УДК 004.056
ББК 67.408

И.Г. СИДОРКИНА, А.Н. САВИНОВ

ТРИ АЛГОРИТМА УПРАВЛЕНИЯ ДОСТУПОМ К КСИИ НА ОСНОВЕ РАСПОЗНАВАНИЯ КЛАВИАТУРНОГО ПОЧЕРКА ОПЕРАТОРА

Ключевые слова: *клавиатурный почерк, идентификация, аутентификация, авторизация, верификация.*

Предложены три алгоритма организации управления доступом к КСИИ, реализующих аутентификацию, идентификацию и верификацию законного оператора по клавиатурному почерку. Методы аутентификации по биометрическим параметрам личности, в том числе и по клавиатурному почерку, ввиду неотъемлемости биометрических характеристик от конкретного человека способны обеспечить повышенную, по сравнению с другими способами проверки соответствия, точность, невозможность отказа от авторства и удобство для операторов автоматизированных систем. Методы постоянного скрытого клавиатурного мониторинга позволяют обнаруживать подмену законного оператора и блокировать ключевую систему от вторжения злоумышленника.

I.G. SIDORKINA, A.N. SAVINOV

THREE ALGORITHMS OF CONTROL ACCESS TO THE KSII ON THE BASIS OF RECOGNITION OF KEYSTROKE DYNAMICS

Key words: *keystroke dynamics, identification, authentication, authorization, verification.*

The paper proposes three algorithms organizations to control access to the KSII implement authentication, identification and verification of the legal operator keystroke dynamics. Authentication Methods for Biometric identity, including the keystroke dynamics, due to the irrevocable biometric characteristics of a particular person, able to provide increased, compared with other methods of checking compliance, accuracy, non-repudiation and convenience for operators of automation systems. Procedures for the permanent hidden keystroke monitoring can detect spoofing legitimate operator and block key system from malicious intrusion.

Методы аутентификации по биометрическим параметрам личности, в том числе и по клавиатурному почерку (КП), ввиду неотъемлемости биометрических характеристик от конкретного человека способны обеспечить повышенную, по сравнению с другими способами проверки соответствия, точность, невозможность отказа от авторства и удобство для операторов автоматизированных систем. Методы постоянного скрытого клавиатурного мониторинга позволяют обнаруживать подмену законного оператора и блокировать ключевую систему (КС) от вторжения злоумышленника. Таким образом, задача исследования моделей, методов и алгоритмов распознавания клавиатурного почерка операторов ключевых систем является актуальной на данный момент. Основными характеристиками КП являются время удержания клавиш и время между нажатиями клавиш. В процессе разработки системы анализа КП операторов ключевой системы необходимо разработать алгоритмы регистрации оператора (обучения), аутентификации и авторизации оператора, а также верификации и идентификации.

Клавиатурный почерк – это набор динамических характеристик работы на клавиатуре. Стандартная клавиатура позволяет измерить следующие временные характеристики: время удержания клавиши нажатой и интервал времени между нажатиями клавиш. Как было выявлено в работе [8], используя алгоритмы распознавания времени удержания клавиш (ВУК) и методы распознавания КП по свободному тексту, можно построить систему постоянного скрытного мониторинга, позволяющую верифицировать законного оператора.

Время удержания клавиши – это период, в течение которого клавиша находится в нажатом состоянии. Программное обеспечение измеряет этот показатель от момента нажатия клавиши (событие «onkeydown») до момента ее отпускания (событие «onkeyup»). Этот параметр, как правило, выражают в миллисекундах. **Среднее**

время удержания клавиши – это математическое ожидание выборки показателей времени удержания конкретной клавиши, собранной за период набора фрагмента текста. Эмпирические исследования базы эталонов КП операторов показали, что время удержания зависит также от наложений, ритмичности и безошибочности. Среднее время удержания может существенно различаться у разных людей при близких скоростях набора из-за разницы в методике, используемой наборщиком.

Наложение движений – одновременные движения нескольких пальцев у наборщиков, уверенно владеющих методом печати. Наложение нажатий клавиш происходит, когда одна клавиша еще не отпущена, а другая уже нажимается. Наблюдается тенденция к повышению количества наложений с повышением скорости набора. Подавляющее большинство наложений происходит, когда клавиши соседних букв в слове нажимаются разными пальцами. Однако при очень быстром наборе скольжениями наложения также возможны.

Наложения происходят по следующим причинам: высокая скорость печати, при которой наборщик не успевает отпускать предыдущие клавиши до нажатия следующих; большое время удержания клавиш нажатыми; сочетание первого и второго факторов.

Установлено, что при одинаковой скорости печати, как правило, на постановке «ЫВАМ ТОЛД» наложений больше, чем на «ФЫВА ОЛДЖ», а при динамическом наборе – больше, чем при наборе по зонам. Возрастание количества наложений обусловлено увеличением числа сочетаний, которые нажимаются разными пальцами. Выявлено, что чем отрывистее удар и ритмичнее печать, тем меньше наложений. Когда клавиши скорее нажимаются, чем ударяются, а ритмичность невысока, наложений получается больше. При очень отрывистой печати время удержания клавиши нажатой может составлять 65 мс и меньше, а при нажатиях с большим числом наложений – 120 мс и больше. Среднее время удержания, как правило, составляет 80-100 мс. Пример набора текста с наложениями можно увидеть на рис. 1.

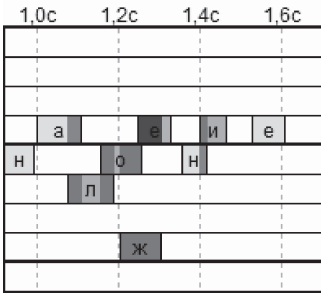

Рис. 1. Набор слова «наложение» с наложениями клавиш

Можно различить три вида наложений:

1. В момент удержания первой клавиши происходит нажатие второй. Клавиша «К1» нажимается. Далее происходит нажатие клавиши «К2», но «К1» ещё не отпущена. Затем происходит отпускание клавиши «К1», далее отпускается клавиша «К2».

2. В момент удержания одной клавиши происходит отпускание другой клавиши, т.е. первая клавиша была нажата в момент удержания второй. «К2» нажата. «К1» нажимается. После этого происходит отпускание клавиши «К2» и затем «К1».

3. Нажатие и отпускание клавиши происходят во время удержания другой клавиши. Нажата клавиша «К2», происходят нажатие и отпускание «К1» и затем отпускание «К2».

Эмпирические исследования клавиатурных почерков операторов [7] показало, что ВУК при наборе текста является бимодальным, а не нормальным распределением. Таким образом методы и модели, предлагаемые для определения ВУК как математического ожидания нормального распределения [8], имеют значительную погрешность и не отображают реальных зависимостей процесса набора текста.

В качестве примера приведём анализ выборки ВУК «А» при наборе случайного фрагмента текста (см. табл. 1). Во время набора текста клавиша «А» была нажата 91 раз. Если предположить, что ВУК подчиняется законам нормального распределения, то можно рассчитать математическое ожидание как среднее ВУК. При таком подходе получим математическое ожидание = 96 мс. Но из табл. 1 видно, что пики распределений приходятся на 35-37 мс, 68-72 мс и 104-108 мс. Инструмент анализа данных PAST version 217b [9] позволяет разделить выборку на 2 пересекающихся нормальных распределения

(рис. 2), математическими ожиданиями которых являются 86 и 106 мс. То есть отклонение от рассчитанного ранее математического ожидания – 10 мс, что составляет от 7% до 30% от измеряемой величины.

В качестве разделителя выборок можно использовать математическое ожидание нормального распределения, но этот подход не отображает причин возникновения бимодального распределения. Из табл. 1 видно, что во второе нормальное распределение элементы попадают в случае, если набор происходил с наложениями клавиш.

Таким образом, используя простой механизм определения наличия наложения клавиш при наборе, можно получить две независимые выборки и две величины математического ожидания ВУК для каждой клавиши, характеризующие КП оператора КСИИ, что, в свою очередь, увеличит точность идентификации оператора, т.е. снизит вероятность возникновения ошибок 1- и 2-го рода [5]. При этом каждая из выборок подчиняется нормальному закону распределения, а значит, к ней применима математическая модель ВУК, разработанная в [6].

Таблица 1

Выборка ВУК клавиши «А»

ВУК клавиши, мс	Количество значений данного ВУК в выборке	Вероятность встречи ВУК в выборке, %	Количество встреченных наложений
32	2	7,14	0
35	2	7,14	0
36	5	17,86	0
37	2	7,14	0
38	1	3,57	0
67	1	3,57	0
68	3	10,71	0
69	1	3,57	0
70	3	10,71	0
71	2	7,14	0
72	4	14,29	0
73	1	3,57	0
74	1	3,57	0
102	1	3,57	1
103	1	3,57	1
104	4	14,29	3
105	13	46,43	13
106	10	35,71	10
107	10	35,71	10
108	8	28,57	8
109	1	3,57	1
110	1	3,57	1
140	1	3,57	1
141	1	3,57	1
142	3	10,71	3
143	3	10,71	3
144	4	14,29	4
145	2	7,14	2

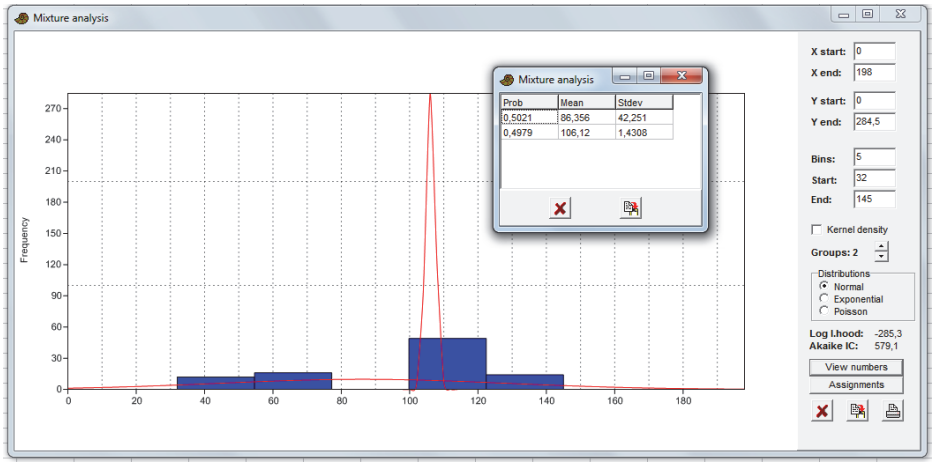

Рис. 2. Анализ выборки ВУК программным инструментом PAST

Алгоритм регистрации клавиатурного почерка. Построение систем биометрической идентификации основано на создании эталонных представлений идентифицируемых лиц. Эталон создается тогда, когда система находится в режиме обучения. Он представляет собой сохраненные в памяти системы, контролирующей доступ, биометрические характеристики человека и используется для сравнения с биометрическими параметрами лиц, претендующих на доступ к ресурсам. В случае, когда измеренные системой значения параметров пользователя отличаются от эталона больше, чем допускается порогом чувствительности, он получает отказ в доступе к ресурсам.

Работу алгоритма можно представить диаграммой деятельности (см. рис. 3). Для выявления усредненных значений времени событий клавиатуры используется вероятностно-статистический метод, поэтому необходим сбор статистики, состоящей из выборки временных значений, где элементом выборки будет являться время удержания клавиши. Алгоритм основан на математической модели, разработанной в [6].

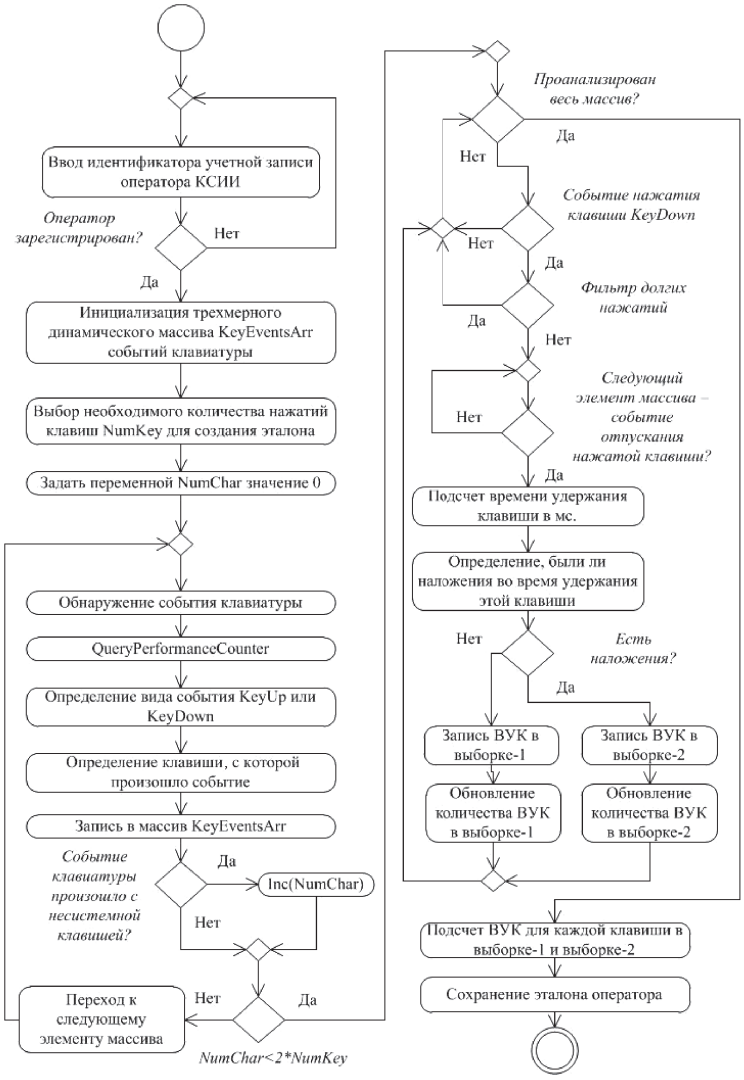

Рис. 3. Диаграмма деятельности алгоритма регистрации КП

В данном алгоритме выполняется процесс получения информационного файла – получения эталона КП оператора. Для этого в начале выполнения производится идентификация оператора по его уникальному идентификатору, например логину.

Инициализируется динамический трехмерный массив KeyEventsArr, в котором сохраняются события клавиатуры в следующем виде – см. табл. 2. Массив заполняется, пока не будет нажато достаточное (задаваемое администратором или установленное по умолчанию) количество клавиш.

Таблица 2

Структура массива KeyEventsArr

Кл1	**…**	**КлN**
Событие KeyUp или KeyDown	…	Событие KeyUp или KeyDown
Тик счетчика высокого разрешения, на котором произошло событие	…	Тик счетчика высокого разрешения, на котором произошло событие

На следующем этапе происходит подсчет времени удержания клавиш. Находится событие нажатия конкретной клавиши. Затем находится событие отпускания этой клавиши. Из тика события отпускания вычитается тик события нажатия и делится на частоту счётчика высокого разрешения для получения значения ВУК в миллисекундах. Тики событий определяются функцией QueryPerformanceCounter, частота счётчика – функцией QueryPerformanceFrequency [7]. На многоядерных системах используется функция SetThreadAffinityMask, чтобы указать родственность процессора для системы.

Алгоритм подразумевает фильтр долгих нажатий на клавишу, применяемых наборщиком для ввода *n*-грамм одинаковых букв, например «мм» в слове «программа» или «ССС» в «СССР». В алгоритме также имеется фильтр нажатия системных клавиш (например, BACKSPACE или ENTER), нажатие которых не сохраняется в эталоне.

В зависимости от наличия или отсутствия наложений при удержании клавиши значение ВУК заносится в выборку первого (без наложений) или второго (с наложениями) нормального распределения. Затем подсчитывается математическое ожидание каждой выборки, и эталон КП сохраняется в учетной записи оператора. Эталонный информационный файл оператора имеет следующую структуру (см. табл. 3).

Таблица 3

Структура информационного файла

Клавиша	**ВУК-1**	**ВУК-2**	**Количество**	**Количество**	**Выборка нормального распределения-1**	**Выборка нормального распределения-2**
А						
Б						
…						
я						

Так как свободный текст, по которому проводится обучение или дальнейшие процедуры аутентификации и идентификации, имеет разную вероятность встречаемости разных букв и символов [6], сбор достаточной для проведения вероятностно-статистического анализа выборки ВУК требует ввода очень большого текста (от 5000 символов и больше). Поэтому предложено ввести в информационный файл поля «количество» и «выборка».

В поле «выборка» сохраняются ВУК, разделяемые знаком «;», в поле «количество» заносится количество элементов в выборке. При достижении достаточного количества элементов в поле «выборка» подсчитывается ВУК, поле «выборка» очищается, в поле «количество» заносится символ «!», который используется для обозначения факта окончания подсчёта ВУК этой клавиши. Дальше элементы в поле «выборка» заноситься не будут. Если в процессе обучения достаточное количество элемен-

тов выборки не собрано, то будет рассчитано временное значение ВУК, а процесс сбора и расчета ВУК продолжится при работе алгоритмов авторизации и мониторинга при подтверждении, что текст набирает законный оператор, которому соответствует данный эталон. При аутентификации и идентификации преимущество будут иметь законченные выборки ВУК клавиш, им будут присвоены большие коэффициенты влияния, предложенные в [5], незаконченным – меньший. Предложено разделить алфавит системы на несколько групп в зависимости от вероятности встречаемости в текстах.

Алгоритм аутентификации и авторизации. Авторизация – предоставление определённому лицу или группе лиц прав на выполнение определённых действий, а также процесс проверки (подтверждения) данных прав при попытке выполнения этих действий. Обычно процессу авторизации предшествует процесс аутентификации – подтверждение подлинности, соответствия оператора предъявленному им идентификатору.

Предложенный алгоритм аутентификации и последующей авторизации представлен на рис. 4. В данном алгоритме сравнение происходит по принципу «один к одному», поэтому загружается эталон конкретного оператора, идентификатором которого он представился. В случае успешной аутентификации происходит процесс авторизации, в противном случае – отказ.

Главное отличие этого алгоритма от предыдущего состоит в том, что здесь для определения почерка используется меньшее количество нажатий клавиш. Обычно пароль состоит из 14-20 символов. Этот факт требует дополнительной настройки порога доступа для уменьшения ошибок первого и второго рода. Порог доступа можно настроить, проанализировав локальную базу эталонов клавиатурных почерков операторов, имеющих доступ к конкретной КСИИ. Сравнение текущего образца КП оператора КСИИ с эталонным происходит путём расчета меры Евклида для каждой клавиши, при этом ВУК-1 и ВУК-2 сравниваются как отдельные элементы:

$$P = \sqrt{\sum_{i=1}^{N} (A_i - B_i)^2} .$$

Полученное значение непохожести, рассчитанное как мера расстояния Евклида, сравнивается с порогом доступа. Если непохожесть меньше порога доступа, то оператор проходит процедуру авторизации, иначе получает отказ.

Алгоритм постоянного скрытного клавиатурного мониторинга (верификации оператора). Парольные и атрибутные методы идентификации и аутентификации имеют ряд существенных недостатков.

Первый из них – неоднозначность идентификации пользователя и возможность обмана системы защиты (например, путем кражи или подделки атрибута или взлома пароля).

Второй по значимости недостаток традиционных методов идентификации и аутентификации – отсутствие возможности обнаружения подмены авторизированного пользователя (например, злоумышленник может воспользоваться ситуацией, когда оператор произвел вход в систему и отлучился, оставив КС без присмотра и не заблокировав ее).

Методы постоянного скрытого клавиатурного мониторинга позволяют обнаруживать подмену законного пользователя и блокировать КС от вторжения злоумышленника.

Диаграмма деятельности алгоритма постоянного скрытного клавиатурного мониторинга и верификации оператора представлена на рис. 5. Массив DynamicKeyEvetsArr имеет небольшие размеры от 10 до 40 элементов и может быть выбран, например, в зависимости от интервала копирования оператора [11]. Интервал копирования – это число символов, которые могут быть напечатаны в точности после одно-

кратного просмотра текста. Солтхаус установил, что интервал копирования в обычной ситуации перепечатки у опытного наборщика составил в среднем 14,6 символа [10]. Текущие характеристики КП определяются по этому массиву. Массив динамически обновляется при вводе текста. Элементы, введенные раньше, удаляются, и ВУК рассчитывается динамически по добавленным новым элементам.

При обнаружении подмены законного оператора измеряется частота счётчика высокого разрешения, и если она изменилась, то ВУК считается снова, и почерки сравниваются ещё раз. Если частота не изменялась или пересчитанный текущий почерк не совпадает – КСИИ блокируется. Система пытается идентифицировать злоумышленника по базе КП, хранящихся в системе. Передается сигнал тревоги в службу безопасности.

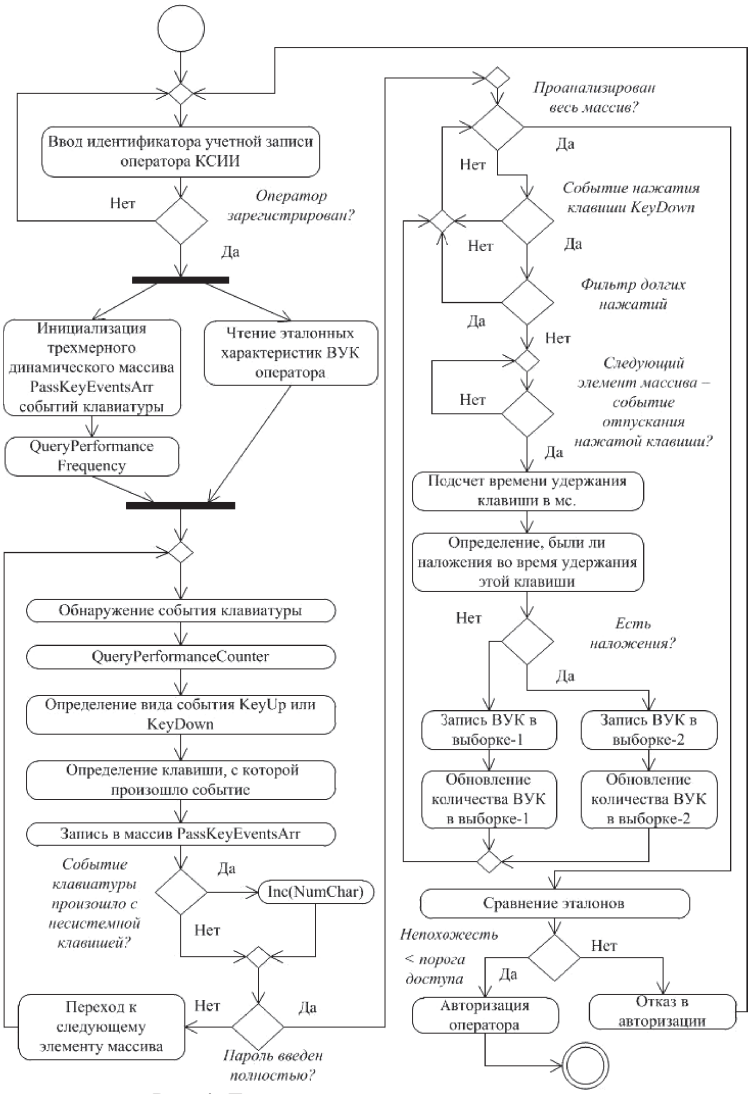

Рис. 4. Диаграмма деятельности алгоритма аутентификации и авторизации

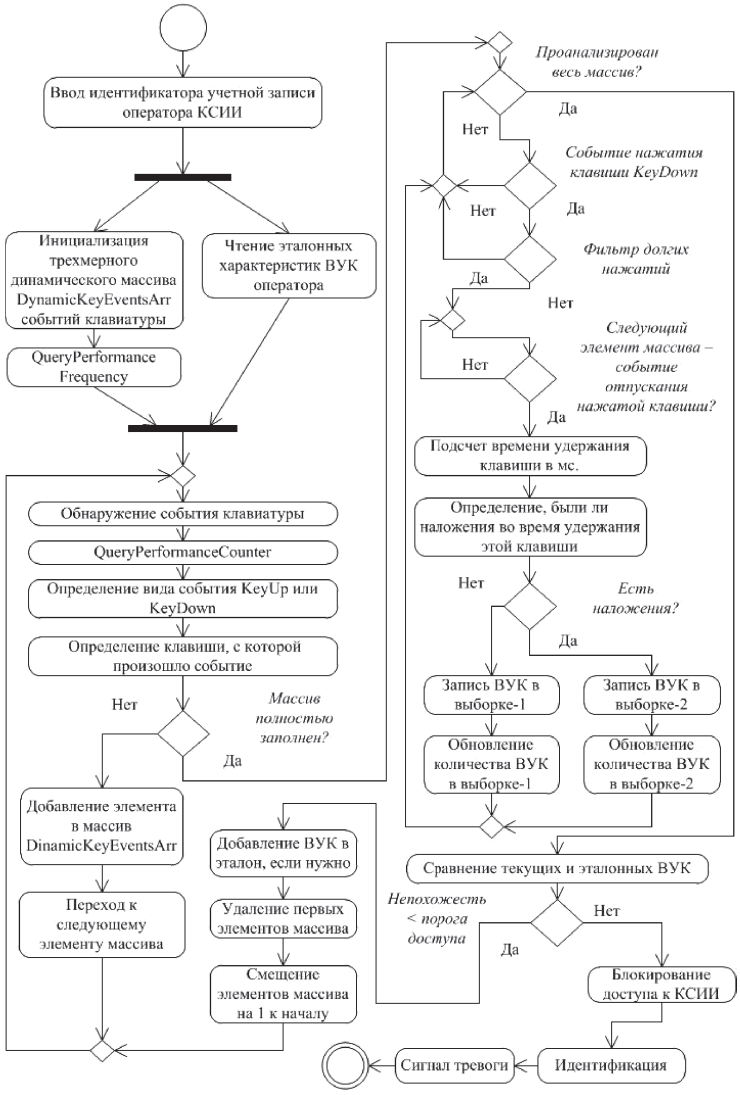

Рис. 5. Диаграмма деятельности алгоритма постоянного скрытного мониторинга

Если почерки совпадают и имеются введенные символы, для которых подсчёт ВУК не закончен, то такие элементы добавляются в соответствующую выборку и происходит перерасчёт ВУК. Если в этот момент набрано достаточное для обучения количество элементов в выборке, то поле «выборка» очищается и в поле «количество» информационного файла ставится метка «!». Данный метод можно также использовать и при переобучении всего эталона почерка оператора в случае изменений клавиатурного почерка, например вызванных совершенствованием оператором техники печати.

Выводы. Предложено использовать в качестве представления ВУК при анализе клавиатурного почерка бимодальное распределение вместо нормального распределе-

ния. Бимодальное распределение предложено разделять на два нормальных, что позволит применять математическую модель клавиатурного почерка, основанную на распределении Гаусса.

Разработаны три алгоритма управления доступом, основанные на распознавании клавиатурного почерка. Алгоритм обучения позволяет сохранить биометрические характеристики клавиатурного почерка как бимодальное распределение. Алгоритм аутентификации и авторизации позволяет распознать оператора КСИИ и проверить его соответствие указанной им учетной записи. Алгоритм постоянного скрытного клавиатурного мониторинга позволяет защитить КСИИ от вторжения злоумышленников путём подмены законного пользователя.

Литература

1. *Гатчин Ю.А., Ермаков Н.В., Коробейников А.Г., Строганов К.В.* Основные аспекты создания системы защиты периметра корпоративной информационной системы // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2007. № 40. С. 279-283.

2. *Гатчин Ю.А., Жаринов И.О., Коробейников А.Г.* Математические модели оценки инфраструктуры системы защиты информации на предприятии // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2012. № 2. С. 92-95.

3. *Коробейников А.Г., Гатчин Ю.А., Липатов А.Л., Осломенко Д.В.* Законодательные требования в области обеспечения информационной безопасности автоматизированных систем // Сборник тезисов IV межвузовской конференции молодых ученых. СПб.: СПбГУ ИТМО, 2007. 165 с.

4. *Коробейников А.Г., Кудрин П.А., Сидоркина И.Г.* Алгоритм распознавания трехмерных изображений с высокой детализацией // Вестник Марийского государственного технического университета. Сер. Радиотехнические и инфокоммуникационные системы. 2010. № 2(9). С. 91-98.

5. *Савинов А.Н., Иванов В.И.* Анализ решения проблем возникновения ошибок первого и второго рода в системах распознавания клавиатурного почерка // Вестник Волжского университета имени В.Н. Татищева. Сер. Информатика. 2011. Вып. 18. С. 115-119.

6. *Савинов А.Н., Сидоркина И.Г.* Нормальное распределение при анализе клавиатурного ввода при разработке математической модели клавиатурного почерка // Автоматизация управления и интеллектуальные системы и среды: материалы III Междунар. конф. (Махачкала. 9-15 октября, Махачкала). Нальчик: Изд-во КБНЦ РАН, 2012. Т. 2. С. 115-119.

7. *Савинов А.Н., Сидоркина И.Г.* Решение проблемы измерения времени удержания клавиш при разработке системы анализа клавиатурного почерка // ИКТ: образование, наука, инновации: труды III Междунар. науч.-практ. конф. Алматы: МУИТ, 2012. С. 328-333.

8. *Савинов А.Н., Сидоркина И.Г., Иванов В.И.* Анализ решения проблемы использования клавиатурного подчерка для обеспечения безопасности ключевой системы предприятия // Труды конгресса по интеллектуальным системам и информационным технологиям «IS&IT'11»: сб.: в 4 т. М.: Физматлит, 2011. Т. 3. С. 40-47.

9. PAST PAlaeontological STatistics [Electronic resource]. URL: http://folk.uio.no/ohammer/past.

10. *Salthouse T.A.* Anticipatory processing in transcription typing // J. Appl. Psychol. 1970. Vol. 2. P. 264-271.

11. *Salthouse T.A.* Perceptual, cognitive, and motoric aspects of transcription typing // Psychol. Bull. 1999. Vol. 3. P. 303-319.

СИДОРКИНА ИРИНА ГЕННАДЬЕВНА. См. с. 292.

САВИНОВ АЛЕКСАНДР НИКОЛАЕВИЧ – аспирант кафедры информационно-вычислительных систем, Поволжский государственный технологический университет, Россия, Йошкар-Ола (Homo-asio-otus@yandex.ru).

SAVINOV ALEXANDR NIKOLAEVICH – post-graduate student of Data-Processing Systems Chair, Volga State University of Technology, Russia, Yoshkar-Ola.