

ГИБРИДНАЯ СИСТЕМА НЕПРЕРЫВНОЙ АУТЕНТИФИКАЦИИ НА ОСНОВЕ АНАЛИЗА РИСКОВ И КЛАВИАТУРНОГО ПОЧЕРКА

Ю. С. Матюшин^{1,}, В. В. Корхов¹*

¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

В ответ на необходимость более надежных методов аутентификации были разработаны новые парадигмы, например непрерывная аутентификация (CA). Представлен гибридный метод CA, сочетающий аутентификацию на основе анализа рисков и машинного обучения (MLE-RBA) и биометрическую аутентификацию по динамике нажатия клавиш. Хотя основной акцент сделан на архитектуре системы, из более ранних экспериментов видно, что MLE-RBA может значительно повысить эффективность обнаружения угроз по сравнению с предыдущими подходами к RBA. Ожидается, что сочетание с биометрией по динамике клавиатуры дополнительно повысит надежность обнаружения атак с перехватом сеанса в реальном времени без ущерба для удобства использования.

In response to the need for more secure authentication methods, new paradigms have been developed, such as continuous authentication (CA). We present a hybrid CA method combining ML-enhanced risk-based authentication (MLE-RBA) and keystroke dynamics. While this paper focuses on architectural design, earlier experiments indicate that MLE-RBA can significantly improve threat detection over earlier RBA approaches. The combination with keystroke biometrics is expected to further strengthen real-time detection of session hijacking attempts without compromising usability.

PACS: 89.20.Ff

ВВЕДЕНИЕ

Аутентификация пользователей остается ключевым составляющим звеном архитектуры безопасности любой распределенной системы. Поскольку значительная часть утечек данных связана с использованием слабых паролей или компрометацией учетных записей, а разработка новых методов атаки на системы аутентификации пользователей идет не прекращаясь, исследование и внедрение надежных методов аутентификации становятся как никогда актуальными [1]. Хотя «традиционные» механизмы аутентификации, такие как использование паролей, по-прежнему широко распространены, все большее внимание уделяется новым подходам к данной проблеме.

* E-mail: y.matyushin@spbu.ru

Одним из таких подходов является модель безопасности Zero Trust, основанная на принципе «никогда не доверяй, всегда проверяй» (never trust, always verify): система не выделяет «периметр безопасности», устройства внутри которого по умолчанию считаются безопасными, а проверяет каждое устройство перед предоставлением доступа к данным [2]. Одна из ключевых составляющих Zero Trust — непрерывная аутентификация (Continuous Authentication, CA), которая предполагает проверку личности пользователя не только в момент входа в систему, но и в течение всего сеанса взаимодействия с системой [3].

Ранее мы провели анализ существующих методов аутентификации, пригодных для реализации концепции CA. Наиболее точными и надежными из них оказались аутентификация по клавиатурному почерку и контекстная аутентификация, также известная как аутентификация на основе оценки рисков (Risk-Based Authentication, RBA) [4].

В то же время проведенный анализ показал, что каждое из этих решений обладает собственными уязвимостями, и уже были разработаны новые методы атаки, основанные на этих уязвимостях. В связи с этим мы считаем, что разработка гибридного метода, совмещающего сильные стороны разных подходов, является перспективной областью исследований.

В данной работе представлен новый подход — комбинация биометрической аутентификации по клавиатурному почерку и улучшенной с помощью методов машинного обучения системы RBA. Эта система ориентирована на обеспечение надежной защиты при сохранении удобства для конечного пользователя. Исходный код проекта доступен по адресу <https://github.com/YuriyM-SPB/Hybrid-auth-project>.

АУТЕНТИФИКАЦИЯ НА ОСНОВЕ АНАЛИЗА РИСКОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ (MLE-RBA)

Подробное описание метода MLE-RBA (Machine Learning-Enhanced Risk-Based Authentication), разработанного на основе работы Фримена и соавторов, было представлено нами в отдельной публикации [5]. В этом разделе приводится краткий обзор предлагаемого подхода.

Аутентификация на основе анализа рисков (RBA) работает за счет сравнения контекстных признаков входа пользователя (например, IP-адреса, устройства, браузера, операционной системы, интервала между входами и др.) с типичными паттернами поведения, характерными для данного пользователя. Чем больше соответствие между ними, тем ниже определяемый системой уровень риска. Если же полученный уровень риска превышает определенный порог, то попытка входа считается подозрительной.

Наш подход вдохновлен методом Фримена и др., основанным на использовании наивного байесовского классификатора [6], и развивает его идею с применением алгоритма машинного обучения (Isolation Forest,

Local Outlier Factor и LightGBM) для повышения точности оценки риска при аутентификации. Модель обучается на исторических данных о входах в систему, что позволяет ей выявлять сложные закономерности поведения, характерные для легитимных и нелегитимных попыток входа.

Мы реализовали и протестировали MLE-RBA на тех же данных, что и алгоритм Фримена. Полученный равный уровень ошибок (Equal Error Rate, EER) составил 19,8%, тогда как у метода Фримена на тех же данных он равен 24%. Показатель AUC-ROC также улучшился с 0,76 до 0,87, что говорит о большей способности модели отличать легитимные попытки входа от атак.

НЕПРЕРЫВНАЯ АУТЕНТИФИКАЦИЯ НА ОСНОВЕ КЛАВИАТУРНОГО ПОЧЕРКА

Аутентификация по клавиатурному почерку представляет собой биометрический метод, использующий уникальные особенности набора текста конкретного пользователя в качестве аутентификационного фактора. Аутентификация по произвольному тексту хорошо подходит для СА, так как работает в фоновом режиме и не требует дополнительных действий со стороны пользователя [7]. Вместе с тем у данного подхода есть и ограничения: например, для сохранения высокого уровня точности может потребоваться создание отдельных профилей для разных языков ввода [8].

Для определения стиля ввода пользователя используются такие параметры, как длительность нажатия клавиш (dwell time) и интервалы между нажатиями (flight time). В нашей системе эти данные собираются в фоновом режиме во время обычного ввода текста. Для обеспечения конфиденциальности на сервер отправляются только временные характеристики и обезличенные коды нажатых клавиш (без передачи самих символов).

На сервере реализован модуль распознавания, который формирует и постепенно уточняет индивидуальный профиль пользователя. В ходе работы системы полученные данные о манере ввода сравниваются с существующим профилем пользователя, на основе чего определяется степень отклонения манеры ввода от нормы. Если отклонения оказываются значительными, система интерпретирует это как возможную попытку подмены пользователя и повышает уровень риска.

Данный подход потенциально может вызывать опасения, связанные с конфиденциальностью, в частности, относительно того, собирает ли метод вводимые пользователем данные. Однако в нашей реализации фиксируются только метаданные, а не сами вводимые символы, по собираемым временным метаданным восстановление введенного текста невозможно. Кроме того, все данные передаются в зашифрованном виде, а на стороне сервера сохраняются только обезличенные статистические профили и оценки риска.

На текущем этапе модуль клавиатурной биометрии реализован в виде прототипа. Разработан скрипт на языке JavaScript, отвечающий за сбор временных характеристик на клиентской стороне, и API на сервере, принимающий и обрабатывающий эти данные. Поведенческие профили пользователей хранятся в базе данных, а логика распознавания аномалий реализована на языке Python.

АРХИТЕКТУРА ГИБРИДНОЙ СИСТЕМЫ

Предлагаемая гибридная система объединяет два описанных ранее подхода и определяет уровень риска на основе информации из двух независимых источников (контекстного анализа и анализа клавиатурного почерка пользователя). Общая логика работы системы представлена на рисунке.

Во время входа в систему модуль контекстной аутентификации формирует начальную оценку риска, используя профиль пользователя, хранящийся в базе данных. Если риск оценивается как низкий, то пользователю предоставляется доступ. Однако система продолжает фоновый мониторинг стиля ввода с помощью модуля клавиатурной аутентификации, который также вносит вклад в итоговую оценку. Если совокупный уровень риска превышает установленный порог, сеанс прерывается и пользователю предлагается пройти дополнительную проверку. В случае успешного прохождения проверки сеанс возобновляется, а уровень риска пересчитывается.

Важной частью архитектуры является механизм объединения рисков, который определяет, как рассчитывается итоговая оценка на основе показаний нескольких источников. Первая версия использовала простой подход по принципу логического «ИЛИ»: если хотя бы один модуль анализирует о высоком риске, попытка входа считается подозрительной. Однако в текущей версии используется взвешенное среднее:

$$R_{\text{combined}} = w_{\text{RBA}} R_{\text{RBA}} + w_{\text{keystroke}} R_{\text{keystroke}},$$

где R — уровень риска, а w — вес, назначенный определенному источнику данных.

Непрерывный мониторинг вводимых данных требует дополнительных ресурсов и может несколько замедлять систему, однако объем передаваемых сведений минимален (в среднем 0,03–0,06 мс на одно нажатие клавиши) и их обработка выполняется асинхронно, поэтому на практике заметного влияния на производительность не происходит.

В настоящее время ведется работа над усовершенствованием механизма объединения рисков, включая динамическую настройку весов и адаптацию порогов на основе статистики предыдущих сессий, поведенческих шаблонов пользователей и особенностей среды. Конечной целью является достижение баланса между уровнем безопасности и удобством использования.

ЗАКЛЮЧЕНИЕ

В работе предложена гибридная система непрерывной аутентификации, объединяющая два подхода — контекстную оценку риска и анализ клавиатурного почерка. Система позволяет выявлять возможные атаки на протяжении всей сессии, опираясь как на поведенческие, так и на контекстные признаки.

Был разработан, реализован и протестирован модуль контекстной аутентификации (MLE-RBA), основанный на алгоритмах машинного обучения, который показал лучшие результаты по сравнению с предыдущими подходами. Также создан прототип модуля клавиатурной аутентификации, действующего в фоновом режиме без вмешательства в работу пользователя.

В дальнейшем планируется доработка механизма объединения рисков, расширение модели клавиатурного анализа и практическая оценка системы.

Благодарности. Работа выполнена при поддержке СПбГУ (шифр проекта 121061000159-6).

СПИСОК ЛИТЕРАТУРЫ

1. 2025 Data Breach Investigations Report. <https://www.verizon.com/business/resources/reports/dbir/> (accessed 01.09.2025).
2. *Barach J.* Towards Zero Trust Security in SDN: A Multi-Layered Defense Strategy // Proc. 26th Intern. Conf. on Distributed Computing and Networking. 2025. P. 331–339.
3. *Oduri S.* Continuous Authentication and Behavioral Biometrics: Enhancing Cybersecurity in the Digital Era // Intern. J. Innov. Res. Sci. Engin. Technol. 2024. V. 13, No. 7. P. 13632–13640.
4. *Matiushin I. S., Korkhov V. V.* Continuous Authentication Methods for Zero-Trust Cybersecurity Architecture // Computational Science and Its Applications. Cham: Springer Nature Switzerland, 2023. P. 334–351.
5. *Matiushin I. S., Korkhov V. V.* MLE-RBA: A Machine Learning-Empowered Risk-Based Authentication Algorithm // Computational Science and Its Applications. Cham: Springer Nature Switzerland, 2025. P. 325–339.
6. *Freeman D. et al.* Who Are You? A Statistical Approach to Measuring User Authenticity // Network and Distributed System Security Symp. 2016. V. 16. P. 21–24.
7. *Shadman R. et al.* Keystroke Dynamics: Concepts, Techniques, and Applications // ACM Comput. Surv. 2025. V. 57, No. 11. P. 1–35.
8. *Altwaijry N.* Authentication by Keystroke Dynamics: The Influence of Typing Language // Appl. Sci. 2023. No. 13(20). 11478.