На правах рукописи

Сав.

Савинов Александр Николаевич

МЕТОДЫ, МОДЕЛИ И АЛГОРИТМЫ РАСПОЗНАВАНИЯ КЛАВИАТУРНОГО ПОЧЕРКА В КЛЮЧЕВЫХ СИСТЕМАХ

Специальность: 05.13.19 – Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени кандидата технических наук

Санкт-Петербург, 2013

Работа выполнена на кафедре информационно-вычислительных систем Поволжского государственного технологического университета.

Научный руководитель:	доктор технических наук, профессор кафедры информационно-вычислительных систем Поволжского государственного технологического университета Сидоркина Ирина Геннадьевна
Официальные оппоненты:	доктор технических наук, профессор, зам. директора по науке Санкт-Петербургского филиала Института земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова РАН Коробейников Анатолий Григорьевич
	кандидат технических наук, начальник отдела программного обеспечения ООО «НИИ мониторинга качества образования» Пылин Владислав Владимирович
Ведущая организация:	ФГБОУ ВПО «Казанский национальный исследовательский технический университет им. А. Н. Туполева – КАИ»

Защита диссертации состоится «20» ноября 2013 года в 15-50 часов на заседании диссертационного совета Д 212.227.05 при Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики по адресу: 197101, г. Санкт-Петербург, Кронверкский пр., д. 49.

Отзывы на автореферат, заверенные печатью, просим направлять по адресу: 197101, Санкт-Петербург, Кронверкский пр., д. 49, СПб НИУ ИТМО, ученому секретарю диссертационного совета Д 212.227.05.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики.

Автореферат разослан «20» октября 2013 года.

Ученый секретарь
диссертационного совета Д 212.227.05
кандидат технических наук, доцент Поляков В.И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Одним из основных факторов, определяющих состояние защищенности той или иной ключевой системы информационной инфраструктуры (КС), является эффективность функционирования подсистемы управления доступом и защиты информации. Парольные и атрибутные методы идентификации и аутентификации имеют ряд существенных недостатков. Главный из них – неоднозначность идентификации оператора ключевой системы (ОКС) и возможность обмана системы защиты, например, путем кражи или имитации атрибута или взлома пароля. Второй недостаток данных методов идентификации и аутентификации – невозможность обнаружения подмены законного авторизированного пользователя. В данном случае злоумышленник может нанести вред обрабатываемой КС информации, когда оператор оставляет без присмотра КС с пройденной процедурой авторизации.

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 КС от вторжения злоумышленника. Таким образом, задача исследования моделей, методов и алгоритмов распознавания клавиатурного почерка операторов ключевых систем является актуальной на данный момент.

В исследованиях по биометрии ряда ученых, таких как А.И. Иванов, М.Н. Десятерик, В.В. Марченко выделены характеристики клавиатурного почерка: время удержания клавиши при нажатии, интервалы между нажатиями клавиш, сила нажатия на клавишу, скорость нажатия на клавишу и др. В работах ученых В.И. Волчихина, А.И. Иванова предложено использовать аппарат искусственных нейронных сетей для математической обработки данных, полученных в результате экспериментов с биометрическими данными человека. А.Н. Лебедев, В.Б. Дорохов, Т.Н. Щукин, Е.В. Луценко в своих трудах доказали наличие зависимости между изменениями клавиатурного почерка оператора и его психофизиологическим состоянием.

Проблема применения клавиатурного почерка в системах идентификации и аутентификации операторов исследовалась в работах таких ученых, как Dawn Song, Peter Venable, Adrian Perrig (Pittsburgh, PA, USA); R. Gaines, W. Lisowski, S. Press, N. Shapiro (Santa Monica, CA, USA); Alen Peacock, J. Leggett, D. Umphress, G. Williams (Texas, USA); M.S. Obaidat, B. Sadoun (New Jersey, USA); С.Н. Расторгуев, Р.Н. Минниханов и др. В их трудах была предложена классическая схема аутентификации операторов КС. Достоверность аутентификации с использованием методов, описанных в трудах вышеперечисленных авторов, имеет допустимые значения вероятности возникновения ошибок первого и второго рода только при определении клавиатурного почерка по ключевой фразе.

Представленные методы применяются в процессе авторизации операторов ключевой системы и не могут быть использованы для скрытого клавиатурного

мониторинга и обнаружения подмены законного пользователя. Это связано с тем, что клавиатурный почерк – динамическая поведенческая биометрическая характеристика человека. Нестабильность почерка операторов объясняется изменением их психофизиологического состояния. Существующие программные реализации методов распознавания клавиатурного почерка характеризуются недостаточной достоверностью идентификации и аутентификации и высокой вероятностью возникновения ошибок первого и второго рода. Вследствие этого актуальна разработка новых моделей, методов, алгоритмов распознавания клавиатурного почерка и их программных реализаций, повышающих точность и качество функционирования систем идентификации и аутентификации.

Таким образом, обоснована необходимость использования аппарата теории вероятностей и математической статистики, в частности теории нормального распределения, для оценки математического ожидания времени удержания клавиш (ВУК) как характеристики клавиатурного почерка оператора. Обоснована необходимость представления ВУК в виде бимодального распределения. Это позволит достичь приемлемого уровня ошибок первого и второго рода при идентификации оператора при малом количестве измерений и даст возможность сократить время как на создание шаблона клавиатурного почерка ОКС, так и на процедуру авторизации. Также доказано, что в связи с использованием метода определения клавиатурного почерка на основе учёта времени удержания клавиш становится возможно определение клавиатурного почерка по свободному смысловому тексту. Это в свою очередь позволяет осуществлять скрытый клавиатурный мониторинг клавиатурного почерка ОКС и обнаруживать подмену законного пользователя.

Предложен метод распознавания клавиатурного почерка по свободному тексту на основе механизма анализа клавиатурного ввода в ключевой системе. Данный метод реализован в алгоритме распознавания клавиатурного почерка по времени удержания клавиш и времени ввода часто употребляемых в языке последовательностей букв (N-грамм). Разработано программное обеспечение постоянного скрытого клавиатурного мониторинга, внедряемое в интерфейс ключевой системы. Разработанная система идентификации оператора ключевой системы информационной инфраструктуры имеет точность в 99% при количестве операторов, зарегистрированных в системе, равном 100.

Целью диссертационной работы является разработка и исследование методов и средств обеспечения поддержки принятия решений о допуске оператора к ключевой системе, основанных на анализе клавиатурного почерка.

Для достижения поставленной цели в работе сформулированы и решены следующие **задачи:**

- анализ и исследование характеристик КП, существующих методов, алгоритмов, моделей и средств определения КП оператора КС;
- разработка математических моделей распознавания КП оператора КС;
- разработка алгоритма распознавания КП по времени удержания клавиш путем постоянного скрытного мониторинга;
- разработка способа хранения и передачи данных о КП;
- разработка метода распознавания КП по свободному тексту;

- разработка и реализация алгоритма распознавания КП по времени удержания клавиш путем постоянного скрытного мониторинга;
- проведение экспериментального исследования подсистемы доступа к КС на основе анализа КП оператора.

Объект исследования – организация и управление доступом к ключевой системе информационной инфраструктуры, осуществляющей управление критически важным объектом (процессом).

Предмет исследования – методы и алгоритмы распознавания КП оператора КС.

Методы исследования. В работе использованы методы теории вероятностей и математической статистики, системного анализа, теории множеств, метрологических методов, методов объектно ориентированного программирования, теории защиты информации.

Достоверность и обоснованность. Теоретические выводы и положения диссертации научно обоснованы и подтверждены результатами экспериментальных исследований автора, актами о внедрении и применении результатов диссертационного исследования.

Научная новизна работы:

1. Предложен метод определения клавиатурного почерка оператора ключевой системы, отличающийся от существующих тем, что распознавание клавиатурного почерка происходит по свободному контрольному тексту и полученный шаблон почерка не зависит от набираемого оператором текста и порядка ввода символов, что обеспечивает возможность применения метода для задач постоянного скрытого клавиатурного мониторинга с целью обнаружения подмены авторизованного законного оператора, определения отклонения психофизиологического состояния оператора ключевой системы от нормального.
2. Разработана математическая модель клавиатурного почерка, отличающаяся от существующих тем, что ВУК представляется в виде бимодального распределения (пересечения двух нормальных распределений), что увеличит до двух раз количество применяемых при распознавании КП характеристик. Разработанная модель применяется в алгоритмах распознавания КП ОКС.
3. Разработана аналитическая модель клавиатурного почерка, позволяющая сравнивать два шаблона клавиатурного почерка.
4. Разработан алгоритм получения шаблона клавиатурного почерка оператора ключевой системы, отличающийся от существующих тем, что при распознавании клавиатурного почерка анализируется время удержания клавиш и время ввода часто встречаемых в языке N-грамм, что обеспечивает возможность определения КП оператора по свободному контрольному тексту. Разработанный алгоритм основан на представлении ВУК в виде пересечения двух нормальных распределений.
5. Разработан алгоритм авторизации оператора КС по КП, представленному в виде бимодального распределения.
6. Разработан алгоритм обнаружения подмены авторизованного ОКС в зависимости от отклонений клавиатурного почерка оператора.
7. Разработан способ представления клавиатурного почерка в ЭВМ.

Положения, выносимые на защиту:

- модель клавиатурного почерка оператора ключевой системы;
- аналитическая модель клавиатурного почерка;
- метод распознавания клавиатурного почерка;
- алгоритм распознавания клавиатурного почерка;
- алгоритм авторизации оператора по клавиатурному почерку;
- алгоритм постоянного скрытного клавиатурного мониторинга с целью обнаружения подмены авторизованного оператора;
- подсистема распознавания клавиатурного почерка, подсистемы принятия решений на основе анализа клавиатурного почерка.

Практическая полезность и реализация результатов работы:

1. Разработан комплекс программ, реализующих алгоритмы и методы распознавания клавиатурного почерка, которые могут использоваться как СКУД к ключевой системе.
2. Разработано программный интерфейс получения доступа к КС реализующего взаимодействие КС и подсистемы защиты доступа, основанной на распознавании клавиатурного почерка.
3. Разработана база шаблонов КП, которая может быть использована в алгоритмах авторизации пользователя и алгоритме скрытного клавиатурного почерка.

Личный творческий вклад автора

Выполнен анализ существующих методов и алгоритмов распознавания клавиатурного почерка, и применяемых в них характеристик КП. Разработана математическая модель КП, основанная на теории нормального распределения. Автором получены все выносимые на защиту положения, сформулированы научные выводы и положения. Исследованы причины возникновения ошибок I и II рода, возникающие при распознавании КП в процессе аутентификации ОКС.

Апробация работы

Основные результаты и положения работы докладывались и обсуждались на следующих конференциях:

1. Научно-техническая конференция «Исследования. Технологии. Инновации» (2010 г., Йошкар-Ола)
2. Первый этап Всероссийского конкурса IT ПРОРЫВ (2010 г., Москва)
3. Четырнадцатые Вавиловские чтения «Россия в глобальном мире: вызовы, потенциалы и перспективы» (2010 г., Йошкар-Ола)
4. Всероссийская научно-практической конференция «Информационные технологии в профессиональной деятельности и научной работе» (Информационные технологии 2011) (2011 г., Йошкар-Ола)
5. Первый Всероссийский фестиваль науки в Республике Марий Эл (2011 г., Йошкар-Ола)
6. ЙО Форум «Форум твоих идей» (2011 г., Йошкар-Ола)
7. Пятнадцатые Вавиловские чтения «Инновационные ресурсы и национальная безопасность в эпоху глобальных трансформаций» (2011 г., Йошкар-Ола)

8. Всероссийская научно-практическая конференция «Информационные технологии в профессиональной деятельности и научной работе» (Информационные технологии 2012) (2012 г., Йошкар-Ола)

9. Конгресс по интеллектуальным системам и информационным технологиям «IS&IT'12» (2012 г., Дивноморское)

10. Третья международная конференция «Автоматизация управления и интеллектуальные системы и среды» (2012 г., Махачкала)

11. Конкурс «Startup Сабантуй!» от Казанского IT-парка (2012 г., Йошкар-Ола)

12. Республиканская ярмарка бизнес-идей и проектов – 2012 (2012 г., Йошкар-Ола)

13. Конгресс по интеллектуальным системам и информационным технологиям «IS&IT'13» (2013 г., Дивноморское)

По итогам исследовательской работы поданы документы для получения Свидетельства государственного образца о регистрации программы для ЭВМ.

Апробация и внедрение результатов диссертационной работы были проведены в ООО «Родэл», Специализированном государственном автономном учреждении Республики Марий Эл «Марийская база авиационной охраны лесов «Авиалесоохрана», ФГБОУ ВПО «Поволжский государственный технологический университет», ФГБОУ ВПО «Марийский государственный университет», ФГБОУ ВПО «Чувашский государственный университет им. И.Н. Ульянова», ФГБОУ ВПО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)». Результаты использовались в проектно-конструкторской деятельности ПГТУ при подготовке и проведении Международной интернет-олимпиады по информатике и программированию (НИР 12.17/12, гос. контракт 12.741.11.0050 от 27 апреля 2012 г.).

Работа выполнена при поддержке программы ФСР МФП НТС «Участник молодежного научно-инновационного конкурса 2012» («У.М.Н.И.К.») № 9955р/14267 от 11 января 2012.

Публикации

Основные результаты диссертационной работы изложены в 16 публикациях, в том числе в 3 статьях в изданиях, рекомендуемых ВАК РФ.

Объем и структура работы

Диссертация состоит из введения, четырех глав, заключения, списка литературы из 120 наименований и одного приложения.

СОДЕРЖАНИЕ РАБОТЫ

Во введении исследовано современное состояние проблемы, обоснована актуальность работы, сформулированы цель и задачи исследования, определены объект и предмет исследования, указаны научная новизна и практические результаты.

В первой главе исследованы известные методы и алгоритмы распознавания клавиатурного почерка, выявлены их основные достоинства и недостатки. Обзор известных работ, а также методик аутентификации пользователей показал недостаточную достоверность аутентификации. Результаты достоверности известных работ не превосходят 90%. Таким образом, возникает задача повышения точности аутентификации операторов по клавиатурному почерку при попытке оператора получить

доступ к ключевой системе и к её ресурсам, которая решается в данном диссертационном исследовании.

В работе дан сравнительный анализ существующих на данный момент систем защиты информации на основании их возможности противодействия подмене законного оператора. В рамках анализа рассматривались парольные, атрибутные, биометрические системы. Выделены основные проблемы указанных систем, главной из которых названа невозможность предотвращения и обнаружения подмены оператора.

Выявлено, что в существующих системах клавиатурный почерк распознается средствами нейронных сетей (B. Sadoun, M.S. Obaidat). При этом не осуществляется обнаружения подмены законного оператора, а происходит формирование шаблона в момент ввода пароля и распознавание шаблона почерка при авторизации в системе и в момент запуска приложений. Доказано, что существующие на данный момент системы контроля и управления доступом не обеспечивают защиты ключевой системы от случайной или намеренной подмены законного оператора злоумышленником. Показано, что для нейронных сетей сложно определить набор входных параметров и архитектуру нейронной сети, которые обеспечивали бы требуемый результат. Также возникают сложности с интерпретацией структуры обученной нейронной сети - как правило, модель используется только в качестве «черного ящика», невозможно определение достаточного для обобщения и последующей успешной работы объема и состава входной выборки.

Доказано, что 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 пользователей автоматизированных систем.

Предложено решение задачи обнаружения подмены оператора и идентификации пользователей путем распознавания клавиатурного почерка. Проведен анализ биометрических параметров характеристик клавиатурного почерка, таких как время удержания клавиши (ВУК) и время между нажатиями клавиш клавиатуры (ВМНК). Предложено использовать непрерывный режим мониторинга работы оператора в течение длительного (потенциально не ограниченного) интервала времени. Предложен метод распознавания КП по ВУК, представленному в виде бимодального распределения, который позволит проводить постоянный скрытый клавиатурный мониторинг и обнаруживать подмены пользователей, что даст возможность блокировать КС при вторжении.

Во второй главе разработаны математическая и аналитическая модели алгоритма распознавания клавиатурного почерка и проведен анализ их адекватности. Показано, что в системах распознавания клавиатурного почерка статистическими данными являются значения времен событий клавиатуры. Признаком клавиатурного почерка выбрано время удержания клавиш, которое соответствует временному интервалу между событиями KeyDown(A) и KeyUp(A), где A – одна из клавиш клавиатуры. Выяснено, что для определения КП оператора при использовании метода постоянного скрытого мониторинга необходим сбор статистики, состоящей из выборки временных значений событий клавиатуры, где элементом выборки будет являться время удержания клавиши. Анализ собранных шаблонов клавиатурных почерков операторов показал, что время удержания клавиш имеет бимодальный закон распре-

деления, т.е. пересечения двух нормальных распределений. Это объясняется тем, что время удержания клавиши увеличивается, если одновременно с ней была нажата другая клавиша (т.е. имеются наложения нажатия клавиш).

Рисунок 1. Бимодальное распределение ВУК

Предложен метод распознавания клавиатурного почерка, основанный на применении вероятностно-статистических методов, в связи с чем возникает необходимость построения среднестатистических шаблонов на основе образцов, предъявленных системе в режиме обучения. Анализ шаблонов КП операторов показал, что ВУК представляет собой бимодальное распределение – пересечение двух нормальных распределений (рис. 1). Выдвинуто предположение, что для каждого из распределений можно применить методы оценки, применяемые при оценке нормальных распределений.

Обоснованием причины возникновения нормального распределения для ВУК является то, что на клавиатурный почерк человека влияет множество случайных факторов: программные и аппаратные задержки (которые являются случайными величинами), движение нервного импульса по нейронам, время отклика мышц человека на сигнал, посланный мозгом, и так далее.

Значит, на клавиатурный почерк влияет множество независимых случайных величин. Эффект их сложения описывается формулой Гаусса. Соответственно с целью уменьшения значения случайных ошибок на выборку необходимо измерять исследуемую величину несколько раз.

Предложено применение в разработанном методе формулы Гаусса в процессе обработки результатов измерения характеристик клавиатурного почерка. Данный подход требует измерения время удержания какой-то конкретной клавиши (обозначим эту величину X) несколько раз. В результате проведенных измерений определяется выборка значений величины (1):

$$X_1, X_2, X_3, \ldots X_N. \quad (1)$$

Для полученный выборки необходимо оценить результат измерений $\bar{X}$, т.е. усредненное значение ВУК клавиши, в которое стремится уложиться «натренированная» рука при нажатии на клавишу. Данное значение не является истинным значением измеряемой величины, поэтому необходимо оценить ошибку измерения. Предположим, что существует возможность определить оценку ошибки ΔX. Тогда результат измерений будет иметь вид (2)

$$\mu = \overline{X} \pm \Delta X. \qquad (2)$$

Так как полученные значения оценки результата измерений и ошибки ΔX не являются точными, результат измерений должен сопровождаться указанием его надежности **P**. Надежностью будем считать вероятность того, что истинное значение ВУК заключено в доверительном интервале выборки, указанном в формуле (2).

Таким образом необходимо, имея выборку (1), найти оценку результата измерений $\overline{X}$, его ошибку ΔX и надежность **P**. Эта задача решается применением теории вероятностей и математической статистики. В задаче измерения ВУК возникающие ошибки подчиняются нормальному закону распределения. Предложено в качестве оценки результатов измерений ВУК рассчитывать среднее значение всех элементов собранной для конкретной клавиши выборки (3)

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad (3)$$

где **N** – число измерений.

Значит, для выборки в **N** измерений времени удержания, наиболее вероятным значением измеряемой величины будет ее среднее значение (арифметическое). Полученное среднее значение ВУК стремится к истинному значению μ измеряемой величины $\overline{X}$ при увеличении числа измерений, т.е. $N \to \infty$.

Среднеквадратичной ошибкой среднего арифметического называется величина

$$S_{\overline{X}} = \sqrt{\frac{\sum (\overline{X} - X_i)^2}{N(N-1)}} = \frac{S}{\sqrt{N}}. \qquad (4)$$

Точность оценки возрастает при увеличении числа измерений. Ошибка $S_{\overline{X}}$ позволяет оценить точность, с которой рассчитано среднее значение ВУК. Обосновано, что предложенный способ расчета ошибок имеет надежность 0,68, когда величина времени удержания клавиш измерялась не менее 30 раз.

При малом количестве измерений вводится специальный коэффициент Стьюдента **t** для расчета абсолютной ошибки, который зависит от надежности **P** и числа измерений **N**,

$$\Delta X = S_{\overline{X}} \cdot t, \qquad (5)$$

где ΔX – абсолютная ошибка для данной вероятности попадания в доверительный интервал.

В ы в о д : величина среднеквадратичной ошибки применима для расчета вероятности, с которой истинное значение ВУК находится в заданном интервале вблизи среднего арифметического. При $N \to \infty$ $S_{\overline{X}} \to 0$, т.е. размер интервала, в котором с заданной доверительной вероятностью находится истинное значение времени удержания клавиши μ, стремится к нулю с увеличением числа измерений.

В результате проведенного исследования показано, что, увеличивая n, можно получить результат с любой степенью точности, когда вероятность возникновения ошибок будет стремиться к нулю. Однако невозможно добиться стопроцентной точности, так как при достижении уровня случайных ошибок, равного уровню систематических ошибок, точность перестанет увеличиваться с увеличением числа измерений. Известно, что последующее увеличение числа измерений не будет иметь результата, т.к. окончательная точность результата будет измеряться уровнем система-

тической ошибки. Определив значение систематической ошибки, выбирают приемлемый уровень случайных ошибок.

Показано, что выбор порога надежности осуществляется исходя из практических соображений той ответственности, с какой делаются выводы о параметрах. Обычно в БСИ используют 99,9%-й порог вероятности попадания в доверительный интервал.

Уровень вероятности попадания в доверительный интервал показывает, какую максимальную вероятность возникновения ошибки первого рода система считает допустимой. Уменьшение уровня вероятности попадания в доверительный интервал, иначе говоря, ужесточение условий тестирования гипотез, увеличивает вероятность ошибок второго рода. Следовательно, выбор уровня вероятности попадания в доверительный интервал должен осуществляться с учетом возможного ущерба от возникновения ошибок первого и второго рода.

Учитывая, что выборочное распределение некоторой статистики, например средней арифметической величины, при достаточно больших объемах выборок (например, 100 и более элементов) имеет нормальную форму, можно записать выражение

$$-t \le \frac{\overline{X}-\mu}{S_{\overline{X}}} \le t. \qquad (6)$$

Это выражение означает, что вероятность того, что средняя $\overline{X}$, найденная по выборке, отклонится случайным образом от центра μ на какую-то долю квадратической ошибки $S_{\overline{X}}$, может быть оценена через нормированное значение по таблицам нормального распределения. Отсюда можно утверждать, что средняя μ находится с этой вероятностью в интервале

$$\overline{X} - tS_{\overline{X}} \le \mu \le \overline{X} + tS_{\overline{X}}. \qquad (7)$$

Для каждой выборки ввода конкретного символа будут иметься два таких уравнения в связи с бимодальностью распределения времени удержания клавиш. Величины $\overline{X}$ и $S_{\overline{X}}$ определяют по выборке, а t зависит только от одного из трех значений вероятности попадания в доверительный интервал 0,95; 0,99; 0,999, принимая величины 1,96; 2,58; 3,29.

Таким образом, задав для выбранного доверительного интервала определенное значение **P**, вычисляют необходимое число измерений времени удержания клавиш, обеспечивающее минимальное влияние случайных ошибок на точность результата.

Следовательно, можно определить, сколько раз должна быть нажата конкретная клавиша, для того чтобы собрать статистику, характеризующую усредненное ВУК. Соответственно, для того чтобы определить клавиатурный почерк оператора КС, нужно собрать данные о средних значениях времени удержания всех используемых клавиш (например, 33 клавиш, соответствующих буквам русского алфавита). В результате, когда система работает в режиме обучения, необходимо получить 33 выборки, состоящие из n элементов. В связи с бимодальностью распределения клавиатурного почерка каждая выборка ВУК разделяется на две подвыборки, в зависимости от присутствия или отсутствия наложений при нажатии клавиши. Средние значения каждой из подвыборок будут сохранены в шаблоне почерка оператора. Но данное предположение соответствует некоторому идеальному случаю с равномерным распределением букв в вводимом оператором тексте.

Проведенный частотный анализ русскоязычных текстов показал, что буквы в русском тексте встречаются с разной вероятностью. Например, самая «встречаемая» буква «О» встречается с вероятностью 0.090, а самая «редкая» буква «Ф» встречается с вероятностью 0.002. В результате проведенных вычислений установлено, что при значении n = 10 для сбора статистических данных по букве «О» нам понадобится ввести текст длиной 111 символов. А для получения выборки ВУК «Ф» нам придется ввести текст, состоящий из 5000 символов, что соответствует 3 страницам текста формата А4 или приблизительно 30 минутам непрерывной печати текста. Можно сделать вывод, что получение шаблона клавиатурного почерка становится долгим и трудоемким процессом, что не является желательным в системах, обеспечивающих безопасность информации. Таким образом может быть проведен частотный анализ для любого языка, с которым будет работать система.

Для решения проблемы изменчивости клавиатурного почерка, а также для уменьшения влияния случайных погрешностей предложено использовать закон распределения Гаусса при построении математической модели клавиатурного почерка.

Также предложено дополнить формулу сравнения текущего и шаблонного почерков коэффициентами веса каждого символа, которые будут устанавливать большую значимость для чаще встречаемых букв. Данные коэффициенты будут устанавливаться в зависимости от частоты встречаемости буквы в свободном смысловом тексте, написанном на русском языке. Это оправданно тем, что при обучении слепому методу печати у человека появляется так называемая «двигательная память», когда он быстрее запоминает расположение тех букв, которые встречаются ему чаще и вводит эти буквы более уверенно.

В третьей главе представлена разработка алгоритмов, реализующих данный метод. Разработан ***алгоритм регистрации*** **КП** (см. диаграмму деятельности – рис. 2). Для выявления усредненных значений времени событий клавиатуры используется вероятностно-статистический метод, требующий сбора статистики, состоящей из выборки временных значений, где элементом выборки будет являться время удержания клавиши. Алгоритм основан на разработанной математической модели.

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

Далее инициализируется динамический трехмерный массив KeyEventsArr, в котором сохраняются события клавиатуры в следующем виде (см. табл. 1). Массив заполняется до тех пор, пока не будет нажато достаточное (задаваемое администратором или установленное по умолчанию) количество клавиш.

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. На многоядерных системах используется функция SetThreadAffinityMask, чтобы указать родственность процессора для системы.

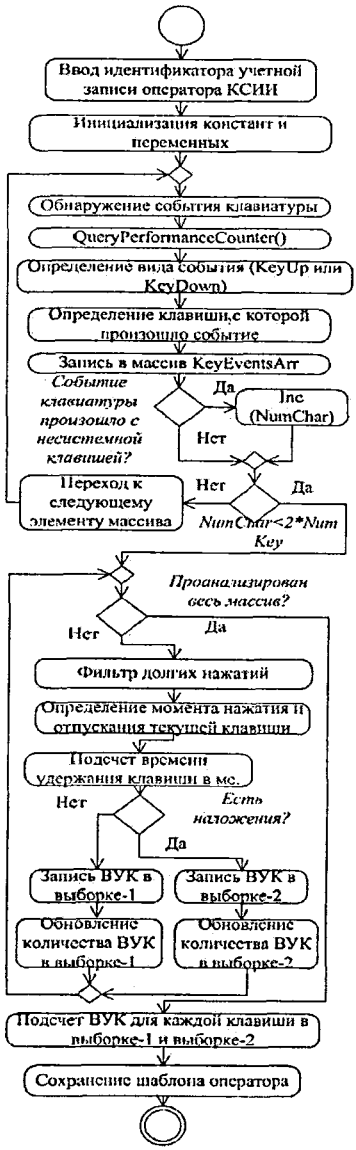

Рисунок 2. Диаграмма деятельности алгоритма регистрации КП

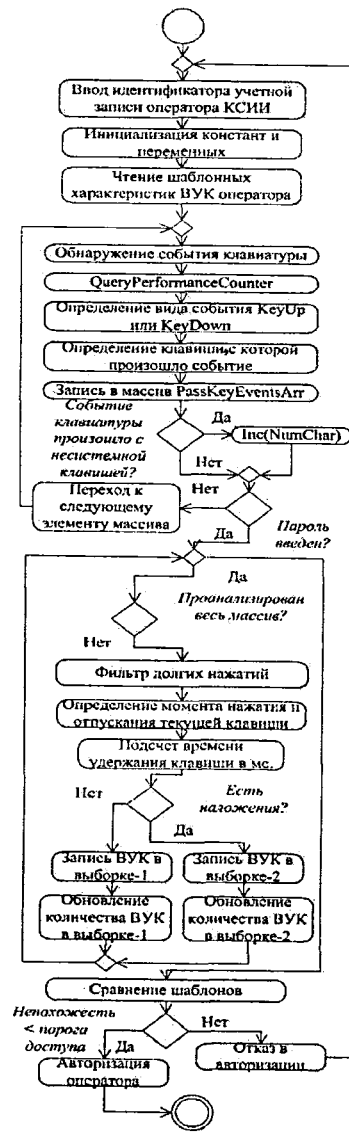

Рисунок 3. Диаграмма деятельности алгоритма аутентификации

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

Таблица 1. **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

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 шаблон КП сохраняется в учетной записи оператора. Шаблонный информационный файл оператора имеет следующую структуру (табл. 2).

Таблица 2. **Структура шаблона КП**

Клавиша	ВУК-1	ВУК-2	Кол-во	Кол-во	Выборка нормального распределения-1	Выборка нормального распределения-2
А	65	104	!	!		
Б	69	106	!	28		100;106;114;106;110; ….
…						
я	81	114	3	2	80;82;81;	112; 116;

Так как свободный текст, по которому проводится обучение или дальнейшие процедуры аутентификации и идентификации, имеет разную вероятность встречаемости разных букв и символов, то 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» и ограничить количество необходимых для обучения символов, с последующей возможностью дообучения.

В поле «выборка» сохраняются ВУК, разделяемые знаком «;», в поле «количество» заносится количество элементов в выборке. При достижении достаточного количества элементов в выборке подсчитывается ВУК, выборка очищается, в поле «кол-во» заносится символ «!», который используется для обозначения факта окончания подсчёта ВУК этой клавиши. Дальше элементы в эту выборку заноситься не будут. Если в процессе обучения достаточное количество элементов 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 шаблон. При аутентификации и идентификации преимущество будут иметь законченные ВУК клавиш, им будут присвоены большие коэффициенты влияния, незаконченным – меньшие. Для разделения предложено разбить алфавит системы на несколько групп, в зависимости от вероятности встречаемости в текстах.

Предложен алгоритм аутентификации и последующей авторизации (см. рис. 3). В данном алгоритме сравнение происходит по принципу «один к одному», поэтому загружается шаблон конкретного оператора, идентификатором которого он представился. В случае успешной аутентификации происходит процесс авторизации, в противном случае – отказ.

Главное отличие данного алгоритма от предыдущего состоит в том, что 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 предложено настраивать путем анализа локальной базы шаблонов клавиатурных почерков операторов, имеющих доступ к конкретной КСИИ.

Сравнение текущего образца КП оператора КСИИ с шаблонным происходит путём расчета меры Евклида для каждой клавиши, при этом ВУК-1 и ВУК-2 сравниваются как отдельные элементы.

$$\mathbf{M} = \sqrt{\sum_{i=1}^{N} (\mathbf{A_i} - \mathbf{B_i})^2}. \qquad (8)$$

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

Предложен **алгоритм постоянного скрытного клавиатурного мониторинга** и обнаружения подмены авторизованного оператора (см. рис. 4). Массив DynamicKeyEvetsArr имеет размеры от 10 до 40 элементов и может быть выбран, например, в зависимости от интервала копирования оператора. Интервал копирования – это число символов, которые могут быть напечатаны в точности после однократного просмотра текста. Солтхаус установил, что интервал копирования в обычной ситуации перепечатки у опытного наборщика составил в среднем 14,6 символов. Текущие характеристики КП определяются по данным, хранящимся в массиве.

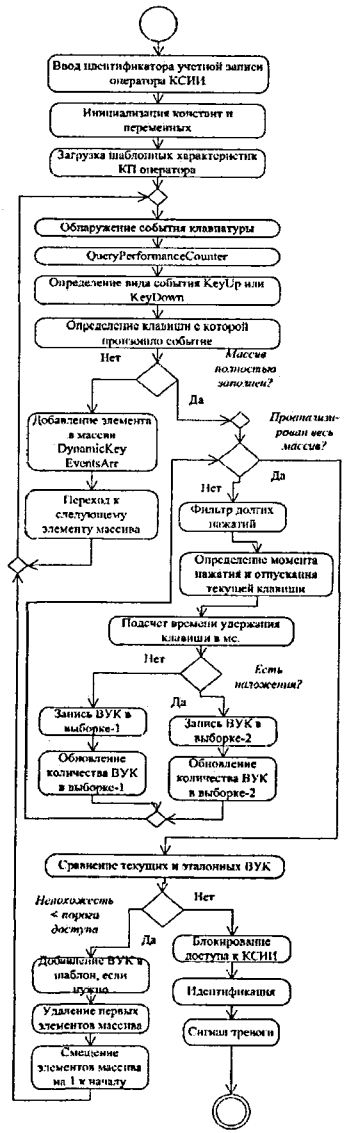

Рисунок 4. Диаграмма деятельности алгоритма постоянного скрытного мониторинга КП

Массив динамически обновляется при вводе текста. Элементы, введенные раньше, удаляются, и ВУК рассчитывается динамически, по добавленным новым элементам. Текущие значения ВУК сравниваются с шаблонными ВУК оператора. При обнаружении подмены законного оператора измеряется частота счётчика высокого разрешения, и если она изменилась, то среднее 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 шаблонов КП.

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 выборка очищается и в поле «количество» информационного файла ставится метка «!». Данный метод можно также использовать и при переобучении всего шаблона почерка оператора, в случае изменений клавиатурного почерка, например, вызванных совершенствованием оператором техники печати.

В четвертой главе разработана архитектура подсистемы распознавания клавиатурного почерка и реализован интерфейс с ключевой системой. Разработана структура ПО, реализующего подсистему распознавания КП.

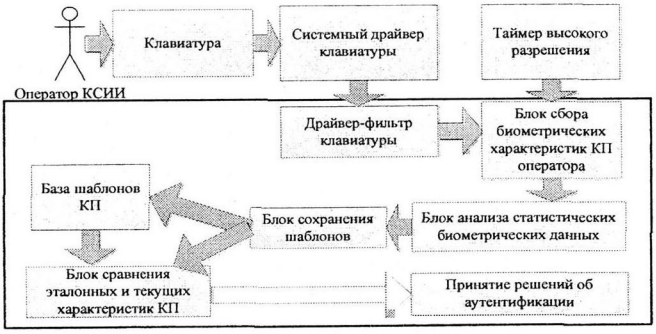

Рисунок 5. Архитектура подсистемы распознавания клавиатурного почерка

Разработанная подсистема состоит:

1) из модифицированного драйвер-фильтра клавиатуры. Применение фильтра позволяет уменьшить задержки, возникающие в ОС Windows;

2) блока сбора биометрических данных – сохраняет выборку времени удержания клавиш, состоящую из времени наступления событий нажатия и отпускания каждой клавиши.

Функция QueryPerformanceCounter возвращает в переменную типа LARGE_INTEGER значение таймера, «набежавшее» на данный момент (в тиках). Для получения времени удержания клавиши разность значений таймера между отпусканием и нажатием клавиши делится на частоту кварцевого генератора;

3) блока анализа статистических данных – на основании выборки строит шаблон клавиатурного почерка, используя математическую модель, основанную на нор-

мальном распределении. На этом этапе вычисляется математическое ожидание времени удержания для каждой клавиши;

4) блока сравнения клавиатурного почерка – сравнивает шаблон почерка аутентифицируемого оператора КС и сравнивает его с шаблоном почерка оператора, хранящимся в базе. В качестве критерия для сравнения шаблонов служит Евклидово расстояние. При совпадении шаблонов в пределах некоторого допустимого порога отклонения система принимает решение о разрешении авторизации оператора КС.

Проведен анализ результатов использования разработанного алгоритма при распознавании операторов ключевых систем на основе собранной базы шаблонов клавиатурных почерков. Планирование экспериментальных исследований заключалось: а) в выборе 1000 операторов, имеющих стаж работы на ЭВМ не менее 5 лет, б) выборе контрольного текста для эксперимента объёмом 1000 символов, в) получении выборки времени удержания клавиш при наборе операторами текста, г) обработке разработанной системой собранных данных и анализе результатов распознавания. Произведено сравнение (499500 операций сравнения) собранных шаблонов операторов друг с другом с различными значениями порога доступа. Был выбран наилучший порог доступа для включения в формулу меры Евклида, при процедурах сравнения шаблонов КП. Порог доступа равен = 0,8.

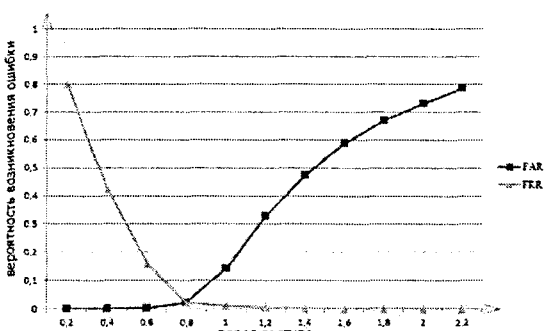

Рисунок 6. Зависимость частоты возникновения ошибок первого и второго рода от порога доступа

Анализ подтверждает, что коэффициентом равной вероятности ошибок 1-го и 2-го рода (ERR) идентификации оператора по клавиатурному почерку в разработанной системе составляет 98%, что имеет лучший показатель по сравнению с известными системами (90%).

Разработанная система распознавания личности по клавиатурному почерку обеспечивает точность идентификации с величиной ошибки первого рода в 1% для 100 операторов одной ключевой системы информационной инфраструктуры.

В **заключении** сформулированы основные преимущества разработанных моделей, методов и алгоритмов распознавания клавиатурного почерка оператора КС.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Проведен анализ характеристик КП, существующих методов, алгоритмов, моделей и средств определения КП оператора КС.

2. Предложен метод определения клавиатурного почерка оператора ключевой системы, отличающийся от существующих тем, что распознавание клавиатурного почерка происходит по свободному контрольному тексту и полученный шаблон почерка не зависит от набираемого оператором текста и порядка ввода символов, что

обеспечивает возможность применения метода для задач постоянного скрытого клавиатурного мониторинга с целью обнаружения подмены авторизованного законного оператора и определения отклонения от нормального психофизиологического состояния оператора ключевой системы.

3. Предложена математическая модель клавиатурного почерка, отличающаяся от существующих тем, что ВУК представляется в виде пересечения двух нормальных распределений, что увеличит в два раза количество применяемых при распознавании КП характеристик.

4. Предложена аналитическая модель клавиатурного почерка, позволяющая сравнить два шаблона клавиатурного почерка.

5. Предложен алгоритм получения шаблона клавиатурного почерка оператора ключевой системы, отличающийся от существующих тем, что при распознавании клавиатурного почерка как характеристика почерка используются время удержания клавиш, представленное в виде пересечения двух нормальных распределений, и время ввода часто встречаемых в языке n-грамм, что обеспечивает возможность определения клавиатурного почерка оператора ключевой системы по свободному контрольному тексту.

6. Предложен алгоритм авторизации оператора ключевой системы по клавиатурному почерку.

7. Предложен алгоритм постоянного скрытного клавиатурного мониторинга с целью обнаружения подмены авторизованного пользователя.

8. Разработан способ представления и хранения клавиатурного почерка в ЭВМ.

9. Разработан комплекс программ, реализующих алгоритмы и методы распознавания клавиатурного почерка, которые могут использоваться как СКУД к ключевой системе.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК РФ

1. Савинов, А.Н. Анализ решения проблем возникновения ошибок первого и второго рода в системах распознавания клавиатурного почерка / А.Н. Савинов, В.И. Иванов // Вестник Волжского университета имени В.Н. Татищева: науч.-теор. журнал. Серия «Информатика». – Тольятти: Волжский университет им. В.Н. Татищева, 2011. – Вып. 18. – С. 115-119.

2. Савинов, А.Н. Математическая модель механизма распознавания клавиатурного почерка на основе Гауссовского распределения / А.Н. Савинов, И.Г. Сидоркина // Известия Кабардино-Балкарского научного центра РАН. Вып. 1. – Нальчик: Кабардино-Балкарский научный центр РАН, 2013. – С. 26-32.

3. Савинов, А.Н. Три алгоритма управления доступом к КСИИ на основе распознавания клавиатурного почерка оператора / А.Н. Савинов, И.Г. Сидоркина // Вестник Чувашского университета. – Чебоксары: ЧГУ им. И.Н. Ульянова, 2013. – № 3. – С. 293-301.

Публикации в иных изданиях

4. Савинов, А.Н. Распознавание клавиатурного почерка пользователя ключевой системы / А.Н. Савинов // Информационные технологии в профессиональной деятельности и научной работе: сб. материалов Всерос. науч.-практ. конф.: в 2 ч. Ч. 2. – Йошкар-Ола: МарГТУ, 2011. – С. 21-25

5. Савинов, А.Н. Признаки, методы и алгоритмы биометрического распознавания личности по клавиатурному почерку / А.Н. Савинов, И.Г. Сидоркина // Четырнадцатые Вавиловские чтения -2011: материалы Всерос. междисципл. науч. конф. с междунар. участием: в 2 ч. / под общ. ред. проф. В.П. Шалаева. – Йошкар-Ола: МарГТУ, 2011. – Ч.2. – С. 342-345.

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

7. Савинов, А.Н. Использование биометрической системы распознавания клавиатурного почерка при выполнении технологических требований обеспечения безопасности ключевой системы / А.Н. Савинов // Инновационные ресурсы и национальная безопасность в эпоху глобальных трансформаций. Пятнадцатые Вавиловские чтения»; постоянно действующая Всерос. междисципл. науч. конф. с междунар. участием: в 2 ч. / редкол.: В.П. Шалаев и др. – Ч. 2. – Йошкар-Ола: МарГТУ, 2012. – С. 282-284.

8. Савинов, А.Н. Распределения Гаусса при анализе времени удержания клавиш при разработке математической модели клавиатурного почерка / А.Н. Савинов // Информационные технологии в профессиональной деятельности и научной работе: сб. материалов Всерос. науч.-практ. конф. с междунар. участием: в 2 ч. – Йошкар-Ола: МарГТУ, 2012. – С. 113-119.

9. Савинов, А.Н. Анализ клавиатурного почерка для ключевой системы с использование закона Гаусса / А.Н. Савинов // Информатика и вычислительная техника: сб. науч. трудов 4-й Всерос. науч.-техн. конф. аспирантов, студентов и молодых ученых: в 2 т. Т. 1 / под ред. Н. Н. Войта. – Ульяновск: УлГТУ, 2012. – С. 211-217.

10. Савинов, А.Н. Анализ причин возникновения ошибок первого и второго рода в системах авторизации, основанных на распознавании клавиатурного почерка / А.Н. Савинов // Программные системы и вычислительные методы. – М.: Nota bene, 2012. – № 1. – С. 54-59.

11. Савинов, А.Н. Распределение Гаусса при анализе времени удержания клавиш при разработке математической модели клавиатурного почерка / А.Н. Савинов, И.Г. Сидоркина // Труды конгресса по интеллектуальным системам и информационным технологиям «IS&IT'12»: в 4 т. – М.: Физматлит, 2012. – Т.2. – С. 145-152.

12. Савинов, А.Н. Нормальное распределение при анализе клавиатурного ввода при разработке математической модели клавиатурного почерка / А.Н. Савинов, И.Г. Сидоркина // Материалы третьей международной конференции «Автоматизация управления и интеллектуальные системы и среды». 9-15 октября, Махачкала, Россия. Т. 2. – Нальчик: Издательство КБНЦ РАН, 2012. – С. 115-119.

13. Савинов, А.Н. Анализ представления клавиатурного почерка человека в виде нормального распределения случайных величин / А.Н. Савинов // Новые информационные технологии и системы: труды X Междунар. науч.-техн. конф. – Пенза: Изд-во ПГУ, 2012. – С. 200-203.

14. Савинов, А.Н. Решение проблемы измерения времени удержания клавиш при разработке системы анализа клавиатурного почерка / А.Н. Савинов, И.Г. Сидоркина // Труды III-й Международной науч.-практ. конф. «ИКТ: образование, наука, инновации». – Алматы: МУИТ, 2013. – С. 328-333.

15. Савинов, А.Н. Анализ способов измерения времени в ОС WINDOWS при разработке системы распознавания клавиатурного почерка / А.Н. Савинов // Информационные технологии в профессиональной деятельности и научной работе: сб. материалов Всерос. науч.-практ. конф. с междунар. участием: в 2 ч. Ч. 2. – Йошкар-Ола: ПГТУ, 2013. – С. 138-144.

16. Савинов, А.Н. Представление времени удержания клавиш в виде бимодального распределения при распознавании клавиатурного почерка / А.Н. Савинов, И.Г. Сидоркина // Труды Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'13»: в 4 т. – М.: Физматлит, 2013. – Т. 1. – С. 93-98.

ПЕРЕЧЕНЬ ИСПОЛЬЗОВАННЫХ СОКРАЩЕНИЙ

ВМНК – время между нажатием клавиш
ВУК – время удержания клавиши
КП – клавиатурный почерк
КС – ключевая система информационной инфраструктуры
ОКС – оператор ключевой системы
СКУД – система контроля и управления доступом

Подписано в печать 18.10.2013. Формат 60×84/16.
Бумага офсетная. Печать офсетная. Усл. печ. л. 1,0. Тираж 100 экз. Заказ № 5208.
Редакционно-издательский центр ПГТУ. 424006 Йошкар-Ола, ул. Панфилова, 17