ПРОГРАММИРОВАНИЕ, 2022, № 6, с. 31–45

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.85.056.5:004.93'1

СКРЫТЫЙ МОНИТОРИНГ ПОЛЬЗОВАТЕЛЯ В ДИСТАНЦИОННОЙ ОБРАЗОВАТЕЛЬНОЙ СИСТЕМЕ НА ОСНОВЕ КЛАВИАТУРНОЙ ДИНАМИКИ

© 2022 г. Е. А. Кочегурова[a,*] (ORCID: 0000-0003-4473-528X),
Р. П. Затеев[a,] (ORCID: 0000-0001-8852-0737)**

[a] Национальный исследовательский Томский политехнический университет
634050, г. Томск, пр. Ленина., д. 30, Россия
**E-mail kocheg@mail.ru*
***E-mail: roma-zateev@mail.ru*

Поступила в редакцию 22.06.2022 г.
После доработки 04.07.2022 г.
Принята к публикации 06.07.2022 г.

Пандемия ускорила развитие дистанционного обучения, важную роль в котором играют онлайн-тесты и экзамены. Особую актуальность при онлайн-тестировании имеет обнаружение подмены личности тестируемого и других фактов академического мошенничества. Способом противодействия несанкционированному доступу может быть непрерывная биометрическая (поведенческая) аутентификация. В работе предложена технология проверки легитимности тестируемого на основе его клавиатурной динамики в режиме скрытого мониторинга. Создано программное приложение для сбора, актуализации образцов клавиатурного почерка пользователей домена и непрерывной аутентификации личности. Показана эффективность сокращения размерности пространства клавиатурных признаков на основе частотности букв алфавита. Традиционные показатели эффективности (FAR, FRR, ERR, ROC, DET) заметно улучшены уже при оценке только метрических расстояний. Например, универсальная ошибка ERR снизилась с 10.1% до 0.79% и сопоставима с оценками kNN-метода для оптимальных значений параметров этого метода.

DOI: 10.31857/S0132347422060048

1. ВВЕДЕНИЕ

Пандемия COVID-19 оказала огромное влияние на функционирование и развитие практически всех сфер жизни. Для снижения инфицирования многие компании и организации ввели удаленный режим работы и другие ограничительные меры. И тем самым ускорили развитие процессов цифровизации, дистанционного обучения, телемедицины, интернет-торговли и других процессов.

Но положительный технологический толчок развития ИТ-отрасли привел к неизбежному росту киберпреступности. В год начала пандемии (2020) и в мире, и в России зафиксирован взрывной рост числа преступлений в сфере компьютерной информации. И согласно данным «АНО “Цифровая экономика” за 2020 год в России зафиксировано 363 тыс. киберпреступлений, что на 77% больше, чем за предыдущий год» [1]. Согласно статистическим данным группы компаний InfoWatch наряду с увеличением количества атак хакерских группировок, возникли новые риски, связанные с удаленной работой [2]. В 2021 г. в мире зафиксирован резкий рост умышленных утечек информации (82% от общего количества) и утечек от действий внешних киберпреступников (до 63%). А общая доля России во всемирном количестве утечек информации довольно внушительна и составляет 16.9%.

В 2021 году по статистике компании Ivideon число кибератак по сравнению с предыдущим годом выросло на 40% в мире и на 54% в России.

И хотя обеспечение информационной безопасности всегда являлось одной из ключевых задач любой организации, в нынешних условиях эта проблема особенно актуальна.

Сфера образования относится к наиболее уязвимым с точки зрения кибербезопасности – ввиду ментальности контингента обучаемых и наличия огромных массивов конфиденциальной информации. Это привлекает университеты для хакерских атак. Только за сентябрь 2021 г. произошло более 10 инцидентов с участием вирусов-вымогателей персональных данных о студентах и преподавателях с последующим предложением о выкупе [3].

С другой стороны, пандемия привела к критическому сбою функционирования традиционных систем университетского и школьного образования. И по данным ЮНЕСКО [4] “кризис затронул почти 1.6 миллиарда учащихся в более чем 190 странах на всех континентах. Закрытие школ и других образовательных учреждений коснулось 94% мирового контингента учащихся”. И хотя в сентябре 2021 года [5] число частично или полностью закрытых образовательных учреждений уменьшилось до 7.5%, но формы и способы подачи знаний значительно трансформировались в сторону дистанционных технологий. Однако готовность университетов к удаленному обучению на сегодня невысокая и функционально ограниченная. В университетском образовании online обучение зачастую сводится к вебинарам посредством видео/аудио конференций и работе на той или иной платформе электронного образования (Moodle, WebTutor, iSpring Learn и др.).

Онлайн-тесты и экзамены имеют большое значение в электронном обучении. Тесты позволяют преподавателю получать обратную связь от студента, оценивать его знания и совершенствовать обучение. Однако иногда студенты используют ряд методов академического мошенничества во время онлайн-тестов, включая выдачу себя за другое лицо [6, 7]. И на сегодня не существуют простых методов обнаружения таких подмен.

2. ОБ АКАДЕМИЧЕСКОЙ ЧЕСТНОСТИ ОНЛАЙН-ОБУЧЕНИЯ

Академическая честность – это проблема не только сферы образования, но и всего общества. Академические нарушения негативно влияют на качество образовательной среды, приобретенные студентами компетенции и общий имидж университета.

Повсеместный перевод вузовского образования в онлайн-формат неизбежно активизировал многие традиционные формы академического мошенничества: плагиат, фальсификация результатов, несанкционированное сотрудничество, подмена личности и пр. Кроме организационных средств, противодействия такому мошенничеству (университетский Кодекс поведения студентов, санкции администрации за академические нарушения), можно выделить методические рекомендации для преподавателей. Для онлайн-занятий рекомендуется:

– повысить индивидуальность и конкретность заданий;

– снизить фактологическую часть заданий и повысить концептуальную для развития мышления;

– увеличить число синхронных проверок ответов в онлайн-режиме;

– увеличить количество мелких и простых заданий, а для больших использовать свободный формат, например эссе.

Предложенные [8] меры частично снижают нарушения образовательной этики, но заметно увеличивают нагрузку на преподавателя.

Способом противодействия несанкционированному сотрудничеству и подмены личности может быть непрерывная аутентификация (НА) студента в скрытом режиме.

3. ВОПРОСЫ КЛАВИАТУРНОЙ АУТЕНТИФИКАЦИИ И ИДЕНТИФИКАЦИИ

Нередко для защиты компьютерной системы от несанкционированного доступа используется двухэтапный процесс верификации:

– первичная идентификация – установление личности пользователя, т.е. подтверждение легитимности авторизованного пользователя;

– динамическая (непрерывная) аутентификация, т.е. непрерывное подтверждение личности легитимного пользователя.

3.1. Методы аутентификации

Аутентификация – это процесс сравнения данных, представленных пользователем, с учетными данными, хранящимися в базе данных служб каталогов. При их совпадении пользователь получает доступ к защищенным ресурсам, при несовпадении – доступ запрещен [9].

Существует несколько методов аутентификации пользователя, рис. 1. Методы можно разделить на три основные категории, исходя из следующих парадигм [10]:

– что вы знаете (например, пароль, PIN-код). Аутентификация на основе знаний;

– чем вы владеете (например, токен, смарт-карта). Аутентификация на основе атрибутов

– кем вы являетесь. Физиологическая и поведенческая биометрия.

Парадигма “кем вы являетесь” связана с биометрическими признаками: физиологическими (отпечаток пальца, лицо, радужная оболочка глаза и пр.) и/или поведенческими (рукописный или клавиатурный почерк, походка, и др.) [19].

Распознавание пользователя на основе клавиатурной динамики довольно привлекательно для организации и имеет более низкую стоимость в сравнении с другими биометрическими методами, поскольку не требуется дополнительного оборудования. Кроме стандартной клавиатуры, требуется высокоэффективное программное приложение. Помимо точности, у этого метода

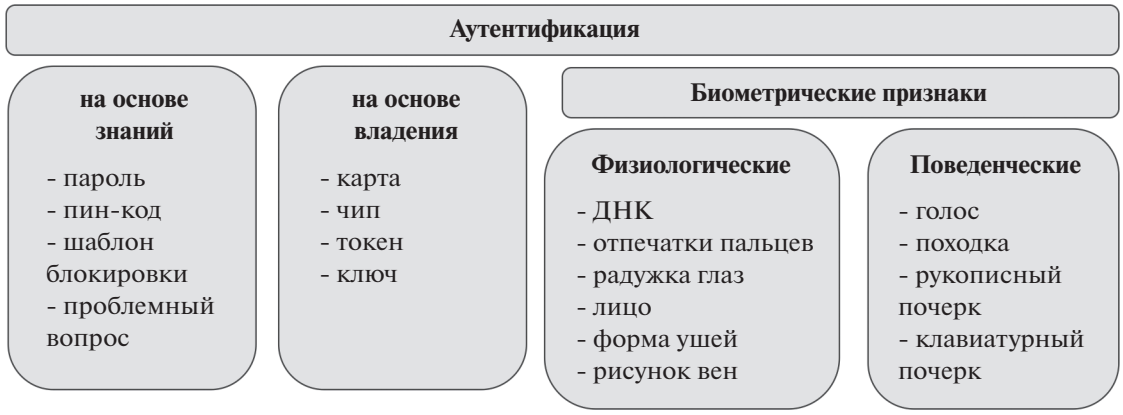

Рис. 1. Методы аутентификации личности.

есть преимущество в виде распознавания пользователей в фоновом, т.е. в скрытом и комфортном для человека режиме. Индивидуальные нажатия клавиш, ритм, скорость набора определяют клавиатурный почерк (КП) пользователя. Все поведенческие характеристики, в том числе и КП, могут постепенно меняться со временем, однако вероятность украсть или имитировать эти данные существенно ниже, чем у физиологических. Поэтому, КП можно использовать для идентификации и аутентификации пользователя [11, 12]. Преимущества, недостатки и примеры вышеописанных методов аутентификации кратко сведены в табл. 1.

3.2. Режимы аутентификации

Наиболее обоснованный для системы распознавания и комфортный для пользователя способ идентификации личности – это постоянный и скрытый мониторинг динамики его работы. И по данным глобального опроса по безопасности IBM (2018 г.) 44% респондентов считают биометрию самым безопасным методом аутентификации, а 65% обнаружили, что биометрия упрощает процесс аутентификации [13].

КП, как поведенческая биометрическая характеристика, является динамической. Она включает условно постоянную и случайную компоненты. Постоянная – обусловлена физиологией человека, его умением, способностями и навыком при работе на клавиатуре. Случайная компонента зависит от психоэмоционального состояния человека.

Для распознавания динамические поведенческие характеристики КП более сложны, чем физиологические. Но именно поведенческие характеристики сложны для подделки и подмены пользователя [10]. Что повышает эффективность для обнаружения самозванцев.

В зависимости от типа создаваемого текста, фиксированного или произвольного, можно выделить два основных режима аутентификации [14]:

– статическая (первичная или по событию);

– динамическая (непрерывная).

Именно НА позволяет организовать скрытую проверку личности пользователя в течение всего сеанса работы в любом программном приложении. Система распознавания фиксирует клавиатурные нажатия и сопоставляет их с уже имеющимся шаблоном пользователя [6, 11, 17–19].

Таблица 1. Характеристики методов идентификации

Метод	Достоинства	Недостатки	Пример
Парольный	1. Простая реализация 2. Однозначное распознавание	1. Может быть забыт или украден	1. Пароль 2. ПИН-код
Атрибутный	1. Простая реализация 2. Не требует затрат	1. Может быть потерян или украден	1. Ключ 2. Смарт-карта 3. Токен
Биометрический	1. Уникальность 2. Невозможно забыть/потерять	1. Стоимость реализации 2. Изменчивость данных независимо от человека	1. Отпечаток пальца 2. Голос 3. Клавиатурный почерк

Таблица 2. Исследования динамической идентификации

Год	Ссылка, автор	Параметр	Метод	Эффективность
2005	[25] Gunetti	FT	Расстояние, R/A	FAR-0.005%, FRR-5%
2010	[32] Shimshon		Кластеризация	FAR 3.47% FRR 0%
2011	[33] Messerman		Статистические	FAR-2.02%, FRR-1.84%
2011	[37] Solami		Кластеризация	Точность 100%
2013	[27] Alsultan	ди-граф	Смешанная (Fusion)	FAR-21%, FRR-17%
2014	[35] Ahmed	ди-граф	Нейронные сети	FAR-0.015%, FRR-4.82%
2014	[39] Antal	DT, FT	Статистические Метод опорных векторов Нейронные сети Дерево решений	Точность 93.04%
2014	[40] Locklear		Статистические	EER 4.55–13.37%
2015	[41] Kang	DT, FT	Кластеризация, Расстояние	EER 3.8%
2015	[42] Matsubara	ди-граф, DT	Расстояние	Точность 99%
2016	[23] Morales	ди-граф, n-граф	kNN, Расстояние	Точность 90%
2017	[31] Alsultan	ди-граф, DT	Метод опорных векторов	FAR 0.169, FRR 0.423
2017	[36] Goodkind	Contextual features	Наивный Байес	Точность 82.2%
2017	[30] Ali		kNN-метод	EER 3.7%
2021	[34] Chang	DT, FT	CNN-GRU	Точность 99% EER 0.0690

Статическая аутентификация может дополнять первичный вход в систему, либо активизируется при возникновении подозрений в злоумышленнике [15, 16].

Оба решения и статическая, и непрерывная аутентификации обеспечивают второй уровень защиты, когда пользователь уже находится в системе. Но при этом динамическая аутентификация нацелена на постоянную проверку легитимности пользователя или его психоэмоционального состояния.

Целью данного исследования является аутентификация личности пользователя на основе непрерывного мониторинга особенностей динамических характеристик его КП в условиях онлайн-тестирования. Для реализации данной цели работа сфокусирована на следующих задачах:

– сбор и актуализация динамических образцов КП для пользователей домена;

– расширение известных подходов статической идентификации пользователей для случая динамического распознавания личности на основе свободных и длинных текстов;

– снижения размерности пространства выделенных клавиатурных признаков для повышения селективных свойств образцов КП и эффективности НА.

3.3. Жизненный цикл аутентификации

НА пользователя на основе КП имеет фазу регистрации и фазу аутентификации, как показано на рис. 2.

На этапе регистрации система получает данные о клавиатурных нажатиях. Далее производится извлечение характеристик нажатий (длительность, паузы и пр.), формируется или модифицируется клавиатурный профиль (шаблон) в базе данных и производится аутентификация пользователя.

Т.о. жизненный цикл НА включает 4 основных этапа:

I. Сбор данных о динамике нажатия клавиш – это непрерывный процесс при работе пользователя на клавиатуре в любом программном приложении. Для ОС Windows задействован механизм пе-

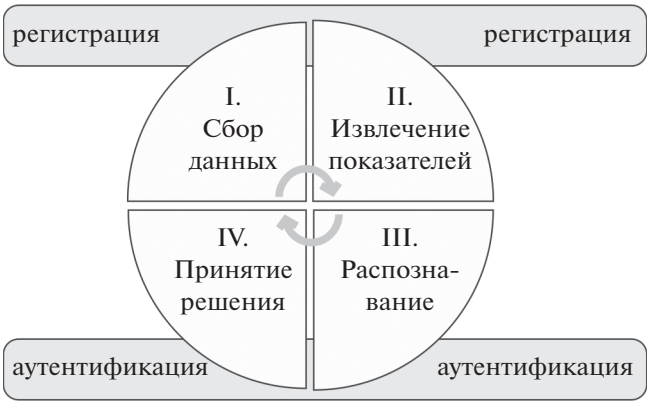

Рис. 2. Жизненный цикл непрерывной аутентификации.

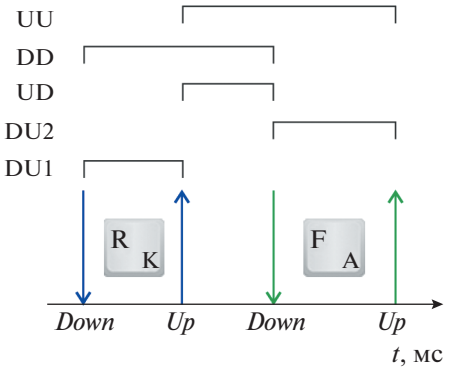

Рис. 3. Показатели нажатия клавиш в нотации Down Time/Up Time.

рехвата сообщений. С помощью Windows-hook можно зафиксировать любое событие использования клавиш [20]. ОС фиксирует код ANSI и временные метки нажатия клавиши (Down/Press) и ее отпускания (Up/Release). Точность измерения клавиатурных нажатий – миллисекунды.

II. Извлечение классификационных признаков Предварительно сырые данные о нажатиях клавиш должны быть очищены от выбросов, недостоверных значений и, в некоторых ситуациях, нормализованы. На основании этих данных можно получить ряд более значимых показателей КП: о ритме, скорости набора, паузах, отражающих уникальные поведенческие характеристики пользователя. Показателей КП довольно много, но наиболее популярны у исследователей ди-граммы (диграфы) – тайминг или временные метки двух состояний клавиши [17, 21–23]. Основные показатели следующие:

– время удержания клавиши;

– паузы между нажатиями;

– скорость набора;

– число ошибок при вводе;

– степень аритмичности при наборе;

– особенности использования служебных клавиш.

На рис. 3 приведены некоторые наиболее часто используемые временные и частотные показатели тайминга.

– DU – время удержания клавиши (ВУК) (Dwell Time, DT) – временной интервал между нажатием (Down, Press) и отпусканием (Up, Release) клавиши.

– UD – время между нажатиями или пауза (Flight Time, F-RP) – временной интервал между отпусканием пользователем предыдущей клавиши и нажатием следующей.

– UU, F-PP или DD, F-RR– интервал между нажатием или отпусканием одной клавиши и нажатием или отпусканием следующей клавиши соответственно

Подсистемы предварительной обработки информации о временных метках и извлечение показателей КП формируют массив требуемых показателей о каждом нажатии клавиши пользователем. Далее на основе этого массива генерируется клавиатурный профиль (шаблон) пользователя для размещения его в базе данных.

Банк профилей – это результат сбора характеристик КП, который требует адаптации. Как поведенческая биометрическая характеристика клавиатурный профиль изменчив и ему необходима актуализации. Коррекция клавиатурного профиля основана на технологии растущего или скользящего окна [49]. Используется банк профилей для обучения классификатора и на этапе распознавания личности пользователя

III. Распознавание пользователя

Нередко для защиты компьютерной системы от несанкционированного доступа используется двухэтапный процесс верификации:

– первичная идентификация, т.е. установление личности пользователя;

– динамическая (непрерывная) аутентификация, т.е. непрерывное подтверждение личности легитимного пользователя.

По типу задачи аутентификация – это задача классификации зарегистрированных в системе пользователей.

Основные методы и алгоритмы классификации (распознавания) пользователей одинаковы для статической и непрерывной (динамической) аутентификации. Их можно разделить на три группы:

– статистические;

– на основе оценки близости;

– методы машинного обучения.

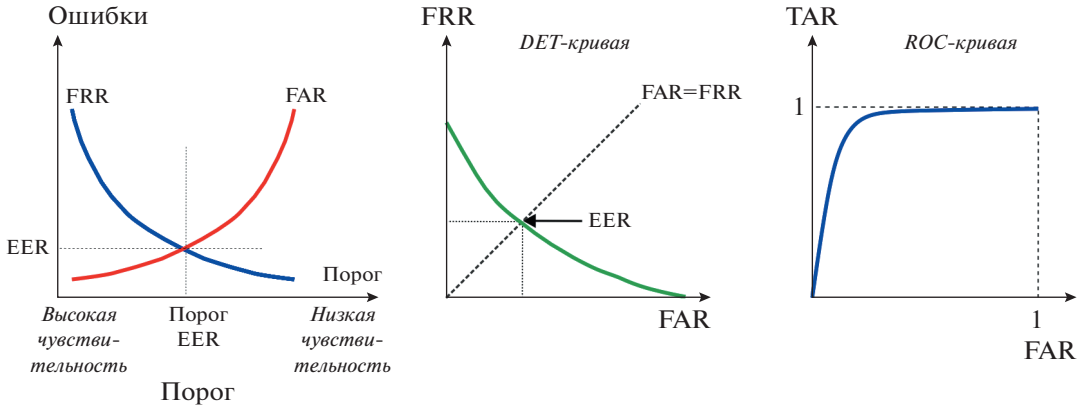

Рис. 4. Показатели эффективности клавиатурной идентификации.

Частота использования методов и основные представители в каждой группе подробно изложены в [20].

Исторически основные работы по распознаванию клавиатурной динамики относились к предопределенным и структурированным текстам, т.е. к статической аутентификации. И по данным разных авторов [22, 24] количество исследований, относящихся к НА по-прежнему невелико, и не превышает 10–15% от всего клавиатурного распознавания. Первым результативным исследованием НА можно считать работу Gunetti 2005 года, точность распознавания в которой превысила 90% [25]. Для контраста точность в самых первых результатах распознавания с использованием свободного текста (1997 г., Monrose) составляла 23%.

Обзорные работы по клавиатурному распознаванию последнего десятилетия позволили обобщить данные об эффективности НА, основные результаты приведены в табл. 2. Кроме библиографических ссылок таблица включает классификационный параметр, метод распознавания и показатели эффективности. Данные получены на основе собственных исследований [20, 26] и адаптированы из обзорных статей [17, 22, 24, 27–33].

В заключение анализа эффективности следует отметить условный характер подходов к распознаванию пользователей на основе его КП. Подходы, как правило, включают модель и методы обучения, но при этом их сочетания могут быть различными.

IV. Принятие решения о легитимности пользователя

Этот этап полностью определяется целями прикладной задачи на основании показателей эффективности распознавания.

При динамической идентификации основная цель непрерывного мониторинга – постоянный доступ к ресурсам сети для зарегистрированного в домене пользователя и предотвращение доступа незарегистрированному.

Следуя поставленной цели, принято оценивать вероятности соответствующих ложных событий: ложного отказа в доступе зарегистрированному и ложного доступа незарегистрированного пользователя. И по аналогии с оценками в статистической радиотехнике в исследованиях клавиатурной динамики чаще других используют следующие ошибки [17, 20, 21, 36, 38]:

– ошибка I рода False Rejection Rate (FRR) – частота ложного отказа в доступе законному (зарегистрированному) пользователю:

$$FRR = \frac{FR}{TA + FA + TR + FR}. \quad (1)$$

– ошибка II рода False Acceptance Rate (FAR) – частота ложного допуска к системе незаконных пользователей:

$$FRR = \frac{FA}{TA + FA + TR + FR}. \quad (2)$$

В (1) и (2) приняты обозначения:

– True Accept (TA) – число верных допусков в систему законным пользователям.

– True Reject (TR) – число верных отказов в доступе незаконным пользователям.

– False Accept (FA) – число ложных допусков незаконным пользователям.

– False Reject (FR) – число ложных отказов в доступе законным пользователям.

В знаменателях (1) и (2) – общее количество попыток.

Показатели FRR и FAR зависят от настраиваемого порога или чувствительности алгоритма, рис. 4.

Показатели FRR, FAR, ERR являются самодостаточными для принятия решения о допуске/отклонении пользователя. И, если цель системы мониторинга, высокая степень защиты, то следу-

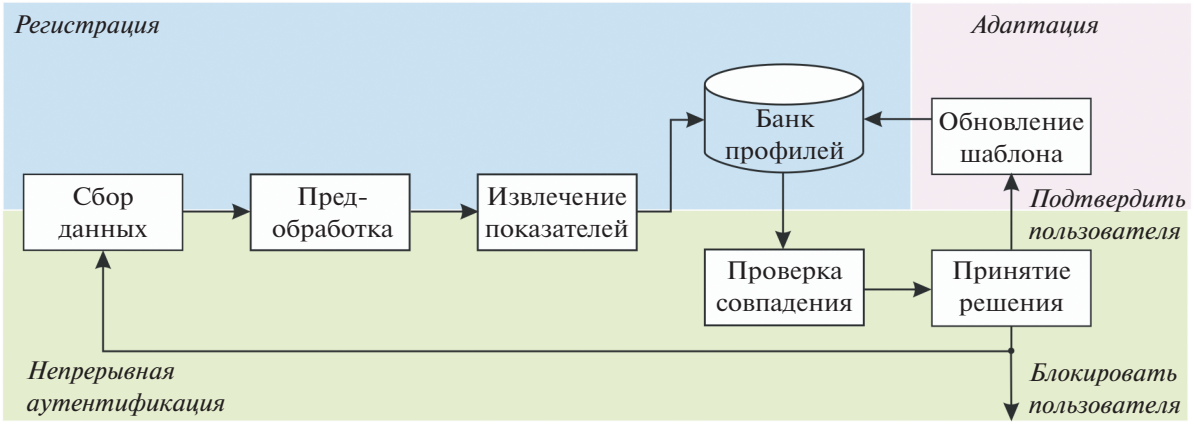

Рис. 5. Архитектура системы непрерывной аутентификации.

ет использовать низкие пороговые значения, которым соответствует большой процент ложно отклоненных (FRR). Большие значения FRR обеспечивают безопасность и более сложный вход в систему для всех – своих и чужих. А при высоком пороге и низкой чувствительности алгоритмов распознавания доступ упрощается, но вместе с тем возрастает FAR и число незаконно проникших пользователей. Вот этот компромисс между FRR и FAR приходится устанавливать индивидуально в каждой прикладной задаче.

Не менее популярным у исследователей является показатель, не зависящий от порогового значения – Equal Error Rate (EER). Значения EER соответствуют равным значениям FRR и FAR, что делает EER универсальным показателем для любой системы аутентификации.

Эти три показателя (FRR, FAR, EER) наиболее популярны при принятии решения в системах мониторинга и аутентификации. В научных исследованиях клавиатурной динамики также часто используется показатель Receiver Operating Characteristic (ROC) – соотношение между TA верно допущенными пользователями и FA ложно допущенными при различных пороговых значениях. ROC отражает предельные возможности алгоритмов, что особенно ценно при исследовании различных классификаторов.

4. СКРЫТЫЙ МОНИТОРИНГ В ДИСТАНЦИОННОЙ ОБРАЗОВАТЕЛЬНОЙ СИСТЕМЕ

Следует отметить, что пандемия ускорила внедрение модели смешанного обучения и многие университеты внедрили LMS – системы управления онлайн-обучением. Основной вклад в итоговую оценку студента в таких системах вносят онлайн-экзамены и тесты. Но именно эти формы онлайн-обучения более всего подвержены академическому мошенничеству, потому что онлайн-среда позволяет студентам работать практически без контроля. И, по мнению ряда исследователей [43–45], успех систем онлайн-экзаменов заключается в использовании биометрических систем контроля и непрерывного мониторинга во время онлайн-экзамена. Клавиатурная динамика позволяет организовать скрытый мониторинг личности легитимного студента (прошедшего первичную идентификацию) в комфортном для него режиме.

Основные вопросы клавиатурной динамики – жизненный цикл, компоненты, эффективность – рассмотрены в разделе 3. Особенности использования этих технологий приведены в последующих разделах 4.1–4.4.

4.1. Жизненный цикл аутентификации

Изложенные выше принципы и особенности НА лежат в основе подтверждения легитимности обучаемого при онлайн-тестировании.

Исследования возможностей НА личности легитимного пользователя выполнены на основе вычислительного эксперимента. Эксперимент проведен в соответствии со структурной схемой системы, представленной на рис. 5, и включает 3 основных подсистемы:

– регистрация;

– адаптация или обновление шаблона;

– аутентификация пользователя.

Причем при дистанционном обучении каждая подсистема работает в непрерывном режиме. Как показано на рис. 5, общими этапами и при регистрации, и при аутентификации являются: сбор данных, их обработка, извлечение клавиатурных признаков.

4.1.1. Сбор данных в эксперименте. Основой для разработки и проверки эффективности системы НА пользователей являются некоторые наборы данных о клавиатурных нажатиях. Это могут

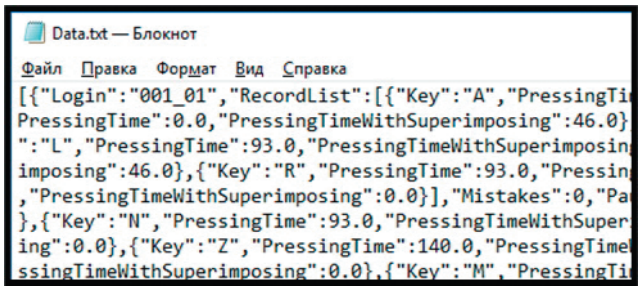

Рис. 6

быть известные статистические наборы (dataset), либо локально собранные.

Существует несколько общедоступных эталонных наборов, сформированных на основе фиксированных (Buffalo, BiosecurID) или свободных текстов (Clarkson II, Villani-2010) [23, 18, 46, 47]. Часть фиксированных наборов (CMU, WEBGREYC) [46] формируют шаблоны на основе логина/пароля, что не приемлемо для динамической аутентификации. Другая особенность подобных наборов – отсутствие шаблонов на основе русскоязычных текстов. Англоязычные шаблоны при распознавании российских пользователей представляют интерес только для первичной идентификации и статической аутентификации на основе парольных данных.

В предлагаемом исследовании осуществлен локальный сбор данных на основе разработанного программного приложения для операционной системы Windows. Архитектура программного приложения имеет клиент-серверную парадигму. На стороне клиента осуществляется локальный сбор клавиатурных данных в фоновом режиме.

Целевая аудитория данного исследования – домен университета, пользователи которого имеют навыки работы на компьютере выше среднего. Используя в ОС Windows механизм для перехвата сообщений от клавиатуры, так называемые hook-ловушки, фиксируются все события воздействия на клавиатуру. По каждому событию формируется массив информации, содержащий следующие сведения:

– логин пользователя в домене;

– код клавиши;

– событие (нажатие/отпускание клавиши);

– временная метка.

И в формате txt собранные данные имеют следующий вид (см. рис. 6).

Остальной функционал приложения реализован на сервере. Такое разделение позволяет снизить риски похищения клавиатурных данных с менее защищенных клиентских машин.

4.1.2. Предобработка данных. Собранные в течение сеанса работы пользователя сырые данные проходят в системе первичную обработку от выбросов, выходов из разумных диапазонов удержания клавиши (30–200 мс), непарных событий и коротких сессий.

Важнейшим элементом первичной обработки является выбор размера сессии (сеанса). Математически это соответствует выбору размера временных окон, на которые разбивается поток событий о клавиатурной динамике. Размер окна может быть различен на этапе регистрации и аутентификации. Методики разбиения на временные окна могут различаться [18]:

– по длительности работы пользователя или по количеству нажатий;

– окно раздвижное или скользящее.

В этом исследовании использовано скользящее окно, размер которого в режиме НА составляет 500 символов. Размер выборки при сборе данных должен быть достаточно большим, чтобы обеспечить репрезентативность отдельных букв в окне, а также статистическую значимость и несмещенность их оценок. Однако, известно, что клавиатурный почерк имеет тенденцию к изменчивости, вследствие психоэмоциональной составляющей. Поэтому временное окно должно быть скользящим. Против раздвижного окна выступает также возможность и процесса обратной биометрии: компрометация длительно незащищенных шаблонов. И объединяя эти два фактора, в работе использован минимально возможный размер выборки для состоятельной оценки средних значений для букв русского алфавита.

4.1.3. Извлечение временных характеристик и формирование клавиатурных шаблонов. Размерность результирующего пространства временных характеристик в задаче НА достаточно высока. Это связано с размером алфавита русского языка и большим числом временных характеристик КП.

Анализ исследований клавиатурных показателей продемонстрировал, что наиболее популярными временными характеристиками являются DU и UD (время удержания клавиши и пауза) и по данным [12, 20] частота использования каждого из них составляет от 30% до 40% в прикладных исследованиях.

В нашем исследовании данные о клавиатурной динамике последнего сеанса конкретного пользователя поступают на сервер в формате txt.

Далее производятся вычисления статистических характеристик DU и UD по каждой букве алфавита конкретной сессии пользователя. Результатом этой части программы является обновленный клавиатурный шаблон пользователя, адаптированный после последней сессии и сохраненный в формате *. json файла.

4.1.4. Формирование векторного показателя. Клавиатурный шаблон хранит обработанную и

достоверную информацию о клавиатурной динамике пользователей в последнем сеансе в соответствии с выбранными временными характеристиками (средние значения удержания каждой клавиши).

Для повышения информативности и достоверности созданных шаблонов при формировании совокупного показателя о клавиатурной динамике можно использовать дополнительные меры и инструменты.

Сокращение пространства на основе выделение стабильных признаков

Сокращение размерности признакового пространства временных характеристик в задаче клавиатурной динамики задача актуальная. Особенно в контексте дистанционного образования и онлайн-тестирования, когда высоки требования к скорости подтверждения легитимности экзаменуемого и обнаружения его подмены.

Выделение стабильных признаков клавиатурной динамики пользователей способствует, с одной стороны, сокращению признакового пространства. А с другой стороны – повышает селективные свойства шаблонов.

При выделении стабильных признаков можно пойти по пути повышения информативности самих временных характеристик [49]. Суть таких способов заключается в сужении диапазонов временных меток по какому-либо правилу, чаще всего с использованием статистических критериев. При этом техника сужения может касаться отдельных сеансов, отдельных пользователей и пр.

Также для повышения информативности временных характеристик могут быть использованы и эвристические способы. В данном исследовании такой способ связан с использованием частоты букв алфавита в текстах. Согласно данным Национального корпуса русского языка ruscorpora.ru частота букв (выраженная в %) уменьшается с 10.98% (буква О) до 0.996% (Х) и 0.037% (буква Ъ), рис. 7*а*. Поэтому для получения состоятельных оценок всех букв алфавита требуется внушительный размер скользящего окна клавиатурных нажатий. Что не допустимо в задаче онлайн-аутентификации. Способом сокращения размера окна является выбор в качестве стабильного признака порог частоты использования букв при наборе текстов. И для принятого в работе 0.5% порога это означает, что в шаблон пользователя не включены редко используемые буквы Ц, Щ, Э, Ф, Ъ, Ё, рис. 7*б*.

Вектор показателей КП включает характеристики (средние значения DU последнего сеанса) 27 букв русского алфавита с весовыми коэффициентами, соответствующими частотам использования букв в текстах, рис. 7*б*.

4.1.5. Распознавание легитимных пользователей. Цель статической и динамической аутентификации одинакова и состоит в подтверждении легитимности пользователя уже прошедшего первичную регистрацию. Разница только в виде создаваемого текста. При динамической (непрерывной) аутентификации – произвольный текст создается пользователем в любом приложении ОС и в произвольное время, при статической – текст и время его создания предопределены системой безопасности.

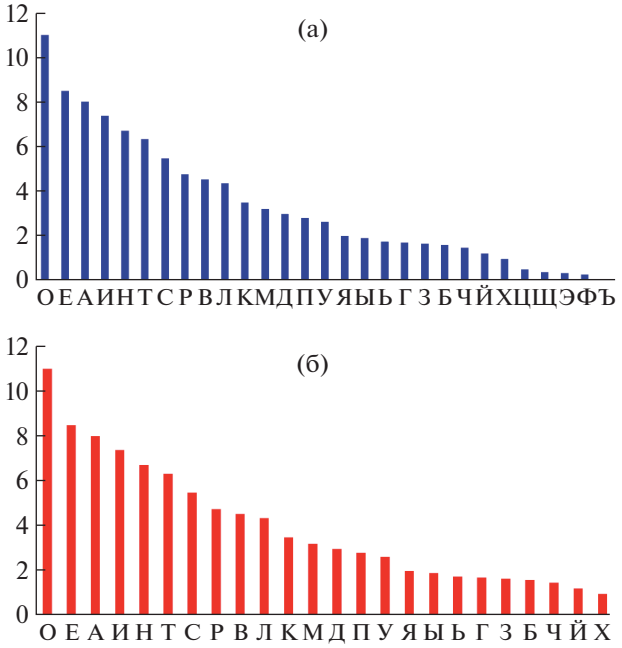

Рис. 7. Частота использования букв русского алфавита.

Следует отметить, что подтверждение легитимности пользователя по динамике его работы на клавиатуре компьютера является задачей одноклассовой классификации. Эта задача более сложна, чем многоклассовая классификация, поскольку данные нелегитимного пользователя системе распознавания неизвестны. Обучение классификатора происходит на объектах одного класса, а при тестировании алгоритм определяет принадлежность нового объекта этому классу. Данные зарегистрированных пользователей хранятся в системе в виде динамически обновляемых шаблонов. Их можно считать объектами одного класса – легитимных пользователей.

В системах онлайн-обучения одноклассовая классификация поможет выявить незарегистрированного пользователя, т.е. злоумышленника.

При онлайн-тестировании – одноклассовая классификация предотвратит несанкционированную подмену личности студента.

И хотя динамическая аутентификация значительно сложнее статической, методы распознавания аналогичны. Результаты краткого обзора ме-

Таблица 3. Сравнительный анализ показателей эффективности

	расстояние				метод	
	Евклидово		Манхэттенское		kNN	SVM
	частотность		частотность			
	–	+	–	+		
ERR, (0–100)%	10.8	0.99	10.1	0.79	0.54	3.72
порог (FAR=FRR), мс	4.8	2.3	4.8	2.2	15	
Accuracy, (0–100)%	90.87	99.20	83.41	99.4	99.45	
Precision, (0–1)	0.83	0.98	0.73	0.98	0.98	
Recall, (0–1)	0.71	0.83	0.74	0.94	0.98	

тодов динамической аутентификации на основании исследований последнего десятилетия представлены в табл. 3. Разделение методов распознавания на классы достаточно условно, но можно выделить методы машинного обучения, статистические методы и основанные на оценке метрических расстояний [20].

При выполнении этой работы в качестве классификатора были выбраны наиболее популярные методы распознавания в каждой из трех групп в рамках задачи одноклассовой классификации:

– метод опорных векторов SVM [17, 34, 21];

– метод k-ближайших соседей [27, 17, 46, 21, 34];

– оценка расстояний с использованием Евклидовой и Манхэттенской метрик.

Основная идея одноклассового метода опорных векторов OCSVM (one-class support vector machine) – выделить границы одного класса, а не разделить объекты нескольких классов, как в многоклассовой задаче [48]. Также одноклассовая классификация известна, как задача обнаружения аномалий и используется для поиска в данных не ожидаемого поведения пользователя [49]. Например, при изменении его психоэмоционального состояния под алкогольным или другим воздействием.

OCSVM отображает векторы признаков в пространство более высокой размерности с помощью функции ядра. И в случае радиального ядра (rbf kernel) находится гиперплоскость, отделяющую от начала координат большинство объектов заданного класса, в данном случае легитимного пользователя. Исключениями являются объекты, которые лежат ближе к началу координат, чем полученная гиперплоскость.

Одноклассовый метод k-ближайших соседей (One-Class KNN k-nearest neighbours) оценивает расстояние между объектами класса. Новый объект считается исключением, если большая часть объектов (например, p-я часть) всего класса находится на расстоянии, большем D, от нового объекта. Расстояние вычисляется согласно выбранной метрике в пространстве признаков, p и D – параметры метода.

Алгоритмы OCSVM и OCkNN достаточно просты в реализации. Их главное требование – это репрезентативность исходного набора данных.

5. РЕЗУЛЬТАТЫ

В соответствии с рис. 2, основными процессами жизненного цикла НА при онлайн-обучении являются этапы регистрации данных и аутентификации личности обучаемого. При регистрации производится непрерывный сбор данных о клавиатурных нажатиях и последующее извлечение показателей клавиатурной динамики, рис. 5.

Шаблоны (профили) пользователей домена динамические, сформированы они на основе непрерывного мониторинга произвольных нажатий на клавиатуру и не связаны с конкретными приложениями ОС. Сбор данных о клавиатурной динамике пользователя домена осуществляется в режиме скользящего окна и формируется шаблон с использованием букв русского и английского алфавитов. Исследования проведены на базе домена национального российского университета. Размер окна включает 500 нажатий, при накоплении которых информация передается в серверный компонент программы для предварительной очистки и извлечения показателей КП. При новых нажатиях окно передвигается, следующий набор данных передается на сервер и так происходит непрерывная регистрация данных в текущем сеансе работы пользователя.

Серверная часть программы производит расчет средних значений ВУК в текущем сеансе для каждого символа и обновляет шаблон в банке.

По каждому пользователю в банке хранятся шаблоны 10 последних сеансов, что позволяет отслеживать изменчивость КП, связанную с усталостью, психоэмоциональным состоянием и пр.

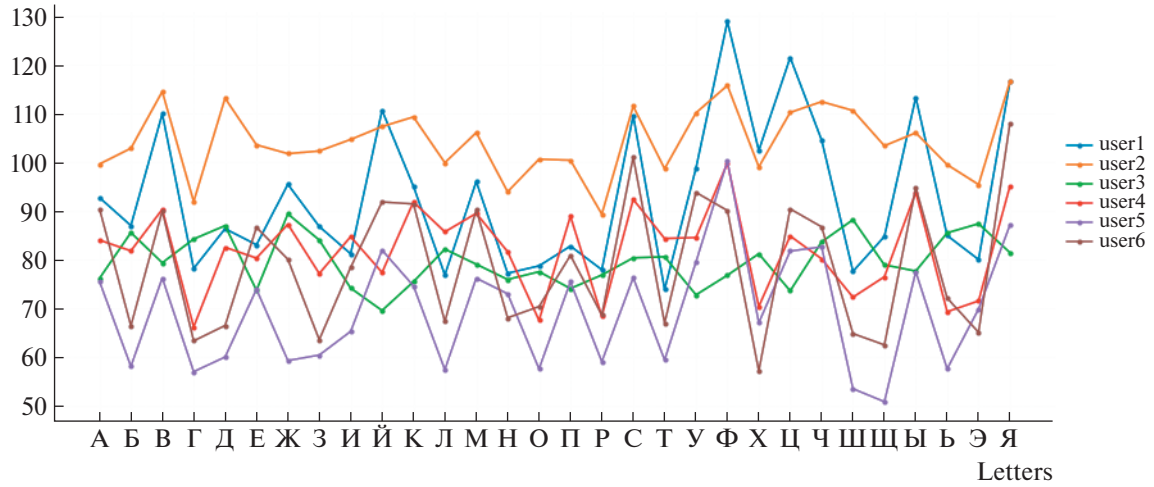

Рис. 8. Визуализация шаблонов пользователей домена.

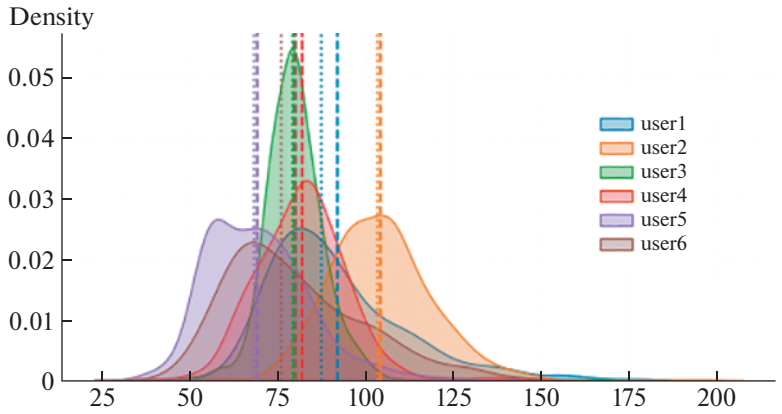

Рис. 9. Плотность распределения показателей КП.

Идентификационные возможности шаблонов разных пользователей могут быть оценены визуально и статистически. На рис. 8 изображены шаблоны шести произвольных пользователей домена университета для букв русского алфавита. Ось ординат соответствует длительности удержания клавиши в мс. Рисунок отдельной линии визуально отображает ритм набора текста и определяет, так называемый, клавиатурный почерк пользователя.

Ритм и скорость набора текстов различны для разных пользователей. Поэтому отличаются и рисунки КП. Это расхождение показателей КП подтверждается и статистически, например, видом и параметрами плотности распределения, рис. 9

Так, гистограмма пользователя User3 отличается малым разбросом и средней скоростью набора (математическое ожидание), что полностью совпадает с картиной на рис. 8 и подтверждается малыми отклонениями между модой и медианой ряда (крупный и мелкий пунктир на графиках).

Из рис. 8, 9 следует, что примерно равны и средние значения для User4 и User1. Однако распределение User1 имеет, так называемые, “тяжелые хвосты”. Мода и медиана для этого ряда значительно различаются между собой, как и для пользователя User06. И при репрезентативной выборке и высокой скорости набора это может являться дискриминантным признаком измененного психоэмоционального состояния.

Высокую скорость набора имеет также User5. При этом плотность распределения имеет бимодальный характер, что соответствует его хорошим навыкам и особенностям работы с клавиатурой. Для этого пользователя имеют место наложения при наборе текста, когда следующая клавиша нажимается, а предыдущая не отжата.

Следующий этап распознавания – подтверждение легитимности онлайн-обучаемого. Для

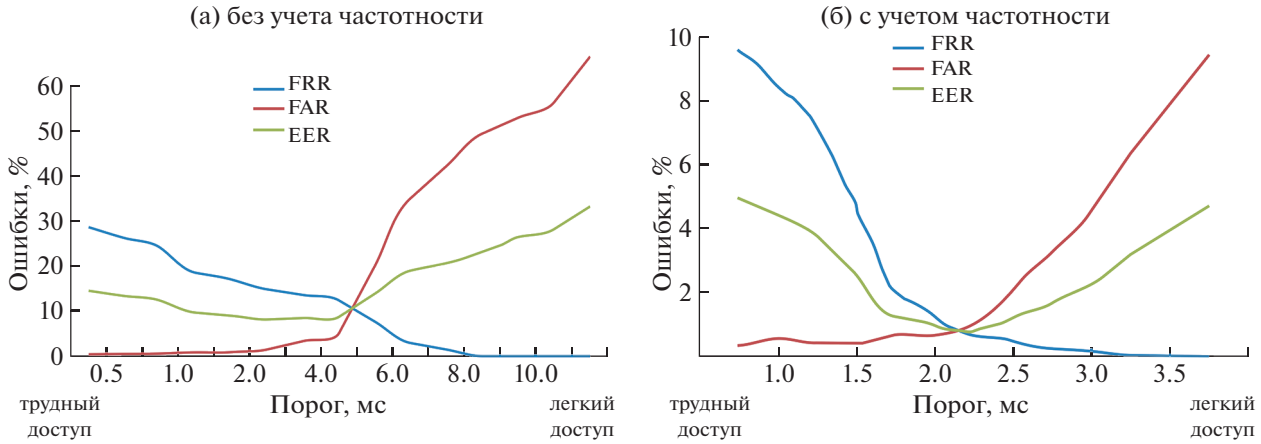

Рис. 10. Оценка эффективности распознавания пользователей.

этого производится сопоставление шаблона зарегистрированного пользователя из банка с его текущим шаблоном, рис. 5. При подтверждении легитимности личности возможны два исхода:

– шаблон последнего сеанса совпадает с шаблоном из банка данного зарегистрированного пользователя;

– шаблоны различны.

Совпадение шаблонов проанализировано в работе с использованием методов опорных векторов SVM, k-ближайших соседей и оценки расстояний на основе Евклидовой и Манхэттенской метрик. Обоснование выбора методов приведено в разделе 4.1.5.

Важнейшей характеристикой в любом методе является порог принятия решения. Пороговое значение выбирается (назначается) системой безопасности исходя из приоритета задач. Небольшие пороговые значения соответствуют малой разнице между базовым и текущем шаблоном и обеспечат сложный доступ в корпоративную сеть для всех, включая легитимных пользователей. Большой порог (низкая чувствительность) – легкий вход для всех. И основные визуальные инструменты, ошибки 1–2 рода, для Манхэттенской метрики приведены на рис. 10 в зависимости от пороговых значений.

На рис. 10*б* FAR, FRR и EER представлены с учетом частотности использования букв, 10*а* – без учета.

Как видно, достигнуто заметное уменьшение ошибок, в среднем на порядок при разных пороговых значениях. Например, для данных в табл. 3 эти значения равны 10.8% и 0.99% для Евклидовой метрики. В табл. 3 для всех выбранных методов ошибки EER приведены для порогового значения, при котором FAR = FRR.

Показатели аутентификации на основе Манхэттенской и Евклидовой метрик практически не отличаются с учетом частотности букв в алфавите. Хотя несколько отличались без использования частотности. Метод kNN показал немного лучшие показатели ERR = 0.54% при несколько большем пороговом значении 15 мс. Но это потребовало длительного подбора параметров алгоритма и не гарантирует необходимости подстройки параметров в реальных условиях. Отсутствием оптимальных значений параметров алгоритма и объясняются невысокие показатели аутентификации SVM метода.

Полезным инструментом для понимания событий при распознавании законных и незаконных пользователей является, так называемая, матрица соответствия (confusion matrix), табл. 4. На пересечении строк и столбцов матрицы показаны возможные верные (T) или ложные (F) исходы распознавания: принять (A) или отклонить (R) пользователя. Столбцы соответствуют исходу распознавания, строки – реальным пользователям.

Указанные в табл. 3 показатели точности, вычисляются следующим образом:

$$Accuracy = \frac{TA + TR}{TA + FA + TR + FR}. \quad (3)$$

$$Precision = \frac{TA}{TA + FA}. \quad (4)$$

$$Recall = \frac{TA}{TA + FR}. \quad (5)$$

Все три показателя отражают точность аутентификации легитимного пользователя. Но акценты каждого показателя разные.

Accuracy – метрика точности верных допусков и отклонений пользователя для всех возможных законных и незаконных пользователей.

Таблица 4. Матрица соответствия ошибок

		Пользователь при распознавании	
		законный	незаконный
Фактический пользователь	законный	TA	FR
	незаконный	FA	TR

Precision показывает отношение верно принятых пользователей ко всем допущенным системой.

Recall – доля допущенных пользователей из всех легитимных. Recall также называют чувствительностью модели для распознавания законных пользователей.

Accuracy принято выражать в процентах, а Precision и Recall изменяются в диапазоне (0–1).

И последний визуальный инструмент качества аутентификации – кривые DET и ROC, представленные на рис. 11.

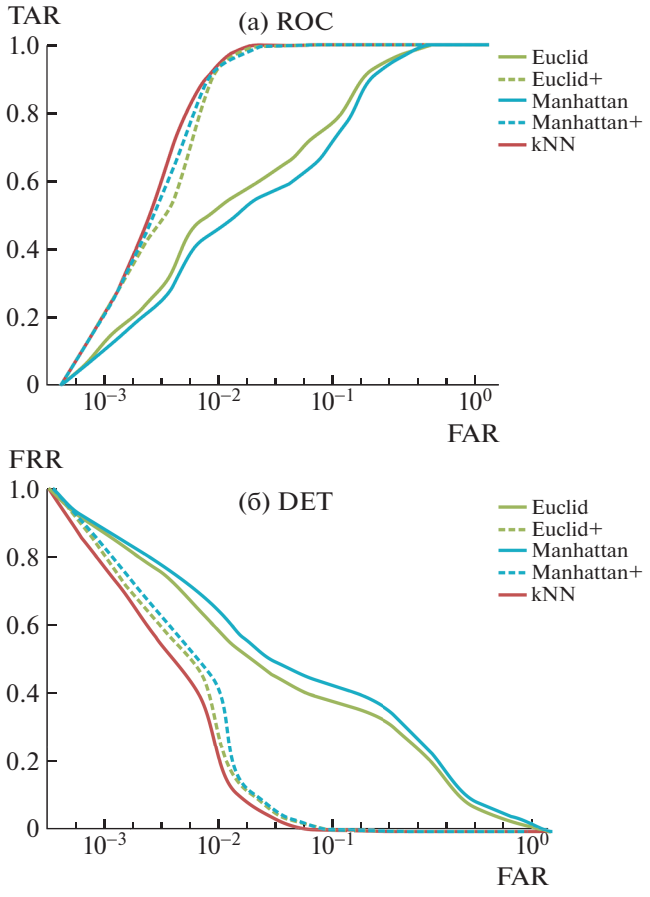

Рис. 11. Оценка эффективности распознавания пользователей.

Кривые ROC и DET подтверждают близкую эффективность методов, основанных на близости шаблонов с учетом частотности и kNN-метода.

ЗАКЛЮЧЕНИЕ

В работе рассмотрен подход к подтверждению легитимности личности при дистанционном обучении на основе мониторинга характеристик его клавиатурного почерка.

Было установлено, что скрытая идентификация обучаемого возможна на основе непрерывного мониторинга его клавиатурных нажатий в любом программном приложении. На основании проведенных исследований были сделаны следующие выводы.

1. Требуется корректировка (адаптация) образцов КП с использованием скользящего окна, позволяющая динамически отслеживать изменчивость КП пользователя и его психоэмоциональное состояние.

2. Выделение стабильных признаков клавиатурной динамики пользователей способствует и сокращению признакового пространства КП, что повышает селективные свойства шаблонов.

3. В данном исследовании таким признаком является частотность использования букв алфавита в текстах в соответствии с данными Национального корпуса русского языка ruscorpora.ru. Исключение из шаблонов шести букв с частотой ниже 0.5% привело к заметному снижению всех показателей эффективности распознавания. Например, ошибка ERR снизилась на порядок, в среднем с 10% до 1%. А показатели точности (Accuracy, Precision, Recall) в среднем повысились на 6–13% и составляют 98% как для Евклидовой, так и Манхэттенской метрики.

4. kNN-метод при оценке легитимности показал для оптимальных настроечных параметров немного лучшие результаты по ошибке ERR = = 0.54% (против 0.79%) при равных оценках точности Accuracy = 99%, Precision = 0.98.

5. Прикладной эффект дополнения простых алгоритмов оценки расстояний частотностью букв алфавита состоит в отсутствии сложных процедур настройки оптимальных параметров, как в

методах машинного обучения (kNN), при равных показателях точности.

А поскольку ключевым фактором непрерывной аутентификации легитимного пользователя является производительность распознавания, то адаптация параметров значительно снижает общую эффективность подтверждения легитимности.

СПИСОК ЛИТЕРАТУРЫ

1. *Ключевская Н.* Информационная безопасность и COVID-19. Рекомендации для бизнеса и граждан. https.//www.garant.ru/article/1421147.
2. Аналитика отрасли информационной безопасности. https.//www. infowatch.ru/analytics/analitik.
3. Хакеры атакуют университеты и колледжи. Дайджест утечек. https.//www.infowatch.ru/analytics/daydzhesty-i-obzory/khakery-atakuyut-universitety-i-kolledzhi-daydzhest-utechek.
4. Образование. От закрытия учебных заведений до возобновления их работы. https.//ru.unesco.org/covid19/educationresponse.
5. Образование: от разрушения к выздоровлению. https.//ru.unesco.org/covid19/educationresponse.
6. *Fenu G., Marras M., Boratto L.* A multi-biometric system for continuous student authentication in e-learning platforms. Pattern Recognition Letters. 2018. V. 113. P. 83–92. doi.org/. https://doi.org/10.1016/j.patrec.2017.03.027
7. *Ngqondi T., Maoneke PB., Mauwa L.* A secure online exams conceptual framework for South African universities. Social Sciences & Humanities Open. 2021. V. 3(1). 100132. DOI.org/10.1016/j.ssaho.2021.100132.
8. *Christine Lee.* How to Maintain Academic Integrity in Distance Learning. https://www.turnitin.com/ru/blog/kak-podderzhivat-akademicheskuyu-chestnost-pri-distantsionnom-obuchenii.
9. *Al-Naji F.H., Zagrouba R.* A survey on continuous authentication methods in Internet of Things environment. Computer Communications. 2020. V. 163. P. 109–133. DOI.org/10.1016/j.comcom.2020.09.006.
10. *Dasgupta D., Roy A., Nag A.* Advances in User Authentication. Springer International Publishing. 2017. DOI.org/10.1007/978-3-319-58808-7.
11. *Stylios I., Kokolakis S., Thanou O., Chatzis S.* Behavioral devices. A survey Information Fusion. 2021. V. 66. P. 76–99. DOI.org/10.1016/j.inffus.2020.08.021.
12. *Toosi R., Akhaee M.A.* Time-frequency analysis of keystroke dynamics for user authentication Future Generation. Computer Systems. 2021. V. 115. P. 438–447. DOI.org/10.1016/j.future.2020.09.027.
13. Future of identity study – IBM security Source. https.//www.ibm.com/downloads/cas/QRBY08NO.
14. *Hazan I., Margalit O., Rokach L.* Supporting unknown number of users in keystroke dynamics models. Knowledge-Based Systems. 2021. V. 221. 106982. DOI.org/10.1016/j.knosys.2021.106982.
15. *Parkinson S., Khan SCrampton A., Xu Q., Xie W., Liu N., Dakin K.* Password policy characteristics and keystroke biometric authentication. IET Biometrics. 2021. V. 10(2). P. 163–178. DOI.org/10.1049/bme2.12017.
16. *Kochegurova E.A., Gorokhova E.S., Mozgaleva A.I.* Development of the Keystroke Dynamics Recognition System. J. Physics. Conf. Ser. 2017. V. 803. 012073. DOI.org/10.1088/1742-6596/803/1/012073.
17. *Kim J., Kim H., Kang P.* Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection. Applied Soft Computing. 2018. V. 62. P. 1077–1087. DOI.org/10.1016/j.asoc.2017.09.045.
18. *Lu X., Zhang S., Hui P., Lio P. Continuous* authentication by free-text keystroke based on CNN and RNN. Computers & Security 2020. V. 96. 01861. DOI.org/10.1016/j.cose.2020.101861.
19. *Dargan S., Kumar M.* A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. Expert Systems with Applications. 2020. V. 143. 113114. DOI. org/. https://doi.org/10.1016/j.eswa.2019.113114
20. *Kochegurova E.A., Martynova Y.A.* Aspects of continuous user identification based on free texts and hidden monitoring. Program Comput Softw. 2020. V. 46(1). P. 12–24. DOI. https://doi.org/10.1134/S036176882001003X
21. *Zaidi A.Z., Chong C.Y. Jin Z., Parthiban R., Sadiq A.S.* Touch-based continuous mobile device authentication. State-of-the-art, challenges and opportunities. J Network Comput Appl. 2021. V. 191. 103162. DOI.org/10.1016/j.jnca.2021.103162.
22. *Teh P.S., Teoh A.B.J., Yue S.* A survey of keystroke dynamics biometrics The Scientific World Journal. 2013. V. 2013. P. 1–24. DOI. https://doi.org/10.1155/2013/408280
23. *Morales A., Fierrez J., Tolosana R., Ortega-Garcia J., Galbally J., Gomez-Barrero M., Anjos A., Marcel S.* KBOC. Keystroke biometrics OnGoing competition. 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). 2016. DOI. https://doi.org/10.1109/BTAS.2016.7791180.
24. *Pisani P.H., Lorena A.C. A* systematic review on keystroke dynamics. J Braz Comput Soc. 2013. V. 19(4). P. 573–587.
25. *Gunetti D., Picardi C.* Keystroke analysis of free text. ACM Trans Inf Syst Secur. 2005. V. 8(3). P. 312–347. DOI. https://doi.org/10.1145/1085126.1085129
26. *Kochegurova E., Luneva E., Gorokhova E.* On continuous user authentication via hidden free-text based monitoring. Adv Intell Sys Comput. 2019. V. 875. P. 66–75. DOI. https://doi.org/10.1007/978-3-030-01821-4_8
27. *Alsultan A., Warwick K.* Keystroke dynamics authentication. a survey of free-text methods. Int. J. Comput. Sci. 2013. V. 10(4). P. 1–10.
28. *Mondal S., Bours P.* A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing. 2017. V. 230. P. 1–22. DOI. https://doi.org/10.1016/j.neucom.2016.11.031
29. *Zhong Y., Deng Y.* A survey on keystroke dynamics biometrics. approaches, advances, and evaluations. Recent Advances in User Authentication Using Keystroke Dy-

namics Biometrics. 2015. V. 2. P. 1–22. DOI. https://doi.org/10.15579/gcsr.vol2.ch1

30. *Ali M.L., Monaco J.V., Tappert C.C. et al.* Keystroke Biometric Systems for User Authentication. J Sign Process Syst. 2017. V. 86. P. 175–190. https://doi.org/10.1007/s11265-016-1114-9

31. *Alsultan A., Warwick K., Wei H.* Non-conventional keystroke dynamics for user authentication. Pattern Recogn Lett 2017. V. 89. № 5. P. 53–59. DOI. https://doi.org/10.1016/j.patrec.2017.02.010

32. *Shimshon T., Moskovitch R., Rokach L., Elovici Y.* Continuous verification using keystroke dynamics. International Conference on Computational Intelligence and Security (CIS'10). 2010. P. 411–415. DOI. https://doi.org/10.1109/CIS.2010.95.

33. *Messerman T., Mustafić T., Camtepe S.A., Albayrak S.* Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. 2011 International Joint Conference on Biometrics (JCB). 2011. P. 1–8. DOI. https://doi.org/10.1109/IJCB.2011.6117552.

34. *Chang H.C., Li J., Wu C., Stamp M.* Machine Learning and Deep Learning for Fixed-Text Keystroke Dynamics. arXiv.2107.07409v1 [cs.LG]. 2021. DOI.org/10.48550/arXiv.2107.07409.

35. *Ahmed A.A., Traore I.* Biometric recognition based on free-text keystroke dynamics/ Cybern. IEEE Trans 2014. V. 44(4). P. 458–472. DOI. https://doi.org/10.1109/TCYB.2013.2257745

36. *Goodkind A., Brizan D.G., Rosenberg A.* Utilizing overt and latent linguistic structure to improve keystroke-based authentication. Image and Vision Computing 2017. V. 58. P. 230–238. DOI. https://doi.org/10.1016/j.imavis.2016.06.003

37. *Al Solami E., Boyd C., Clark A., Ahmed I.* User-representative feature selection for keystroke dynamics. 5th International Conference on Network and System Security (NSS'11) 2011. P. 229–233. DOI. https://doi.org/10.1109/ICNSS.2011.6060005.

38. *Eberz S., Rasmussen K.B., Lenders V., Martinovic I.* Evaluating behavioral biometrics for continuous authentication. challenges and metrics. 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17). 2017. P. 386–399. DOI. https://doi.org/10.1145/3052973.3053032.

39. *Antal M., Szabó L.Z., Laszlo I.* Keystroke dynamics on Android platform. Procedia Technology. 2015. V. 19. P. 820–826. DOI. https://doi.org/10.1016/j.protcy.2015.02.118

40. *Locklear H., Govindarajan S., Sitova Z. etc.* Continuous authentication with cognition-centric text production and revision features. IEEE/IAPR international joint conference on biometrics (IJCB 2014). 2014. DOI. https://doi.org/10.1109/BTAS.2014.6996227

41. *Kang P., Cho S.* Keystroke dynamics-based user authentication using long and free text strings from various input devices. Inf Sci. 2015. V. 308. P. 72–93. DOI. https://doi.org/10.1016/j.ins.2014.08.070

42. *Matsubara Y., Samura T., Nishimura H.* Keyboard Dependency of Personal Identification Performance by Keystroke Dynamics in Free Text Typing. Journal of Information Security. 2015. V. 6. P. 229–240. DOI. https://doi.org/10.4236/jis.2015.63023

43. *Wang X., Yan Z., Zhang R., Zhang P.* Attacks and defenses in user authentication systems. A survey. Journal of Network and Computer Applications. 2021. V. 188. 103080. DOI. https://doi.org/10.1016/j.jnca.2021.103080

44. *Muzaffar A.W., Tahir M., Anwar M.W., Chaudry Q., Mir S.R., Rasheed Y.* A systematic review of online exams solutions in e-learning. Techniques, tools, and global adoption. IEEE Access. 2021. V. 9. P. 32689–32712. DOI. https://doi.org/10.1109/ACCESS.2021.3060192

45. *Jagadamba G., Sheeba R., Brinda K.N., Rohini K.C., Pratik S.K.* Adaptive E-Learning Authentication and Monitoring. 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). 2020. 277–283. DOI. https://doi.org/10.1109/ACIMIA48430.2020.9074955.

46. *Iapa A., Cretu V.* Shared Data Set for Free-Text Keystroke Dynamics Authentication Algorithms. Preprints. 2021. 2021050255. DOI. https://doi.org/10.20944/preprints202105.0255.v1

47. *González N., Calot E.P., Ierache J.S., Hasperué W.* On the shape of timings distributions in free-text keystroke dynamics profiles. Heliyon 2021. V. 7(11). e08413. DOI. https://doi.org/10.1016/j.heliyon.2021.e08413

48. *Mhenni A., Cherrier E., Rosenberger C., Essoukri Ben Amara N.* Analysis of Doddington zoo classification for user dependent template update. Application to keystroke dynamics recognition. Future Gener Comput Syst. 2019. V. 97. P. 210–218. DOI. . https://doi.org/10.1016/j.future.2019.02.039

49. *Казачук М.А.* Динамическая аутентификация пользователей на основе анализа работы с клавиатурой компьютера. Дисс. на соискание уч. степени к.ф.-м.н. Москва. 2019. 155 с.

50. *Alpar O.* Biometric keystroke barcoding. A next-gen authentication framework. Expert Sys Appl. 2021. V. 177. 114980. DOI. https://doi.org/10.1016/j.eswa.2021.114980

51. *Yang Y., Guo B., Wang Z., Li M., Yu Z., Zhou X.* BehaveSense. Continuous authentication for security-sensitive mobile apps using behavioral biometrics. Ad Hoc Networks. 2019. V. 84. P. 9–18. DOI. https://doi.org/10.1016/ j.adhoc.2018.09.015