ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ
LOGICAL DESIGN

УДК 004.33.054
https://doi.org/10.37661/1816-0301-2024-21-2-54-72

Оригинальная статья
Original Article

Меры различия, основанные на применении расстояния Хэмминга, для генерирования управляемых вероятностных тестов

В. Н. Ярмолик[1]✉, В. В. Петровская[1], Н. А. Шевченко[2]

[1]Белорусский государственный университет информатики и радиоэлектроники, ул. П. Бровки, 6, Минск, 220013, Беларусь
✉E-mail: yarmolik10ru@yahoo.com

[2]Дармштадтский технический университет, Каролиненплац, 5, Дармштадт, 64289, Германия

Аннотация

Цели. Решается задача построения мер различия, основанных на применении расстояния Хэмминга, для генерирования управляемых вероятностных двоичных тестовых наборов. Целью настоящей статьи является развитие методов определения расстояния Хэмминга для нахождения различия между тестовыми наборами при их совпадении по оценкам других мер различия.

Методы. На базе расстояния Хэмминга, используемого в теории и практике формирования управляемых вероятностных тестов, предлагаются новые меры различия для сравнения двух двоичных n-разрядных тестовых наборов. Основой предлагаемых мер различия является формирование множества расстояний Хэмминга для исходных наборов, представляемых в виде последовательностей символов различных алфавитов.

Результаты. Показывается неразличимость пар двоичных тестовых наборов при использовании меры различия, основанной на применении расстояния Хэмминга. В этом случае отличающиеся пары наборов могут иметь совпадающие значения расстояния Хэмминга. Для построения новых мер различия исходные двоичные тестовые наборы представляются в виде последовательностей, состоящих из символов, принадлежащих различным алфавитам. Предлагаются различные стратегии применения новых мер различия, основанных на использовании одного из трех правил, при генерировании управляемых вероятностных тестов. Показано, что во всех трех случаях новых мер различия информативными являются только несколько первых их компонент, как правило, не более двух или трех. Соответственно, вычислительная сложность для всех трех вариантов сравнима и не превышает $3n$ операций сравнения. Проведенные экспериментальные исследования подтверждают эффективность предложенных мер различия и их невысокую вычислительную сложность.

Заключение. Предложенные меры различия расширяют возможности генерирования тестовых наборов при формировании управляемых вероятностных тестов. Показывается, что тестовые наборы, неразличимые при использовании в качестве меры различия расстояния Хэмминга, имеют отличающиеся значения предложенных мер различия. Это позволяет более точно классифицировать формируемые случайным образом наборы, которые являются кандидатами в тестовые наборы.

Ключевые слова: тестирование вычислительных систем, управляемые вероятностные тесты, двоичный тестовый набор, мера различия символьных наборов, расстояние Хэмминга

Для цитирования. Ярмолик, В. Н. Меры различия, основанные на применении расстояния Хэмминга, для генерирования управляемых вероятностных тестов / В. Н. Ярмолик, В. В. Петровская, Н. А. Шевченко // Информатика. – 2024. − Т. 21, № 2. – С. 54–72.
https://doi.org/10.37661/1816-0301-2024-21-2-54-72

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 25.03.2024
Подписана в печать | Accepted 15.04.2024
Опубликована | Published 28.06.2024

Dissimilarity measures based on the application of Hamming distance to generate controlled probabilistic tests

Vyacheslav N. Yarmolik[1]✉, Vita V. Petrovskaya[1], Nikolai A. Shevchenko[2]

[1]Belarusian State University of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220013, Belarus
✉E-mail: yarmolik10ru@yahoo.com

[2]Technical University of Darmstadt,
Karolinenplatz, 5, Darmstadt, 64289, Germany

Abstract.
Objectives. The problem of constructing dissimilarity measures based on the application of the Hamming distance to generate controlled random binary test sets is solved. The main goal of this article is to develop methods for determining the Hamming distance for the achievability of finding the difference between test sets when they coincide according to estimates of other difference measures.
Methods. Based on the Hamming distance used in the theory and practice of generating controlled random tests, new dissimilarity measures are proposed for two binary test *n*-bit patterns. The basis of the proposed dissimilarity measures is the formation of sets of Hamming distances for initial sets, represented as sequences of characters from different alphabets.
Results. The indistinguishability of pairs of binary test sets T_i and T_k is shown using a dissimilarity measure based on the application of the Hamming distance. In this case, different pairs of sets may have identical Hamming distance values. To construct new measures of difference, the original binary test sequences are represented as sequences consisting of characters belonging to different alphabets. Various strategies are proposed for applying new measures of difference based on the use of one of three rules in generating controlled probability tests. It is shown that in all three cases of dissimilarity measures, only the first few of their components areinformative, as a rule, no more than two or three. Accordingly, the computational complexity for all three options is comparable and does not exceed $3n$ comparison operations. The experimental studies carried out confirm the effectiveness of the proposed dissimilarity measures and their low computational complexity.
Conclusion. The proposed dissimilarity measures expand the possibilities of generating test sets when forming controlled random tests. It is shown that test sets that are indistinguishable when using the Hamming distance as a dissimilarity measure have different values of the proposed dissimilarity measures, which makes it possible to more accurately classify randomly generated sets that are candidate test cases.

Keywords: computer systems testing, controlled random tests, binary test patterns, character patterns difference measure, Hamming distance

For citation. Yarmolik V. N., Petrovskaya V. V., Shevchenko N. A. *Dissimilarity measures based on the application of Hamming distance to generate controlled probabilistic tests*. Informatika [*Informatics*], 2024, vol. 21, no. 2, pp. 54−72 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-2-54-72

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

Введение. Традиционно тестирование современных вычислительных систем заключается в использовании тестовых наборов, сформированных случайным образом из множества всех возможных входных данных. Подобная процедура тестирования систем называется вероятностным тестированием (Random Testing) [1–3]. Вероятностное тестирование является основополагающей технологией тестирования, основанной на методе черного ящика, которая, как правило, не учитывает особенности тестируемого объекта [3–5]. Вероятностный тест задается количеством q тестовых наборов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$, $i \in \{0, 1, \ldots, q-1\}$, данных $t_{i,j}$, $j \in \{0, 1, \ldots, n-1\}$, из заданного алфавита данных и их числом n в наборе. С целью увеличения эффективности вероятностных тестов используются их модификации, так называемые управляемые вероятностные тесты (Controlled Random Tests) [5–8]. Множество управляемых вероятностных тестов включает различные их разновидности, которые в англоязычной литературе встречаются под названиями быстрое антивероятностное тестирование (Fast Antirandom Testing, FAR), адаптивное вероятностное тестирование (Adaptive Random Testing), эволюционное вероятностное тестирование (Evolutionary Random Testing), эффективное вероятностное тестирование (Good Random Testing), ограниченное вероятностное тестирование (Restricted Random Testing), зеркальное вероятностное тестирование (Mirror Random Testing), упорядоченное вероятностное тестирование (Orderly Random Testing), гибридное адаптивное вероятностное тестирование (Hybrid Adaptive Random Testing), расширенное адаптивное вероятностное тестирование (Enhanced Adaptive Random Testing) и др. [8–10].

Процедура управляемого генерирования вероятностных тестовых наборов основывается на информации, которая извлекается в виде некоторых характеристик (мер) из ранее сгенерированных тестовых наборов и используется для формирования следующего набора [6]. Очередной тестовый набор T_i управляемого вероятностного теста формируется максимально удаленным (различным) от ранее сгенерированных наборов T_0, T_1,..., T_{i-1} в терминах заранее выбранных мер различия. Таким образом, принимается гипотеза, что для двух тестовых наборов T_i и T_k, имеющих максимальное отличие, количество обнаруживаемых неисправностей будет максимальным [5–10]. Главная проблема управляемого вероятностного тестирования состоит в нахождении численных значений мер различия для тестовых наборов T_i и T_k. Вычисление мер различия для символьных тестовых последовательностей, в свою очередь, сводится к задаче их сравнения [11].

На сегодня широко известны следующие меры сравнения для символьных последовательностей: расстояние Хэмминга [12], расстояние Левенштейна [13], расстояние Дамерау – Левенштейна [14], сходство Джаро – Винклера [15], метрика Нидлмана – Вунша [16], метрика Смита – Вотермана [17] и др. Отметим, что все известные авторам алгоритмы, применяемые для решения задачи определения расстояния между последовательностями символов, т. е. их отличия, так или иначе основаны именно на этих метриках. Расстояние Хэмминга является одной из универсальных мер близости последовательностей символов одинаковой размерности.

При формировании очередного тестового набора первоначально генерируется фиксированное количество кандидатов в тесты, представляющих собой, как правило, равномерно распределенные случайные наборы [6, 10, 18]. Для каждого из них вычисляются метрики различия, с учетом которых и выбирается наилучший из кандидатов в тесты в качестве очередного тестового набора T_i [6, 18–20].

Главная задача управляемого вероятностного тестирования состоит в нахождении меры различия для тестовых наборов T_i и T_k, которая максимально адекватно показывает их отличие и характеризуется невысокой вычислительной сложностью [5, 6, 19]. Вычисление мер различия тестовых наборов, в общем случае представляющих собой символьные последовательности, в свою очередь, сводится к задаче их сравнения. Большинство известных подходов генерирования управляемых вероятностных тестов основано на применении расстояния Хэмминга [4–8]. Это позволяет значительно уменьшить вычислительную сложность формирования тестов, что достигается за счет использования достаточно общей и зачастую неточной меры различия. Более точные меры различия символьных последовательностей имеют существенно большую вычислительную сложность [19–21].

Результаты, представленные в настоящей работе, направлены на решение задачи поиска новых эффективных мер различия тестовых наборов при формировании управляемых вероятностных тестов. Основываясь на применении в качестве меры различия расстояния Хэмминга, авторы предлагают ряд модификаций его определения в качестве мер различия, используемых для генерирования управляемых вероятностных тестов.

1. Мера различия для построения управляемых вероятностных тестов. Как уже отмечалось, формально мера различия последовательностей символов существует – это расстояние Хэмминга $D(T_i, T_k)$. Расстояние Хэмминга между последовательностями $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ и $T_k = t_{k,0}, t_{k,1}, ..., t_{k,n-1}$, включающими по n символов, равняется числу их несовпадающих значений $t_{i,j}$ и $t_{k,j}$ [11, 12]:

$$D(T_i, T_k) = \sum_{j=0}^{n-1} \mathrm{I}(t_{i,j} \neq t_{k,j}). \quad (1)$$

Выражение $\mathrm{I}(t_{i,j} \neq t_{k,j})$ представляет собой индикаторную функцию, равную единице при $t_{i,j} \neq t_{k,j}$ и нулю в противном случае. Минимальное значение $\min D(T_i, T_k)$ равняется нулю при совпадении всех символов последовательностей T_i и T_k, а максимальное $\max D(T_i, T_k) = n$ – при несовпадении всех n символов.

Расстояние Хэмминга как метрика малоэффективно, так как позволяет различать лишь полностью совпадающие последовательности при $D(T_i, T_k) = 0$ и все остальные несовпадающие [11]. Аргументом для подтверждения неразличимости несовпадающих последовательностей являются наборы двоичных символов T_i и $T_k = \overline{T_i}$, расстояние Хэмминга $D(T_i, \overline{T_i})$ для которых всегда неизменно и равняется n, например D(11111110, 00000001) = D(10101000, 01010111) = D(11001100, 00110011) = 8. Видно, что расстояние Хэмминга $D(T_i, T_k)$ во всех рассмотренных выше примерах равняется восьми. Это свидетельствует об одинаковом максимальном отличии T_i от $\overline{T_i}$ во всех рассмотренных парах наборов, хотя структуры пар последовательностей существенно отличаются. Еще большим отличием характеризуются пары последовательностей символов T_i, T_k: 00000000, 11110000; 11111111, 00001111 и 01011010, 11010100, для каждой из которых $D(T_i, T_k) = 4$. Приведенные примеры показывают необходимость использования новых более эффективных мер сравнения последовательностей символов, позволяющих более полно оценивать схожесть их структур.

Исследуем расширение возможностей применения расстояния Хэмминга для сравнения конечных последовательностей символов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ и $T_k = t_{k,0}, t_{k,1}, ..., t_{k,n-1}$ как объектов, представляющих упорядоченные тестовые наборы T_i и T_k и состоящих из n символов (элементов) $t_{i,j}$ и $t_{k,j}$, где $j \in \{0, 1, …, n-1\}$. Алфавит символов $t_{i,j}$ и $t_{k,j}$ может быть произвольным, так же как и их количество n в наборах T_i и T_k.

Не нарушая общности дальнейшего изложения, предположим, что тестовый набор T_i является двоичным, т. е. символы $t_{i,j} \in \{0, 1\}$ для $j \in \{0, 1, …, n-1\}$, а $n = 2^w$, где w – целое. Отметим, что приведенные ограничения для T_i достаточно часто выполняются на практике при решении задач тестового диагностирования современных вычислительных систем. Такие ограничения позволяют сформировать $w + 1$ отображений исходной двоичной последовательности $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ в виде $T_i(0) = t_{i,0}(0), t_{i,1}(0), ..., t_{i,n-1}(0)$; $T_i(1) = t_{i,0}(1), t_{i,1}(1), ..., t_{i,n/2-1}(1)$; $T_i(2) = t_{i,0}(2), t_{i,1}(2), ..., t_{i,n/4-1}(2)$; $T_i(3) = t_{i,0}(3), t_{i,1}(3), ..., t_{i,n/8-1}(3)$; …; $T_i(w-1) = t_{i,0}(w-1), t_{i,1}(w-1)$ и $T_i(w) = t_{i,0}(w) = T_i$. В общем случае $T_i(v)$ состоит из 2^{w-v} символов, принадлежащих алфавиту, включающему (2^{2^v}) символов, каждый из которых получен на основании двух символов последовательности $T_i(v-1)$ для $v \in \{0, 1, 2, …, w\}$. Соответственно, для $T_i(1)$ имеем $t_{i,0}(1) = t_{i,0}(0), t_{i,1}(0)$; $t_{i,1}(1) = t_{i,2}(0), t_{i,3}(0)$; …; $t_{i,n/2-1}(1) = t_{i,n-2}(0), t_{i,n-1}(0)$, а для остальных значений $T_i(v) - t_{i,j}(v) = t_{i,2j}(v-1), t_{i,2j+1}(v-1)$, где $j = 0, 1, 2, …, n/2^v-1$. Видно, что каждая из последовательностей $T_i(1)$, $T_i(2)$, $T_i(3)$, …, $T_i(w)$ состоит из символов, алфавиты которых различны. Так, алфавит символов последовательности $T_i(1)$ включает четыре (2^{2^1}) символа, а алфавит последовательности

$T_i(2)$ – 16 (2^{2^2}) символов и так далее вплоть до последовательности $T_i(w)$, которая состоит из одного символа, принадлежащего алфавиту, включающему (2^{2^w}) символов.

Пример 1. В качестве примера двоичных тестовых наборов рассмотрим T_i = 01100011 и T_k = 01011011, для которых выполняется условие $n = 8 = 2^3$. Для каждого набора двоичных символов $T_i = 01100011_{(2)}$ и $T_k = 01011011_{(2)}$ в соответствии с приведенными выше определениями существует $w + 1 = 4$ их представлений в виде последовательностей символов, принадлежащих различным алфавитам. Следовательно, $T_i(0) = 01100011_{(2)}$; $T_i(1) = 1203_{(4)}$; $T_i(2) = 63_{(16)}$; $T_i(3) = \text{c}_{(256)}$ и $T_k(0) = 01011011_{(2)}$; $T_k(1) = 1123_{(4)}$; $T_k(2) = 5\text{B}_{(16)}$; $T_k(3) = [_{(256)}$. Видно, что $T_i(0)$ и $T_k(0)$ представлены в виде наборов двоичных символов 0 и 1, а $T_i(1)$ и $T_k(1)$ состоят из символов четверичного алфавита, включающего символы 0, 1, 2, 3; $T_i(2)$ и $T_k(2)$ используют шестнадцатеричный алфавит; $T_i(3)$ и $T_k(3)$ включают по одному символу из алфавита, состоящего из 2^8 символов. В данном примере для представления $T_i(3)$ и $T_k(3)$ использованы символы кодов ASCII.

Следует отметить, что предложенная интерпретация последовательностей двоичных символов T_i и T_k в виде последовательностей $T_i(0)$, $T_i(1)$, $T_i(2)$, …, $T_i(w)$ и $T_k(0)$, $T_k(1)$, $T_k(2)$, …, $T_k(w)$ позволяет применять для их сравнения расстояние Хэмминга (1). При этом рассматривается пара символов из двух наборов одного и того же алфавита, причем каждый набор состоит из одинакового количества символов. Основываясь на предыдущем примере (T_i = 01100011 и T_k = 01011011) и используя (1), получим D(01100011, 01011011) = 3, D(1203, 1123) = 2, D(63, 5B) = 2 и D(c, [) = 1.

Приведенный пример определения расстояния Хэмминга показывает возможность получения на основании равенства (1) нескольких численных оценок соотношения исходных наборов двоичных символов. Определим новую меру различия двоичных тестовых наборов T_i и T_k, которая состоит из множества численных характеристик, представляющих собой расстояния Хэмминга.

Определение 1. Мера различия $HD(T_i, T_k)$ двоичных тестовых наборов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ и $T_k = t_{k,0}, t_{k,1}, ..., t_{k,n-1}$, где $t_{i,j}, t_{k,j} \in \{0, 1\}$; $j \in \{0, 1, …, n-1\}$ и $n = 2^w$, а w – целое, состоит из (w + 1) значений расстояния Хэмминга (1) $HD_0, HD_1, ..., HD_v, …, HD_w$, полученных для T_i и T_k, которые представлены символами, задаваемыми их 2^v последовательными битами, где $v \in \{0, 1, …, w\}$.

Анализируемые символы $t_{i,j}$ и $t_{k,j}$ тестовых наборов T_i и T_k согласно определению 1 представляются одним, двумя, четырьмя и так далее вплоть до $n = 2^w$ двоичными битами. Соответственно, применяя равенство (1), формируем численные значения компонент $HD_0, HD_1, HD_2, …, HD_w$ меры различия $HD(T_i, T_k)$. В табл. 1 приведены примеры вычисления меры различия для различных пар тестовых наборов T_i и T_k в случае, когда $n = 2^3 = 8$.

Таблица 1
Численные значения компонент HD_0, HD_1, HD_2, HD_3 меры различия $HD(T_i, T_k)$ для тестовых наборов T_i и T_k

Table 1
Numerical values of the difference measure $HD(T_i, T_k)$ components HD_0, HD_1, HD_2, HD_3 for test patterns T_i and T_k

T_i, T_k		HD_0	T_i, T_k		HD_1	T_i, T_k		HD_2	T_i, T_k		HD_3
T_i	00000000	–	T_i	0000	–	T_i	00	–	T_i	0	–
T_k	11110000	4	T_k	3300	2	T_k	F0	1	T_k	р	1
	00110011	4		0303	2		33	2		3	1
	11100010	4		3202	3		E2	2		в	1
	01010110	4		1112	4		56	2		V	1

Например, для T_i = 00000000 и T_k = 00110011 значения компонент $HD(T_i, T_k)$ сформированы путем сравнения их различий согласно (1): сначала по одному биту, соответственно, имеем HD_0(00000000, 00110011) = 4, затем по двум битам – HD_1(0000, 0303) = 2, по четырем битам –

$HD_2(00, 33) = 2$, по восьми битам – $HD_3(0, 3) = 1$. Отметим, что для вычисления HD_0 использовался двоичный алфавит, для HD_1 и HD_2 – четверичный и шестнадцатеричный алфавиты, а для HD_3 – алфавит расширенных символов кодов ASCII.

Приведенные в табл. 1 примеры показывают неразличимость всех четырех наборов T_k по отношению к набору T_i при использовании классической меры различия – расстояния Хэмминга, так как во всех четырех случаях $HD_0 = 4$. В то же время новая мера различия (см. определение 1) показывает разную степень различия наборов T_k от T_i, выраженную в отличающихся величинах компонент HD_1 и HD_2 указанной меры (табл. 1).

Мера различия $HD(T_i, T_k)$ двоичных тестовых наборов T_i и T_k, соответствующая определению 1, имеет следующие очевидные свойства.

Свойство 1. *Минимальное численное значение компонент HD_0, HD_1, HD_2, …, HD_{w-2}, HD_{w-1}, HD_w меры различия $HD(T_i, T_k)$ равняется 0, а максимальные значения определяются количеством символов в сравниваемых наборах и равняются n, n/2, n/4, …, 4, 2, 1.*

Все компоненты HD_0, HD_1, HD_2, …, HD_w меры различия равняются 0 при совпадении тестовых наборов, т. е. когда $T_k = T_i$, а максимальные их значения достигаются для случаев, когда $T_k = \overline{T_i}$. Отметим, что $n/2^w = n/n = 1$. Соответственно, компонента HD_w принимает только два значения, а именно 0 при $T_k = T_i$ и 1 при $T_k \neq T_i$.

Свойство 2. *Численные значения компонент меры различия $HD(T_i, T_k)$ связаны соотношениями $HD_0 \geq HD_1 \geq ... \geq HD_v \geq … \geq HD_w$.*

Выполнение данного свойства объясняется тем, что при вычислении HD_{v+1} используются символы наборов $T_i(v+1)$ и $T_k(v+1)$, каждый из которых состоит их двух символов наборов $T_i(v)$ и $T_k(v)$, т. е. $t_{i,j}(v+1) = t_{i,2j}(v), t_{i,2j+1}(v)$ и $t_{k,j}(v+1) = t_{k,2j}(v), t_{k,2j+1}(v)$, где $j = 0, 1, 2, …, n/2^v–1$. Таким образом, результатом сравнения $\mathrm{I}(t_{i,j}(v+1) \neq t_{k,j}(v+1))$ символов наборов $T_i(v+1)$ и $T_k(v+1)$ могут быть только значения 0 или 1, используемые для получения HD_{v+1} (1), а результатом сравнения двух пар символов последовательностей $T_i(v)$ и $T_k(v)$ могут быть 0, 1 и 2. Следует отметить, что при выполнении неравенства $t_{i,j}(v+1) \neq t_{k,j}(v+1)$ значение HD_{v+1} (1) увеличится только на 1, а HD_v может быть увеличено как на 1, так и на 2, но как минимум на 1.

Из свойства 2 следует, что, как уже отмечалось ранее, при выполнении равенства $HD_0 = 0$ все остальные компоненты меры различия HD_1, HD_2, …, HD_w также равняются 0. Кроме того, при $HD_0 \neq 0$ значения всех остальных компонент меры различия также принимают ненулевые значения. Например, это видно в случае $T_i = 00000000$ и $T_k = 10000000$, для которых $HD_0 = 1$, $HD_1 = 1$, $HD_2 = 1$ и $HD_3 = 1$.

Как упоминалось в работах [5, 6], идея управляемых вероятностных тестов заключается в том, что очередной тестовый набор T_i формируется максимально отличным (удаленным) от ранее сгенерированных наборов T_0, T_1,..., T_{i-1} в терминах заранее определенных мер различия. Для этого на каждом шаге формирования очередного тестового набора осуществляется его выбор из множества кандидатов в тестовые наборы [5, 6, 10, 18]. Основная операция процедуры выбора заключается в определении численного значения используемой меры различия между двумя наборами T_i и T_k, один из которых, например первый, является тестовым набором, а другой – одним из кандидатов в тесты. В результате в качестве очередного тестового набора выбирается тот кандидат в тестовый набор, для которого величины мер (меры) различия принимают максимальные значения.

Поясним процедуру генерирования управляемого вероятностного теста на простейшем примере, представленном в табл. 1, для случая использования расстояния Хэмминга в качестве меры различия.

Предположив, что первым набором управляемого вероятностного теста является $T_i = 00000000$, случайным образом генерируются четыре кандидата в тестовые наборы $T_k = 11110000$, $T_k = 00110011$, $T_k = 11100010$ и $T_k = 01010110$. Затем для каждого кандидата в тесты T_k вычисляется значение меры различия (1) по отношению к тестовому набору T_i. Как видно из табл. 1, значения HD_0 во всех четырех случаях равняются четырем. Классическая методика формирования управляемых вероятностных тестов предполагает использование любого из четырех кандидатов в тесты в качестве следующего тестового набора.

В случае получения максимального значения HD_0 для нескольких кандидатов в тесты введенная авторами новая мера различия $HD(T_i, T_k)$ (см. определение 1) позволяет более полно учитывать различия кандидатов в тесты T_k по отношению к тестовому набору T_i. Для этого необходимо проанализировать значения следующей компоненты HD_1 предложенной авторами меры различия. Из рассматриваемого примера видно, что максимальное значение $HD_1 = 4$ достигается для набора $T_k = 01010110$, который в дальнейшем можно использовать как тестовый набор управляемого вероятностного теста.

В качестве интегральной меры различия двоичных тестовых наборов T_i и T_k размерностью $n = 2^w$ бит можно применить арифметическую сумму компонент HD_0, HD_1, ..., HD_v, …, HD_w предложенной меры:

$$HD_{Total}(T_i, T_k) = \sum_{v=0}^{w} HD_v(T_i, T_j),\ w = \log_2 n. \tag{2}$$

Критерием включения набора в тест может быть максимальное значение $HD_{Total}(T_i, T_k)$. Для примера, рассмотренного выше (см. табл. 1), с четырьмя кандидатами в тесты HD_{Total}(00000000, 11110000) = 8, HD_{Total}(00000000, 00110011) = 9, HD_{Total}(00000000, 11100010) = 10, HD_{Total}(00000000, 01010110) = 11. Максимальное значение интегральной меры различия достигается для набора $T_k = 01010110$.

2. Анализ меры различия. Рассматриваемая мера различия $HD(T_i, T_k)$, ориентированная на применение в управляемых вероятностных тестах, может быть использована и для сравнения двух символьных наборов T_i и T_k любой природы. Как указывалось ранее, новая мера различия основана на сравнении наборов T_i и T_k, представленных символами из различных алфавитов. Основная операция при сравнении наборов символов состоит в определении значения индикаторной функции $\mathrm{I}(t_{i,j} \neq t_{k,j})$ как результата сравнения двух символов $t_{i,j}$ и $t_{k,j}$ (1). При вычислении HD_v, $v \in \{0, 1, …, w\}$, согласно равенству (1), используется алфавит символов, каждый из которых состоит из 2^{2^v} значений. Принимая во внимание, что в случае генерирования управляемых вероятностных тестов тестовые наборы и кандидаты в тестовые наборы представляют собой случайные и независимые последовательности символов, можно оценить вероятность получения единичного значения индикаторной функции $\mathrm{I}(t_{i,j} \neq t_{k,j})$.

Учитывая, что символы $t_{i,j}$ и $t_{k,j}$ набора $T_i(v)$ формируются равновероятно из их алфавита, состоящего из 2^{2^v} символов, вероятность выполнения равенства $\mathrm{I}(t_{i,j} \neq t_{k,j}) = 1$ определяется как $P(t_{i,j} \neq t_{k,j}) = (1 - 1/2^{2^v})$, где выражение $1/2^{2^v}$ определяет вероятность выполнения равенства $t_{i,j} = t_{k,j}$.

Например, для двоичного случая ($v = 0$) $P(t_{i,j} \neq t_{k,j}) = (1 – 1/2)$, а для случая, когда T_i и T_k представлены одним символом из алфавита, включающего $2^{2^w} = 2^n$ символов ($v = w$), $P(t_{i,j} \neq t_{k,j}) = (1 – 1/2^n)$. Для произвольного значения $HD_0(T_i, T_k) = h$ (1), где $h \in \{0, 1, …, n\}$, а $n = 2^w$, вероятность равенства расстояния Хэмминга для компонент новой метрики, а именно HD_0, HD_1, HD_2, …, величине h определяется выражением

$$P(HD_v = 0) = (1/2^{2^v})^{2^{w-v}};$$
$$P(HD_v = h) = \binom{n/2^v}{h} \cdot \left[(1/2^{2^v})^{2^{w-v}-h} \cdot (1 - 1/2^{2^v})^h \right]; v = 1, ..., \lfloor \log_2(2^w/h) \rfloor; 0 < h \le 2^{w-v}. \tag{3}$$

Численные значения вероятностей $P(HD_v = h)$ для $n = 2^w = 2^3$ и различных значений $h \in \{0, 1, 2, …, 8\}$ приведены в табл. 2.

Таблица 2
Значения вероятностей компонент HD_0, HD_1, HD_2, HD_3 меры различия двоичных тестовых наборов

Table 2
Probability values of components HD_0, HD_1, HD_2, HD_3 of the difference measure for binary test patterns

h	$P(HD_0 = h)$	$P(HD_1 = h)$	$P(HD_2 = h)$	$P(HD_3 = h)$
0	0,00391	0,00391	0,00391	0,00391
1	0,03125	0,04688	0,11718	0,99609
2	0,10937	0,21094	0,87891	–
3	0,21875	0,42188	–	–
4	0,27344	0,31640	–	–
5	0,21875	–	–	–
6	0,10937	–	–	–
7	0,03125	–	–	–
8	0,00391	–	–	–

Указанные значения вероятностей подтверждают данные примера, представленного в табл. 1. Действительно, например, HD_3 принимает только значения 0 и 1, причем вероятность $P(HD_3 = 1)$ практически равняется единице, что в том числе определяется и свойством 1.

Согласно свойству 2 соотношение значений $HD_0 \geq HD_1 \geq ... \geq HD_v \geq … \geq HD_w$ компонент меры различия $HD(T_i, T_k)$ и их максимальных величин указывает на зависимость компонент меры различия двоичных тестовых наборов. Определяющей является наиболее информативная компонента HD_0, значение которой накладывает ограничения на величины остальных компонент.

Все множество ненулевых значений $HD_0 \in \{1, 2, …, n\}$, $n = 2^w$, представим в виде диапазонов, каждый из которых определяется w представлениями T_i и T_k в виде последовательностей символов $T_i(v)$ и $T_k(v)$ различных алфавитов. Соответственно, в качестве минимальной границы minHD_0 используем значения 1, 3, 5, 9, …, $2^{w-1}+1$, а в качестве максимальной maxHD_0 – значения 2, 4, 8, …, 2^w. Диапазоны значений компонент HD_0, HD_1, HD_2, …, HD_w меры различия $HD(T_i, T_k)$ для случая $n = 64$ ($w = 6$) представлены в табл. 3.

Таблица 3
Диапазоны значений HD_0, HD_1, HD_2, HD_3, HD_4, HD_5 и HD_6 меры различия двоичных тестовых наборов

Table 3
Value ranges HD_0, HD_1, HD_2, HD_3, HD_4, HD_5 and HD_6 of the difference measures between binary test patterns

HD_0		HD_1		HD_2		HD_3		HD_4		HD_5		HD_6	
min	max	min	max	min	max	min	max	min	max	min	max	min	max
1	2	1	2	1	2	1	2	1	2	1	2	1	1
3	4	2	4	1	4	1	4	1	4	1	2	1	1
5	8	3	8	2	8	1	8	1	4	1	2	1	1
9	16	5	16	3	16	2	8	1	4	1	2	1	1
17	32	9	32	5	16	3	8	2	4	1	2	1	1
33	64	17	32	9	16	5	8	3	4	2	2	1	1

В первых двух столбцах табл. 3 приведены значения minHD_0 и maxHD_0, определяющие границы диапазонов величины HD_0 для случая $n = 64$. Численные значения границ диапазонов для остальных компонент HD_1, HD_2, …, HD_6 меры различия, приведенные в табл. 3, получены на основании величин minHD_0 и maxHD_0. Например, для $3 \leq HD_0 \leq 4$ значение minHD_1 не может быть меньше двух в связи с тем, что три отличающихся бита в наборах T_i и T_k приведут к отличию как минимум двух символов в наборах $T_i(1)$ и $T_k(1)$, так как каждый символ в этих наборах представлен двумя битами. В то же время для тех же значений $3 \leq HD_0 \leq 4$ величина min$HD_2 = 1$ в силу того, что три отличающихся бита в наборах T_i и T_k могут оказаться в четырехбитном коде одного и того же символа наборов $T_i(2)$ и $T_k(2)$ (см. свойство 2). Значения maxHD_v, приведенные в табл. 3, повторяют величину maxHD_0 для всех диапазонов, для которых max$HD_0 \le 2^{w-v}$, а в остальных случаях равняются 2^{w-v}.

В общем случае минимальное значение minHD_v компоненты HD_v для T_i и T_k, состоящих из $n = 2^w$ двоичных символов, а также максимальная ее величина maxHD_v определяются в соответствии со следующими выражениями:

$$\min HD_v = \lceil \min HD_0 / 2^v \rceil,\ v = 0,1,...,w, \text{ для } \min HD_0 = 0,1,...,2^w; \tag{4}$$

$$\begin{aligned} &\max HD_v = \max HD_0,\ v = 0,1,...,w - \lceil \log_2(\max HD_0) \rceil, \text{ для } \max HD_0 = 1,2,...,2^w, \\ &\max HD_v = 2^{w-v},\ v = w - \lceil \log_2(\max HD_0) \rceil + 1,...,w-1,w, \text{ для } \max HD_0 = 1,2,...,2^w. \end{aligned} \tag{5}$$

Как видно из табл. 3 и соотношений (4) и (5), с ростом $v \in \{0, 1, \dots, w\}$ множество значений HD_v заметно уменьшается. Для $n = 2^3 = 8$ численные величины компонент HD_0, HD_1, HD_2 и HD_3 принадлежат диапазонам [0 ÷ 8], [0 ÷ 4], [0 ÷ 2] и [0 ÷ 1], что подтверждается данными табл. 1.

Анализ приведенных выше соотношений значений компонент HD_0, HD_1, HD_2, …, HD_w меры различия $HD(T_i, T_k)$ позволяет сформулировать три основные стратегии применения данной меры различия для генерирования управляемых вероятностных тестов. Напомним, что критерием для выбора одного из кандидатов в тесты в качестве следующего тестового набора является его максимальное отличие от ранее сгенерированных наборов. Рассмотрим предлагаемые стратегии на примере сравнения тестового набора T_i с кандидатами в тесты T_k, множество которых может составлять от десятков наборов до их тысяч [10, 18]. Для двоичных наборов T_i и T_k, состоящих из n бит, выполняются ранее введенные ограничения, т. е. $n = 2^w$, где w – целое. Для случая классической стратегии в качестве тестового набора случайным образом выбирается один из кандидатов в тесты, имеющий максимальное значение HD_0, полученное согласно равенству (1). Допускается, что максимальное значение HD_0 могут иметь несколько кандидатов в тесты. Каждая из стратегий в рамках предложенной меры различия $HD(T_i, T_k)$ основана на применении правила, определяющего критерий различия.

Правило 1. В качестве тестового набора T_k выбирается кандидат в тесты, который только один из всего их множества имеет максимальное значение HD_v для минимального $v \in \{0, 1, \dots, w\}$ меры различия $HD(T_i, T_k)$.

Как уже отмечалось, для примера, представленного в табл. 1, согласно данной стратегии, из четырех кандидатов в тесты выбирается T_k = 01010110, который для $v = 1$ имеет максимальное значение HD_1, равное четырем, при этом остальные кандидаты в тесты имеют меньшие значения HD_1. Заметим, что уже для $v = 2$ и $v = 3$ кандидаты в тесты T_k = 01010110 и T_k = 11100010 неразличимы (см. табл. 1).

Правило 2. В качестве критерия различия двоичных тестовых наборов T_i и T_k размерностью $n = 2^w$ используется их интегральная мера различия $HD_{Total}(T_i, T_k)$ (2), согласно которой в качестве тестового набора выбирается один из кандидатов в тесты T_k, имеющий ее максимальное значение.

В соответствии с данным правилом, как показывалось ранее, в случае примера, представленного в табл. 1, также будет выбран набор T_k = 01010110.

Для формулировки правила 3 обратим внимание на максимальные значения maxHD_v (5) компонент новой меры различия и вероятность равенства расстояния Хэмминга этим величинам для произвольных двоичных тестовых наборов T_i и T_k (3). В случае когда для T_i и T_k выполняется неравенство $HD_w \ne$ maxHD_w, имеем полное совпадение анализируемых наборов. При выполнении неравенства $HD_{w-1} \ne$ maxHD_{w-1} наборы T_i и T_k либо полностью совпадают, либо совпадают их первые $n/2$ бит или последующие $n/2$ бит. Как видно из примера, представленного в табл. 1, HD_{w-1}(00000000, 11110000) = HD_2(00000000, 11110000) = 1 $\ne$ maxHD_{w-1} = = max$HD_2(T_i, T_k)$ = 2. Соответственно, имеем $n/2$ = 4 совпадающих бит в указанных наборах. Ранее уже показывалось, что максимальные значения для всех компонент HD_v, $v \in \{0, 1, \dots, w\}$, которые интерпретируются как максимальные отличия, достигаются для двоичного случая, когда $T_k = \overline{t_{i,0}}, \overline{t_{i,1}}, ..., \overline{t_{i,n-1}}$.

Правило 3. В качестве тестового набора T_k выбирается один из кандидатов в тесты, для которого выполняются равенства $HD_w = \max HD_w$, …, $HD_{v+1} = \max HD_{v+1}$, $HD_v = \max HD_v$ при минимальном значении $v \in \{0, 1, …, w\}$.

Применяя данное правило для того же примера, получаем, что только один кандидат в тесты, а именно T_k = 01010110, соответствует условию правила 3, так как только для этого набора из четырех рассматриваемых $HD_3 = \max HD_3 = 1$, $HD_2 = \max HD_2 = 2$ и $HD_1 = \max HD_1 = 4$ для минимального $v = 1$ (см. табл. 1). Если рассматривать только три кандидата в тесты, а именно 11110000, 00110011 и 11100010, то для двух из них (00110011 и 11100010) будут выполняться равенства $HD_3 = \max HD_3 = 1$ и $HD_2 = \max HD_2 = 2$ при минимальном значении $v = 2$. Соответственно, один из указанных наборов может быть использован в качестве очередного тестового набора.

Необходимо отметить уточнение правила 3 в части принятия к рассмотрению расстояния HD_{v-1} и расстояний HD_{v-2}, HD_{v-3}, … в случае, когда условия этого правила выполняются более чем для одного кандидата в тесты. Для трех кандидатов в тесты предыдущего примера в качестве тестового набора будет выбран кандидат в тесты 11100010, так как для него HD_{v-1}(00000000, 11100010) = HD_1(000000000, 11100010) = 3, а для 00110011, соответственно, HD_1(00000000, 00110011) = 2. Такое решение принимается на базе основного свойства расстояния Хэмминга, заключающегося в том, что чем больше значение этого расстояния, тем больше различий между базовым набором T_i и набором T_k, для которого получено это значение.

В результате при использовании правил 1 и 3 на примере двоичных наборов, приведенных в табл. 1, в обоих случаях был выбран один и тот же кандидат в тесты T_k = 01010110 в качестве тестового набора. Этот факт свидетельствует о близости, но неэквивалентности правил 1 и 3 по результату их применения для генерирования управляемых вероятностных тестов. Их неэквивалентность подтверждается следующим примером.

Пример 2. В качестве примера двоичных тестовых наборов рассмотрим T_i = 0000000000000000 как базовый набор и два кандидата в тесты: T_k = 0101010111110000 и T_k = 0011010111001100, для которых выполняется условие $n = 16 = 2^4$. Для обоих кандидатов в тесты значение $HD_0 = 8$, что предопределяет использование остальных компонент новой меры различия HD_1, HD_2, HD_3, HD_4 и одного из правил 1, 2 или 3. Для первого кандидата в тесты T_k = 0101010111110000 компоненты меры различия принимают значения $HD_1 = 6$, $HD_2 = 3$, $HD_3 = 2$, $HD_4 = 1$, а для второго кандидата T_k = 0011010111001100 значения $HD_1 = 5$, $HD_2 = 4$, $HD_3 = 2$, $HD_4 = 1$. Применяя первое правило, в качестве тестового набора будет выбран первый набор, так как для него значение $HD_1 = 6$ больше чем для второго кандидата в тесты, для которого $HD_1 = 5$. Третье правило определит в качестве тестового набора второго кандидата в тесты, так как для него $HD_2 = \max HD_2 = 4$ для минимального $v = 2$. Оба кандидата в тесты являются неразличимыми при применении правила 2, так как в обоих случаях $HD_{Total}(T_i, T_k)$ (2) равняется 12.

3. Практические модификации меры различия *HD*(T_i, T_k). Первоначально рассмотрим ограничение, в рамках которого была определена новая мера различия $HD(T_i, T_k)$, соответствующая определению 1. Требование к размерности n двоичного набора T_i о том, что $n = 2^w$, где w – целое, может не всегда выполняться на практике. Соответственно, для $n \neq 2^w$ при отображении исходного набора T_i в последовательности $T_i(1)$, $T_i(2)$, $T_i(3)$, …, $T_i(w)$ для последнего символа последовательности $T_i(v)$, $v \in \{1, 2, …, w\}$, может отсутствовать необходимое количество бит, равное 2^v. Например, в силу того что для набора $T_i = 01100_{(2)}$ ($n = 5$) $w = \lceil \log_2 n \rceil = 3$, возможно его представление в виде последовательностей $T_i(1)$, $T_i(2)$, $T_i(3)$. Однако во всех трех случаях, а именно $T_i(1)$, $T_i(2)$ и $T_i(3)$, для последнего символа соответствующего алфавита отсутствует необходимое количество бит, для $T_i(1)$ не хватает одного бита, а для $T_i(2)$ и $T_i(3)$ – трех бит. Очевидным решением для устранения данного ограничения является циклическая интерпретация исходного набора $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$. Подобная интерпретация предполагает, что следующим битом набора T_i за последним битом $t_{i,n-1}$ будет первый его бит $t_{i,0}$. Таким образом, для получения необходимого количества бит для последнего символа набора $T_i(v)$ используются первые биты набора. В случае набора T_i = $01100_{(2)}$ подобная интерпретация позволяет получить $T_i(1) = 01100\underline{\mathbf{0}}_{(2)} = 120_{(4)}$, $T_i(2) = 01100\underline{\mathbf{011}}_{(2)} = 63_{(16)}$ и $T_i(3) = 01100\underline{\mathbf{011}} = c_{(256)}$. В приведенном примере выделены последние биты набора T_i, повторяющие его первые биты, которые использовались для расширения T_i и получения $T_i(1)$, $T_i(2)$ и $T_i(3)$. Возможны и другие подходы доопределения исходного набора T_i для получения необходимого количества бит, представляющих последний символ набора $T_i(v)$. Число дополнительных бит определяется соотношением величины n и числом бит 2^v, необходимых для представления символа. Для заданного n число недостающих бит для последнего символа набора $T_i(v)$ определяется как $2^v - n \bmod 2^v$. Эти биты необязательно доопределяются рассмотренным выше способом. Например, они могут быть определены постоянными нулевыми или единичными значениями либо случайными двоичными величинами.

Снятие ограничения на размерность n двоичного набора T_i путем его доопределения до нужного количества бит позволяет расширить количество алфавитов для иных отображений исходного набора. Напомним, что согласно определению 1 для вычисления компонент HD_0, HD_1, ..., HD_v, …, HD_w используются следующие алфавиты: двоичный, четверичный, шестнадцатеричный, …, (2^{2^v})-й, …, (2^{2^w})-й. Общее их количество равняется величине $w+1=\lceil \log_2 n \rceil+1$. Очевидно, что с учетом возможности расширения исходного двоичного набора до нужного количества бит число алфавитов может быть увеличено до n. Эти алфавиты состоят из символов, задаваемых одним, двумя, тремя, четырьмя битами и т. д., вплоть до алфавита, в котором каждый символ определяется n последовательными битами. Рассматривая пример исходного набора $T_i = 01100_{(2)}$ и его циклические расширения, представим его в виде последовательностей, полученных для n = 5 различных алфавитов. Соответственно, $T_i(0) = 01100_{(2)}$, $T_i(1) = 01100\underline{\mathbf{0}}_{(2)} = 120_{(4)}$, $T_i(2) = 01100\underline{\mathbf{0}}_{(2)} = 30_{(8)}$, $T_i(3) = 01100\underline{\mathbf{011}}_{(2)} = 63_{(16)}$, $T_i(4) = 01100_{(2)} = C_{(32)}$. Отметим, что в зависимости от соотношения величины n и количества бит, используемых для задания символа в заданном алфавите, исходный набор не всегда требует своего расширения (см., например, $T_i(4)$).

Определим значения компонент модифицированной меры различия $MD(T_i, T_k)$ как MD_0, MD_1, ..., MD_v, …, MD_{n-1} для исходных двоичных наборов T_i и T_k, представленных в виде последовательностей $T_i(0)$, $T_i(1)$, …, $T_i(n-1)$ и $T_k(0)$, $T_k(1)$, …, $T_k(n-1)$. При отображении наборов T_i и T_k в последовательности $T_i(1)$, $T_i(2)$, …, $T_i(n-1)$ и $T_k(1)$, $T_k(2)$, …, $T_k(n-1)$ будем использовать их циклическую интерпретацию. Значения MD_v вычисляются согласно соотношению (1). Отметим, что каждая из последовательностей будет состоять из фиксированного количества символов, определяемого числом бит для представления символа из заданного алфавита. Для восьмиразрядных двоичных наборов T_i и T_k, используемых в табл. 1, $T_i(0)$ и $T_k(0)$ представляются восемью двоичными символами, $T_i(1)$ и $T_k(1)$ – четырьмя четверичными символами, $T_i(2)$ и $T_k(2)$ – тремя восьмеричными символами; $T_i(3)$, $T_i(4)$, $T_i(5)$ и $T_i(6)$, а также $T_k(3)$, $T_k(4)$, $T_k(5)$ и $T_k(6)$ включают по два символа из соответствующего алфавита, $T_i(7)$ и $T_k(7)$ – по одному символу.

В свою очередь, число символов в последовательностях $T_i(v)$ и $T_k(v)$, $v \in \{0, 1, …, n-1\}$, определяет максимальные значения компонент MD_v меры различия $MD(T_i, T_k)$, вычисляемые согласно равенству (1):

$$\max MD_v = \lceil n/(v+1) \rceil, v = 0,1,...,n-1. \quad (6)$$

В табл. 4 представлены результаты вычислений на основании (1) компонент MD_0, MD_1, MD_2, MD_3 и MD_4 меры различия $MD(T_i, T_k)$ для двоичных наборов T_i и T_k, приведенных в табл. 1. Значение $MD_2(T_i, T_k) = MD_2(00000000, 11100010) = 2$ (табл. 4) получено как результат циклического расширения восьмибитовых наборов T_i = 00000000 и T_k = 11100010 до наборов, состоящих из девяти бит $T_i = 00000000\underline{\mathbf{0}}$ и $T_k = 11100010\underline{\mathbf{1}}$. Значение MD_2(**<u>000</u>**000**<u>000</u>**,

111000**101**) = 2 равняется числу несовпадающих восьмеричных символов (см. формулу (1)), выделенных подчеркиванием.

Таблица 4
Численные значения компонент MD_0, MD_1, MD_2, MD_3 и MD_4 меры различия $MD(T_i, T_k)$ для наборов T_i и T_k

Table 4
Numerical values for the difference measure $MD(T_i, T_k)$ components MD_0, MD_1, MD_2, MD_3 and MD_4 for patterns T_i and T_k

		MD_0	MD_1	MD_2	MD_3	MD_4
T_i	00000000	–	–	–	–	–
T_k	11110000	4	2	3	1	2
	00110011	4	2	3	2	2
	11100010	4	3	2	2	2
	01010110	4	4	3	2	2

Остальные значения компонент MD_5, MD_6 и MD_7 модифицированной меры различия $MD(T_i, T_k)$, так же как и MD_4, равняются их максимальным значениям.

Приведенный анализ модифицированной меры различия показывает, что по мере увеличения v значимость компоненты MD_v, так же как и HD_v, существенно уменьшается. Это объясняется тем, что для $v > n/2$ все компоненты MD_v, кроме последней MD_{n-1}, принимают только три возможных значения, а именно значение 0, если $T_i = T_k$, и значение 1 или 2, если $T_i \neq T_k$. Если $HD_v = \max HD_v$, то все последующие компоненты HD_{v+1}, HD_{v+2}, …, HD_{n-1} также принимают максимальные значения, как это видно из примера, приведенного в табл. 3. В обоих случаях новой меры различия, а именно $HD(T_i, T_k)$, описанной в определении 1, и ее модификации $MD(T_i, T_k)$, представленной в настоящем разделе, значимость компонент HD_v и MD_v с ростом v уменьшается из-за уменьшения количества символов при их вычислении согласно равенству (1).

Для более полного сопоставления сравниваемых последовательностей T_i и T_k, состоящих из n бит, расширим диапазон значений компонент новой меры различия за счет представления их отображений n символами в иных алфавитах. В этом случае исходный двоичный набор T_i представляется последовательностями $T_i(1)$, $T_i(2)$, $T_i(3)$, …, $T_i(n-1)$, каждая из которых будет состоять из n символов соответствующего алфавита. При построении $T_i(v)$, $v \in \{1, 2,\ldots, n-1\}$, для $T_i = t_{i,0}, t_{i,1}, \ldots, t_{i,n-1}$ используем его циклическую интерпретацию, в которой каждый бит $t_{i,j}$, $j \in \{0, 1,\ldots, n-1\}$, набора T_i является первым битом двоичного кода символа, состоящего из $v + 1$ бит. В качестве примера подобного отображения рассмотрим получение $T_i(2)$ для $T_i = 01100_{(2)}$. Каждый из $n = 5$ символов $T_i(2)$ состоит из $2 + 1 = 3$ бит T_i. Процедура получения символов восьмеричного алфавита последовательности $T_i(2)$ для исходного двоичного набора $T_i = 01100_{(2)}$ показана на рисунке. Выделены биты исходного набора $T_i = 01100$, используемые для каждого из пяти восьмеричных символов его отображения в последовательность $T_i(2) = 36401$.

01100	0**110**0	01**100**	**0**11**00**	**01**10**0**
3	6	4	0	1

Формирование символов $T_i(2)$
Character formation of $T_i(2)$

Двоичный тестовый набор $T_i = 000\ldots0$, состоящий из n нулей, в любом алфавите будет иметь вид, аналогичный $T_i = 000\ldots0$, а набор $T_i = 111\ldots1$, состоящий из n единиц, будет состоять из n старших символов алфавита. Например, для восьмеричного алфавита имеем 777…7, а шестнадцатеричного – FFF…F.

С учетом рассмотренных модификаций и уточнений новая мера различия $MHD(T_i, T_k)$ между двоичными тестовыми наборами T_i и T_k соответствует следующему определению.

Определение 2. Мера различия $MHD(T_i, T_k)$ двоичных тестовых наборов $T_i = t_{i,0}, t_{i,1}, ..., t_{i,n-1}$ и $T_k = t_{k,0}, t_{k,1}, ..., t_{k,n-1}$, где $t_{i,j}, t_{k,j} \in \{0, 1\}$, $j \in \{0, 1, …, n-1\}$, а n – произвольное целое, состоит из n компонент MHD_0, MHD_1, …, MHD_{n-1}, полученных согласно соотношению

$$MHD_v(T_i, T_k) = \sum_{j=0}^{n-1} \mathrm{I}(t_{i,j}\, t_{i,(j+1)\bmod n} \dots t_{i,(j+v)\bmod n} \neq t_{k,j}\, t_{k,(j+1)\bmod n} \dots t_{k,(j+v)\bmod n}),\; v = \overline{0, 1, ..., (n-1)}. \quad (7)$$

Подобная мера различия, основанная на использовании нескольких значений расстояния Хэмминга, выполняемых для $v = 2$, 4 и 8, применяется в работе [22] для анализа близости анализируемой последовательности к идеальному белому шуму. Анализ расстояния Хэмминга для восьмибитных фрагментов двоичной последовательности с эталонными последовательностями с целью анализа их качества как криптографических паролей приведен в статье [23].

С помощью соотношения (7) формируются численные значения MHD_0, MHD_1, MHD_2, …, MHD_{n-1}. В табл. 5 приведены примеры вычисления меры различия $MHD(T_i, T_k)$ для разных пар тестовых наборов T_i и T_k для случая, когда $n = 8$.

Таблица 5
Численные значения компонент меры различия MHD_0, MHD_1, MHD_2, MHD_3 для тестовых наборов T_i и T_k
Table 5
Numerical values of the difference measure components MHD_0, MHD_1, MHD_2, MHD_3 for test patterns T_i and T_k

T_i, T_k		MHD_0	T_i, T_k		MHD_1	T_i, T_k		MHD_2	T_i, T_k		MHD_3
T_i	00000000	–	T_i	00000000	–	T_i	00000000	–	T_i	00000000	–
T_k	11110000	4	T_k	33320001	5	T_k	77640013	6	T_k	FEC80137	7
	00110011	4		01320132	6		13641364	8		36C936C9	8
	11100010	4		33200121	6		76401253	7		EC8125B7	8
	01010110	4		12121320	7		25253641	8		5A5B6C92	8

Из приведенных данных следует, что в качестве следующего тестового набора будет выбран $T_k = 01010110$ в результате применения любого из трех ранее рассмотренных правил. Действительно, согласно правилу 1 $T_k = 01010110$ имеет максимальное значение MHD_1, равное семи, при этом остальные кандидаты в тесты имеют меньшие значения MHD_1. В соответствии с правилом 2 аналогично уравнению (2) определяется $MHD_{Total}(T_i, T_k)$ как сумма всех компонент $MHD(T_i, T_k)$. В результате получим, что наибольшее значение $MHD_{Total}(00000000, 01010110) = 27$ достигается для набора $T_k = 01010110$. И правило 3 предопределяет выбор в тестовые наборы того же кандидата в тесты. Согласно этому правилу два кандидата в тесты 00110011 и 01010110 имеют максимальное значение MHD_2, равное восьми, однако следующая компонента MHD_1 имеет большее значение, равное семи, для набора 01010110 (см. табл. 5).

Для случая, когда $T_i = 000…0$, наличие нулевых символов в представлении T_k в последовательности, представленной в ином алфавите, позволяет существенно упростить процедуру вычисления компонент рассмотренных мер различия, включая последнюю $MHD(T_i, T_k)$. Действительно, результатом вычислений согласно (6) значения MHD_v является разность величины n и количества нулевых символов в представлении $T_i(v)$, $v \in \{0, 1, 2,…, n-1\}$. Например, в последовательности четверичных символов $T_k(1) = 33320001$ (см. табл. 5) число нулевых символов равняется трем. Соответственно, $MHD_1 = 8 - 3 = 5$.

Для определения вычислительной сложности трех вариантов, а именно $HD(T_i, T_k)$, $MD(T_i, T_k)$ и $MHD(T_i, T_k)$, рассмотренной меры различия в качестве базовой используем операцию сравнения двух символов исходных наборов T_i и T_k, которые представлены в виде последовательностей $T_i(0)$, $T_i(1)$, …, $T_i(n-1)$ и $T_k(0)$, $T_k(1)$, …, $T_k(n-1)$. Анализ ранее рассмотренных примеров показывает, что определение расстояния Хэмминга можно свести к подсчету количества s нулевых двоичных символов поразрядной суммы по модулю два T_s исходных наборов T_i и T_k, где $t_{s,j} = t_{i,j} \oplus t_{k,j}$, $j \in \{0, 1, 2, …, n-1\}$. Тогда, например, HD_0, MD_0 и MHD_0 будут определяться величиной $n - s$. Аналогичным образом путем подсчета нулевых символов для других алфавитов рассчитываются остальные компоненты $HD(T_i, T_k)$, $MD(T_i, T_k)$ и $MHD(T_i, T_k)$.

Для $HD(T_i, T_k) = \{HD_0, HD_1, ..., HD_v, …, HD_w\}$ количество операций сравнения символов определяется как $O(HD) = n + n/2 + n/4 + … + 1 = 2n - 1$, где $n = 2^w$, а w – целое. В случае $MD(T_i, T_k) = \{MD_0, MD_1, ..., MD_v, …, MD_{n-1}\}$ число операций сравнения двух символов определяется выражением $O(MD) = n + \lceil n/2 \rceil + \lceil n/3 \rceil + \lceil n/(n-1) \rceil$.

Наибольшей вычислительной сложностью характеризуется мера различия $MHD(T_i, T_k) = \{MHD_0, MHD_1, ..., MHD_v, …, MHD_{n-1}\}$, для которой $O(MHD) = n + n + … + n = n^2$.

Как показывалось ранее, во всех трех случаях информативными являются только несколько первых компонент мер различия, как правило не более двух или трех. Соответственно, вычислительная сложность для всех трех вариантов сравнима и не превышает $3n$.

4. Экспериментальные оценки меры различия. Рассмотренные меры позволяют оценить степень различия двух тестовых наборов T_i и T_k, которые могут быть неразличимыми при использовании других мер различия, например расстояния Хэмминга. Для подтверждения факта неразличимости тестовых наборов был реализован следующий эксперимент. Для заданного тестового набора T_i, полученного случайным образом, формировалось множество из 1000 кандидатов в тесты T_k, которые также формировались случайным образом по равномерному закону распределения. Затем рассчитывались расстояния Хэмминга $D = HD_0$ между T_i и всеми остальными двоичными тестовыми наборами из списка кандидатов T_k и определялось подмножество кандидатов в тесты, которые имели максимальное значение HD_0. Отметим, что HD_0 является первой компонентой всех трех предложенных мер различия $HD(T_i, T_k)$, $MD(T_i, T_k)$ и $MHD(T_i, T_k)$. Эксперименты проводились для различных величин разрядности n наборов T_i и T_k. В качестве примера в табл. 6 приведены результаты вычисления трех компонент HD_0, HD_1 и HD_2 меры различия $HD(T_i, T_k)$ для $n = 16$. Для каждого из семи приведенных в табл. 6 наборов T_i формировались 1000 кандидатов в тесты T_k, и среди них находились наборы с максимальным значением HD_0.

Таблица 6
Результаты вычисления меры различия $HD(T_i, T_k)$ для $n = 16$

Table 6
Results of calculating the difference measure $HD(T_i, T_k)$ for $n = 16$

Номер эксперимента *Experiment number*	T_i	$\max HD_0(T_i, T_k)$, T_k	$\max HD_1(T_i, T_k)$, T_k	$\max HD_2(T_i, T_k)$, T_k
1	1001100010001110	$\max HD_0 = 15$ 0110011001110001	–	–
2	0101000000010010	$\max HD_0 = 14$ 1000101111101101 1000111111100101	$\max HD_1 = 8$ 1000101111101101 1000111111100101	$\max HD_2 = 4$ 1000101111101101 1000111111100101
3	0000111111000001	$\max HD_0 = 14$ 1111000001110110 1111000000010110	$\max HD_1 = 8$ 1111000001110110 1111000000010110	$\max HD_2 = 4$ 1111000001110110 1111000000010110
4	0101011100111111	$\max HD_0 = 14$ 1010010011000000 1010100010100000	$\max HD_1 = 8$ 1010100010100000	–
5	1100110110001110	$\max HD_0 = 14$ 0011001011110000	–	–
6	1110101000000110	$\max HD_0 = 13$ 0001011111011011 0001011010111001 0001010110100001 0001000110111101 1001111111111001 0011011110111001 1001010101110001 1001100111111001 0001010101110011	$\max HD_1 = 8$ 0001011111011011 0001010110100001 0001000110111101 1001111111111001 0011011110111001 1001010101110001 0001010101110011	$\max HD_2 = 4$ 0001011111011011 0001010110100001 0001000110111101 1001111111111001 0011011110111001 1001010101110001 0001010101110011
7	0100000010100001	$\max HD_0 = 14$ 1011110101010110	–	–

Представленные в табл. 6 результаты подтверждают выводы, которые следуют из проведенных авторами экспериментальных исследований:

1. Результаты 1, 5 и 7 доказывают эффективность расстояния Хэмминга, позволившего выбрать единственного из 1000 кандидатов в тесты, максимально отличающегося от T_i.

2. Данные экспериментов 2–4 и 6 показывают, что больше одного из 1000 кандидатов в тесты имеют максимальное значение расстояния Хэмминга. Например, в эксперименте 2 $\max HD_0 = 14$ имеют два кандидата в тесты, а именно 1000101111101101 и 1000111111100101.

3. Эксперимент 4 иллюстрирует работоспособность новой меры $HD(T_i, T_k)$, позволившей выбрать один набор T_k, максимально отличный от T_i. Эксперимент 6 также свидетельствует о целесообразности применения данной меры различия, так как за счет использования компоненты HD_1 множество потенциальных кандидатов в тесты уменьшилось с 9 до 7.

4. Как видно из всех представленных в табл. 6 результатов экспериментов, достаточным оказалось вычисление только трех компонент HD_0, HD_1 и HD_2, что свидетельствует о невысокой вычислительной сложности определения $HD(T_i, T_k)$, превышающей сложность вычисления расстояния Хэмминга не более чем в три раза.

5. Данные экспериментов 2 и 3 показывают неразличимость в обоих случаях двух кандидатов в тесты, имеющих одинаковые значения как максимального расстояния Хэмминга, так и компонент меры $HD(T_i, T_k)$. Данный вывод констатирует тот факт, что, так же как и в случае расстояния Хэмминга, мера $HD(T_i, T_k)$ не всегда позволяет определить единственного кандидата в тесты T_k, максимально отличного от T_i. Однако подмножество T_k максимально отличных от T_i наборов в случае $HD(T_i, T_k)$ меньше либо равно подмножеству, полученному на основании расстояния Хэмминга.

Суммарные значения количеств кандидатов в тесты T_k из 1000, 100 и 10, сгенерированных случайным образом и имеющих максимальное значение HD_0, приведены в табл. 7. Эти значения получены для трех величин общего количества кандидатов в тесты, из которых и выбирается лучший.

Таблица 7

Суммарные значения количеств кандидатов в тесты T_k с максимальным значением HD_0 для $n = 16$

Table 7

Total values of the numbers of test candidates T_k with maximum value of HD_0 for $n = 16$

Общее количество T_k *Total value of T_k*	Количество T_k с максимальным значением HD_0 *Number of T_k with maximum HD_0 value*															
1000	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	424	227	141	67	27	13	25	14	13	19	14	6	3	3	1	1
100	1	2	3	4	5	6	7	8	9	10	11					
	535	219	111	63	42	12	10	3	2	2	1					
20	1	2	3	4	5	6	7	8								
	623	229	99	23	17	4	4	1								

Данные табл. 7 во всех случаях получены на основании 1000 повторяющихся экспериментов. Например, для общего количества T_k, равного 1000, значения 1 и 424 в первом столбце означают, что в 424 экспериментах был найден только один кандидат с максимальным значением HD_0. Следующий столбец показывает, что в 227 экспериментах получено по два набора, имеющих максимальное HD_0. Во всех экспериментах значения T_k и T_i формировались случайным образом по равномерному закону распределения. Данные, представленные в табл. 7, получены для $n = 16$.

Анализ данных табл. 7 позволяет сделать вывод о часто встречающейся неразличимости кандидатов в тесты, имеющих максимальное значение расстояния Хэмминга. Действительно, в среднем в половине случаев более чем один кандидат в тесты T_k имеет максимальное значение расстояния Хэмминга по отношению к T_i (табл. 7). Эта тенденция сохраняется как для различных разрядностей n тестовых наборов, так и для разного числа кандидатов в тесты, из которых выбирается очередной тестовый набор. Действительно, выбрав предельно малое значение общего количества кандидатов в тесты, равное пяти, и проведя 1000 аналогичных экспериментов, были получены 270 результатов с двумя и более неразличимыми наборами T_k на основании HD_0. Данный вывод обосновывает необходимость применения других мер различия, позволяющих более эффективно и однозначно выбирать очередной тестовый набор из списка кандидатов в тесты.

При неразличимости нескольких кандидатов в тесты на основании расстояния Хэмминга можно применять предложенные авторами новые меры различия. При этом важным является количество компонент, вычисляемых для каждой из таких ситуаций при выборе одного из кандидатов. Это количество зависит от состава множества кандидатов в тесты T_k и набора T_i. При тех же условиях, что и в предыдущих экспериментальных исследованиях, была проведена оценка количества компонент новых мер различия, необходимых для выбора одного кандидата. В табл. 8 представлено количество экспериментов с определенным числом итераций расчета компонент $HD(T_i, T_k)$ и $MHD(T_i, T_k)$, которые были проведены для обнаружения наиболее удаленного от T_i тестового набора T_k. Рассматривался случай 32-битных наборов T_i и T_k (списки из 10 кандидатов в тесты), который в целом повторяет результаты авторов, полученные для других разрядностей тестовых наборов и разных значений размерности списка кандидатов в тесты.

В первом и четвертом столбцах табл. 8 перечислены компоненты мер различия $HD(T_i, T_k)$ и $MHD(T_i, T_k)$, которые вычислялись для выбора одного кандидата. В столбцах *Количество* приведено число случаев из 1000 экспериментов, для которых вычислялась более чем одна компонента. Например, в первой строке и первых трех столбцах показано, что в 175 экспериментах из 1000 вычислялись две компоненты – HD_0 и HD_1. В процентном исчислении это составляет 69,4 % от всех случаев вычислений более одной компоненты.

Таблица 8
Количество экспериментов с определенным числом итераций вычисления компонент мер отличия для n = 32
Table 8
Number of experiments with a certain number of iterations of calculating the components of difference measures for n = 32

$HD(T_i, T_k)$			$MHD(T_i, T_k)$		
Компоненты *Components*	Количество *Quantity*	%	Компоненты *Components*	Количество *Quantity*	%
HD_0, HD_1	175	69,4	MHD_0, MHD_1	220	78,3
HD_0, HD_1, HD_2	75	29,8	MHD_0, MHD_1, MHD_2	47	16,7
HD_0, HD_1, HD_2, HD_3	2	0,8	MHD_0, MHD_1, MHD_2, MHD_3	12	4,3
–	–	–	MHD_0, MHD_1, MHD_2, MHD_3, MHD_4	2	0,7

Анализ данных табл. 8 подтверждает невысокую вычислительную сложность новых мер различия. В подавляющем числе случаев для $HD(T_i, T_k)$ и $MHD(T_i, T_k)$ необходимо вычисление только двух компонент, заметно реже – трех и в исключительно редких случаях – больше трех.

Заключение. В работе рассмотрена мера различия, основанная на применении модификаций определения расстояния Хэмминга и отображении двоичных тестовых наборов в виде последовательностей символов, представленных в различных алфавитах. Предложенная мера различия расширяет возможности определения тестовых последовательностей при генерировании управляемых вероятностных тестов. В среднем в половине случаев возникает вопрос о выборе единственного тестового набора из множества наборов, для которых расстояние Хэмминга принимает максимальное значение относительно предыдущего тестового набора. Показано, что тестовые наборы, неразличимые при использовании в качестве меры различия расстояния Хэмминга, имеют отличающиеся значения компонент мер различия $HD(T_i, T_k)$, $MD(T_i, T_k)$ и $MHD(T_i, T_k)$. Это позволяет более точно классифицировать формируемые случайным образом наборы, которые являются кандидатами в тесты. Введенные меры не всегда дают возможность определить единственного кандидата в тесты, однако подмножество потенциальных наборов после применения мер различия меньше исходного, полученного на основании расстояния Хэмминга. В редких случаях подмножество не изменяется. Вычислительные сложности предложенных модификаций определения расстояния Хэмминга для всех трех вариантов сравнимы и превышают сложность вычисления расстояния Хэмминга не более чем в три раза. Проведенные экспериментальные исследования показали невысокую временную сложность вычисления предложенных мер различия. Дальнейшие исследования целесообразно расширить в части исследования свойств новой меры различия и ее применимости для различных прикладных задач.

Вклад авторов. *В. Н. Ярмолик* предложил меру различия для тестовых наборов, основанную на применении расстояния Хэмминга. *В. В. Петровская* провела экспериментальные исследования. *Н. А. Шевченко* принял участие в обобщении и анализе полученных результатов.

Список использованных источников

1. Duran, J. W. An evaluation of random testing / J. W. Duran, S. C. Ntafos // IEEE Transactions on Software Engineering. – 1984. – Vol. SE-10, no. 4. – P. 438–444.
2. Arcuri, A. Random testing: Theoretical results and practical implications / A. Arcuri, M. Z. Iqbal, L. Briand // IEEE Transactions on Software Engineering. – 2011. – Vol. 38, no. 2. – P. 258–277.
3. An orchestrated survey on automated software test case generation / S. Anand [et al.] // J. of Systems and Software. – 2014. – Vol. C-39, no. 4. – P. 582–586.
4. An empirical comparison of combinatorial testing, random testing and adaptive random testing / H. Wu [et al.] // IEEE Transactions on Software Engineering. – 2020. – Vol. 46, no. 3. – P. 302–320.
5. Ярмолик, В. Н. Контроль и диагностика вычислительных систем / В. Н. Ярмолик. – Минск : Бестпринт, 2019. – 387 с.
6. A survey on adaptive random testing / R. Huang [et al.] // IEEE Transactions on Software Engineering. – 2021. – Vol. 47, no. 10. – P. 2052–2083.
7. A preliminary study of adaptive random testing techniques / M. S. Roslina [et al.] // Intern. J. of Information Technology & Computer Science. – 2015. – Vol. 19, no. 1. – P. 116–127.
8. Ярмолик, С. В. Управляемое случайное тестирование / С. В. Ярмолик, В. Н. Ярмолик // Информатика. – 2011. – № 1(29). – С. 79–88.
9. Nikravan, E. Hybrid adaptive random testing / E. Nikravan, S. Parsa // Intern. J. of Computing Science & Mathematics. – 2020. – Vol. 11, no. 3. – P. 209.
10. Zhibo, Li. An enhanced adaptive random testing by dividing dimensions independently / Li. Zhibo, Li. Qingbao, Yu Lei // Mathematical Problems in Engineering. – 2019. – Vol. 2019. – P. 1–15.
11. Садовский, М. Г. О сравнении символьных последовательностей / М. Г. Садовский // Вычислительные технологии. – 2005. – № 3(10). – С. 106–116.
12. Hamming, R. W. Error detecting and error correcting codes / R. W. Hamming // The Bell System Technical J. – 1950. – Vol. 29, no. 2. – P. 147–160.
13. Левенштейн, В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов / В. И. Левенштейн // Доклады Академии наук СССР. – 1965. – Т. 163, № 4. – С. 845–848.
14. Алгоритмы: построение и анализ : пер. с англ. / Т. Кормен [и др.]. – 3-е изд. – М. : Изд. дом «Вильямс», 2013. – 1328 с.
15. Tannga, M. J. Comparative analysis of Levenshnein distance algorithm and Jaro Winkler for text plagiarism detection application / M. J. Tannga, S. Rahman, Hasniati // J. of Technology Research in Information System and Engineering. – 2017. – Vol. 4, no. 2. – P. 44–54.
16. Needleman, S. A general method applicable to the search for similarities in the amino acid sequence of two proteins / S. Needleman, C. Wunsch // J. of Molecular Biology. – 1970. – Vol. 48, no. 3. – P. 443–453.
17. Smith, T. F. Identification of common molecular subsequences / T. F. Smith, M. S. Waterman // J. of Molecular Biology. – 1981. – Vol. 147. – P. 195–197.
18. Candidate test set reduction for adaptive random testing: An overheads reduction technique / R. Huang [et al.] // J. of Molecular Biology. – 2021. – Vol. 214, no. C. – P. 102730.
19. Ярмолик, В. Н. Мера различия для тестовых наборов при генерировании управляемых вероятностных тестов / В. Н. Ярмолик, В. В. Петровская, И. Мрозек // Информатика. – 2022. – Т. 19, № 4. – С. 7–26.
20. Ярмолик, В. Н. Мера различия для управляемых вероятностных тестов / В. Н. Ярмолик, Н. А. Шевченко, В. В. Петровская // Доклады БГУИР. – 2022. – Т. 20, № 6. – С. 52–60.
21. Неразрушающие тесты с четным повторением адресов для запоминающих устройств / В. Н. Ярмолик [и др.] // Информатика. – 2021. – Т. 18, № 3. – С. 18–35.

22. Многомерный портрет цифровых последовательностей идеального «белого шума» в свертках Хэмминга / В. И. Волчихин [и др.] // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 4(44). – С. 4–13.

23. Условия корректного вычисления энтропии осмысленных длинных паролей в пространстве сверток Хэмминга с эталонными текстами на русском и английском языках / В. И. Волчихин [и др.] // Измерение. Мониторинг. Управление. Контроль. – 2019. – № 3(29). – С. 33–38.

References

1. Duran J. W., Ntafos S. C. An evaluation of random testing. *IEEE Transactions on Software Engineering*, 1984, vol. SE-10, no. 4, pp. 438–444.

2. Arcuri A., Iqbal M. Z., Briand L. Random testing: Theoretical results and practical implications. *IEEE Transactions on Software Engineering*, 2011, vol. 38, no. 2, pp. 258–277.

3. Anand S., Burke E. K., Chen T. Y., Clark J., Cohen M. B., …, Zhu H. An orchestrated survey on automate software test case generation. *Journal of Systems and Software*, 2014, vol. C-39, no. 4, pp. 582–586.

4. Wu H., Nie C., Petke J., Jia Y., Harman M. An empirical comparison of combinatorial testing, random testing and adaptive random testing. *IEEE Transactions on Software Engineering*, 2020, vol. 46, no. 3, pp. 302–320.

5. Yarmolik V. N. Control' i diagnostika vuchislitel'nuch system. *Computer Systems Testing and Diagnoses.* Minsk, Bestprint, 2019, 387 p. (In Russ.).

6. Huang R., Sun W., Xu Y., Chen H., Towey D., Xia X. A survey on adaptive random testing. *IEEE Transactions on Software Engineering*, 2021, vol. 47, no. 10, pp. 2052–2083.

7. Roslina M. S., Ghani A. A. A., Baharom S., Zulzazil H. A preliminary study of adaptive random testing techniques. *International Journal of Information Technology & Computer Science*, 2015, vol. 19, no. 1, pp. 116–127.

8. Yarmolik S. V., Yarmolik V. N. *Controlled random testing*. Informatika [*Informatics*], 2011, no. 1(29), pp. 79–88 (In Russ.).

9. Nikravan E., Parsa S. Hybrid adaptive random testing. *International Journal of Computing Science and Mathematics*, 2020, vol. 11, no. 3, p. 209.

10. Zhibo Li., Qingbao Li., Lei Yu. An enhanced adaptive random testing by dividing dimensions independently. *Mathematical Problems in Engineering*, 2019, vol. 2019, pp. 1–15.

11. Sadovskii M. G. *About symbolical sequences comparigion*. Vuchislitel'nue technologii [*Computational Technologise*], 2005, no. 3(10), pp. 106–116 (In Russ.).

12. Hamming R. W. Error detecting and error correcting codes. *The Bell System Technical Journal*, 1950, vol. 29, no. 2, pp. 147–160.

13. Levenshtein V. I. *Binary codes with correction of deletions, insertions and substitutions of characters*. Doklady Akademii nauk SSSR [*Proceedings of the USSR Academy of Sciences*], 1965, vol. 163, no. 4, pp. 845–848 (In Russ.).

14. Cormen T. H., Leiserson C. E., Rivest R. L., Stein C. *Introduction to Algorithm*, 3rd edition. MIT Press, 2009, 1292 p.

15. Tannga M. J., Rahman S., Hasniati. Comparative analysis of Levenshnein distance algorithm and Jaro Winkler for text plagiarism detection application. *Journal of Technology Research in Information System and Engineering*, 2017, vol. 4, no. 2, pp. 44–54.

16. Needleman S., Wunsch C. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 1970, vol. 48, no. 3, pp. 443–453.

17. Smith T. F., Waterman S. Identification of common molecular subsequences. *Journal of Molecular Biology*, 1981, vol. 147, pp. 195–197.

18. Huang R., Chen H., Sun W., Towey D. Candidate test set reduction for adaptive random testing: An overheads reduction technique. *Journal of Molecular Biology*, 2021, vol. 214, no. C, p. 102730.

19. Yarmolik V. N., Petrovskaya V. V., Mrozek I. *Distance measure for controlled random tests*. Informatika [*Informatics*], 2022, vol. 19, no. 4, pp. 7–26 (In Russ.).

20. Yarmolik V. N., Shauchenka M. A., Petrovskaya V. V. *Distance measure for controlled random tests*. Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioèlektroniki [*Doklady BGUIR*], 2022, vol. 20, no. 6, pp. 52–60 (In Russ.).

21. Yarmolik V. N., Mrozek I., Levantsevich V. A., Demenkovets D. V. *Transparent memory tests with even repeating addresses for storage devices*. Informatika [*Informatics*], 2021, vol. 18, no. 3, pp. 18–35 (In Russ.).

22. Volchikhin V. I., Ivanov A. I., Yunin A. P., Malygina E. A. *A multidimensional picture of numerical sequences of the ideal "white noise" in Hamming convolutions*. Izvestija vysshih uchebnyh zavedenij. Povolzhskij region. Tehnicheskie nauki [*University Proceedings. Volga Region. Engineering Sciences*], 2017, no. 4(44), pp. 4–13 (In Russ.).

23. Volchikhin V. I., Ivanov A. I., Karpov A. P., Yunin A. P. Conditions for the correct calculation of the entropy of meaningful long passwords in the Hamming convolution space with reference texts in Russian and English. Izmerenie. Monitoring. Upravlenie. Kontrol' [*Measuring. Monitoring. Management. Control*], 2019, no. 3(29), pp. 33–38 (In Russ.).

Информация об авторах

Ярмолик Вячеслав Николаевич, доктор технических наук, профессор, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: yarmolik10ru@yahoo.com

Петровская Вита Владленовна, магистр технических наук, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: vita.petrovskaya@gmail.com

Шевченко Николай Алексеевич, студент, Дармштадтский технический университет.
E-mail: nik.sh.de@gmail.com

Information about the authors

Vyacheslav N. Yarmolik, D. Sc. (Eng.), Prof., Belarusian State University of Informatics and Radioelectronics.
E-mail: yarmolik10ru@yahoo.com

Vita V. Petrovskaya, M. Sc. (Eng.), Belarusian State University of Informatics and Radioelectronics.
E-mail: vita.petrovskaya@gmail.com

Nikolai A. Shevchenko, Student, Darmstadt Technical University.
E-mail: nik.sh.de@gmail.com