УДК 004.852

**И.С. Лошак, Е.Ю. Костюченко**

# Продлённая аутентификация на основе анализа журналов пользователя в ОС

Статья посвящена систематизации современных методов извлечения признаков и выявления аномалий на основе анализа журналов операционной системы для решения задачи продлённой аутентификации. Рассматриваются и классифицируются подходы к обработке и структурированию системных логов, включая извлечение количественных, индексных, семантических, временных, параметрических и графовых признаков. Проведен обзор открытых наборов данных для анализа логов и выполнен сравнительный анализ эффективности различных методов извлечения признаков и алгоритмов обнаружения аномалий, включая статистические методы, классическое машинное обучение, нейронные сети и гибридные модели. Эффективность оценивалась с точки зрения показателей качества работы классификаторов, решающих итоговую задачу. Определены наиболее перспективные направления для разработки систем продлённой аутентификации. Результаты исследования могут быть применены для повышения безопасности информационных систем за счёт разработки адаптивных механизмов аутентификации на основе мониторинга пользовательской активности.

**Ключевые слова:** продлённая аутентификация, машинное обучение, отбор признаков, информационная безопасность.

**DOI:** 10.21293/1818-0442-2025-28-4-39-49

С развитием цифровых технологий защита информации становится приоритетной задачей, так как традиционные методы аутентификации – действий по проверке подлинности субъекта доступа и/или объекта доступа, а также по проверке принадлежности субъекту доступа и/или объекту доступа предъявленного идентификатора доступа и аутентификационной информации [1], такие как пароли, подвержены компрометации. Биометрическая продлённая аутентификация (далее – продлённая аутентификация) представляет собой метод, при котором личность пользователя подтверждается на основе его биометрических характеристик или поведенческих данных на протяжении всего времени работы с системой [2].

Если поведение отклоняется от установленного шаблона, доступ может быть заблокирован или отправлено уведомление администратору. Ключевое преимущество продлённой аутентификации заключается в том, что уровень доверия, т.е. выполнения соответствующих действий или процедур для обеспечения уверенности в том, что оцениваемый объект соответствует своим целям безопасности [3], поддерживается и подтверждается на протяжении всего сеанса работы с системой, а не единожды при входе в неё. В качестве традиционных направлений для аутентификации на основе его биометрических характеристик или поведенческих данных можно выделить следующие:

1) текст [4, 5];
2) динамика подписи [6];
3) голос [7];
4) клавиатурный почерк [8].

Использование системных журналов в продлённой аутентификации позволяет фиксировать действия пользователя и анализировать их в реальном времени. При этом следует отметить, что использование динамических биометрических характеристик не является панацеей, поскольку практически все они имеют существенную вероятность ошибки первого и второго рода, что не позволяет говорить об их самостоятельном использовании [9].

Основная цель данной работы – систематизация основных методов по извлечению признаков на основе событий (логов) из журналов системы и выявлению на их основе аномалий при решении задачи продлённой аутентификации. Объект исследования – процедура проведения продлённой аутентификации. Предмет исследования – используемые признаки журналов системы и методов выявления аномалий для задачи продлённой аутентификации. Основные задачи, поставленные в данном исследовании, включают в себя:

– классифицировать методы извлечения признаков из системных журналов;

– классифицировать методы обнаружения аномалий для продлённой аутентификации;

– определить перспективные направления развития систем продлённой аутентификации.

**Стратегия поиска и анализа**

Поиск публикаций осуществлялся в базах данных Google Scholar и eLibrary за период с 2010 по ноябрь 2025 г. по формализованным запросам на русском и английском языках. В обзор включались исключительно первичные научные исследования (статьи и материалы конференций), содержащие оригинальные методы или экспериментальные результаты; обзорные работы и вторичные исследования исключались. Отбор публикаций выполнялся по критериям тематической релевантности: рассматривались только работы, связанные с продлённой аутентификацией либо поведенческим анализом и выявлением аномалий на основе журналов событий операционной системы, а не отдельных приложений. Для каждого компонента рассматривались наиболее цитируемые первичные публикации за два временных интервала (2016–2021 и 2022–2025 гг.).

Ввиду существенных различий между датасетами, метриками качества и сценариями применения, представленными в анализируемых работах, прямое

сравнение количественных показателей не рассматривалось. Вместо этого агрегирование осуществлялось на уровне используемых классов моделей и типов извлекаемых признаков. Процедура агрегирования заключалась в приведении результатов анализируемых работ к единой структурной модели и выявлении повторяющихся и устойчивых решений, подтверждённых несколькими независимыми исследованиями. Количественные метрики (F-мера, AUC) рассматривались как индикаторы эффективности целостных систем выявления аномалий, а не отдельных признаков.

Существующие обзорные работы, посвящённые анализу журналов событий в операционных системах, в основном рассматривают выявление аномалий с точки зрения устойчивости и безопасности системы. В исследованиях [10, 11] аномалии интерпретируются преимущественно как сбои, атаки или нетипичные состояния системы, тогда как поведенческий аспект пользователя либо не выделяется явно, либо рассматривается фрагментарно как второстепенный источник признаков. В результате методы анализа логов оцениваются преимущественно в контексте обнаружения инцидентов, а не как инструмент моделирования индивидуального поведения пользователя. В отличие от указанных подходов в данной работе анализ журналов событий рассматривается именно в контексте продлённой аутентификации, где ключевым объектом является динамика действий пользователя, а аномалии интерпретируются как отклонения от его индивидуального поведенческого профиля.

**Проблемы продленной аутентификации по журналам системы**

Реализация системы продленной аутентификации по журналам системы обнаруживает ряд проблем на различных компонентах. В первую очередь, сбор данных о действиях пользователя предполагает выбор источников логов, которые содержат информацию о действиях пользователя в системе. Основная задача – выбрать те журналы (или логи), из которых извлекаются данные о пользовательской активности, минимизируя при этом шумовые события, связанные с внутренними процессами операционной системы.

Обработка событий необходима для стандартизации данных, удаления избыточной информации и приведения их к унифицированному формату. Это важно, поскольку данные из журналов часто имеют разрозненный формат и могут содержать повторяющиеся события. Разные подходы к обработке данных позволяют добиться баланса между точностью анализа и производительностью системы.

Выделение признаков играет ключевую роль в продленной аутентификации, так как признаки являются основой для анализа поведения пользователя. На этом этапе события преобразуются в информативные показатели, которые описывают характерные паттерны действий, последовательности событий и временные зависимости. Различные методы выделения признаков предлагают разные преимущества в зависимости от задач и доступных данных.

Компонент обнаружения аномалий выполняет финальную задачу системы – анализ текущих признаков поведения и сравнение их с эталоном. Здесь применяются методы, которые могут варьировать от простых статистических моделей до сложных алгоритмов машинного обучения и нейронных сетей. Выбор метода зависит от сложности данных, требований к производительности и уровня адаптивности системы.

**Обзор наборов данных**

Среди наиболее известных наборов данных, применяемых в анализе логов и системного мониторинга, особое значение имеют HDFS, BGL, Thunderbird, OpenStack и Loghub. HDFS [12] стал одним из первых масштабных источников событий распределённой файловой системы Hadoop, позволив исследователям разрабатывать и тестировать алгоритмы анализа больших данных в условиях высокой нагрузки и распределённой архитектуры. Он стал основой для формирования подходов к структурированию неупорядоченных событий и поиску закономерностей в их потоках.

Набор данных BGL [13], созданный на базе суперкомпьютера Blue Gene/L, сосредоточен на исследовании надежности вычислительных систем. Он способствовал формированию целого направления в анализе событий, связанного с прогнозированием отказов и анализом взаимосвязей между процессами в масштабных вычислительных кластерах. Thunderbird [13] предоставил исследователям богатую базу для изучения поведения систем в долгосрочных наблюдениях, включая сетевые сбои, аппаратные ошибки и аномалии производительности.

OpenStack [14], в свою очередь, направлен на анализ событий облачных инфраструктур и применение для разработки методов мониторинга распределённых сервисов и виртуализированный сред. Наиболее обобщающим из них является Loghub [15] – репозиторий, созданный для систематизации и объединения разнообразных логов из различных источников, включая HDFS, BGL, Thunderbird и OpenStack. Он стал универсальной платформой для тестирования и сравнения алгоритмов анализа логов, способствуя стандартизации подходов в области интеллектуального анализа системных событий. Для сравнения в табл. 1 представлены основные открытые наборы данных.

Таблица 1

**Сравнение открытых наборов данных**

| Название | Кол-во событий | Размер, ГБ | Тип |
|---|---|---|---|
| HDFS | 11 175 629 | 2,412 | Системные логи Hadoop |
| BGL | 4 747 963 | 1,2 | Логи суперкомпьютера IBM |
| Thunderbird | 211 212 192 | 27,3 | Кластерные вычисления суперкомпьютера |
| OpenStack | 1 335 318 | 0,058 | Логи облачных платформ |
| HPC | 433 490 | 0,032 | Логи вычислительного кластера |
| LogHub | 494 765 193 | 83,47 | 16 различных систем |

Несмотря на то, что большинство открытых наборов данных напрямую не связаны с действиями пользователей в операционной системе, их также можно применять при реализации методов продлен-

ной аутентификации. Открытые данные могут быть полезны в задачах выявления аномалий. Многие из них содержат последовательности событий, которые можно анализировать для обнаружения отклонений от нормального поведения системы. Например, наборы данных, такие как BGL, широко используются для тестирования моделей машинного обучения и выявления аномалий, связанных с поведением системных процессов. Эти подходы могут быть адаптированы для анализа пользовательской активности.

**Методы извлечения признаков**

Набор событий, поступающий на вход системы выявления аномалий, не подходит для прямого анализа, так как в этом случае система анализирует не поведение пользователя, а отдельные, разрозненные данные о действиях в системе. Такой подход не только затрудняет понимание реального поведения пользователя, но и значительно снижает производительность системы, особенно при обработке больших объемов данных. Кроме того, данный формат данных не адаптирован для применения методов машинного обучения, так как модели требуют структурированных и информативных признаков, которые отражают ключевые аспекты поведения, поэтому необходимо преобразовать события в набор признаков.

*Количественные признаки.* Эти признаки отражают частоту событий в заданном временном окне, что делает их простыми для извлечения и полезными при обнаружении аномалий. Так, события [16–18] группировались по фиксированным, скользящим и сессионным окнам, после чего формировалась матрица частот событий, где каждая строка представляла последовательность, а столбцы – количество различных событий. Также стоит учитывать, что между различными шаблонами логов существуют числовые взаимосвязи, указывающие на коррелированные изменения частот [19, 20].

*Индекс событий.* В отличие от количественных признаков индексные не зависят от частоты событий, а кодируют каждый шаблон события уникальным номером. Эти номера, упорядоченные по времени, позволяют сохранять позиционные отношения между событиями. В [14] предложено представлять журналы как последовательность индексов, где предсказание следующего события осуществляется на основе истории, а отклонения от нормальной последовательности сигнализируют об аномалиях. Аналогичный подход использовали [21, 22], назначая уникальные индексы лог-ключам – фиксированным частям сообщений в логах, что упрощает обработку

*Семантические признаки.* Традиционные статистические признаки, основанные на количестве или индексах событий, не всегда сохраняют эффективность при изменениях в структуре журналов и не отражают семантику текста. При помощи данного признака можно анализировать как одно событие [23–30], так и их последовательность [31, 32].

Авторы работы [23] предложили использовать алгоритм Word2Vec для преобразования слов логов в векторные представления: сначала текст очищается от небуквенно-цифровых символов, затем каждое событие представляется как среднее арифметическое векторов слов. Несмотря на это, Word2Vec ограничен в интерпретации сложных зависимостей между событиями.

Чтобы решить данные проблемы, в работе [24] предложен метод Template2Vec, использующий базу синонимов и антонимов из WordNet для векторизации шаблонов логов. В работе [25] применяется векторизация слов с помощью TF-IDF и предобученных векторов FastText для представления каждого события в виде семантического вектора. Исследование [26, 27] предлагает метод для прямого кодирования лог-событий, учитывающего локальные зависимости между событиями. Также популярно применение языковой модели (BERT, GPT-2) для создания устойчивых к изменениям в логах векторов фиксированной размерности [28–30].

Признаки отдельных событий логов не всегда содержат достаточно информации, особенно о предыдущих событиях. Поэтому важно учитывать временные связи между событиями. Но учитывание всех слов занимает много места, поэтому в [31] незначимые слова в событиях фильтруются, а важные – кодируются в векторы, затем объединяются в вектор последовательности.

В [32] использовали взаимодействие слов в логах для создания признаков последовательности. Хотя признаки последовательности агрегируют информацию о событиях, их применение ограничено из-за сложности обработки длинных последовательностей и вариативности данных.

*Временные характеристики.* Эти признаки основаны на временной информации каждой записи лога и позволяют анализировать распределение событий во времени для выявления аномалий, например, чрезмерно длинных интервалов или резких всплесков [33]. В [34] предложили два метода временного кодирования для этого: временная маскировка встраивания событий и совместное встраивание событийного времени. Оба метода позволяют повысить точность прогнозирования временных рядов, что создает предпосылки для применения временного кодирования в логарифмическом анализе.

В [35] предложен метод обнаружения аномалий в логах по временных характеристикам, не влияющий на производительность системы. Исследователи [36] вычисляли разницу во времени между событиями, формируя последовательность временных интервалов. Однако это одномерный временной ряд с ограниченной информативностью. Чтобы улучшить его, в [37] применили one-hot кодирование и линейный слой. Однако в распределённых системах логи генерируются разными процессами, усложняя анализ временных зависимостей.

*Параметры событий.* Помимо временных характеристик, важны и параметры событий (например, IP-адреса или имена приложений), содержащие значимую системную информацию. В [14] предложен метод, объединяющий анализ временных интервалов с контекстом параметров логов для обнаружения аномалий. Такой подход позволяет учитывать последовательность событий вместе с дополнительными данными, однако встречается проблема: числовые значения некоторых параметров могут существенно отличаться, оставаясь семантически схо-

жими (например, «загрузка завершена за 1 с» и «загрузка завершена за 800 мс»). Применение простых средних значений или фиксированных порогов не всегда эффективно, поэтому модель обучается учитывать контекстное взаимодействие параметров для более точной интерпретации логов.

*Графовые признаки.* Анализ графов широко используется в различных областях исследований, и методы встраивания графов привлекли большое внимание исследователей [38–40]. Представление данных в виде графа подразумевает, что объекты (например, события, процессы) рассматриваются как узлы, а зависимости или последовательности между ними – как ребра [41–43].

В работе [44] для извлечения признаков из логов применяется алгоритм Node2Vec, который с помощью случайных блужданий изучает локальную и глобальную структуру графа, обрабатывая последовательности узлов аналогично предложениям в естественном языке.

Алгоритм Log2Vec, описанный в работе [45], преобразует журналы в гетерогенный граф, где узлы представляют записи логов, а ребра отражают причинно-следственные, временные и логические связи между ними. Для векторизации узлов используются случайное блуждание и метод Word2Vec, что позволяет выделить их поведенческие характеристики.

Для сравнения в табл. 2 представлены основные методы извлечения с указанием набора данных, а также значение F-мера, полученное, различными исследователями за счет применение представленных методов. В таблицу были внесены только лучшие результаты, которые смогли достичь исследователи. Стоит отметить, что явным недостатком такого сравнения является то, что сравниваются метрики, отражающие результаты систем выявления аномалий, базирующиеся на данных признаках, а не метрики эффективности признаков в определении пользователя, такие как информативность. Это связано с тем, что большая часть исследований была посвящена разработке системы выявления аномалий, а извлечение признаков было лишь одним из этапов. Поэтому для оценки эффективности проделанной работы использовались метрики, отражающие результаты всей системы.

Таблица 2

**Сравнение методов извлечения признаков**

| Вид признака | Работа | Модель | Набор данных | F-мера |
|---|---|---|---|---|
| Количественный | [16] | Дерево решений | HDFS | 0,99 |
| | | | BGL | 0,85 |
| | [18] | SVM с объединением с наивным Байесом | 2 закрытых набора данных | Среднее значение – 0,97 |
| | [19] | LSTM | 10 собственных закрытых наборов | Среднее значение – 0,66 |
| | [20] | MLP | HDFS | 0,99 |
| Количественный и временные | [17] | OCSVM с объединением с PSO | Закрытый набор данных | 0,98 |
| Индексный, временные и параметры | [14] | LSTM | HDFS | 0,96 |
| | | | OpenStack | 0,97 |
| Индексный | [21] | CNN | HDFS | 0,99 |
| | [22] | MLP и случайный лес | 3 собственных закрытых наборов | Среднее значение – 0,90 |
| Семантический | [24] | LSTM | BGL | 0,96 |
| | | | HDFS | 0,95 |
| | [25] | LSTM | HDFS | 0,99 |
| | [26] | Случайный лес | BGL | 0,89 |
| | [27] | Автоэнкодер | CERT | 0,90 |
| | [28] | Transformer | BGL | 0,92 |
| | | | HDFS | 0,98 |
| | | | OpenStack | 0,88 |
| | [29] | LSTM | OpenStack | 0,81 |
| | [30] | Transformer | BGL | 0,99 |
| | | | HDFS | 0,99 |
| | | | OpenStack | 0,99 |
| | [31] | LSTM | HDFS | 0,98 |
| | [32] | MLP | BGL | 0,99 |
| Семантический и временные | [36] | LSTM | BGL | 0,99 |
| | | | HDFS | 0,99 |
| Семантический, параметры и временные | [37] | Transformer (BERT) и LSTM | HDFS | 0,99 |
| Графовые | [45] | Конечный автомат | CERT | AUC – 0,93 |
| | | | LANL | AUC – 0,91 |

**Методы выявления аномалий**

Поведение пользователей может быть как статичным (регулярное выполнение однотипных задач), так и динамичным (действия зависят от контекста, времени суток и других факторов). Статистические методы эффективны для предсказуемых сценариев, в то время как для сложных и изменчивых шаблонов требуются алгоритмы машинного обучения или

нейронные сети, способные адаптироваться к изменениям в поведении и структуре данных.

*Конечные автоматы.* Этот метод основан на определении типичных состояний поведения пользователя и построении графа переходов между ними. В данной модели [46, 47] события инициируют переходы между состояниями конечного автомата, которые представляют этапы процессов, например, «начало транзакции», «обработка данных» или «завершение операции». Автомат обучается на нормальном поведении системы, определяя допустимые переходы, а любые отклонения от заданных правил считаются аномалией.

*Нейронные сети.* Наибольшее применение для выявления аномалий находят методы с применением нейронных сетей. Среди них можно выделить несколько основных архитектур. В первую очередь, это рекуррентные нейронные сети (RNN), а также их вариации: долгая краткосрочная память (LSTM) [12, 14, 17] и управляемые рекуррентные блоки (GRU) [48]. RNN позволяют учитывать предыдущие входные данные при анализе текущих, что позволяет рассматривать событие в системе в контексте предыдущих. Данная архитектура используется для обработки последовательных данных, поскольку способна учитывать временные зависимости между событиями [49].

Также к данной задаче применима архитектура свёрточных нейронных сетей (CNN), которые подходят для обработки событий, представленных в виде матриц или векторных последовательностей. CNN позволяет объединять некоторую последовательность событий в одно число, что позволяет проанализировать данную последовательность и связи событий внутри нее [22, 50].

Помимо этого, стоит отметить, что на данный момент популярно использование архитектуры трансформеров, а также предобученных моделей на данной архитектуре BERT и GPT. Данная архитектура позволяет анализировать логи с учетом их контекста, обеспечивая высокоточную классификацию событий. Модели BERT [51, 52] и GPT [53] использовались для векторизации логов и анализа последовательностей событий. Также некоторые исследователи [53–57] строили свои модели на данной архитектуре и получили аналогичные результаты, как и при использовании предобученных моделей. Такой подход позволяет выявлять аномалии, учитывая семантические связи между событиями.

Также для решения данной задачи используют автокодировщики [27, 58]. В контексте данной задачи данная архитектура предназначена для уменьшения размерности данных и поиска аномалий на основе ошибки реконструкции записи событий. Эти модели обучаются на нормальных данных и затем определяют аномалии по высокой ошибке восстановления.

Некоторые исследования комбинируют различные архитектуры для повышения точности обнаружения аномалий. В исследованиях [59, 49] использовались модели, объединяющие CNN, и графовых нейронных сетей (GNN). Такой подход позволяет использовать сильные стороны каждой из моделей. Так, RNN анализирует последовательности, а CNN анализирует связи между объектами.

В работе [60] предложена гибридная модель, которая объединяет несколько архитектур: LSTM для извлечения признаков, вариационные автокодировщики (VAE) для оценки аномалии через ошибку реконструкции, а также алгоритмы обнаружения выбросов типа Isolation Forest. Так, разные блоки покрывают разные аспекты задачи: LSTM определяет поведение системы, автокодировщик даёт возможность выявлять аномалии на основе того, что модель не научилась «восстанавливать» подобные случаи, а алгоритм Isolation Forest добавляет устойчивости и учитывает остаточные разбросы вне пространства, где нейронная сеть может быть переобучена. Данное объединение архитектур позволяет повысить точность, устойчивость к шуму и способность обнаружения неожиданных изменений в системах.

*Методы классического машинного обучения.* Классическое машинное обучение в данной статье – это набор алгоритмов, которые обучаются на данных событий для автоматического определения нормального поведения системы и выявления аномалий, которые не включают в себя нейронные сети. Они не требуют больших вычислительных ресурсов и могут быть интерпретируемыми, что важно для анализа и объяснения решений модели. Аналогичные результаты нейронным сетям показали методы опорных векторов [61, 17], случайный лес [62] и деревья решений [63]. Также в работе [17] было показано, что метод наивного Байеса не подходит для данной задачи и дает точность выявления ниже 40%.

Для сравнения в табл. 3 представлены основные методы выявления аномалий с указанием набора данных, а также значение F-мера и наименование модели, которая выдала данный результат. В таблицу были внесены только лучшие результаты, которые смогли достичь исследователи. Также не были включены методы, результаты работы которых не оценивались по F-мере, что не позволяет их сравнить с другими методами.

**Заключение**

В результате обзора показано, какие классы методов выявления аномалий и по каким признакам являются наиболее подходящими для продленной аутентификации по данным ОС. Это позволяет читателю использовать результаты обзора при проектировании архитектуры системы продлённой аутентификации и выборе методов анализа. В отличие от ранее опубликованных обзоров, рассматривающих анализ логов преимущественно в контексте выявления сбоев и атак на систему, в данной статье практические выводы сформулированы с точки зрения поведения пользователя как основного объекта анализа.

Проведенное исследование позволило систематизировать современные подходы к продлённой аутентификации на основе анализа системных журналов. Отметим, что за последние 8 лет существенно

увеличилось количество технических решений, построенных с использованием алгоритмов глубокого обучения, что позволяет с высокой точностью определять аномалии в журналах, достигая значений F-меры 0,95–1,00 на различных наборах данных. При этом важно отметить, что классические методы машинного обучения в ряде случаев показывают сопоставимые с нейронными сетями результаты, что свидетельствует о сохранении их практической значимости.

Таблица 3

**Сравнение методов выявления аномалий**

| Метод обнаружения аномалий | Работа | Модель | Набор данных | F-мера |
|---|---|---|---|---|
| Нейронная сеть | [48] | GRU | BGL | 0,98 |
| | | | HDFS | 0,96 |
| | [49] | GGNN | TrainTicket | 0,95 |
| | [50] | CNN (TCN) | BGL | 0,98 |
| | | | HDFS | 0,97 |
| | [51] | Transformer (BERT) | BGL | 0,90 |
| | | | HDFS | 0,82 |
| | | | Thunderbird | 0,96 |
| | [52] | Transformer (BERT) | BGL | 0,91 |
| | | | HDFS | 0,95 |
| | | | Thunderbird | 0,94 |
| | [53] | Transformer (GPT) | HDFS | 0,99 |
| | [54] | Transformer | BGL | 0,83 |
| | | | HDFS | 0,99 |
| | [55] | Transformer | BGL | 0,98 |
| | | | HDFS | 0,99 |
| | | | Thunderbird | 0,99 |
| | [56] | Transformer | BGL | 0,65 |
| | | | Thunderbird | 0,99 |
| | [57] | Transformer | BGL | 0,92 |
| | | | HDFS | 0,89 |
| | | | Thunderbird | 0,90 |
| | [58] | Автоэнкодеры | BGL | 0,95 |
| | [59] | LSTM и GNN | HDFS | 0,96 |
| | | | OpenStack | 0,96 |
| Классические методы машинного обучения | [60] | Случайный лес | Собственный закрытый набор с 4 видами атак | 0,97 |
| | [63] | Дерево решений, случайный лес, SVM | HDFS | 0,99 |
| | | Дерево решений | BGL | 0,97 |
| | | Дерево решений | Thunderbird | 0,99 |

Анализ методов извлечения признаков выявил, что наиболее перспективным направлением является использование индексных, семантических, временных характеристик и параметров событий. Эти типы признаков оптимально сочетают информативность для описания поведения пользователя и применимость в алгоритмах машинного обучения. Однако следует констатировать, что большинство существующих прикладных исследований в данной области ориентировано на анализ работоспособности системы, а не на решение задач аутентификации пользователей.

Перспективы дальнейших исследований заключаются в адаптации рассмотренных методов и инструментов для создания специализированной системы биометрической продлённой аутентификации на основе журналов операционной системы Windows. Такой подход позволит реализовать непрерывный контроль подлинности пользователя в течение всего сеанса работы, что существенно повысит уровень безопасности информационных систем.

Данная работа выполнялась в рамках Программы развития ТУСУРа на 2025–2036 годы, Программы стратегического академического лидерства «Приоритет 2030».

*Литература*

1. ГОСТ Р 58833–2020. Национальный стандарт Российской Федерации. Защита информации. Идентификация и аутентификация. – М.: Стандартинформ, 2020. – 32 с.

2. European Data Protection Supervisor. Biometric Continuous Authentication [Электронный ресурс]. – URL: https://www.edps.europa.eu/press-publications/publications /techsonar/biometric-continuous-authentication_en (дата обращения: 10.12.2024).

3. ГОСТ Р ИСО/МЭК 15408-1–2012. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. – М.: Стандартинформ, 2012. – 75 с.

4. Романов А.С. Обобщенная методика идентификации автора неизвестного текста / А.С. Романов, А.А. Шелупанов, С.С. Бондарчук // Доклады ТУСУР. – 2010. – № 3-1 (21). – С. 108–112.

5. Fedotova A. Authorship attribution of social media and literary Russian-language texts using machine learning methods and feature selection / A. Fedotova, A. Romanov, A. Kurtukova, A. Shelupanov // Future Internet. – 2021. – Vol. 14, No. 1. – P. 4. DOI: 10.3390/fi14010004.

6. Дорошенко Т.Ю. Система аутентификации на основе динамики рукописной подписи / Т.Ю. Дорошенко, Е.Ю. Костюченко // Доклады ТУСУР. – 2014. – № 2 (32). – С. 219–223.

7. Рахманенко И.А. Автоматическая верификация диктора по произвольной фразе с применением свёрточных глубоких сетей доверия / И.А. Рахманенко, А.А. Шелупанов, Е.Ю. Костюченко // Компьютерная оптика. – 2020. – Т. 44, № 4. – С. 596–605.

8. Затеев С. В. Продленная аутентификация на основе анализа клавиатурного почерка // II Всерос. науч.-практ. конф. «Теория и практика обеспечения информационной безопасности». – М.: МТУСИ, 2023. – С. 116–124.

9. Актуальные направления развития методов и средств защиты информации / А.А. Шелупанов, О.О. Евсютин, А.А. Конев, Е.Ю. Костюченко, Д.В. Кручинин, Д.С. Никифоров // Доклады ТУСУР. – 2017. – Т. 20, № 3. – С. 11–24.

10. Ma J. Automatic parsing and utilization of system log features in log analysis: A survey / J. Ma, Y. Liu, H. Wan, G. Sun // Applied Sciences. – 2023. – Vol. 13, No. 8. – P. 4930.

11. A survey on automated log analysis for reliability engineering / S. He, P. He, Zh. Chen, T. Yang, Yu. Su, M.R. Lyu // ACM computing surveys (CSUR). – 2021. – Vol. 54, No. 6. – P. 1–37.

12. Detecting large-scale system problems by mining console logs / W. Xu, L. Huang, A. Fox, D. Patterson, M.I. Jordan // Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles. – Big Sky: Association for Computing Machinery, 2009. – P. 117–132.

13. Oliner A. What supercomputers say: A study of five system logs / A. Oliner, J. Stearley // 37th Annual IEEE/IFIP international conference on dependable systems and networks (DSN'07). – Washington: IEEE, 2007. – P. 575–584.

14. Du M. Deeplog: Anomaly detection and diagnosis from system logs through deep learning / M. Du, F. Li, G. Zheng, V. Srikumar // Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. – Dallas: Association for Computing Machinery, 2017. – P. 1285–1298.

15. Loghub: A large collection of system log datasets for ai-driven log analytics / J. Zhu, Sh. He, P. He, J. Liu, M.R. Lyu // 2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE). – Florence: IEEE, 2023. – P. 355–366.

16. He Sh. Experience report: System log analysis for anomaly detection / Sh. He, J. Zhu, P. He, M.R. Lyu // 2016 IEEE 27th International Symposium on software reliability engineering (ISSRE). – Ottawa: IEEE, 2016. – P. 207–218.

17. Abin A.A. Continuous user authentication using a combination of operation and application-related features / A.A. Abin, P. Hosseini, R.A. Torabian // Journal of Innovations in Computer Science and Engineering (JICSE). – 2023. – Vol. 1. – P. 11–22.

18. Pokhrel R. Anomaly-based–intrusion detection system using user profile generated from system logs / R. Pokhrel, P. Pokharel, A.K. Timalsina // International Journal of Scientific and Research Publications (IJSRP). – 2019. – Vol. 9. – P. 8631.

19. Zhao N. An empirical investigation of practical log anomaly detection for online service systems / N. Zhao, H. Wang, Z. Li, X. Peng // Proceedings of the 29th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering. – Athens: Association for Computing Machinery, 2021. – P. 1404–1415.

20. Logdp: Combining dependency and proximity for log-based anomaly detection / Y. Xie, H. Zhang, B. Zhang, M.A. Babar, Sh. Lu // International Conference on Service-Oriented Computing. – Dubai: Cham: Springer International Publishing, 2021. – P. 708–716.

21. Lu S. Detecting anomaly in big data system logs using convolutional neural network / S. Lu, X. Wei, Y. Li, L. Wang // 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). – Athens: IEEE, 2018. – P. 151–158.

22. Yen S. Causalconvlstm: Semi-supervised log anomaly detection through sequence modeling / S. Yen, M. Moh, T.S. Moh // 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA). – Boca Raton: IEEE, 2019. – P. 1334–1341.

23. Bertero C. Experience report: Log mining using natural language processing and application to anomaly detection / C. Bertero, M. Roy, C. Sauvanaud, G. Trédan // 2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE). – Toulouse: IEEE, 2017. – P. 351–360.

24. Loganomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs / W. Meng, Y. Liu, Y. Zhu, Sh. Zhang, D. Pei, Y. Liu, Y. Chen, R. Zhang, Sh. Tao, P. Sun, R. Zhou // IJCAI. – 2019. – Vol. 19, No. 7. – P. 4739–4745.

25. Robust log-based anomaly detection on unstable log data / X. Zhang, Y. Xu, Q. Lin, B. Qiao, H. Zhang, Y. Dang et al. // Proceedings of the 2019 27th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering. – Tallinn: Association for Computing Machinery, 2019. – P. 807–817.

26. LogEvent2vec: LogEvent-to-vector based anomaly detection for large-scale logs in internet of things / J. Wang, Y. Tang, Sh. He, Ch. Zhao, P.K. Sharma, O. Alfarraj, A. Tolba // Sensors. – 2020. – Vol. 20, No. 9. – P. 2451.

27. Doc2vec-based insider threat detection through behaviour analysis of multi-source security logs / L. Liu, Ch. Chen, J. Zhang, O. De Vel, Y. Xiang // 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). – Guangzhou: IEEE, 2020. – P. 301–309.

28. Hitanomaly: Hierarchical transformers for anomaly detection in system log / Sh. Huang, Y. Liu, C. Fung, R. He, Y. Zhao, H. Yang, Z. Luan // IEEE transactions on network and service management. – 2020. – Vol. 17, No. 4. – P. 2064–2076.

29. Robust and transferable anomaly detection in log data using pre-trained language models / H. Ott, J. Bogatinovski, A. Acker, S. Nedelkoski, O. Kao // 2021 IEEE/ACM international workshop on cloud intelligence (CloudIntelligence). – Madrid: IEEE, 2021. – P. 19–24.

30. Unsupervised Log Anomaly Detection Method Based on Multi-Feature / Sh. He, T. Deng, B. Chen, R.S. Sherratt, J. Wang // Computers, Materials & Continua. – 2023. – Vol. 76, No. 1. – P. 517–541.

31. Lv D. ConAnomaly: Content-based anomaly detection for system logs / D. Lv, N. Luktarhan, Y. Chen // Sensors. – 2021. – Vol. 21, No. 18. – P. 6125.

32. Ryciak P. Anomaly detection in log files using selected natural language processing methods / P. Ryciak, K. Wasielewska, A. Janicki // Applied Sciences. – 2022. – Vol. 12, No. 10. – P. 5089.

33. Corney M. Detection of anomalies from user profiles generated from system logs / M. Corney, G. Mohay, A. Clark // Proceedings of the Ninth Australasian Information Security Conference. – Perth: Australian Computer Society, 2011. – P. 23–31.

34. Li Y. Time-dependent representation for neural event sequence prediction / Y. Li, N. Du, S. Bengio // arXiv preprint arXiv:1708.00065. – 2017. – P. 1–11.

35. Rak T. Using Data Mining techniques for detecting dependencies in the Outcoming Data of a web-based system / T. Rak, R. Żyła // Applied Sciences. – 2022. – Vol. 12, No. 12. – P. 6115.

36. Swisslog: Robust and unified deep learning based log anomaly detection for diverse faults / X. Li, P. Chen, L. Jing, Z. He, G. Yu // 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE). – Coimbra: IEEE, 2020. – P. 92–103.

37. AllInfoLog: Robust diverse anomalies detection based on all log features / R. Xiao, H. Chen, J. Lu, W. Li, Sh. Jin // IEEE Transactions on Network and Service Management. – 2022. – Vol. 20, No. 3. – P. 2529–2543.

38. Backes M. Walk2friends: Inferring social links from mobility profiles / M. Backes, M. Humbert, J. Pang, Y. Zhang // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. – Dallas: Association for Computing Machinery, 2017. – P. 1943–1957.

39. Dai H. Discriminative embeddings of latent variable models for structured data / H. Dai, B. Dai, L. Song // International conference on machine learning. – New York: PMLR, 2016. – P. 2702–2711.

40. Neural network-based graph embedding for cross-platform binary code similarity detection / X. Xu, Ch. Liu, Q. Feng, H. Yin, L. Song, D. Song // Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. – Dallas: Association for Computing Machinery, 2017. – P. 363–376.

41. Non-Intrusive performance profiling for entire software stacks based on the flow reconstruction principle / X. Zhao, K. Rodrigues, Y. Luo, D. Yuan, M. Stumm // 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). – Savannah: USENIX Association, 2016. – P. 603–618.

42. Holmes: real-time apt detection through correlation of suspicious information flows / S.M. Milajerdi, R. Gjomemo, B. Eshete, R. Sekar, V.N. Venkatakrishnan // 2019 IEEE symposium on security and privacy (SP). – San Francisco: IEEE, 2019. – P. 1137–1152.

43. RShield: A refined shield for complex multi-step attack detection based on temporal graph network / W. Yang, P. Gao, H. Huang, X. Wei, W. Liu, Sh. Zhu & W. Luo // International Conference on Database Systems for Advanced AP.lications. – Hyderabad: Cham: Springer International Publishing, 2022. – P. 468–480.

44. Grover A. node2vec: Scalable feature learning for networks / A. Grover, J. Leskovec // Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. – San Francisco: Association for Computing Machinery, 2016. – P. 855–864.

45. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise / F. Liu, Y. Wen, D. Zhang, X. Jiang, X. Xing, D. Meng // Proceedings of the 2019 ACM SIGSAC conference on computer and communications security. – London: Association for Computing Machinery, 2019. – P. 1777–1794.

46. A real-time anomaly detection method for industrial control systems based on long-short period deterministic finite automaton / X. Lin, Y. Yao, B. Hu, W. Yang, X. Zhou, G. Li, W. Zhang // IEEE Internet of Things Journal. – 2025. – Vol. 12, No. 10. – P. 14599–14621.

47. LogLens: A real-time log analysis system / B. Debnath, M. Solaimani, M.A. Gulzar Gulzar, N. Arora et al. // 2018 IEEE 38th international conference on distributed computing systems (ICDCS). – Vienna: IEEE, 2018. – P. 1052–1062.

48. Semi-supervised log-based anomaly detection via probabilistic label estimation / L. Yang, J. Chen, Z. Wang, W. Wang, J. Jiang, X. Dong, W. Zhang // 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE). – Madrid: IEEE, 2021. – P. 1448–1460.

49. Deeptralog: Trace-log combined microservice anomaly detection through graph-based deep learning / Ch. Zhang, X. Peng, Ch. Sha, K. Zhang, Zh. Fu, X. Wu, Q. Lin, D. Zhang // Proceedings of the 44th international conference on software engineering. – Pittsburgh: Association for Computing Machinery, 2022. – P. 623–634.

50. LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge / Z. Wang, J. Tian, H. Fang, L. Chen, J. Qin // Computer Networks. – 2022. – Vol. 203. – P. 108616.

51. Guo H. Logbert: Log anomaly detection via bert / H. Guo, S. Yuan, X. Wu // 2021 international joint conference on neural networks (IJCNN). – Shenzhen: IEEE, 2021. – P. 1–8.

52. Almodovar C. LogFiT: Log anomaly detection using fine-tuned language models / C. Almodovar, F. Sabrina, S. Karimi, S. Azad // IEEE Transactions on Network and Service Management. – 2024. – Vol. 21, No. 2. – P. 1715–1723.

53. Hadadi F. LLM meets ML: Data-efficient Anomaly Detection on Unseen Unstable Logs / F. Hadadi, Q. Xu, D. Bianculli, L. Briand // ACM Transactions on Software Engineering and Methodology. – 2025. DOI: 10.1145/3771283.

54. Logformer: Cascaded Transformer for System Log Anomaly Detection / F. Hang, W. Guo, H. Chen, L. Xie, Ch. Zhou, Y. Liu // Computer Modeling in Engineering & Sciences (CMES). – 2023. – Vol. 136, No. 1. – P. 517–529.

55. Translog: A unified transformer-based framework for log anomaly detection / H. Guo, X. Lin, J. Yang, Y. Zhuang, J. Bai, T. Zheng, B. Zhang, Z. Li // arXiv preprint arXiv:2201.00016. – 2022. – P. 1–7.

56. Self-attentive classification-based anomaly detection in unstructured logs / S. Nedelkoski, J. Bogatinovski, A. Acker, J. Cardoso, O. Kao // 2020 IEEE International Conference on Data Mining (ICDM). – Sorrento: IEEE, 2020. – P. 1196–1201.

57. TraLogAnomaly: A microservice system anomaly detection approach based on hybrid event sequences / X. Wei, Ch.-ai Sun, P. Yang, X.-Y. Zhang, D. Towey // Science of Computer Programming. – 2025. – Vol. 245. – P. 103303.

58. Catillo M. AutoLog: Anomaly detection by deep autoencoding of system logs / M. Catillo, A. Pecchia, U. Villano // Expert Systems with Applications. – 2022. – Vol. 191. – P. 116263.

59. SecEncoder: Logs are All You Need in Security / M.F. Bulut, Y. Liu, N. Ahmad, M. Turner, S.A. Ouahmane, C. Andrews, L. Greenwald // arXiv preprint arXiv:2411.07528. – 2024. DOI: 10.48550/arXiv.2411.07528.

60. Bereketoglu A.B. Hybrid Meta-Learning Framework for Anomaly Forecasting in Nonlinear Dynamical Systems via Physics-Inspired Simulation and Deep Ensembles // arXiv preprint arXiv:2506.13828. – 2025. DOI: 10.48550/arXiv.2506.13828.

61. Husselman L. Anomaly Detection with Windows Event Logs: A comparative study between traditional and ML based approaches: Master's thesis. – University of Zurich, 2024. – 184 p.

62. Кечеджиев А.С. Методика выявления аномалий в данных оценки кибератак с использованием Random Forest

и градиентного бустинга в машинном обучении / А.С. Кечеджиев, О.Л. Цветкова, А.И. Дубровина // Вестник Дагестанского гос. техн. ун-та. Технические науки. – 2024. – Т. 51, № 3. – С. 72–85.

63. Wu X. On the effectiveness of log representation for log-based anomaly detection / X. Wu, H. Li, F. Khomh // Empirical Software Engineering. – 2023. – Vol. 28, No. 6. – P. 137.

---

**Лошак Иван Сергеевич**
Студент факультета безопасности
Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Ленина пр-т, 40, г. Томск, Россия, 634050
Тел.: +7 (382-2) 70-18-17
Эл. почта: lis@fb.tusur.ru

**Костюченко Евгений Юрьевич**
Канд. техн. наук, доцент каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050
ORCID: 0000-0001-8000-2716
Тел.: +7 (382-2) 70-15-29
Эл. почта: key@fb.tusur.ru

Поступила в редакцию: 06.10.2025.
Принята к публикации: 27.02.2026.

Loshak I.S., Kostyuchenko E.Y.

**Extended authentication based on user log analysis in the operating system**

The paper is devoted to the systematization of modern methods for feature extraction and anomaly detection based on the analysis of operating system logs to address the problem of extended authentication. Approaches to processing and structuring system logs are reviewed and classified, including the extraction of quantitative, index, semantic, temporal, parametric, and graph features. An overview of open datasets for log analysis is provided. The authors performed a comparative analysis of the effectiveness of various feature extraction methods and anomaly detection algorithms, encompassing statistical methods, classical machine learning, neural networks, and hybrid models. Effectiveness was evaluated in terms of the performance metrics of classifiers solving the final task. The most promising areas for developing extended authentication systems are identified. The research results can be applied to to enhance the security of information systems through the development of adaptive authentication mechanisms based on user activity monitoring.

**Keywords:** extended authentication, machine learning, feature extraction, information security.
**DOI:** 10.21293/1818-0442-2025-28-4-39-49

*References*

1. GOST R 58833–2020. *Natsional'nyj standart Rossijskoi Federatsii. Zashchita informatsii. Identifikatsiya i autentifikatsiya* [National standard of the Russian Federation. Information protection. Identification and authentication]. Moscow, Standartinform, 2020, 32 p.

2. European Data Protection Supervisor. Biometric Continu-ous Authentication. Available at: https://www.edps.europa.eu/press-publications/publications /techsonar/biometric-continuous-authentication_en (Accessed: 10 December 2024).

3. GOST R ISO/IEC 15408-1–2012. *Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya bezopasnosti. Kriterii otsenki bezopasnosti informatsionnyh tekhnologii* [Information technology. Security techniques. Evaluation criteria for IT security]. Part 1: Introduction and general model. Moscow, Standartinform, 2012, 75 p.

4. Romanov A.S., Shelupanov A.A., Bondarchuk S.S. [Generalized authorship identification technique]. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki*, 2010, No. 3-1 (21), pp. 108–112 (in Russ.)

5. Fedotova A., Romanov A., Kurtukova A., Shelupanov A. Authorship attribution of social media and literary Russian-language texts using machine learning methods and feature selection. *Future Internet*, 2021, vol. 14, no. 1, pp. 4. DOI: 10.3390/fi14010004.

6. Doroshenko T.Y., Kostyuchenko E.Yu. [The authentication system based on dynamic handwritten signature]. *Doklady TUSUR*, 2014, no. 2(32), pp. 219–223 (in Russ.)

7. Rakhmanenko I.A., Shelupanov A.A., Kostyuchenko E.Yu. [Automatic text-independent speaker verification using convolutional deep belief network]. *Computer Optics*, 2020, vol. 44, no. 4, pp. 596–605 (in Russ.).

8. Zateev S.V. *Prodlenaya autentifikatciya na ocnove klaviaturnogo pocherka* [Continuous authentication based on keystroke dynamics analysis]. II Vserossijskaya nauchno-prakticheskaya konferentsiya «Teoriya i praktika obespecheniya informatsionnoi bezopasnosti» [*2nd All-Russian scientific and practical conference «Theory and practice of information security»*], M.: MTUCI, 2023, pp. 116 (in Russ.)

9. Shelupanov A.A., Evsyutin O.O., Konev A.A., Kostyuchenko E.Yu., Kruchinin D.V., Nikiforov D.S. [Modern trends in development of methods and means for information protection]. *Doklady TUSUR*, 2017, vol. 20, no. 3, pp. 11–24 (in Russ.).

10. Ma J., Liu Y., Wan H., Sun G. Automatic parsing and utilization of system log features in log analysis: A survey. *Applied Sciences*, 2023, vol. 13, no. 8, pp. 4930.

11. He S., He P., Chen Zh., Yang T., Su Yu., Lyu M.R. A survey on automated log analysis for reliability engineering. *ACM computing surveys (CSUR)*, 2021, vol. 54, no. 6, pp. 1–37.

12. Xu W., Huang L., Fox A., Patterson D., Jordan M.I. Detecting large-scale system problems by mining console logs. Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles, Big Sky, Association for Computing Machinery, 2009, pp. 117–132.

13. Oliner A., Stearley J. What supercomputers say: A study of five system logs. 37th annual IEEE/IFIP international conference on dependable systems and networks (DSN'07), Washington, IEEE, 2007, pp. 575–584.

14. Du M., Li F., Zheng G., Srikumar V. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, Dallas, Association for Computing Machinery, 2017, pp. 1285–1298.

15. Zhu J., He Sh., He P., Liu J., Lyu M.R. Loghub: A large collection of system log datasets for ai-driven log analytics. 2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE), Florence, IEEE, 2023, pp. 355–366.

16. He Sh., Zhu J., He P., Lyu M.R. Experience report: System log analysis for anomaly detection. 2016 IEEE 27th international symposium on software reliability engineering (ISSRE), Ottawa, IEEE, 2016, pp. 207–218.

17. Abin A.A., Hosseini P., Torabian R.A. Continuous user authentication using a combination of operation and application-related features. *Journal of Innovations in Computer Science and Engineering (JICSE)*, 2023, vol. 1, pp. 11–22.

18. Pokhrel R., Pokharel P., Timalsina A.K. Anomaly-based–intrusion detection system using user profile generated from system logs. *International Journal of Scientific and Research Publications (IJSRP)*, 2019, vol. 9, pp. 8631.

19. Zhao N., Wang H., Li Z., Peng X. An empirical investigation of practical log anomaly detection for online service systems. Proceedings of the 29th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering, Athens, Association for Computing Machinery, 2021, pp. 1404–1415.

20. Xie Y., Zhang H., Zhang B., Babar M.A., Lu Sh. Logdp: Combining dependency and proximity for log-based anomaly detection. International Conference on Service-Oriented Computing, Dubai, Cham: Springer International Publishing, 2021, pp. 708–716.

21. Lu S., Wei X., Li Y., Wang L. Detecting anomaly in big data system logs using convolutional neural network. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, IEEE, 2018, pp. 151–158.

22. Yen S., Moh M., Moh T.S. Causalconvlstm: Semi-supervised log anomaly detection through sequence modeling. 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, IEEE, 2019, pp. 1334–1341.

23. Bertero C., M. Roy, Sauvanaud C., Trédan G. Experience report: Log mining using natural language processing and application to anomaly detection. 2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE), Toulouse, IEEE, 2017, pp. 351–360.

24. Meng W., Liu Y., Zhu Y., Zhang Sh., Pei D., Liu Y., Chen Y., Zhang R., Tao Sh., Sun P., Zhou R. Loganomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs. *IJCAI*, 2019, vol. 19, no. 7, pp. 4739–4745.

25. Zhang X. Xu Y., Lin Q., Qiao B., Zhang H., Dang Y. et al. Robust log-based anomaly detection on unstable log data. Proceedings of the 2019 27th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering, Tallinn Association for Computing Machinery, 2019, pp. 807–817.

26. Wang J., Tang Y., He Sh., Zhao Ch., Sharma P.K., Alfarraj O., Tolba A. LogEvent2vec: LogEvent-to-vector based anomaly detection for large-scale logs in internet of things. *Sensors*, 2020, vol. 20, no. 9, pp. 2451.

27. Liu L., Chen Ch., Zhang J., De Vel O., Xiang Y. Doc2vec-based insider threat detection through behaviour analysis of multi-source security logs .2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, IEEE, 2020, pp. 301–309.

28. Huang Sh., Liu Y., Fung C., He R., Zhao Y., Yang H., Luan Z. Hitanomaly: Hierarchical transformers for anomaly detection in system log. *IEEE Transactions on network and service management*, 2020, vol. 17, no. 4, pp. 2064–2076.

29. Ott H., Bogatinovski J., Acker A., Nedelkoski S., Kao O. Robust and transferable anomaly detection in log data using pre-trained language models. 2021 IEEE/ACM International workshop on cloud intelligence (CloudIntelligence), Madrid, IEEE, 2021, pp. 19–24.

30. He Sh., Deng T., Chen B., Sherratt R.S., Wang J. Unsupervised Log Anomaly Detection Method Based on Multi-Feature. *Computers, Materials & Continua*, 2023, vol. 76, no. 1, pp. 517–541.

31. Lv D., Luktarhan N., Chen Y. ConAnomaly: Content-based anomaly detection for system logs. *Sensors*, 2021, vol. 21, no. 18, pp. 6125.

32. Ryciak P., Wasielewska K., Janicki A. Anomaly detection in log files using selected natural language processing methods. *Applied Sciences*, 2022, vol. 12, no. 10, pp. 5089.

33. Corney M., Mohay G., Clark A. Detection of anomalies from user profiles generated from system logs. Proceedings of the Ninth Australasian Information Security Conference, Perth, Australian Computer Society, 2011, pp. 23–31.

34. Li Y., Du N., Bengio S. Time-dependent representation for neural event sequence prediction. *arXiv preprint arXiv:1708.00065*, 2017, pp. 1–11.

35. Rak T., Żyła R. Using Data Mining techniques for detecting dependencies in the Outcoming Data of a web-based system. *Applied Sciences*, 2022, vol. 12, no. 12, pp. 6115.

36. Li X., Chen P., Jing L., He Z., Yu G. Swisslog: Robust and unified deep learning based log anomaly detection for diverse faults. 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE), Coimbra, IEEE, 2020, pp. 92–103.

37. Xiao R., Chen H., Lu J., Li W., Jin Sh. AllInfoLog: Robust diverse anomalies detection based on all log features. *IEEE Transactions on Network and Service Management*, 2022, vol. 20, no. 3, pp. 2529–2543.

38. Backes M., Humbert M., Pang J., Zhang Y. Walk2friends: Inferring social links from mobility profiles. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, Association for Computing Machinery, 2017, pp. 1943–1957.

39. Dai H., Dai B., Song L. Discriminative embeddings of latent variable models for structured data. International conference on machine learning, New York, PMLR, 2016, pp. 2702–2711.

40. Xu X., Liu Ch., Feng Q., Yin H., Song L., Song D. Neural network-based graph embedding for cross-platform binary code similarity detection. Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, Dallas, Association for Computing Machinery, 2017, pp. 363–376.

41. Zhao X., Rodrigues K., Luo Y., Yuan D., Stumm M. Non-Intrusive performance profiling for entire software stacks based on the flow reconstruction principle. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, USENIX Association, 2016, pp. 603–618.

42. Milajerdi S.M., Gjomemo R., Eshete B., Sekar R., Venkatakrishnan V.N. Holmes: real-time apt detection through correlation of suspicious information flows. 2019 IEEE symposium on security and privacy (SP), San Francisco IEEE, 2019, pp. 1137–1152.

43. Yang W., Gao P., Huang H., Wei X., Liu W., Zhu Sh. & Luo W. RShield: A refined shield for complex multi-step attack detection based on temporal graph network. International Conference on Database Systems for Advanced AP.lications, Hyderabad, Cham: Springer International Publishing, 2022, pp. 468–480.

44. Grover A., Leskovec J. node2vec: Scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco: Association for Computing Machinery, 2016, pp. 855–864.

45. Liu F., Wen Y., Zhang D., Jiang X., Xing X., Meng D. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise. Proceedings of the 2019 ACM SIGSAC conference on computer and

communications security, London, Association for Computing Machinery, 2019, pp. 1777–1794.

46. Lin X., Yao Y., Hu B., Yang W., Zhou X., Li G., Zhang W. A real-time anomaly detection method for industrial control systems based on long-short period deterministic finite automaton. *IEEE Internet of Things Journal*, 2025, vol. 12, no. 10, pp. 14599–14621.

47. Debnath B., Solaimani M., Gulzar Gulzar M.A., Arora N. et al. LogLens: A real-time log analysis system .2018 IEEE 38th international conference on distributed computing systems (ICDCS), Vienna, IEEE, 2018, pp. 1052–1062.

48. Yang L., Chen J., Wang Z., Wang W., Jiang J., Dong X., Zhang W. Semi-supervised log-based anomaly detection via probabilistic label estimation. 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), Madrid, IEEE, 2021, pp. 1448–1460.

49. Zhang C., Peng X., Sha Ch., Zhang K., Fu Zh., Wu X., Lin Q., Zhang D. Deeptralog: Trace-log combined microservice anomaly detection through graph-based deep learning. Proceedings of the 44th international conference on software engineering, Pittsburgh, Association for Computing Machinery, 2022, pp. 623–634.

50. Wang Z., Tian J., Fang H., Chen L., Qin J. LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge. *Computer Networks*, 2022, vol. 203, pp. 108616.

51. Guo H., Yuan S., Wu X. Logbert: Log anomaly detection via bert. 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, IEEE, 2021, pp. 1–8.

52. Almodovar C., Sabrina F., Karimi S., Azad S. LogFiT: Log anomaly detection using fine-tuned language models. *IEEE Transactions on Network and Service Management*, 2024, vol. 21, no. 2, pp. 1715–1723.

53. Hadadi F., Xu Q., Bianculli D., Briand L. LLM meets ML: Data-efficient Anomaly Detection on Unseen Unstable Logs. *ACM Transactions on Software Engineering and Methodology*, 2025. DOI: 10.1145/3771283.

54. Hang F., Guo W., Chen H., Xie L., Zhou Ch., Liu Y. Logformer: Cascaded Transformer for System Log Anomaly Detection. *Computer Modeling in Engineering & Sciences (CMES)*, 2023, vol. 136, no. 1, P. 517–529.

55. Guo H., Lin X., Yang J., Zhuang Y., Bai J., Zheng T., Zhang B., Li Z. Translog: A unified transformer-based framework for log anomaly detection. *arXiv preprint arXiv:2201.00016*, 2022, pp. 1–7.

56. Nedelkoski S., Bogatinovski J., Acker A., Cardoso J., Kao O. Self-attentive classification-based anomaly detection in unstructured logs. 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, IEEE, 2020, pp. 1196–1201.

57. Wei X. Sun Ch.-ai, Yang P., Zhang X.-Y., Towey D. TraLogAnomaly: A microservice system anomaly detection approach based on hybrid event sequences. *Science of Computer Programming*, 2025, vol. 245, pp. 103303.

58. Catillo M., Pecchia A., Villano U. AutoLog: Anomaly detection by deep autoencoding of system logs. *Expert Systems with Applications*, 2022, vol. 191, pp. 116263.

59. Bulut M.F., Liu Y., Ahmad N., Turner M., Ouahmane S.A., Andrews C., Greenwald L. SecEncoder: Logs are All You Need in Security. *arXiv preprint arXiv:2411.07528*, 2024. DOI: 10.48550/arXiv.2411.07528.

60. Bereketoglu A.B. Hybrid Meta-Learning Framework for Anomaly Forecasting in Nonlinear Dynamical Systems via Physics-Inspired Simulation and Deep Ensembles. *arXiv preprint arXiv:2506.13828*, 2025. DOI: 10.48550/arXiv.2506.13828.

61. Husselman L. Anomaly Detection with Windows Event Logs: A comparative study between traditional and ML based approaches. Master's thesis, University of Zurich, 2024, 184 p.

62. Kechedzhiev A.S., Tsvetkova O.L., Dubrovina A.I. Terskikh M., Tishina E. [Methodology for detecting anomalies in cyber attack assessment data using Random Forest and Gradient Boosting in machine learning]. *Herald of Daghestan State Technical University. Technical Sciences*, 2024, vol. 51, no. 3, pp. 72–85 (in Russ.)

63. Wu X., Li H., Khomh F. On the effectiveness of log representation for log-based anomaly detection. *Empirical Software Engineering*, 2023, vol. 28, no. 6, pp. 137.

---

**Ivan S. Loshak**
Student, Faculty of Security
Tomsk State University of Control Systems
and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 70-18-17
Email: lis@fb.tusur.ru

**Evgeniy Y. Kostyuchenko**
Doctor of Engineering Sciences, Associate Professor
Department of Complex Information Security of Computer Systems, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0001-8000-2716
Phone: +7 (382-2) 70-15-29
Email: key@fb.tusur.ru

Received: 06.10.2025.
Accepted: 27.02.2026.