДШСП длину подпоследовательности целесообразно выбирать равной 64 битам. При этом количество 1 и 0 в каждой подпоследовательности должно быть в среднем равным 32 элементам:

$$n_1 \approx n_0 \approx n/2 \approx 32.$$

Согласно постулатам Голомба, в идеальном варианте количество 1 в каждом периоде должно отличаться от количества 0 не более чем на единицу.

Диаграммы распределения среднего количества единиц и нулей (n_1 и n_0 соответственно) в 64-битных подпоследовательностях тестируемых последовательностей для разной степени насыщенности изображений приведены на рис. 2.

Рис. 2. Диаграмма значений величин n_1 и n_0

Из анализа диаграмм на рис. 2 видно, что количество 1 в каждой 64-битной подпоследовательности отличаться от количества 0 в среднем на два, что превышает на единицу эталонное значение и удовлетворяет постулатам Голомба.

3. Тест серий для проверки "дырок" в последовательностях. Цель теста – проверка равновероятности распределения нулей и единиц в тестируемых последовательностях на основе анализа количества серий единичных и нулевых бит (так называемых "дырок" в последовательностях). Для случайной последовательности ожидается приблизительно n/2 серий нулей и единиц.

Рассчитывается значение статистики теста серий, состоящей из общего количества единичных и нулевых серий по всей последовательности:

$$s_{test} = \sum_{k=1}^{n-1} r_k + 1, \text{ где } r_k = \begin{cases} 0, \text{если } \varepsilon_k = \varepsilon_{k+1}, \\ 1, \text{если } \varepsilon_k \neq \varepsilon_{k+1}. \end{cases}$$

В частном случае тест определяет монотонность колебаний между нулями и единицами в битовых последовательностях. Большие значения статистики теста говорят о быстрых колебаниях между сериями 0 и 1, а ее малые значения говорят о медленных колебаниях. В идеальном варианте значение статистики теста должно принимать значение равное $s_{test} = n/2$.

На рис. 3 представлены диаграммы значений статистики теста s_{test} в процентах от общего объема для разной степени насыщенности изображений.

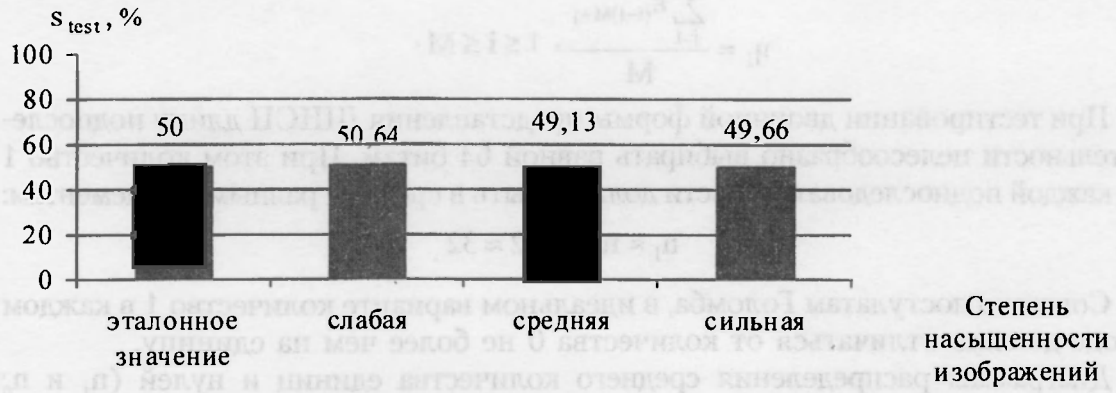


Рис. 3. Диаграмма значений величин статистики теста серий в процентах от общего объема

Из анализа диаграмм на рис. 3 видно, что значение статистики теста отличается от оптимального значения менее 1 %.

4. Тест серий пар и троек бит. Цель теста – проверка равновероятности распределения серий-пар (00, 01, 10 и 11) и серий-троек (000, 001, 010, 011, 100, 101, 110 и 111) в тестируемых двоичных последовательностях.

Исходная двоичная последовательности $\varepsilon = \varepsilon_1\varepsilon_2...\varepsilon_n$ длиной n бит разбивается на двойки и тройки элементов и определяется их количество.

Последовательности является случайной, если равновероятно появление нулевых и единичных бит, серий-пар бит и серий-троек бит, которое определяется по формулам соответственно:

$$P(f) = \frac{\sum_{i=1}^{n} \varepsilon_i}{n} \to \frac{1}{2}, \text{ где } f = \{0,1\},$$

$$P(j) = \frac{\sum_{i=1}^{n/2} \varepsilon(j)_i}{n} \to \frac{1}{4}, \text{ где } j = \{00,01,10,11\},$$

$$P(k) = \frac{\sum_{i=1}^{n/3} \varepsilon(k)_i}{n} \to \frac{1}{8}, \text{ где } k = \{000,001,010,011,100,101,110,111\}.$$

Диаграммы вероятностей появления серий-пар в битовых последовательностях кодов-номеров для разной степени насыщенности изображений представлены на рис. 4, а для серий-троек – на рис. 5.

Рис. 4. Диаграмма вероятности появления серий-пар бит

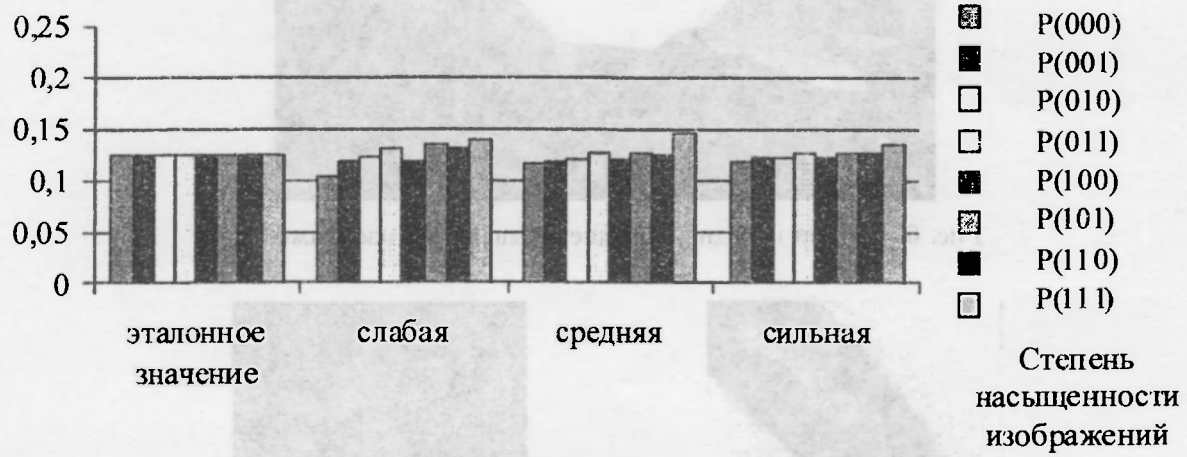


Рис. 5. Диаграмма вероятности появления серий-троек бит

Из анализа диаграмм на рис. 4 и 5 видно, что вероятность распределения серий-пар в тестируемых двоичных последовательностях находится в пределах 0,223–0,272 при расчетном значении в 0,25, а серий-троек – в пределах 0,103–0,146 при расчетном значении в 0,125. Отклонение от равновероятного распределения составляет до 10 % для серий-пар элементов и до 18 % для серий-троек. Наилучшие результаты получены для сильнонасыщенных реалистических

изображений, где отклонения от равновероятного распределения составляет до 5 % для серий-пар и до 9 % для серий-троек элементов.

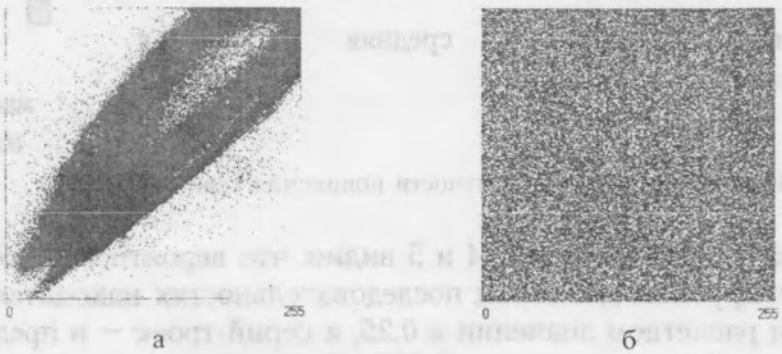
5. Распределение элементов последовательностей на плоскости. Определение зависимости между элементами тестируемой последовательности предлагается путем построения распределения на плоскости (на поле) размером $(2^R-1)\times(2^R-1)$, где R – разрядность исследуемой последовательности. Координаты точек на плоскости определяются как $(\varepsilon_i;\varepsilon_{i+1})$, где ε_i – элементы тестируемой последовательности ε, $i=\overline{1,(n-1)}$, n – длина последовательности.

Для построения распределения видеоданных разрядность тестируемой последовательности принимается равной R = 8 бит, следовательно распределение строится на плоскости размером 255×255 бит. Распределение последовательностей кодов-номеров предлагается строить с учетом разрядности R-8 бит и R=64 бит.

Если между элементами последовательности отсутствует зависимость, то точки на поле расположены хаотично. В противном случае на поле наблюдается узор. Для последовательностей большой длины хорошим результатом является "черный квадрат".

Для исходного средненасыщенного изображения, представленного на рис. 6, распределения элементов изображения и кодов-номеров ДШСП (в динамическом диапазоне изображений) на плоскости представлены на рис. 7.

Рис. 6. Пример исходного средненасыщенного изображения



а б

Рис. 7. Пример распределения элементов на плоскости:
а) изображения; б) кодов-номеров ДШСП

Из анализа распределений на плоскости видно, что между элементами ДШСП (рис. 7б) отсутствует зависимость, а оно стремится к "черному квадрату". При этом распределение элементов исходного изображения (рис. 7а) формирует четко выраженый узор.

6. Гистограмма распределения элементов последовательности. С помощью гистограммы распределения элементов определяется частота появления (использования) определенного цвета (номера цвета) в изображении или кода-номера в кодограммах. Гистограмма предназначена для оценки равномерности распределения элементов в исследуемой последовательности.

Исходная тестируемая последовательность байтов ε преобразуется в десятичную форму ASCII-значений элементов, в которой подсчитывается количество появлений каждого значения. Данные представляются в виде графика зависимости числа появления значений элементов в последовательности. Последовательность считается случайной, если в ней присутствуют все возможные элементы рассматриваемой разрядности, а разброс частот появления элементов стремится к нулю.

Для исходного средненасыщенного изображения, представленного на рис. 6, гистограммы распределения элементов изображения и кодов-номеров ДШСП (в динамическом диапазоне изображений) представлены на рис. 8.

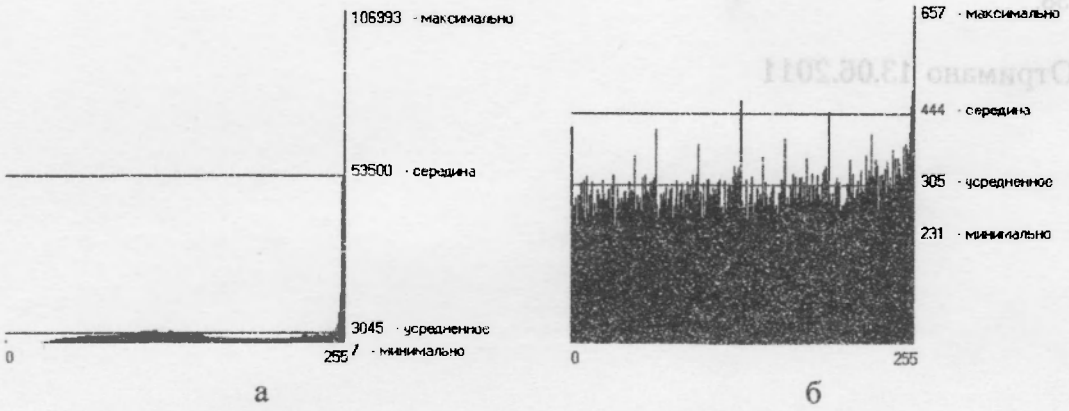


Рис. 8. Пример гистограммы распределения элементов:
а) изображения; б) кодов-номеров ДШСП

Из анализа гистограмм на рис. 8 видно, что в последовательности кодов-номеров ДШСП присутствуют все возможные элементы рассматриваемой разрядности, а разброс частот появления элементов значительно меньше, чем в исходном изображении.

Выводы. Из результатов статистического тестирования последовательностей ДШСП видно, что:

– количество 1 в последовательностях больше количества 0 в среднем на 2–5 %, а вероятность появления единиц отклоняется от 1/2 всего на 1–2,5 %;

– количество 1 в каждой 64-битной подпоследовательности отличаться от количества 0 в среднем на два, что превышает на единицу эталонное значение и удовлетворяет постулатам Голомба;

– в последовательностях кодов-номеров ДШСП наблюдается одинаковое количество серий 0 и 1, отличающееся от оптимального значения менее 1 %;

– вероятность распределения серий-пар в тестируемых двоичных последовательностях находится в пределах 0,223–0,272 при расчетном значении в 0,25, а серий-троек – в пределах 0,103–0,146 при расчетном значении в 0,125. Наилучшие результаты получены для сильнонасыщенных реалистических изображений;

– между элементами ДШСП отсутствует зависимость;

– в последовательности кодов-номеров ДШСП присутствуют все возможные элементы исходного динамического диапазона изображения, а разброс частот появления элементов значительно уменьшен по сравнению с исходным изображением.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. A statistical test suite for random and pseudorandom number generators for cryptographic applications // NIST Special Publication 800-22 [Електронний ресурс] – Режим доступа: http://csrc.nist.gov/publications/PubsSPs.html.

2. *Баранник В.В.* Метод криптосемантического представления изображений на основе комбинированного подхода / В.В. Баранник, С.А. Сидченко, В.В. Ларин // Сучасна спеціальна техніка. – 2010. – № 3 (22). – С. 33–38.

3. *Баранник В.В.* Метод дешифрируемо-стойкого представления изображений / В.В. Баранник, С.А. Сидченко, В.В. Ларин // Сучасна спеціальна техніка. – 2011. – № 1 (24). – С. 33–38.

Отримано 13.06.2011