УДК 004.056.5:004.89

Козачок Александр Васильевич, Доктор технических наук, профессор кафедры КБ-4, «Интеллектуальные системы информационной безопасности», МИРЭА – Российский Технологический Университет, г. Москва

Худяков Дмитрий Андреевич, магистрант, МИРЭА – Российский Технологический Университет, г. Москва

ПРИМЕНЕНИЕ ГРАФОВОГО АВТОЭНКОДЕРА ДЛЯ ВЫЯВЛЕНИЯ АНОМАЛЬНЫХ СОБЫТИЙ АУТЕНТИФИКАЦИИ В КОРПОРАТИВНОЙ ИНФРАСТРУКТУРЕ

Аннотация

В статье рассматривается подход к выявлению аномальных событий аутентификации в корпоративной инфраструктуре на основе графового представления журналов безопасности. Актуальность темы связана с тем, что сложные целевые атаки APT часто сопровождаются использованием легитимных учётных записей, штатных удалённых протоколов и постепенным перемещением злоумышленника между узлами сети. В таких условиях отдельное событие входа может не содержать очевидных признаков вредоносной активности, однако его подозрительность проявляется в контексте других взаимодействий. Предлагается представлять пользователей и хосты как вершины графа, а события аутентификации и удалённого доступа — как рёбра между ними. Для анализа таких данных используется графовый автоэнкодер с кодировщиком на основе механизма графового внимания. Оценка аномальности формируется на основе ошибки восстановления признаков взаимодействий. Разработан программный прототип на языке

Python с использованием библиотек PyTorch и PyTorch Geometric. Экспериментальная проверка на модельном наборе данных показала возможность ранжирования событий по степени подозрительности и выделения нетипичных взаимодействий, связанных с появлением новых связей, удалёнными подключениями и использованием привилегированных учётных записей.

Annotation

The article considers an approach to detecting anomalous authentication events in corporate infrastructure based on a graph representation of security logs. The relevance of the topic is related to the fact that advanced persistent threats often involve the use of legitimate accounts, standard remote access protocols, and gradual movement of an attacker between network nodes. In such conditions, a single authentication event may not contain obvious signs of malicious activity, while its suspicious nature becomes visible only in the context of other interactions. The proposed approach represents users and hosts as graph nodes, while authentication and remote access events are represented as edges between them. A graph autoencoder with an attention-based graph encoder is used to analyze such data. The anomaly score is calculated using the reconstruction error of interaction features. A software prototype was developed in Python using PyTorch and PyTorch Geometric libraries. Experimental evaluation on a model dataset demonstrated the possibility

of ranking events by their suspiciousness and identifying atypical interactions related to new links, remote connections, and the use of privileged accounts.

Ключевые слова: APT-атаки, журналы аутентификации, графовые нейронные сети, графовый автоэнкодер, GAT, UEBA, обнаружение аномалий, информационная безопасность.

Keywords: APT attacks, authentication logs, graph neural networks, graph autoencoder, GAT, UEBA, anomaly detection, information security.

Сложные целевые атаки Advanced Persistent Threat, или APT, остаются одной из наиболее серьёзных угроз для корпоративных информационных систем. Их отличие от массовых атак заключается не только в более высокой технической подготовке злоумышленника, но и в самой логике развития инцидента. APT-кампания, как правило, не ограничивается одним действием. После первичного доступа атакующий стремится закрепиться в инфраструктуре, собрать дополнительные учётные данные, изучить внутренние ресурсы и перейти к более ценным узлам сети. На практике это означает, что признаки атаки распределены по времени и по разным источникам событий [2; 6].

Особую сложность создаёт использование легитимных учётных записей и штатных механизмов доступа. В базе MITRE ATT&CK техника Valid Accounts описывает сценарии, при которых злоумышленники используют существующие локальные, доменные, облачные или VPN-учётные записи для получения доступа, закрепления и перемещения внутри сети [8]. В этом случае

событие входа может выглядеть корректным: пользователь существует, протокол разрешён, целевой хост доступен. Подозрительность возникает не из самого факта аутентификации, а из контекста: пользователь обращается к непривычному серверу, привилегированная запись используется с нетипичного хоста, а несколько отдельных подключений складываются в цепочку горизонтального перемещения [9].

Именно поэтому журналы аутентификации представляют интерес не только как набор отдельных записей, но и как источник данных о структуре доступа внутри инфраструктуры. События входа отражают, какие пользователи, хосты и сервисные учётные записи взаимодействуют между собой. При табличном анализе такая информация частично теряется, поскольку каждая запись рассматривается изолированно. Графовое представление, напротив, сохраняет связи между сущностями и позволяет анализировать изменение этих связей во времени [1].

В предлагаемом подходе журнал аутентификации преобразуется в граф взаимодействий. Пользователи и хосты рассматриваются как вершины, а события аутентификации и удалённого доступа — как рёбра между ними. В пределах временного окна граф можно описать следующим образом:

$$G_t = (V_t, E_t, X_t, F_t),$$

где V_t— множество вершин, E_t— множество рёбер, X_t— матрица признаков вершин, F_t— признаки рёбер. Такой формат удобен для анализа сценариев, в которых подозрительность связана не с одним параметром события, а с изменением карты взаимодействий. Например, новая связь между пользователем и сервером может быть важнее самого факта успешного входа.

Для формирования графа исходные события проходят нормализацию. Имена пользователей, хостов и других сущностей приводятся к единому виду,

после чего каждой уникальной сущности присваивается внутренний идентификатор вершины. Рёбра строятся на основе фактов взаимодействия. Если пользователь выполнил вход на целевой хост, между соответствующими вершинами создаётся ребро. В качестве признаков ребра могут использоваться тип входа, результат аутентификации, признак привилегированности, активность в нерабочее время, частота взаимодействия и признак новой связи.

Таблица 1 — Примеры признаков, используемых при построении графа событий аутентификации

Группа признаков	**Примеры признаков**	**Назначение**
Признаки пользователя	количество входов, число уникальных хостов, доля успешных входов	описание поведения учётной записи
Признаки хоста	число пользователей, входящие и исходящие связи	оценка роли узла в инфраструктуре
Признаки ребра	тип входа, успешность, привилегированность, нерабочее время	описание конкретного события доступа
Исторические признаки	новая связь, изменение частоты взаимодействий	выявление отклонений от привычного поведения

Для анализа построенного графа используется графовый автоэнкодер. Автоэнкодерная схема выбрана из-за специфики задачи: для APT-активности трудно получить полный и достоверный набор размеченных примеров, тогда как данные о нормальной или преимущественно легитимной активности доступны чаще. Модель обучается восстанавливать типичные взаимодействия. Если некоторое событие восстанавливается хуже остальных, оно получает более высокую оценку аномальности [7].

Кодировщик автоэнкодера строит скрытые представления вершин с учётом их признаков и локального окружения. В качестве кодировщика используется Graph Attention Network, или GAT. Механизм внимания важен для данной задачи, поскольку большая часть событий аутентификации относится к регулярной легитимной активности. Простое усреднение соседних связей может сгладить редкие, но значимые взаимодействия. GAT,

напротив, обучаемым образом назначает разный вес соседним вершинам и тем самым помогает выделять связи, которые выбиваются из привычного окружения пользователя или хоста [10].

Декодировщик восстанавливает признаки взаимодействий между вершинами. В прототипе основной акцент сделан на ошибке восстановления признаков рёбер. Такая постановка удобна для журналов аутентификации: аномалия не всегда связана с появлением новой связи. Иногда пользователь действительно имеет доступ к серверу, но конкретное событие отличается временем, типом входа, частотой обращений или результатом аутентификации. Поэтому модель должна учитывать не только факт связи, но и параметры события [7].

Итоговая оценка аномальности может быть представлена как комбинация структурной и признаковой ошибки:

$$s_{ij} = \alpha \ell_{ij}^{struct} + (1 - \alpha)\ell_{ij}^{feat},$$

где s_{ij}— оценка аномальности взаимодействия между вершинами v_iи v_j, ℓ_{ij}^{struct}— ошибка восстановления структуры, ℓ_{ij}^{feat}— ошибка восстановления признаков ребра, α— коэффициент, определяющий вклад каждой составляющей. В рамках реализованного прототипа использовалась признаковая ошибка восстановления, а структурная составляющая рассматривается как направление дальнейшего расширения модели.

Программный прототип разработан на языке Python. Для построения и обучения модели использовались библиотеки PyTorch и PyTorch Geometric [8, 9]. Обработка табличных данных выполнялась с использованием Pandas, расчёт метрик — с помощью Scikit-learn, визуализация результатов — с помощью Matplotlib. Структура проекта включала модули генерации

модельных логов, построения графа, описания модели, обучения и оценки результатов.

Экспериментальная проверка выполнялась на модельном наборе событий аутентификации. Структура набора была спроектирована с учётом открытых наборов событий безопасности, прежде всего LANL Comprehensive Multi-Source Cyber-Security Events, который содержит события аутентификации, процессы, DNS-запросы, сетевые потоки и red team events [5]. В рамках прототипа моделировались нормальные события входа пользователей, обращения к хостам, а также аномальные сценарии: появление новой связи между пользователем и хостом, горизонтальное перемещение, использование привилегированной учётной записи в нерабочее время и серия неуспешных попыток с последующим успешным входом.

В ходе эксперимента был сформирован набор из 15 000 событий. После построения графа получено 208 вершин и 15 000 рёбер. Размерность признаков вершин составила 7, размерность признаков рёбер — 10. Доля аномальных событий в наборе составила около 3,5 %, что отражает характерный для задач информационной безопасности дисбаланс между нормальной и подозрительной активностью.

Обучение графового автоэнкодера проводилось в течение 80 эпох. Значение ошибки реконструкции снизилось с 0,328009 на первой эпохе до 0,020335 на последней эпохе. После обучения для каждого ребра была рассчитана оценка аномальности. В качестве подозрительных выбирались верхние 4 % событий с наибольшим anomaly score.

По результатам оценки были получены следующие значения метрик: ROC-AUC — 0,9234, PR-AUC — 0,3755, F1-score — 0,4071. Полученное значение ROC-AUC показывает, что модель достаточно хорошо ранжирует аномальные события выше нормальных. Более умеренные значения PR-AUC и F1-score объясняются сильным дисбалансом классов и ограниченным

характером модельного набора данных. Для задач безопасности такая ситуация является ожидаемой: увеличение полноты обнаружения часто сопровождается ростом числа ложных срабатываний.

Матрица ошибок имела следующий вид:

$$\begin{bmatrix} 14104 & 371 \\ 296 & 229 \end{bmatrix}$$

Модель корректно определила 14 104 нормальных события и 229 аномальных событий. При этом 371 нормальное событие было ошибочно отмечено как подозрительное, а 296 аномальных событий не попали в выбранный порог. Эти результаты показывают, что разработанный прототип способен выделять часть нетипичных взаимодействий, однако требует дальнейшей настройки признаков, порога и архитектуры для уменьшения числа ложных срабатываний и пропусков.

Полученные результаты следует рассматривать как проверку работоспособности предложенного конвейера, а не как промышленную оценку готовой системы обнаружения APT-атак. Основная ценность прототипа состоит в демонстрации полного цикла обработки: подготовка событий аутентификации, построение графа, обучение графового автоэнкодера и ранжирование событий по степени подозрительности. Для практического применения метод необходимо проверять на более крупных открытых или корпоративных журналах безопасности.

Таким образом, графовое представление журналов аутентификации является перспективным направлением для выявления аномальных взаимодействий в корпоративной инфраструктуре. В отличие от табличного анализа, графовая модель сохраняет структуру связей между пользователями и хостами. Использование графового автоэнкодера даёт возможность выделять события, которые плохо согласуются с изученным поведением, и

передавать их на дальнейшую проверку специалисту по информационной безопасности.

Дальнейшее развитие работы связано с применением предложенного подхода к открытым наборам событий безопасности, включая LANL Comprehensive Multi-Source Cyber-Security Events, а также с расширением графа за счёт сетевых потоков, событий процессов и данных средств защиты конечных узлов. Кроме того, отдельного внимания требует переход от однородного графа к гетерогенному представлению, где пользователи, хосты, сервисные учётные записи и сетевые адреса обрабатываются как сущности разных типов.

Список литературы

1. Лаборатория Los Alamos National Laboratory. Comprehensive Multi-Source Cyber-Security Events [Электронный ресурс]. – URL: https://csr.lanl.gov/data/cyber1/ (дата обращения: 14.05.2026).
2. Mandiant. M-Trends 2025 [Электронный ресурс]. – URL: https://cloud.google.com/blog/topics/threat-intelligence/m-trends-2025 (дата обращения: 14.05.2026).
3. PyTorch Documentation [Электронный ресурс]. – URL: https://pytorch.org/docs/stable/index.html (дата обращения: 14.05.2026).
4. PyTorch Geometric Documentation [Электронный ресурс]. – URL: https://pytorch-geometric.readthedocs.io/ (дата обращения: 14.05.2026).
5. Scikit-learn Documentation [Электронный ресурс]. – URL: https://scikit-learn.org/stable/ (дата обращения: 14.05.2026).
6. Solar 4RAYS. Отчёт по расследованиям инцидентов за первое полугодие 2025 года [Электронный ресурс]. – URL: https://rt-solar.ru/upload/iblock/1df/6ciryrpphqtguxi3u5hxob07auawu376/DFIR-report-H1-2025_final-_2_.pdf (дата обращения: 14.05.2026).
7. Kipf T. N., Welling M. Variational Graph Auto-Encoders [Электронный ресурс] // arXiv. – 2016. – URL: https://arxiv.org/abs/1611.07308 (дата обращения: 14.05.2026).
8. MITRE ATT&CK. Valid Accounts [Электронный ресурс]. – URL: https://attack.mitre.org/techniques/T1078/ (дата обращения: 14.05.2026).
9. MITRE ATT&CK. Lateral Movement [Электронный ресурс]. – URL: https://attack.mitre.org/tactics/TA0008/ (дата обращения: 14.05.2026).
10. Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph Attention Networks [Электронный ресурс] // arXiv. – 2017. – URL: https://arxiv.org/abs/1710.10903 (дата обращения: 14.05.2026).

References

1. Los Alamos National Laboratory. Comprehensive Multi-Source Cyber-Security Events. Available at: https://csr.lanl.gov/data/cyber1/ (accessed: 14.05.2026).
2. Mandiant. M-Trends 2025. Available at: https://cloud.google.com/blog/topics/threat-intelligence/m-trends-2025 (accessed: 14.05.2026).
3. PyTorch Documentation. Available at: https://pytorch.org/docs/stable/index.html (accessed: 14.05.2026).
4. PyTorch Geometric Documentation. Available at: https://pytorch-geometric.readthedocs.io/ (accessed: 14.05.2026).
5. Scikit-learn Documentation. Available at: https://scikit-learn.org/stable/ (accessed: 14.05.2026).
6. Solar 4RAYS. Incident Response Report for the First Half of 2025. Available at: https://rt-solar.ru/upload/iblock/1df/6ciryrpphqtguxi3u5hxob07auawu376/DFIR-report-H1-2025_final-_2_.pdf (accessed: 14.05.2026).
7. Kipf T. N., Welling M. Variational Graph Auto-Encoders. arXiv, 2016. Available at: https://arxiv.org/abs/1611.07308 (accessed: 14.05.2026).
8. MITRE ATT&CK. Valid Accounts. Available at: https://attack.mitre.org/techniques/T1078/ (accessed: 14.05.2026).
9. MITRE ATT&CK. Lateral Movement. Available at: https://attack.mitre.org/tactics/TA0008/ (accessed: 14.05.2026).
10. Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y. Graph Attention Networks. arXiv, 2017. Available at: https://arxiv.org/abs/1710.10903 (accessed: 14.05.2026).