http://dx.doi.org/10.35596/1729-7648-2022-20-6-52-60

Оригинальная статья / Original paper

УДК 004.415.533

МЕРА ОТЛИЧИЯ ДЛЯ УПРАВЛЯЕМЫХ ВЕРОЯТНОСТНЫХ ТЕСТОВ

В.Н. ЯРМОЛИК[1], Н.А. ШЕВЧЕНКО[2], В.В. ПЕТРОВСКАЯ[1]

[1]*Белорусский государственный университет информатики и радиоэлектроники (г. Минск, Республика Беларусь)*

[2]*Дармштадский технический университет (г. Дармштадт, Германия)*

Поступила в редакцию 26 марта 2022

Аннотация. Исследуется задача построения характеристик различия тестовых последовательностей. Обосновывается ее актуальность для генерирования управляемых вероятностных тестов и сложность нахождения мер отличия для символьных тестов. Показывается ограниченность применения расстояния Хэмминга и Дамерау – Левенштейна для получения меры отличия тестовых наборов. Для произвольного случая определяется новая мера различия двух символьных тестовых наборов на основе интервала, используемого в теории строя цепи последовательных событий. Расстояние $D(T_i, T_k)$ между тестовыми наборами T_i и T_k, использующее характеристику интервала, основано на определении независимых пар одинаковых (тождественных) символов, принадлежащих двум наборам, и вычислении интервалов между ними. Показывается комбинаторный характер вычисления предложенной меры отличия для символьных тестовых наборов произвольного алфавита и размерности. Приводится пример вычисления данной меры и показываются возможные ее модификации и ограничения. Рассматривается применение меры различия для случая многократного тестирования запоминающих устройств на основе адресных последовательностей pA с четным p повторением адресов. Для случая $p = 2$ приводятся математические соотношения вычисления интервалов и расстояния $D(T_i, T_k)$ для последовательностей адресов $2A$, используемых для управляемого вероятностного тестирования запоминающих устройств. Приводятся экспериментальные результаты, подтверждающие эффективность предложенной меры отличия.

Ключевые слова: расстояние, расстояние Хэмминга, расстояние Левенштейна, тест, тестовый набор.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Ярмолик В.Н., Шевченко Н.А., Петровская В.В. Мера отличия для управляемых вероятностных тестов. Доклады БГУИР. 2022; 20(6): 52-60.

DISTANCE MEASURE FOR CONTROLLED RANDOM TESTS

VYACHESLAV N. YARMOLIK[1], MIKALAI A. SHAUCHENKA[2], VITA V. PETROVSKAYA[1]

[1]*Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)*

[2]*Darmstadt Technical University* (*Darmstadt, Germany*)

Submitted 26 March 2022

Abstract. The problem of constructing characteristics of the difference between test sequences is investigated. Its relevance for generating controlled random tests and the complexity of finding difference measures for symbolic tests are substantiated. The limitations of using the Hamming and Damerau–Levenshtein distances to obtain a measure of the difference between test patterns are shown. For an arbitrary case, a new measure of the difference between two symbolic test sets is determined based on the interval used in the theory of the chain of successive events. The distance $D(T_i, T_k)$ between test patterns T_i and T_k, using the interval characteristic, is based on determining independent pairs of identical (equal) symbols belonging to two patterns and calculating the intervals between them. The combinatorial nature of the calculation, the proposed difference measure for symbolic test patterns of an arbitrary alphabet and dimension, is shown. An example of calculating this measure is given and its possible modifications and limitations are shown. The application of the measure of difference is considered for the case of multi-run testing of memory devices based on address sequences pA with even p repetition of addresses. For the case $p = 2$, mathematical relations are given for calculating intervals and distances $D(T_i, T_k)$ for address sequences $2A$ used for controlled random testing of memory devices. Experimental results are presented confirming the effectiveness of the proposed difference measure.

Keywords: distance, Hamming distance, Levenshtein distance, test, test pattern.

Conflict of interests. The authors declare no conflict of interests.

For citation. Yarmolik V.N., Shauchenka M.A., Petrovskaya V.V. Distance Measure for Controlled Random Tests. Doklady BGUIR. 2022; 20(6): 52-60.

Введение

Вероятностное тестирование (random testing) является традиционной технологией тестирования, основанной на методе черного ящика, которая не учитывает особенности объекта тестирования [1–3]. Формально вероятностный тест определяется количеством q тестовых наборов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$, $i \in \{0, 1, …, q-1\}$ данных $t_{i,j}$, $j \in \{0, 1, …, n-1\}$, их пространством, определяемым заданным алфавитом данных, и числом n в наборе. Чаще всего в качестве тестовых данных рассматриваются двоичные значения, когда тестовый набор $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ представляет собой n-разрядный двоичный вектор. Однако в общем случае T_i представляет собой последовательность из n символов алфавита, определяемого объектом тестирования. Для повышения эффективности вероятностных тестов используют их модификации, которые получили общее название «управляемые вероятностные тесты» (Controlled Random Tests) [1–3]. Под управляемыми вероятностными тестами понимают вероятностные тесты, в которых очередной тестовый набор формируется с учетом ранее сгенерированных наборов и которые соответствуют следующему определению [1, 2].

Определение 1. Управляемым вероятностным тестом $CRT = \{T_0, T_1, ..., T_{q-1}\}$ является тест, состоящий из q наборов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ тестовых данных $t_{i,j}$, $j \in \{0, 1, …, n-1\}$, сгенерированных случайным образом так, что очередной тестовый набор T_i удовлетворяет заданным критериям, полученным на основании ранее сгенерированных наборов $T_0, T_1, ..., T_{i-1}$.

Согласно определению 1 ключевой особенностью управляемого генерирования вероятностных тестовых наборов T_i является информация, которая извлекается в виде некоторых характеристик (метрик) из ранее сгенерированных тестовых наборов и используется для формирования очередного набора [1]. Основная идея управляемых вероятностных тестов заключается в том, что очередной тестовый набор T_i формируется максимально удаленным от ранее сгенерированных наборов $T_0, T_1,..., T_{i-1}$ в терминах заранее определенных мер отличия. Таким образом, принимается гипотеза, что для двух тестовых наборов T_i и T_k, имеющих минимальное различие, количество обнаруживаемых неисправностей будет минимальным и, наоборот, для максимально различных тестовых наборов обнаруживающая способность будет максимальной [1−3]. Задача управляемого вероятностного тестирования состоит в нахождении мер отличия для тестовых наборов T_i и T_k. Вычисление мер отличия символьных последовательностей, в свою очередь, сводится к задаче их сравнения.

Меры отличия символьных последовательностей

Проблема сравнения символьных последовательностей актуальна для различных областей науки. Основная трудность заключается в том, что в пространстве символьных последовательностей сложно ввести метрику отличия [4]. Формально подобная метрика в таком пространстве существует – это *расстояние Хэмминга* [2, 4]. Первоначально отметим, что расстояние Хэмминга $HD(T_i, T_k)$ между двумя наборами T_i и T_k равняется числу несовпадающих их компонент $t_{i,j}$ и $t_{k,j}$.

$$HD(T_i, T_k) = \sum_{j=0}^{n-1} I_{t_{i,j} \neq t_{k,j}}. \tag{1}$$

Выражение $I_{t_{i,j} \neq t_{k,j}}$ представляет собой индикаторную функцию, равную единице при $t_{i,j} \neq t_{k,j}$ и нулю в противном случае [5]. Содержательно эта метрика малопродуктивна, так как позволяет лишь различать полностью совпадающие последовательности при $HD(T_i, T_k) = 0$ и все остальные несовпадающие [4]. Наиболее распространенным в настоящее время методом сравнения символьных последовательностей является метод выравнивания, или *редакционного расстояния*, и его модификации [4, 5]. Чаще всего в литературе упоминается *расстояние Дамерау – Левенштейна* (*Damerau – Levenshtein distance*). Этот метод заключается в сравнении одной последовательности символов с другой с помощью операций вставки, замены (либо удаления) и транспозиции так, чтобы эти две последовательности совпали. Метод выравнивания требует выбора системы штрафных (весовых) функций и выбора опорной последовательности, относительно которой проводится выравнивание, кроме того, его эффективность снижается экспоненциально при увеличении длины строк символов [5].

В [6] предлагается подход, который предназначен для формального описания и анализа *строя*, представляющего собой последовательность данных (символов) любой природы. Под *строем* цепи событий (символов) понимают упорядоченное множество, в котором каждому символу цепи поставлено в соответствие натуральное число, причем идентичные по выбранному признаку компоненты отображены одним и тем же числом, т. е. нумеруются элементы собственного алфавита данной символьной последовательности по порядку их встречаемости [6]. Подобное отображение строя унифицирует его представление независимо от его алфавита и является необязательным атрибутом. Для определения строя отдельной цепи при обычном (естественном) способе ее чтения поэлементно подряд используется вторая нумерация, представляющая собой сквозную нумерацию всех компонентов строя от начала до конца, что, по сути, является их индексами. Пример строя для случая символьных данных $t_{i,j}$, который по своей структуре повторяет тестовый набор T_i, приведен на рис. 1.

V	**N**	**A**	**B**	**J**	**K**	**T**	**T**	**B**	**T**	**A**	**A**	**T**	**V**	**T**	**A**
1	2	3	4	5	6	7	7	4	7	3	3	7	1	7	3
0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*

Рис. 1. Пример строя цепи событий для случая символьных данных
Fig. 1. An example of order a chain of events for the case of character data

Важным результатом теории строя является формализация получения компактных числовых характеристик как результата разложения цепи, подобно используемым для описания случайных величин, которые являются необходимыми при идентификации строев цепей и определении степени их различия. В общем случае цепь может представлять собой последовательность символов любого алфавита, произвольной структуры и длины, а ее разложение осуществляться по различным правилам [6]. Исходной числовой характеристикой является *интервал*, определяемый как расстояние от выделенного в цепи символа до другого ближайшего, отмеченного в направлении просмотра, такого же символа. Величина интервала – это натуральное число, определенное как модуль разности индексов двух выделенных символов, а направление просмотра может быть справа налево, слева направо, циклическим либо произвольным [6]. Для примера строя, приведенного на рис. 1, при его циклическом анализе для символа **V** существует два интервала, равные 13 и 3, для символа **N** – один интервал величиной 16, а символ **A** имеет 4 интервала, величины которых равняются 8, 1, 4 и 3.

Идея использования характеристики интервала оказалась весьма продуктивной для получения числовых характеристик строя цепи событий любой природы [4−7]. Важно отметить, что структура строя и его особенности, определяемые природой его формирования, могут быть весьма различны. Обобщенной моделью строя является случайная и независимая последовательность символов произвольной длины из заданного алфавита. Однако во многих приложениях размерность строя и его алфавит фиксированы, так же как и сама их структура может характеризоваться разнообразными зависимостями. Примером могут быть тестовые последовательности, состоящие из тестовых наборов фиксированной длины. Для случая строя, представляющего собой тестовую последовательность адресов, характеристика интервала применялась для оценки эффективности адресных последовательностей с точки зрения обнаруживающей способности тестов запоминающих устройств [7].

Определение 2. Под последовательностью адресов понимают упорядоченную последовательность из 2^m m-битовых векторов $A(i) = a_{m-1}(i)\, a_{m-2}(i) \ldots a_0(i)$, $a_r(i) \in \{0, 1\}$, $r \in \{0, 1, \ldots, m-1\}$ и $i \in \{0, 1, \ldots, 2^m-1\}$, каждый из которых принимает одно из 2^m значений.

Данное определение можно интерпретировать как определение последовательности произвольных, отличающихся друг от друга символов, т. е. $A(i) \neq A(k)$, $i \neq k \in \{0, 1, \ldots, n\}$ фиксированной длины $n = 2^m$. Новым развитием адресных последовательностей являются pA последовательности для произвольного четного p, удовлетворяющие определению [7].

Определение 3. Адресной последовательностью pA, состоящей из $p2^m$ адресов, называется упорядоченная последовательность адресов $A(i) = a_{m-1}(i)\, a_{m-2}(i) \ldots a_0(i)$, $i \in \{0, 1, \ldots, p2^m-1\}$, где $a_r \in \{0, 1\}$ для $r \in \{0, 1, \ldots, m-1\}$, состоящая из всех возможных 2^m m-разрядных двоичных комбинаций $a_{m-1}(i)\, a_{m-2}(i) \ldots a_0(i)$, каждая из которых формируется ровно p раз.

Для тестовых последовательностей адресов, соответствующих определению 3, с использованием характеристики интервала, была определена метрика расстояния $D(A, pA)$ [7]. Показано, что в общем случае произвольная пара из возможных пар повторяющихся адресов $A \in \{0, 1, 2, \ldots, 2^m-1\}$ в последовательности pA имеет два значения метрики расстояния $D(A, pA)$, а именно r и $p2^m - r$. Для оценки каждого адреса A в последовательности pA использовалось минимальное значение расстояния $MD(A, pA)$ из всех возможных расстояний $D(A, pA)$. В качестве характеристики последовательности pA была введена метрика среднего значения $AD(pA)$ минимальных расстояний $MD(A, pA)$ между повторяющимися адресами A последовательности pA. Эта метрика для $p = 2$ определяется согласно выражению

$$AD(2A) = \frac{1}{2^m} \sum_{A=0}^{2^m-1} \min \left[D(A, 2A), (2^{m+1} - D(A, 2A)) \right]. \quad (2)$$

Рассмотренные метрики $MD(A, pA)$ и $AD(pA)$ для последовательностей pA являются их характеристиками, что позволило сформулировать условие максимальной эффективности теста *March_pA_2* запоминающих устройств в терминах указанных характеристик. Эти характеристики $MD(A, pA)$ и $AD(pA)$ идентифицируют последовательность pA, которая, по своей сути, представляет строй цепи событий. Эффективность этой характеристики показана в [7] для случая однократного тестирования запоминающих устройств.

Расстояние отличия для символьных тестовых наборов

При построении управляемых вероятностных тестов, как отмечалось ранее, очередной тестовый набор T_i формируется максимально отличающимся от ранее сгенерированных наборов. Соответственно, нахождение такого набора сводится к задаче сравнения двух тестовых наборов T_i и T_k. Для решения этой задачи введем метрику расстояния $D(T_i, T_k)$, основанную на характеристике интервала [6−7]. Первоначально, используя теорию строя, уточним определение 1 тестового набора T_i для произвольного случая. Будем считать, что тестовые данные $t_{i,j}$, $j \in \{0, 1, \ldots, n_i-1\}$ набора T_i принадлежат алфавиту с заданной мощностью L, а их количество $n_i > 0$ в наборе принимает произвольное значение из допустимого диапазона. Примером тестовых наборов T_i и T_k, в которых данные представлены символами английского алфавита (L = 26), являются два набора, приведенные на рис. 2. Отметим, что для T_i величина n_i равняется 9, а для T_k – n_k = 7.

T_i	**V**	**N**	**A**	**B**	**J**	**K**	**T**	**T**	**B**
j	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
T_k	**T**	**A**	**A**	**T**	**V**	**T**	**A**		
j	*0*	*1*	*2*	*3*	*4*	*5*	*6*		

Рис. 2. Примеры символьных тестовых наборов T_i и T_k
Fig. 2. Examples of symbolic test sets T_i and T_k

Мера отличия $D(T_i, T_k)$ тестовых наборов T_i и T_k, использующая характеристику интервала, основана на определении независимых пар одинаковых (тождественных) данных, принадлежащих двум наборам. Независимость пар означает участие каждого значения данных $t_{i,j}$ и $t_{k,j}$ тестовых наборов T_i и T_k только в одной паре. Процедура формирования подобных пар носит комбинаторный характер и заключается в нахождении такого их сочетания, для которого сумма их минимальных положительных разностей индексов (интервалов) также минимальна. Эти значения для всех данных наборов T_i и T_k суммируются, и в результате формируется численное значение расстояния $D(T_i, T_k)$. При отсутствии пары для очередного значения данных в наборе T_i разность величин индексов принимается равной $\min(n_i, n_k)$. Такое же значение задается и данным в наборе T_k, для которых отсутствует пара в T_i.

В упрощенном варианте получение сочетания пар одинаковых данных тестовых наборов T_i и T_k может быть реализовано путем анализа тестовых наборов, например, начиная с T_i, слева направо, справа налево или циклически. При нахождении очередной пары данные, участвующие в ней, не рассматриваются при дальнейшем поиске пар.

Минимальное значение метрики $D(T_i, T_k)$ определяется соотношением $\min D(T_i, T_k) = |n_i - n_k| \cdot \min(n_i, n_k)$, что свидетельствует о максимальном совпадении двух наборов T_i и T_k. При выполнении равенства $n_i = n_k$ значение $\min D(T_i, T_k)$ равняется нулю, что свидетельствует о тождественности сравниваемых наборов. Максимальное значение $D(T_i, T_k)$ определяется как $\max D(T_i, T_k) = (n_i + n_k) \cdot \min(n_i, n_k)$, что свидетельствует о полном (максимальном) отличии наборов T_i и T_k. Среднее значение $AD(T_i, T_k)$ определяется как взвешенная сумма интервалов $AD(T_i, T_k)$, аналогично соотношению (2), исследованному ранее [7].

Рассмотрим пример вычисления $D(T_i, T_k)$ для тестовых наборов T_i и T_k, приведенных на рис. 2, для которых $\min(n_i, n_k) = \min(9, 7) = 7$. Анализ данных в указанных наборах будем производить слева направо, начиная с набора T_i. Значение первого данного $t_{i,0}$ = **V** набора T_i равняется такому же данному $t_{k,4}$ набора T_k, образуя пару данных $(t_{i,0}, t_{k,4})$ с интервалом, равным $|0 - 4| = 4$, т. е. $(t_{i,0}, t_{k,4}) = 4$. Следующее значение данных $t_{i,1}$ не имеет пары $(t_{i,1}, -)$, так как в наборе T_k отсутствует значение **N**, тогда $(t_{i,1}, -) = \min(n_i, n_k) = 7$. Последовательно анализируя данные набора T_i, находим остальные пары и соответствующие им величины интервалов: $(t_{i,2}, t_{k,2}) = 0$; $(t_{i,3}, -) = 7$; $(t_{i,4}, -) = 7$; $(t_{i,5}, -) = 7$; $(t_{i,6}, t_{k,5}) = 1$; $(t_{i,7}, t_{k,0}) = 2$ и $(t_{i,8}, -) = 7$. На следующем шаге определяются данные набора T_k, не участвующие в ранее определенных парах данных, и для них задается значение интервала, равное $\min(n_i, n_k)$, т. е. $(t_{k,3}, -) = 7$; $(t_{k,1}, -) = 7$ и $(t_{k,6}, -) = 7$. Сумма значений всех полученных интервалов и определяет численное значение $D(T_i, T_k)$, которое в данном примере принимает значение 63, а $\max D(T_i, T_k) = 16{\cdot}7 = 112$. Соотношение величины $D(T_i, T_k) = 63$ с $\max D(T_i, T_k) = 112$ и $\min D(T_i, T_k) = 14$ показывает количественную степень отличия тестовых наборов T_i и T_k.

Метрика расстояния $D(T_i, T_k)$, описанная выше, определена для общего случая тестовых наборов T_i и T_k. При наличии ограничений на размерность наборов и их структуру расстояние $D(T_i, T_k)$ позволяет большую его формализацию. Например, при многократном тестировании памяти на принципах управляемого вероятностного тестирования важным является использование данной метрики $D(B, C)$, показывающей степень отличия тестовых адресных последовательностей B и C. Обе адресные последовательности соответствуют определению 2, а метрика расстояния $D(B, C)$ определяется согласно следующему соотношению:

$$D(B,C) = \sum_{i=0}^{2^m-1} \sum_{k=0}^{2^m-1} I_{B(i)=C(k)} \cdot \min\left[\,|i-k|,\, 2^m - |i-k|\,\right]. \tag{3}$$

Аналогично, как и в (1), выражение $I_{B(i)=C(k)}$ представляет собой индикаторную функцию, равную нулю при $B(i) \neq C(k)$ и единице в противном случае. Минимальное значение $D(B, C)$ равняется 0 при совпадении последовательностей B и C, а максимальное $\max D(B, C)$ для отличающихся последовательностей принимает значение 2^{2m-1}.

Для случая адресных последовательностей pB и pC, соответствующих определению 3, вычисление метрики $D(pB, pC)$ сопряжено с определением минимальной суммы минимальных расстояний для всех возможных пар одинаковых значений адресов $A \in \{0, 1, \ldots, 2^m–1\}$, входящих в pB и pC. Основная сложность определения данной метрики заключается в получении $p!$ сочетаний пар для каждого адреса в pB и pC. Для случая последовательностей $2B$ и $2C$ количество подобных сочетаний равняется двум. Рассмотрим задачу вычисления $D(2B, 2C)$ для случая последовательностей $2B$ и $2C$, представленных на рис. 3.

$2B=$	000	001	011	010	110	111	101	100	100	101	111	110	010	011	001	000
$=B(i)$	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$2C=$	110	111	101	100	101	000	001	011	010	000	010	011	100	110	001	111
$=C(k)$	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*

Рис. 3. Примеры тестовых последовательностей адресов $2B$ и $2C$
Fig. 3. Examples of test sequences for addresses $2B$ and $2C$

В последовательностях $2B$ и $2C$ каждый из 8 адресов повторяется дважды и каждый из них входит в два возможных сочетания пар адресов. Например, адрес 011 входит в последовательность $2B$ как два ее элемента $B(2)$ и $B(13)$, а в последовательность $2C$ соответственно, как $C(7)$ и $C(11)$. Таким образом, первое сочетание пар адресов состоит из двух пар $B(2)$, $C(7)$ и $B(13)$, $C(11)$, а второе сочетание – из пар $B(13)$, $C(7)$ и $B(2)$, $C(11)$.

В общем случае для каждого значения адреса $\{0, 1, \ldots, 2^m–1\}$ в адресных последовательностях $2B$ и $2C$ используются четыре одинаковых значения $B(i) = B(l) = C(k) = C(g)$, образующие два сочетания пар адресов. Соответственно, $\{(B(i), C(k)), (B(l), C(g))\}$ представляет собой первое сочетание, а $\{(B(l), C(k)), (B(i), C(g))\}$ – второе. Каждое сочетание пар адресов характеризуется суммой минимальных интервалов для каждой пары сочетания. Далее для каждого адреса $A \in \{0, 1, 2, \ldots, 2^m–1\}$ определяется минимальное значение суммы минимальных интервалов в соответствии с выражением:

$$D(A) = \min \begin{Bmatrix} \min[\,|i-k|, 2^{m+1}-|i-k|\,] + \min[\,|l-g|, 2^{m+1}-|l-g|\,], \\ \min[\,|l-k|, 2^{m+1}-|l-k|\,] + \min[\,|i-g|, 2^{m+1}-|i-g|\,] \end{Bmatrix}. \tag{4}$$

Значение 2^{m+1} в выражении (4) представляет собой размерность последовательностей $2B$ и $2C$, которая для примера, приведенного на рис. 3, равняется $2^{3+1} = 16$. Соотношение $\min[|i-k|, 2^{m+1}–|i-k|]$ в выражении (4) представляет собой расстояние $D(B(i), C(k))$ между адресами $B(i)$ и $C(k)$ при циклическом их анализе.

Как показывалось ранее, адрес A = 011 образует в последовательностях $2B$ и $2C$ два сочетания адресов $\{(B(2), C(7)), (B(13), C(11))\}$ и $\{(B(13), C(7)), (B(2), C(11))\}$. Соответствующие расстояния $D(B(i), C(k))$ между адресами пары $(B(i), C(k))$ первого сочетания принимают значения: $D(B(2), C(7)) = \min(|2-7|, 16–|2–7|) = \min(5, 11) = 5$; $D(B(13), C(11)) = \min(|13-11|, 16–|13–11|) = \min(2, 14) = 2$. Для второго сочетания получим: $D(B(13), C(7)) = \min(|13-7|, 16–|13–7|) = \min(6, 10) = 6$; $D(B(2), C(11)) = \min(|2-11|, 16–|2–11|) = \min(9, 7) = 7$. Далее для каждого сочетания, в соответствии с (4), вычисляется сумма полученных расстояний, которая для первого сочетания адресов 011 равняется $D(B(2), C(7)) + D(B(13), C(11)) = 5 + 2 = 7$, а для второго – $D(B(13), C(7)) + D(B(2), C(11)) = 6 + 7 = 13$. Окончательно для адреса 011, согласно (4), имеем значение $D(011) = 7$. Для адреса $A = 000$ существуют два сочетания $\{(B(0), C(5)), \{(B(15), C(9))\}$ и $\{(B(15), C(5)), \{(B(0), C(9))\}$ (см. рис. 3). Соответствующие интервалы для каждой из пар двух сочетаний принимают значения: $D(B(0), C(5)) = 5$; $D(B(15), C(9)) = 6$; $D(B(15), C(5)) = 6$; $D(B(0), C(9)) =7$. Тогда, в соответствии с (4), $D(000) = 11$.

В общем случае для последовательностей $2B$ и $2C$, соответствующих определению 3, метрика расстояния $D(2B, 2C)$, показывающая их отличие друг от друга, определяется как сумма сумм минимальных расстояний $D(A)$ (4) для всех значений адресов $A \in \{0, 1, \ldots, 2^m–1\}$.

Для рассмотренного выше примера значение $D(2B, 2C) = D(000) + D(001) + D(010) + D(011) + + D(100) + D(101) + D(110) + D(111) = 11 + 5 + 7 + 7 + 8 + 9 + 6 + 9 = 62$. Метрика расстояния $D(2B, 2C)$ для произвольных последовательностей $2B$ и $2C$, удовлетворяющих условиям определения 3, по аналогии со схожей метрикой, рассмотренной в [7], принимает значения в диапазоне $0 \le D(2B, 2C) \le 2^{2m}$. Для рассмотренного выше примера $D(2B, 2C) = 62$, а $\max D(2B, 2C) = 2^{2\cdot 3} = 64$, что свидетельствует о большой степени отличия адресных тестовых последовательностей $2B$ и $2C$, представленных на рис. 3.

Экспериментальные результаты

Для подтверждения эффективности предложенной меры отличия при реализации управляемого вероятностного тестирования был проведен ряд вычислительных экспериментов. В качестве объекта тестирования рассматривалось запоминающее устройство, в котором моделировалась кодочувствительная неисправность PNPSFk для $k = 5$. Во время эксперимента случайным образом задавалась неисправность PNPSF5, после чего применялся однократно, двукратно и трехкратно тест *MATS++* и фиксировалось обнаружение неисправности либо ее необнаружение. После проведения 1 000 000 таких экспериментов вычислялась полнота покрытия *FC* как процентное отношение количества обнаруженных тестом кодочувствительных неисправностей PNPSF5 к общему их числу. В качестве тестовой последовательности использовалась случайная адресная последовательность $A(i)$, удовлетворяющая определению 2, для $m = 3$ и 4. При повторных процедурах тестирования применялись модифицированные последовательности $A(i+s)$, представляющие собой сдвинутые на s позиций (индексов) копии исходной последовательности адресов $A(i)$. Для каждого сочетания адресных последовательностей вычислялось значение меры отличия $D(A(i), A(i+s))$. Экспериментальные результаты для двукратного теста *MATS++* представлены в табл. 1 и 2 для $m = 3$ и 4.

Таблица 1. Покрывающая способность *FC* двукратного применения теста *MATS++* для $m = 3$, %
Table 1. Coverage *FC* of two run *MATS++* tests for $m = 3$, %

s	0	1	2	3	4	5	6	7
$D(A(i), A(i+s))$	0	8	16	24	**32**	24	16	8
FC	6,20	10,1	11,8	12,3	**12,4**	12,3	11,8	10,1

Таблица 2. Покрывающая способность *FC* двукратного применения теста *MATS++* для $m = 4$, %
Table 2. Coverage *FC* of two run *MATS++* tests for $m = 4$, %

s	0	1	2	3	4	5	6	7
$D(A(i), A(i+s))$	0	16	32	48	64	80	96	112
FC	6,12	8,10	9,50	10,60	11,30	11,70	12,00	12,20
s	8	9	10	11	12	13	14	15
$D(A(i), A(i+s))$	**128**	112	96	80	64	48	32	16
FC	**12,33**	12,20	12,0	11,70	11,30	10,60	9,50	8,10

Из табл. 1, 2 видно, что наибольшая эффективность двукратного применения теста *MATS++* достигается в том случае, когда мера отличия $D(A(i), A(i+s))$ принимает максимальные значения, соответственно равные 32 и 128. При трехкратном применении теста *MATS++*, кроме исходной случайной адресной последовательности $A(i)$, удовлетворяющей определению 2, использовались две ее циклические копии $A(i+s_1)$ и $A(i+s_2)$, сдвинутые соответственно на s_1 и s_2 позиций относительно $A(i)$.

Как видно из табл. 3, максимальная полнота покрытия *FC* достигается при максимальном отличии каждой последовательности по отношению друг к другу. Действительно, максимальное значение 17,7 % обеспечивается для $A(i)$, $A(i+2)$ и $A(i+5)$, для которых $D(A(i), A(i+2)) = 16$, $D(A(i), A(i+5)) = 24$ и $D(A(i+2), A(i+5)) = 24$. Аналогично в двух остальных случаях максимального значения *FC* имеем: $D(A(i), A(i+3)) = 24$, $D(A(i), A(i+5)) = 24$ и $D(A(i+3), A(i+5)) = 16$; $D(A(i), A(i+3)) = 24$, $D(A(i), A(i+6)) = 16$ и $D(A(i+3), A(i+6)) = 24$.

Таблица 3. Покрывающая способность *FC* трехкратного применения теста *MATS*++ для $m = 3$, %
Table 3. Coverage *FC* of three run *MATS*++ tests for $m = 3$, %

s_1	1	1	1	1	1	1	2	2	2	2	2
s_2	2	3	4	5	6	7	3	4	5	6	7
FC	14,00	15,67	16,13	16,13	15,60	14,00	15,60	17,33	**17,70**	17,33	15,60
s_1	3	3	3	3	4	4	4	5	5	6	
s_2	4	5	6	7	5	6	7	6	7	7	
FC	16,20	**17,70**	**17,70**	16,20	16,20	17,30	16,20	15,60	15,60	14,00	

Для оценки эффективности предложенной меры отличия при управляемом тестировании запоминающих устройств с применением неразрушающих тестов *March*_2*A*_2 рассматривались неисправности PNPSF3. В качестве тестовых последовательностей для случая $m = 3$ применялись последовательности $2A(i)$, $2B(i)$ и $2C(i)$, соответствующие определению 3, приведенные в табл. 4 [7].

Таблица 4. Последовательности $2A(i)$, $2B(i)$ и $2C(i)$ для $m = 3$
Table 4. Sequences $2A(i)$, $2B(i)$ and $2C(i)$ for $m = 3$

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$2A(i)$	000	001	010	011	100	101	110	111	000	001	010	011	100	101	110	111
$2B(i)$	100	101	110	111	000	001	010	011	100	101	110	111	000	001	010	011
$2C(i)$	001	000	011	010	101	100	111	110	001	000	011	010	101	100	111	110

Первоначально использовалась тестовая последовательность $2A(i)$, которая обеспечивала полноту покрытия 50 % для неисправностей PNPSF3. Повторное тестирование с применением тестовой адресной последовательности $2B(i)$, для которой $D(2A(i), 2B(i)) = 64$, увеличило полноту покрытия *FC* неисправностей PNPSF3 до 78,57 %. В то же время использование в качестве второй адресной последовательности $2C(i)$ с минимальным значением меры отличия $D(2A(i), 2C(i)) = 16$ обеспечило только 64,29 % покрытия.

Результаты, приведенные в табл. 5, для больших размеров памяти $m = 8$ и одинаковых по структуре последовательностей $2A(i)$ и $2B(i)$ подтверждают эффективность предложенной меры отличия для управляемого на ее основе тестирования запоминающих устройств.

Таблица 5. Покрывающая способность *FC* двукратного применения теста *March*_2*A*_2 для $m = 8$, %
Table. 5. Coverage *FC* of two run *March*_2*A*_2 tests for $m = 8$, %

$A(i)$, $B(i)$	*Counter*				*Gray Code*		*Pseudorandom*			
$D(A(i), B(i))$	768	7936	32512	65280	512	65536	33082	33272	40846	44566
FC	25,63	30,96	45,54	58,54	25,5	49,46	58,21	58,25	58,42	59,04

Как видно из табл. 5, во всех парах $2A(i)$, $2B(i)$ адресных последовательностей, а именно счетчиковых (*Counter*), кода Грея (*Gray Code*) и псевдослучайных (*Pseudorandom*), для бóльших значений новой меры отличия обеспечивается большее значение *FC* для PNPSF3.

Заключение

Предложена мера отличия для построения управляемых вероятностных тестов, основанная на применении интервала, используемого в теории строя цепи последовательных событий. Показана ее эффективность для случая тестирования запоминающих устройств при применении различных видов тестовых наборов управляемых вероятностных тестов. Дальнейшие исследования целесообразно расширить в части свойств новой меры отличия и ее применимости для различных прикладных задач. Наиболее интересным является применение данной меры отличия в современных поисковых приложениях и оценка вычислительной сложности ее получения.

Список литературы / References

1. Huang R., Sun W., Xu Y., Chen H., Towey D., Xia X. A Survey on Adaptive Random Testing. *IEEE Transactions on Software Engineering*. 2021;47(10):2052-2083. DOI: 10.1109/tse.2019.2942921.
2. Yarmolik V.N, Mrozek I., Yarmolik S.V. Controlled Method of Random Test Synthesis. *Automatic Control and Computer Sciences*. 2015;49(6):395-403.
3. Леванцевич В.А., Ярмолик В.Н. Многократное управляемое вероятностное тестирование. *Доклады БГУИР*. 2019;121(3):65-69. / Levantsevich V.A., Yarmolik V.N. [Multiple controlled random testing]. *Doklady BGUIR = Doklady BGUIR*. 2019;121(3):65-69. (in Russ.)
4. Sadovsky M.G. Comparison of symbol sequences: no editing, no alignment. *Open Systems & Information Dynamics*. 2002;09(01):19-36. DOI: 10.1023/A:1014278811727.
5. Bard G.V. Spelling-error tolerant, order-independent pass-phrases via the Damerau – Levenshtein string-edit distance metric. *Proc. of the Fifth Australasian Symposium on ACSW Frontiers.* 2007:117-124.
6. Gumenyuk A.S., Skiba A.A., Pozdnichenko N.N., Shpynov S.N. About Similarity Measures of Components Arrangement of Naturally Ordered Data Arrays. *SPIIRAS Proceedings*. 2019;18(2):471-503. DOI: 10.15622/sp.18.2.471-503. / Gumenjuk A.S, Skiba A.A., Pozdnichenko N.N., Shpunov S.N. [On the measures of similarity of the arrangement of components in arrays of naturally ordered data]. *Proc. SPIIRAS.* 2019;18(2):471-503. DOI: 10.15622/sp.18.2.471-503. (In Russ.)
7. Mrozek I., Yarmolik V.N. Transparent Memory Tests Based on the Double Address Sequences. *Entropy*. 2021;23:894. DOI: 10.3390/e23070894.

Вклад авторов

Ярмолик В.Н. предложил меру отличия для управляемых вероятностных тестов.
Шевченко Н.А. принял участие в обобщении результатов и проведении экспериментов.
Петровская В.В. приняла участие в анализе результатов и проведении экспериментов.

Authors' contribution

Yarmolik V.N. proposed a distance measure for controlled random tests.
Shauchenka M.A. took part in the generalization of the results and conduct of experiment.
Petrovskaya V.V. took part in the analysis of the results and experiments.

Сведения об авторах

Ярмолик В.Н., д.т.н., профессор Белорусского государственного университета информатики и радиоэлектроники.

Шевченко Н.А., студент Дармштадского технического университета.

Петровская В.В., магистр технических наук Белорусского государственного университета информатики и радиоэлектроники.

Information about the authors

Yarmolik V.N., Dr. of Sci. (Eng.), Professor at the Belarusian State University of Informatics and Radioelectronics.

Shauchenka M.A., Student at the Darmstadt Technical University.

Petrovskaya V.V., M. Sci. at the Belarusian State University of Informatics and Radioelectronics.

Адрес для корреспонденции

220013, Республика Беларусь,
г. Минск, ул. П. Бровки, 6,
Белорусский государственный университет
информатики и радиоэлектроники;
тел. +375 29 769-96-77;
e-mail: yarmolik10ru@yahoo.com
Ярмолик Вячеслав Николаевич

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovka St., 6,
Belarusian State University
of Informatics and Radioelectronics;
tel. +375 29 769-96-77;
e-mail: yarmolik10ru@yahoo.com
Yarmolik Vyacheslav Nikolaevich