Труды научно-технической конференции кластера пензенских предприятий,
обеспечивающих **БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ.**
Том 10, Пенза-2016, с. 52-57 (http://пниэи.рф/activity/science/BIT/T10-p52.pdf)

УДК: 519.2, 612.087, 621.319.7

Гончаров С.М., Боршевников А.Е., Половинко А.С.
(г. Владивосток)

Генератор синтетических образов электроэнцефалограмм активности головного мозга, используемый для увеличения размеров тестовых и обучающих выборок биометрических данных

Современные информационные угрозы, исходящие от различных террористических и экстремистских организаций, а также других источников, являются важной проблемой для развития и функционирования информационного общества [1]. Решение данных проблем требует развития средств информационной безопасности, в частности технологий биометрической аутентификации. Особый интерес представляют технологии высоконадежной биометрической аутентификации, использующие биометрические характеристики высокой конфиденциальности. Одной из таких характеристик является электроэнцефалограмма (ЭЭГ).

Исследования в области высоконадежной биометрической аутентификации на основе ЭЭГ с использованием больших нейронных сетей (нейросетевых преобразователей биометрия-код доступа) авторами проводятся в течение пяти лет. Однако, в силу сложности сбора естественных биометрической образов электроэнцефалограмм активности головного мозга, исследования обычно проводятся на малых выборках. Было решено дополнить существующую базу естественных биометрических образов синтетическими образами. С этой целью был разработан программный генератор синтетических образцов ЭЭГ, соответствующий требованиям стандарта ГОСТ Р 52633.2-2010 [2].

Стимуляция ЭЭГ образов потенциала головного мозга

Для генерации синтетических образов была использована база ЭЭГ, полученная ранее [3]. Данная база содержит данные 10 пользователей, для каждого из которых было снято 100 примеров ЭЭГ. Используемая стимуляция выглядит, как поочередно меняющиеся цифры от «0» до «9». Фрагмент стимуляции изображен на рисунке (рис. 1).

Пользователи выбирают один, два или четыре символа, и при их появлении на экране концентрируются на них. Данные символы являются "мысленным паролем".

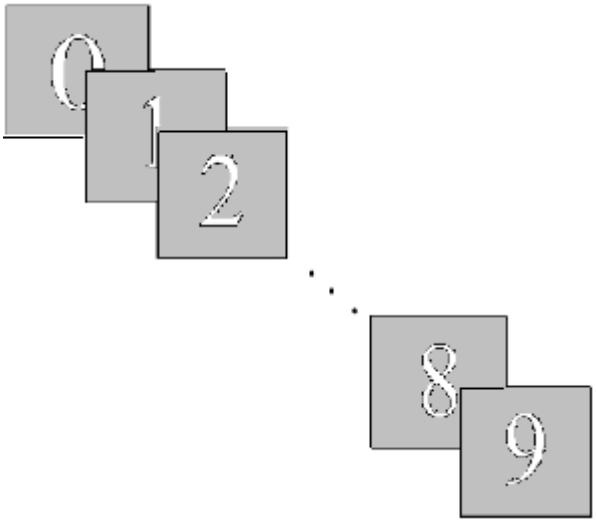

Рис. 1. Схема визуальной стимуляции

Снятие ЭЭГ производится в течение 10 секунд (20 секунд для четырех символов). Для односимвольного «мысленного пароля» пользователь концентрируется во время появления ключевого символа на экране в течение всей процедуры снятия ЭЭГ. Для случая, когда пользователь запоминает два символа в качестве пароля, съем ЭЭГ разбивается на два этапа по 5 секунд. В течение первого этапа пользователь концентрируется на первом символе, а в течение второго – на втором символе. В случае, когда пользователь использует четыре символа в качестве пароля, все время процедуры снятия ЭЭГ разбивается на четыре фрагмента по 5 секунд, в каждом из которых пользователь поочередно концентрируется на текущем символе пароля.

В результате проведенной стимуляции у пользователя формируется потенциал P300, который изображен на рисунке (Рис.2).

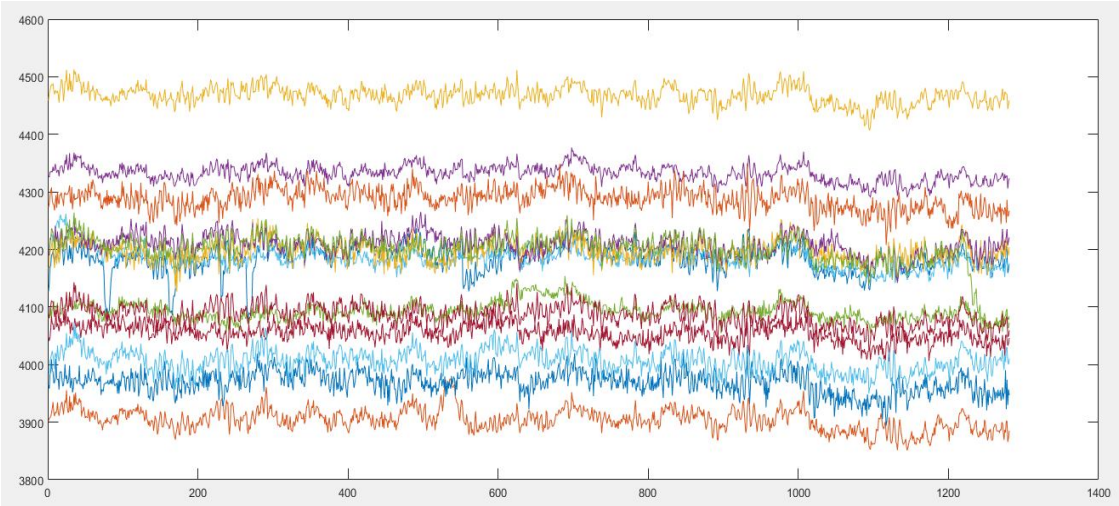

Рис. 2. Пример естественного образа ЭЭГ (14 каналов)

Генерация синтетических образов ЭЭГ

Рассмотрим алгоритмы, применяемые в исследовании. Для синтезирования образов принято применять следующие типы преобразований: синтез, мутацию, морфинг и перестановку фрагментов.

Для начала опишем процедуру синтеза случайных биометрических примеров.

1. По всем имеющимся примерам образа вычисляются математические ожидания каждого *i*-го контролируемого биометрического параметра $E_{образ}(v_i)$.

2. Вычисляется стандартное отклонение каждого из контролируемых биометрических параметров $\sigma_{образ}(v_i)$.

3. Генератором нормальных случайных чисел со статистическими характеристиками, определенными в пунктах 1 и 2, формируются векторы параметров заранее заданного числа биометрических примеров.

Такой метод размножения биометрических примеров не учитывает корреляционные связи между параметрами биометрических примеров.

Опишем процедуру мутации биометрических примеров.

1. По всем имеющимся примерам вычисляют математические ожидания каждого *i*-го контролируемого биометрического параметра $E_{образ}(v_i)$.

2. Вычисляют стандартное отклонение каждого из контролируемых биометрических параметров $\sigma_{образ}(v_i)$.

3. Каждый параметр v_i исходного (мутируемого) биометрического примера изменяется на величину Δv_i, полученную от случайного генератора с нормальным законом распределения значений с математическим ожиданием v_i и стандартным отклонением, равным трети $\Delta v_{i\max}$, где $\Delta v_{i\max}$ – расстояние до ближайшей границы $E_{образ}(v_i)\pm 3\sigma_{образ}(v_i)$ от значения Δv_i.

Такой метод размножения биометрических примеров ослабляет естественные корреляционные связи между параметрами биометрических примеров.

Опишем процедуру морфинга биометрических примеров одного биометрического образа. Суть морфинга заключается в нахождении промежуточных значений для каждого из параметров пары биометрических примеров-родителей. Такие значения и будут являться примерами-потомками.

Если исходные биометрические примеры имеют различный набор параметров, либо их смысл различается, необходимо преобразовать такие примеры, чтобы набор параметров совпадал. В случае, если преобразование невозможно, такие биометрические примеры нельзя использовать при морфинге.

Оценка количества биометрических примеров образов-потомков для каждой пары биометрических примеров-родителей происходит по следующему алгоритму:

1. Выбирают N пар биометрических примеров-родителей (N> 2).

2. Рассчитывается необходимое число синтетических биометрических примеров.

3. Для каждой пары биометрических примеров-родителей *A* и *B* рассчитывают расстояние между их биометрическими параметрами:

$$S_{AB} = \frac{1}{n}\sum_{i=1}^{n}\left|v_{i,A} - v_{i,B}\right| \qquad (1)$$

,

где $v_{i,A}$ – *i*-й параметр биометрического примера *A* и $v_{i,B}$ – *i*-й параметр биометрического примера *B*; *n* – общее число параметров биометрического примера.

4. Для каждой пары *A* и *B* биометрических примеров-родителей рассчитывают среднее значение расстояния между их биометрическими параметрами *E(s)*.

5. Расчет количества примеров-потомков *k* для пары примеров-родителей *A*, *B* происходит по формуле:

$$k_{AB} = \frac{2N_{синт}}{N}P(S_{AB}), \qquad (2)$$

где $P(S_{AB})$ – вероятность появления расстояния между параметрами S_{AB} из множества всех возможных расстояний между параметрами; k_{AB} округляется до ближайшего целого числа.

Морфинг конкретной пары биометрических примеров-родителей происходит по следующему алгоритму:

Значения каждого i-го биометрического параметра каждого из биометрических примеров биометрических образов-потомков вычисляют по формуле:

$$v_{i,j} = \frac{(k_{AB}+1)-j}{k_{AB}+1} v_{i,A} + \frac{j}{k_{AB}+1} v_{i,B}, \quad (3)$$

где j – порядковый номер потомка ($j = 1,2, \ldots, k_{AB}$); k_{AB} – количество потомков для пары примеров-родителей A и B.

Такой метод размножения биометрических примеров сохраняет естественные корреляционные связи, присутствующие у биометрических примеров-родителей.

Опишем процедуру размножения биометрических примеров перестановкой фрагментов.

Для использования метода размножения биометрических примеров-родителей перестановкой фрагментов необходимо разделить исходные биометрические примеры на фрагменты. Существует два варианта такого разделения:

– по естественной фрагментации исходных биометрических примеров;

– по фрагментации биометрических примеров, заложенной производителем средства при вычислении биометрических параметров и связанной с тем, что каждый из множества полученных параметров сам по себе является некоторым фрагментом исходного примера.

При размножении N исходных биометрических примеров, как при использовании естественной фрагментации, так и при фрагментации биометрических параметров, необходимо для формирования каждого синтетического биометрического примера использовать приблизительно $1/N$ фрагментов каждого из исходных биометрических примеров [2].

Используя выше перечисленные методы, был разработан программный генератор синтетических образов для нейросетевого преобразователя биометрия-код доступа на основе ЭЭГ.

Количество выбранных файлов предпочтительно равно 10, так как в нашем случае это составляет один образ-родитель. Количество примеров потомков задается вручную, вводом в соответствующее поле. Каждый последующий алгоритм генерации использует в качестве основы предыдущий пример-потомок.

Полученные результаты

Для проведения тестирования данного генератора был использован нейросетевой преобразователь, описанный в работе [4]. Было решено использовать данную конструкцию преобразователя из-за лучших показателей по сравнению с другими исследуемыми преобразователями. При тестировании использовались следующие параметры преобразователя. Количество электродов – 14. Количество выбираемых коэффициентов Фурье для одного электрода – 15. Количество нейронов первого слоя – 320. Размер восстанавливаемого ключа был выбран – 256, что означает использование во втором слое нейронной сети 256 нейронов. Количество входов на нейрон было взято 4. Преобразователь обучался по стандарту ГОСТ Р 52633.5 [5].

Размер сгенерированной базы образов «Чужой» для образов, полученных в ходе эксперимента, составляет 10^4. Пример сгенерированного синтетического образа приведен на рисунке (Рис. 3).

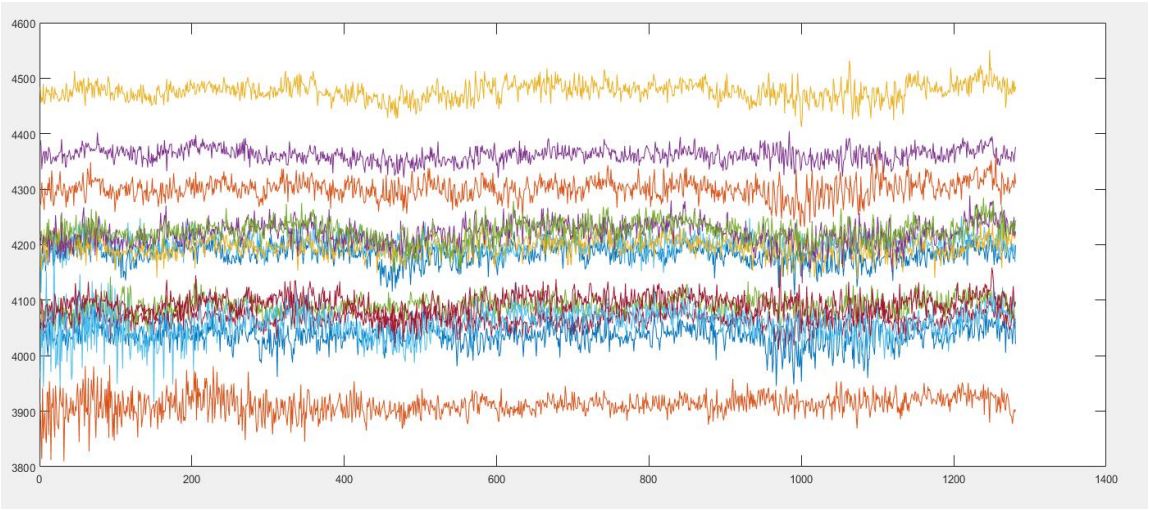

Рис. 3. Пример синтетического образа ЭЭГ (14 каналов)

В таблице 1 приведены результаты опыта по восстановлению ключа из естественных образов «Чужой» с дополненной синтетическими образами обучающей базой данных. Для тестирования работы генератора один естественный образ был выбран в качестве образа "Свой", остальные девять естественных образов сформировали базу "Чужой".

Таблица 1. Расстояния Хэмминга до исходного ключа в случае, если злоумышленник угадал символ

Номер пользователя	1 символ	2 символа	4 символа
1	173	99	186
2	98	172	70
3	147	64	179
4	177	157	179
5	147	76	95
6	97	189	175
7	89	143	167
8	162	144	114
9	85	58	188

Результаты по сравнению с предыдущими экспериментами улучшились – минимальное и среднее расстояние Хэмминга по всем экспериментам увеличилось и стало больше соответствовать требованиям, предъявляемым стандартами ГОСТ Р 52633. Стоит отметить, что для образа «Свой» ключ, как и в проведенных ранее экспериментах, безошибочно восстанавливается.

Также был проведены эксперименты по восстановлению секретного ключа из данных синтетического образа «Чужой». В результате расстояния Хемминга до секретного ключа колебались от 101 до 150 (для ключа длиной 256 бит). Данные результаты удовлетворяют требованиям стандартов.

Разработан программный генератор синтетических биометрических образов ЭЭГ, соответствующий требованиям стандарта ГОСТ Р 52633.2-2010. Требуется исследование качества работы нейросетевого преобразователя

биометрия-код доступа на основе ЭЭГ на более обширных базах синтетических образов, а также расчет численных показателей качества параметров сгенерированных образов.

Литература

1. Доктрина информационной безопасности Российской Федерации [утв. Указом Президента Российской Федерации от 5 декабря 2016 г. N 646]. – М. : Кремль, 2016. – 8 с.

2. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации: ГОСТ Р 52633.2-2010. – Введен впервые 30.09.10 – М.: Стандартинформ, 2011. – 17 с.

3. Гончаров С.М., Боршевников А.Е. Построение нейросетевого преобразователя "Биометрия-код доступа" на основе параметров визуального вызванного потенциала электроэнцефалограммы / С.М. Гончаров, А.Е. Боршевников // Доклады Томского государственного университета систем управления и радиоэлектроники: Научный журнал. – Томск: Изд-во ТУСУР, 2014. – № 2. – С. 51–55.

4. Гончаров С.М. Восстановление секретного ключа на основе электроэнцефалограммы при движении глаз с закрытыми веками. / Гончаров С.М., Боршевников А.Е., Михайлов А.Г., Апальков А.Ю. // Журнал «Информация и безопасность». Том. 19, часть 1. Воронеж: ВГТУ, 2016. – С. 114–117.

5. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа: ГОСТ Р 52633.5-2011. – Введен впервые 01.12.2011. – М.: Стандартинформ, 2012. – 20 с.

Статья поступила 13.11.2016, опубликована 27.11.2016
по положительной рецензии д.т.н. Иванова А.И.