УДК 519.24 DOI: 10.24151/1561-5405-2019-24-3-301-308

Идентификация бернуллиевских последовательностей криптографическими методами

И.П. Кобяк

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь

IPKobyak2012@mail.ru

В представляемой работе рассмотрен метод синтеза оценок вероятности регистрации двоичных (01)-переходов при решении задач шифрования и идентификации сообщений. Получено соотношение для вероятности пропуска ошибки, соответствующее данному методу при наблюдении спецсистемами априори неопределенной асимптотической выборки. Определена мода распределения и выполнен сравнительный анализ вероятностей пропуска ошибки, характеризующих предлагаемый метод и известные алгоритмы свертки в точке моды. Показано, что метод наблюдения переходов в вероятностных процессах является более точным алгоритмом идентификации, чем сигнатурный анализ или счет бернуллиевских событий.

Ключевые слова: оценка вероятности; двоичный переход; асимптотическая выборка; методы свертки; мода распределения; сигнатурный анализ

Для цитирования: Кобяк И.П. Идентификация бернуллиевских последовательностей криптографическими методами // Изв. вузов. Электроника. – 2019. – Т. 24. – № 3. – С. 301–308. DOI: 10.24151/1561-5405-2019-24-3-301-308

Identification of Bernoulli's Sequences by Cryptographic Special Computers

I.P. Kobyak

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarusia

IPKobyak2012@mail.ru

Abstract: For real applications at the stage of preparing the information for piping an algorithm, having small probability of mistake admission for concrete realization of the formed message, has been chosen. In the paper the method of synthesis for assessment of the probability of registration of the binary (1)-transitions in solving the problems of coding and identification of messages has been considered. The ratio for the probability of a mistake admission, corresponding to the specified method at supervision a priori of uncertain asymptotic selection has been obtained. The fashion of distribution has been defined and the comparative analysis of the probabilities of a mistake admission, characterizing the offered method and the known algorithms of convolution at the given

point has been defined. The proposed method of supervision in the probability processes is a more precise algorithm of identification, than the signature analysis.

Keywords: probability assessment; binary transition; asymptotic selection; convolution methods; fashion of distribution; signature analysis

For citation: Kobyak I.P. Identification of Bernoulli's sequences by cryptographic special computers. *Proc. Univ. Electronics*, 2019, vol. 24, no. 3, pp. 301–308. DOI: 10.24151/1561-5405-2019-24-3-301-308

Введение. Использование значений выборочных функций в качестве контрольных кодов в настоящее время широко применяется при решении задач контроля и диагностики, в системах шифрования и передачи данных, при хранении информации в памяти компьютеров. Основным параметром, характеризующим конкретно выбранный метод синтеза контрольного кода, является вероятность пропуска ошибки (*eo*-error omit) P_{eo}, которая фактически указывает на число последовательностей, составляющих класс эквивалентностей нулевой гипотезы. В настоящей работе под нулевой гипотезой понимается достоверное сообщение, принимаемое или передаваемое некоторой системой наблюдения, регистрации и идентификации последовательностей. При этом, чем меньше мощность класса эквивалентностей (МКЭ), тем больше вероятность перехода в случае появления ошибок текущей статистической оценки в другие классы сигнатур. Следовательно, для реальных приложений на стадии подготовки информации к пересылке должен быть выбран алгоритм, имеющий минимальную вероятность пропуска ошибки для конкретной реализации сформированного сообщения.

В большинстве задач формирования контрольных кодов используются два основных алгоритма: синтез линейных сигнатур с вероятностью пропуска ошибки (*sa*-signature analysis) P_{sa} и метод формирования оценок бернуллиевских событий с вероятностью ошибки (*cvc*-condition vector counting) P_{cvc}. Недостатком данных алгоритмов является большое число последовательностей, принадлежащих МКЭ достоверных оценок, особенно при длине выборки $n \to \infty$.

Наиболее совершенным и эффективным методом синтеза контрольных кодов с точки зрения минимальности вероятности ошибки следует считать алгоритм формирования оценок на основе наблюдения двоичных переходов вида 01 с вероятностью ошибки (*wfc*-wave front counting) P_{wfc}. Данное предположение основывается, во-первых, на том, что формируемый код оценки имеет длину меньшую на единицу по отношению к квазистандартным сигнатурам, во-вторых, на факте уменьшения числа различных перестановок с повторениями заданных пар событий на *n* местах размещения.

Цель настоящей работы – сопоставительный анализ вероятностей P_{sa} и P_{cvc} при формировании и выборе кодов идентификаторов в системах шифрования и передачи данных.

Вероятность пропуска ошибки при наблюдении (01)-переходов. С учетом эквивалентности методов наблюдения (01)-переходов и событий в автокорреляционной функции для решения поставленной задачи будем использовать известное соотношение для классической функции ковариации двух бернуллиевских событий:

$$\operatorname{cov}\left(X, \tau = 1\right) = M\left[\left(x_t - M_X\right)\left(x_{t+1} - M_X\right)\right]. \qquad (1)$$

Рассмотрим теперь статистическую выборку достаточно большой длины ($n \to \infty$), эквивалентную стационарному и эргодическому процессу с математическим ожиданием $M_X = 0,5$ и выберем в качестве регистрируемых элементарных событий x_t значения «0», а в качестве регистрируемых отсчетов x_{t+1} значения «1». Выполнив замену пары событий x_t, x_{t+1} на их произведение $\overline{x}_t x_{t+1}$ в соотношении (1), можно выделить слагаемое

$$f\left(\overline{x}_t x_{t+1}\right) = M\left[-0,5\overline{x}_t - 0,5x_{t+1} + 0,25\right] = -0,25. \qquad (2)$$

В двоичных последовательностях при $\tau = 1$ произведение событий $\overline{x}_t\, x_{t+\tau} = 1$ характеризует переход информационного объекта из состояния «0» в состояние «1». На практике такие события могут быть реализованы с использованием дифференциального звена, позволяющего синтезировать соответствующие объекты с математическим ожиданием:

$$M_\omega = M_X\left(1 - M_X\right) + \operatorname{cov}\left(X, \tau = 1\right). \qquad (3)$$

Очевидно, что для реального эксперимента в функции (3) выполняется замена теоретического значения M_ω на оценку числа элементарных событий заданного вида $\hat{k}_\omega = \hat{p}_\omega n$. Однако для больших значений n имеем сумму событий $\overline{x}_t$ и $x_{t+\tau}$ в (2), примерно равную n. Тогда для рассматриваемых последовательностей будет справедливо соотношение, аналогичное (3):

$$M_\omega = 0,25 + \left(\frac{1}{n}\sum_{t=1}^{n} \overline{x}_t x_{t+\tau} - 0,25\right). \qquad (4)$$

При $n \to \infty$ равенство (4) позволяет вычислить число двоичных переходов вида 01, соответствующее математическому ожиданию указанных событий в наблюдаемой последовательности. Действительно, учитывая, что

$$\frac{1}{n}\sum_{t=0}^{n-1} \overline{x}_t x_{t+\tau} - 0,25 \to 0,$$

имеем

$$\hat{k}_\omega = \sum_{t=0}^{n-1} \overline{x}_t x_{t+\tau} \xrightarrow{n\to\infty} 0,25n. \qquad (5)$$

Таким образом, с учетом (5) теоретическое значение $M_\omega n$ равно $0,25n$.

Определим вероятность пропуска ошибки при наблюдении в последовательностях случайных событий переходов вида 01. Современные комбинаторные подходы к определению вероятности ошибки P_{wfc} при регистрации переходов в двоичной выборке основываются на анализе множества единичных серий $\hat{k}_\psi$ длиной $\psi = \overline{1, n-1}$ и приводят к необходимости анализа пограничных вопросов в теории многочленов разбиений.

Однако существует более простой аналитический подход [1, 2] к решению поставленной задачи. При этом каждая серия из единичных символов длиной $\psi > 1$ рассматривается как событие единичной длины в последовательности с $n - \psi + 1$ отсчетами.

При общем числе серий $\hat{k}_\omega = \hat{p}_\omega n$ длина такой реорганизованной выборки или число мест для различных перестановок с повторениями (01)-переходов будет равна $n_1 = n - g$ [2], где $g = \hat{k}_\omega + \sum_{\psi=2}^{n-1} (\psi - 1)\hat{k}_\psi$.

Тогда в зависимости от длины и числа серий количество реализаций выборки, составляющих МКЭ оценки $\hat{k}_\omega$, может быть определено по формуле

$$Q = \sum_{g=\hat{k}_\omega}^{n-\hat{k}_\omega} C_{n-g}^{\hat{k}_\omega} C_{g-1}^{\hat{k}_\omega - 1}. \quad (6)$$

Из соотношения (6) следует, что вероятность пропуска ошибки при счете двоичных переходов будет определяться формулой

$$P_{wfc} = \frac{1}{2^n}\left[\sum_{g=\hat{k}_\omega}^{n-\hat{k}_\omega} C_{n-g}^{\hat{k}_\omega} C_{g-1}^{\hat{k}_\omega - 1} - 1\right], \quad (7)$$

где $0 \le \hat{k}_\omega \le 0,5n.$

Использование зависимостей (6) и (7) в общем случае затруднено в связи со сложностью представления параметра g. Поэтому для определения уровня ошибки, характеризующего алгоритм счета (01)-переходов, преобразуем равенство (7) с использованием асимптотического свойства равновероятности для элементарных событий в случайных последовательностях. При этом можно показать, что $g \to 0,5n$, и равенство (7) может быть приведено к виду

$$P_{w.f.c} = \frac{1}{2^n}\left[2p_{0,\omega}\left(C_{0,5n}^{np_{0,\omega}}\right)^2 - 1\right], \quad (8)$$

где $\hat{p}_\omega = p_{0,\omega}$ при $n = \infty$.

Очевидно, что формула (8) будет верна в асимптотике. Однако полученное соотношение может использоваться для приближенного поиска общих точек на плоскости ошибки P_{eo} при сравнении уровней вероятностных параметров, порождаемых конкурирующими в задачах идентификации алгоритмами. При этом решение $np_{0,\omega}$ в (8) будет соответствовать моде функции распределения вероятностей P_{wfc} [3].

Сравнительный анализ методов синтеза оценок числа двоичных переходов и линейных сигнатур. Определим преимущество метода счета (01)-векторов переходов по отношению к алгоритму формирования линейных сверточных кодов при $n \to \infty$. С этой целью приравняем МКЭ, соответствующие двум рассматриваемым методам, и составим гипотетическое равенство вида

$$2p_{0,\omega}\left(C_{0,5n}^{np_{0,\omega}}\right)^2 = 2^{n-l}, \quad (9)$$

где $l = \log_2 n$ – степень примитивного полинома, описывающего обратную связь теоретического регистра сдвига сигнатурного анализатора.

Равенство (8) базируется на биномиальной теореме с разрядностью случайного процесса $r = 1$. Поэтому очевидно, что уравнение (9) может иметь либо два решения, либо ни одного.

Теорема. Если длина выборки элементарных событий n стремится к бесконечности, то $P_{sa} > P_{wfc}$ во всем диапазоне значений аргумента p_{ω}.

Для доказательства теоремы проведем преобразование (9) с учетом формул Стирлинга для факториальных составляющих. При этом получим

$$2p_{0,\omega}\frac{n^n}{2^n}\frac{\exp 2\left(\theta_1-\theta_2-\theta_3\right)}{\left[np_{0,\omega}\right]^{2np_{0,\omega}}\left[\frac{n}{2}-np_{0,\omega}\right]^{n-2np_{0,\omega}}}\left(4\pi np_{0,\omega}\left[\frac{1}{2}-p_{0,\omega}\right]\right)^{-1}=2^{n-l}. \qquad (10)$$

В соотношении (10) считаем, что значение $\exp 2\left(\theta_1-\theta_2-\theta_3\right)\approx 1$, так как $n>>1$. Учитывая, что

$$\frac{n^n}{2^n}\frac{1}{n^{2np_{0,\omega}}n^{n-2np_{0,\omega}}}=\frac{1}{2^n},$$

рассматриваемое уравнение можно привести к виду

$$2p_{0,\omega}\frac{2^{-n}}{p_{0,\omega}^{\,2np_{0,\omega}}\left[\frac{1}{2}-p_{0,\omega}\right]^{n-2np_{0,\omega}}}\left(4\pi np_{0,\omega}\left[\frac{1}{2}-p_{0,\omega}\right]\right)^{-1}=\frac{2^n}{2^l}.$$

Прологарифмировав данное соотношение и поделив на длину выборки n, получим уравнение, позволяющее найти решение в гипотетическом равенстве (9) в асимптотике:

$$\ln 2+p_{0,\omega}\ln p_{0,\omega}+\left[\frac{1}{2}-p_{0,\omega}\right]\ln\left[\frac{1}{2}-p_{0,\omega}\right]=0. \qquad (11)$$

Для нахождения решения в (9) запишем производную от его левой части

$$\frac{\partial}{\partial k_{0,\omega}}\left[\left(C_{0,5n}^{k_{0,\omega}}\right)^2\frac{2k_{0,\omega}}{n}\right]=\frac{2}{n}\left(C_{0,5n}^{k_{0,\omega}}\right)^2\left[1+2k_{0,\omega}\ln\frac{n-2k_{0,\omega}}{2k_{0,\omega}}-\frac{n-4k_{0,\omega}}{n-2k_{0,\omega}}\right].$$

При асимптотических значениях n и $k_{\omega}=k_{0,\omega}$ отношение

$$\frac{n-4k_{0,\omega}}{n-2k_{0,\omega}}\to 0,$$

так как $n-4k_{0,\omega}$ □ $n-2k_{0,\omega}$.

Таким образом, приравнивая производную к нулю, можно записать

$$1+2k_{0,\omega}\ln\frac{n-2k_{0,\omega}}{2k_{0,\omega}}=0. \qquad (12)$$

Из равенства (12) имеем

$$\ln\frac{n-2k_{0,\omega}}{2k_{0,\omega}}=-\frac{1}{2k_{0,\omega}}\ln e. \qquad (13)$$

Отсюда, учитывая затухание экспоненты при $k_{0,\omega} >> 1$, для $n \to \infty$ имеем

$$n - 2k_{0,\omega} = 2k_{0,\omega} e^{-\frac{1}{2k_{0,\omega}}} \to 2k_{0,\omega}.$$

Переходя от логарифма к самим статистическим параметрам, а также прибавляя к левой и правой частям полученного формального следствия из (13) единицу, имеем

$$1 + \frac{n - 2k_{0,\omega}}{2k_{0,\omega}} = 1 + e^{-\frac{1}{2k_{0,\omega}}} \text{ или } n = 2k_{0,\omega}\left(1 + e^{-\frac{1}{2k_{0,\omega}}}\right).$$

Тогда

$$k_{0,\omega} = \frac{n}{2}\left(1 + e^{-\frac{1}{2k_{0,\omega}}}\right)^{-1} \approx \frac{n}{4}. \quad (14)$$

Полученный результат является модой Mo_{ω} распределения и численно совпадает с вычисленным выше $M_{\omega} n = 0,25n$ математическим ожиданием (01)-событий в асимптотической последовательности. При этом проверка корня (14) подстановкой в уравнение (11) приводит к требуемому равенству левой и правой частей.

Определим границы преимущественного использования методов сигнатурного анализа и счета двоичных векторов переходов в асимптотике, применяя постановку задачи (9). При этом с учетом соотношения для биномиального коэффициента в центре распределения $C_{0,5n}^{k_{0,\omega}}$ [4, 5] для точки Mo_{ω} можно записать соотношение

$$C_{0,5n}^{k_{0,\omega}} = 2^{\frac{n}{2}} \frac{1}{\sqrt{0,25\pi n}}.$$

Соответственно, из (9) при асимптотическом $k_{0,\omega}$ следует

$$\left(C_{0,5n}^{k_{0,\omega}}\right)^2 \frac{2k_{0,\omega}}{n} = \frac{2^n}{0,5\pi n}.$$

С учетом соотношения для мощности классов эквивалентностей, порождаемых сигнатурным анализатором, из постановки задачи (9) имеем

$$\frac{2}{\pi}\frac{2^n}{n} \neq \frac{2^n}{n}.$$

Отсюда следует, что точка экстремума функции P_{wfc} находится ниже любой точки графика функции ошибок P_{sa} и

$$P_{wfc} \leq \frac{2}{\pi} P_{sa}. \quad (15)$$

Таким образом, $P_{wfc} < P_{sa}$ при всех теоретических k_{ω}. Теорема доказана.

Соотношение, аналогичное (9), сформируем и при сравнении методов счета двоичных переходов и счета состояний «0» или «1» при $n = \infty$

$$P_{wfc} \le \frac{1}{\sqrt{0,5\pi n}} P_{cvc}\,. \qquad (16)$$

Равенство (16) свидетельствует в пользу метода наблюдения (01)-переходов.

Заключение. Двоичные переходы могут отождествляться с событиями автокорреляционной функции в соотношении (4) заданного вида. При этом графики функций вероятностей пропуска ошибки, соответствующие наблюдению переходов и формированию линейных сигнатур, не имеют в асимптотике точек пересечения (15). Следовательно, метод наблюдения переходов в вероятностных процессах является более точным алгоритмом идентификации, чем сигнатурный анализ. Аналогичный вывод в пользу автокорреляционной функции специального вида для двух заданных событий может быть сделан и при сравнении исследуемого метода с алгоритмом наблюдения бернуллиевских 0 или 1 элементарных состояний (16).

Следует ожидать, что улучшение показателей, характеризующих методы идентификации двоичной выборки в системах шифрования и передачи данных, может быть достигнуто в рамках синтеза других более сложных событий на базе ряда заданных состояний процесса [6]. При этом следует учитывать динамику формирования отсчетов для различных сдвигов τ, а также возможно иные, формально не определенные на сегодняшний день свойства объектов с технической, информационной или другой дискретной природой.

Литература

1. ***Кобяк И.П.*** Сравнительная оценка достоверности методов сигнатурного анализа и счета состояний // Электронное моделирование. – 1996. – Т. 18. – № 1. – С. 58–62.

2. ***Кобяк И.П.*** Сравнительный анализ вероятностей пропуска ошибки при синтезе сигнатур и оценок числа векторов переходов // АВТ. – 2005. – № 6. – С. 60–68.

3. ***Кобяк И.П.*** Производящая функция для распределения вероятностей наблюдения векторов переходов // АВТ. – 2006. – № 6. – С. 60–67.

4. ***Риордан Дж.*** Комбинаторные тождества. – М.: Наука, 1982. – 255 с.

5. ***Стенли Р.*** Перечислительная комбинаторика. – М.: Мир, 1990. – 440 с.

6. ***Кобяк И.П.*** О границах вероятностных аргументов при синтезе линейных сигнатур и статистических аргументов // Информационные технологии и системы 2017: материалы междунар. науч. конф. (Республика Беларусь, Минск, 25 окт. 2017 г.). – Минск: БГУИР, 2017. – С. 216–217.

Поступила в редакцию 11.10.2018 г.; после доработки 10.01.2019 г.; принята к публикации 19.03.2019 г.

Кобяк Игорь Петрович – кандидат технических наук, доцент кафедры электронных вычислительных машин Белорусского государственного университета информатики и радиоэлектроники (Беларусь, 220600, г. Минск, ул. П. Бровки, д. 6), IPKobyak2012@mail.ru

References

1. Kobiak I.P. Comparative assessment of reliability of methods of the signature analysis and account of states. *Engineering Simulation = Elektronoye modelirovaniye*, 1996, vol. 18, no. 1, pp. 58–62. (In Russian).

2. Kobiak I.P. Comparative analysis of error missing probabilities by synthesis of signatures and estimates of transfer vector number. *AVT*, 2005, no. 6, pp. 60–68. (In Russian).

3. Kobiak I.P. Course-of-value function for probability distribution of wave front vectors. *AVT*, 2006, no. 6, pp. 60–67. (In Russian).

4. Riordan J. *Combinatory identities*. Moscow, Nauka Publ., 1982. 255 p. (In Russian).

5. Stenli R. *Enumerative combination theory*. Moscow, Mir Publ., 1990. 440 p. (In Russian).

6. Kobyak I.P. About borders of probabilistic arguments at synthesis of linear signatures and statistical arguments. Information technologies and systems 2017. *Proc. of the International scientific conference.* Minsk, 2017, pp. 216–217. (In Russian).

Received 11.10.2018; Revised 10.01.2019; Accepted 19.03.2019.

Information about the author:

Igor P. Kabiak – Cand. Sci. (Eng.), Assoc. Prof. of the Electronic Computing Machines Department, Belarusian State University of Informatics and Radioelectronics (Belarus, 220600, Minsk, P. Brovki st., 6), ipkobyak2012@mail.ru