НЕПРЕРЫВНЫЙ МОНИТОРИНГ ПОЛЬЗОВАТЕЛЕЙ УДАЛЕННОЙ ОБРАЗОВАТЕЛЬНОЙ ПЛАТФОРМЫ С ИСПОЛЬЗОВАНИЕМ КЛАВИАТУРНОЙ ДИНАМИКИ

Барышев М.В.[1], Кочегурова Е.А [2]
[1] Томский политехнический университет, Инженерная школа информационных технологий и робототехники, гр.8И21, e-mail: mvb41@tpu.ru
[2] Томский политехнический университет, Инженерная школа информационных технологий и робототехники, доцент, e-mail: kocheg@tpu.ru

Введение

Современные технологии дистанционного обучения становится все более актуальным в образовании разного уровня. Это связано с доступностью дистанционных платформ обучения, а также возможностью получать знания в асинхронном режиме, в комфортной обстановке и с учётом индивидуального темпа обучения.

Однако, в плане кибербезопасности дистанционное образование и вся образовательная сфера являются одной из наиболее уязвимых. Это обусловлено как менталитетом обучаемых, так и значительными объемами конфиденциальных данных. Что делает университеты и другие образовательные учреждения привлекательными целями для хакерских атак. И согласно аналитическим данным ЭАК InfoWatch только за первое полугодие 2024 г. было скомпрометировано около 1 млрд записей персональных данных и это на 10 % больше, чем за полугодие в 2023 году.

Доля университетских дистанционных технологий обучения довольно велика – более 48 % [1], часть из которых реализуется в партнерстве с частными EdTech-компаниями. Но формы онлайн обучения по-прежнему функционально ограничены. Кроме вебинаров на платформах электронного образования, для повышения интерактивности и контроля знаний используются онлайн экзамены и тесты. Однако именно при онлайн-тестировании проявляются различные формы академического мошенничества. В том числе подмена личности, когда другой человек подменяет тестируемого в реальном масштабе времени [2]. Кроме организационных методов противодействия такой подмене, может быть предложена низкозатратная биометрическая идентификация обучаемого на основе его клавиатурного почерка (КП). Такой способ идентификации относится к поведенческой биометрической и может являться классификационным признаком личности.

Целью данной работы является идентификация личности пользователя на основе непрерывного и скрытого мониторинга динамических характеристик его КП при онлайн тестировании.

Вопросы биометрической аутентификации

Идентификация и аутентификация пользователя – это ключевые понятия в области информационной безопасности [3]. Идентификация – первый этап верификации, на котором пользователь заявляет свои учетные данные. На втором этапе аутентификации пользователь подтверждает подлинность личности либо на основе знаний (пароля, пин-кода), либо владения (ключом, смарт-картой).

Существуют также биометрические методы аутентификации: физиологические (отпечатки пальцев, радужка глаза) и поведенческие (рукописный и клавиатурный почерк, голос, движения мышью). Аутентификация по физиологическим признакам имеет высокую точность, но затратна с позиции оснащения техническими средствами. Поведенческая аутентификация имеет низкую стоимость, не требует дополнительного оборудования, но требуется разработка высокоэффективного программного приложения для распознавания личности.

Скрытый мониторинг наиболее комфортен для пользователя, а эффективность мониторинга повышается в непрерывном (динамическом) режиме. Однако и КП, являясь поведенческой биометрической характеристикой, также изменяется во времени. Поэтому логично использование динамической аутентификации, которая обеспечивает сбор актуальных данных о нажатиях клавиш во всех приложениях операционной системы на протяжении всего времени работы пользователя за компьютером.

Жизненный цикл аутентификации пользователя включает две стадии (рис. 1):

– регистрация,

– аутентификация.

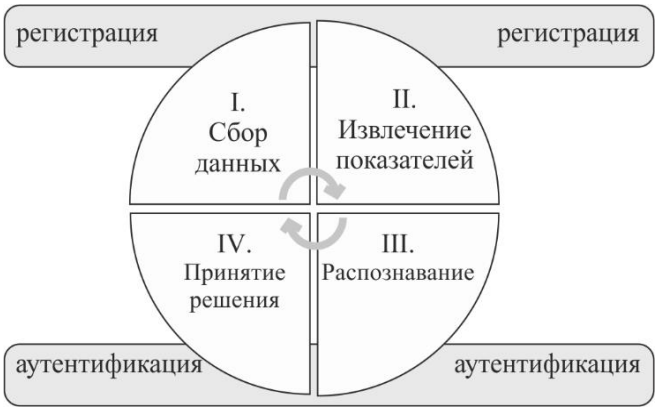


Рис. 1. Этапы динамической аутентификации

Регистрация включает сбор данных и извлечение показателей, отражающих ритм набора текста на клавиатуре. У всех пользователей ритм и темп индивидуальны.

Данные о клавиатурных нажатиях фиксируются при перехвате операционной системой любого события использования клавиатуры. Из полученной информации предлагаемая система извлекает данные, необходимые для создания шаблона пользователя. В самом общем случае кроме логина пользователя и кода клавиши, записывается время нажатия (Down) и отжатия клавиши (Up).

Затем, если собрано достаточное количество данных о КП для репрезентативных оценок, текущий шаблон сравнивается с шаблоном из БД. Так происходит распознавание пользователя. А на следующем этапе принимается решение о легитимности пользователя.

В случае легитимности личности система обновляет его пользовательский шаблон в БД.

Вычислительный эксперимент и результаты

Значительная часть систем аутентификации основаны на статическом подходе с использованием предопределенных текстов. Динамическая аутентификация сложнее и затратнее по нагрузке системы, но многие аспекты (жизненный цикл, временные показатели КП, алгоритмы и методы распознавания, эффективность аутентификации) в основном одинаковы. Различны только этапы сбора временных показателей и актуализация шаблонов пользователей.

В данной работе использована концепция скользящего окна при сборе информации, т.к. имеет смысл анализировать не весь поток данных, а только его актуальную часть. Минимальный объем данных для начала аутентификации включает 200 символов, а размер скользящего окна – 600 символов. Такие объемы данных обеспечивают состоятельные оценки показателей КП. Кроме того, размерность окна гарантирует репрезентативность всех букв алфавита, которая согласуется с частотностью букв русского алфавита (рис. 2).

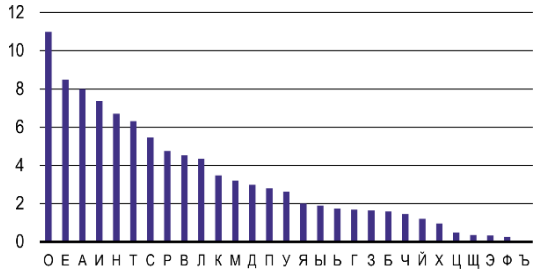


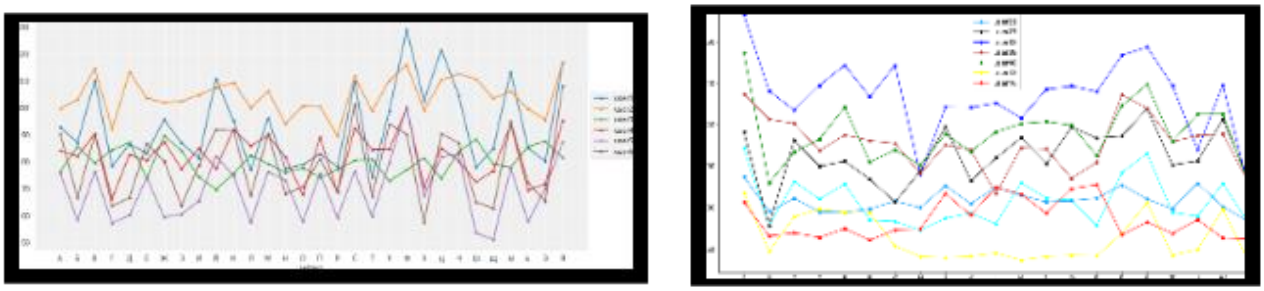
Рис. 2. Этапы динамической аутентификации

Существует несколько временных показателей клавиатурной динамики, отражающие моменты нажатия Down (D) и отпускания Up (U) клавиши. Однако самым популярным показателем является показатель время удержания клавиши (ВУК / DU), который и используется в данном исследовании.

Тестирование задачи динамической аутентификации проведено с использованием личных и общедоступных наборов данных, в том числе для набора КМ [4]. Следует отметить, что все общедоступные датасеты содержат информацию только на иностранных языках. Локальный набор данных на русском языке был собран в домене университета. Он включает шаблоны 20 пользователей с различным

числом сессий от 1 до 10, что недостаточно для полноценного исследования. Для увеличения и выравнивания количества сессий различных пользователей потребовались дополнительные сессии, которые получены на основе модели Бокса-Мюллера. Это позволило получить шаблоны пользователей на русском и английском языке на основе усреднения показателя ВУК.

Графически шаблоны представляют собой временные засечки для 6 пользователей русского алфавита и 8 – английского, рис. 3, а и 3, б соответственно.



а) *б)*

Рис. 3. Визуальное представление шаблонов пользователя

Визуально каждый шаблон уникален и с позиции скорости набора, и с позиции ритма. Соответственно различаются и статистические показатели ВУК. Например, из приведенной на рис. 4 гистограммы плотности распределения ВУК отчетливо различаются пользователи с низкой (user 4) и высокой (user 5, user 6) скоростью набора; равномерным темпом (user 3) и крайне неравномерным темпом (user 1). Приведенные выводы подтверждают также моды и медианы (пунктирные линии), которые могут быть формальным классификационным признаком. Здесь и далее представлено описание только для локального датасета и 6 пользователей.

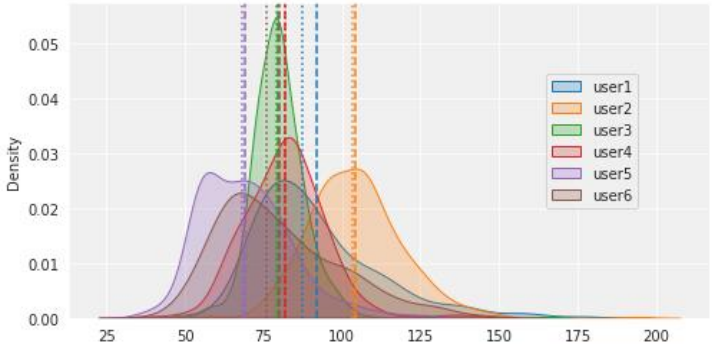


Рис. 4. Плотность распределения ВУК

Задача распознавания пользователя для подтверждения его легитимности решена в работе, как задача одноклассовой классификации [5]. Результат решения подтверждает / опровергает, что текущий образец КП и шаблон в БД являются объектами одного класса, а именно зарегистрированного пользователя домена. В отличие от многоклассовой классификации - цель которой выяснить кому из пользователей домена принадлежит текущий образец КП. И, применительно к системе дистанционного обучения и тестирования, одноклассовая классификация позволяет выявить нелегетимного пользователя, т.е. выявлять подмену личности.

Эффективность подтверждения легитимности студента может быть оценена традиционными ошибками I и II рода: FRR – ложный отказ в доступе легитимному пользователю и FAR – ложный доступ нелигитимного.

Вычислительный эксперимент проведен с целью получения FRR/ FAR ошибок, отражающих эффективность классификации пользователей. Для этого для каждого пользователя было смоделировано 200 сессий и 1200 – для всего эксперимента.

Ошибки распознавания определены значением порога для входа в систему. Пороговое значение

измеряется в тех же единицах, что и значения ВУК, т.е. в мс. Малые пороговые значения соответствуют трудному доступу в систему, соответственно многие пользователи будут отклонены и ошибка FRR – велика. И напротив, легкий доступ, соответствующий большим пороговым значениям, обеспечивает вход большого числа пользователей, в том числе незарегистрированных. В этом случае возрастает ошибка FAR.

На рис. 5 изображены показатели эффективности распознавания пользователей, полученные в предлагаемом вычислительном эксперименте.

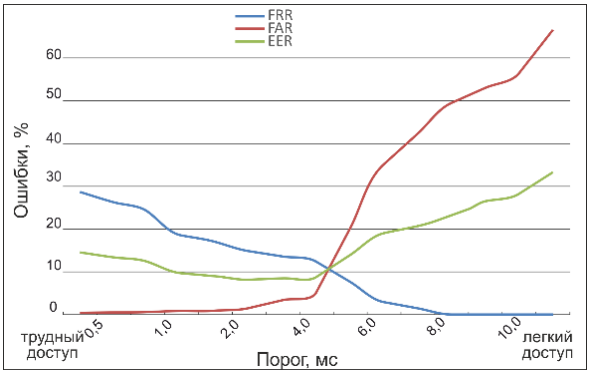


Рис. 5. Плотность распределения ВУК

Кроме ошибок FRR и FAR на рисунке представлена равная частота ошибок EER, отражающая общую эффективность системы. EER не монотонна, в отличие от FRR и FAR, и определяет общую ошибку системы аутентификации при различных пороговых значениях. В частности, для данного набора данных минимальное значение EER чуть меньше 10 % при пороге 4.1 мс. Это означает, что за 4.1 мс система установит легитимность пользователя со средней ошибкой 10 %. А учитывая, что среднее время ВУК равно 100 мс, величина порога довольно низкая.

Заключение

Полученные показатели эффективности при установлении легитимности пользователя вполне приемлемы и по точности, и по производительности, основной вклад в которую вносит величина порога. Однако показатели эффективности могут быть существенно улучшены при использовании дополнительных инструментов распознавания. Например, использование частотности букв русского языка, приведенной на рис.2, позволило уменьшить ошибку EER с 10 % до 2.5 % при пороговом значении 2.4 мс. А именно производительность распознавания пользователя является главным фактором при подтверждении легитимности в образовательной сфере, поскольку оперативно пресекает факты академического мошенничества.

Работа выполнена при поддержке гранта РНФ (№ 23-21-00259).

Список использованных источников

1. Гресева И. Онлайн-образование: объем рынка и основные тенденции // Softline. Тренды в цифре: сайт. – 2024. – URL: http:// https://slddigital.com/article/onlajn-obrazovanie-obem-rynka-i-osnovnye-tendencii/#sotrudnichestvo-s-vuzami-i-gosudarstvom.

2. Иоголевич Н.И., Лободенко Е.И. Академическая недобросовестность студентов технического вуза: масштабы проблемы и пути решения // Педагогика. Вопросы теории и практики. – 2020. – Т. 5, №. 1. – С. 99–106.

3. Kaspersky daily: URL:https://www.Kaspersky.ru/blog/identification-authentication-authorization-difference/29123/ (дата обращения: 10.10.2024). Текст: электронный.

4. González N, Calot EP, Ierache JS, Hasperué W. On the shape of timings distributions in free-text keystroke dynamics profiles // Heliyon. – 2021. – Vol. 7 (11). – DOI: 10.1016/j.heliyon.2021.e08413.

5. Кочегурова Е.А., Затеев Р.П. Скрытый мониторинг пользователя в дистанционной образовательной системе на основе клавиатурной динамики // Программирование. – 2022. – № 6. – С. 31–45.