ОШ МАМЛЕКЕТТИК УНИВЕРСИТЕТИНИН ЖАРЧЫСЫ

ВЕСТНИК ОШСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

BULLETIN OF OSH STATE UNIVERSITY

ISSN 1694-7452 e-ISSN: 1694-8610
№3/2026, 170-182

ИНФОРМАТИКА

УДК: 004.056
DOI: 10.52754/16948610_2026_3_14

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ АНАЛИЗА МЕТАДАННЫХ TCP-ПАКЕТОВ И МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ВОССТАНОВЛЕНИЯ СОЦИАЛЬНЫХ ГРАФОВ ПОЛЬЗОВАТЕЛЕЙ МЕССЕНДЖЕРОВ

МЕССЕНДЖЕРЛЕРДИН КОЛДОНУУЧУЛАРЫНЫН СОЦИАЛДЫК ГРАФТАРЫН КАЛЫБЫНА КЕЛТИРҮҮ ҮЧҮН TCP-ПАКЕТТЕРДИН МЕТАМААЛЫМАТТАРЫН ЖАНА ЖАСАЛМА ИНТЕЛЛЕКТ МЕТОДДОРУН ТАЛДООНУН НЕГИЗИНДЕ МААЛЫМАТТЫК КООПСУЗДУК КОРКУНУЧТАРЫН МОДЕЛДӨӨ ЖАНА БОЛЖОЛДОО

MODELING AND FORECASTING INFORMATION SECURITY THREATS BASED ON TCP PACKET METADATA ANALYSIS AND ARTIFICIAL INTELLIGENCE METHODS FOR RECONSTRUCTING SOCIAL GRAPHS OF MESSENGER USERS

Асилбеков Тынчтыкбек Майрамбекович
Асилбеков Тынчтыкбек Майрамбекович
Tynchtykbek Mairambekovich Asilbekov

преподаватель, Ошский государственный университет
окутуучу, Ош мамлекеттик университети
Lecturer, Osh State University
mir.titan.90@gmail.com
ORCID: 0009-0002-4292-1580

Орозов Максатбек Омурбекович
Орозов Максатбек Омурбекович
Maksatbek Omurbekovich Orozov

к.ф.-м.н., доцент, Ошский государственный университет
ф.-м.и.к., доцент, Ош мамлекеттик университети
Candidate of Physical and Mathematical Sciences, Associate Professor, Osh State University
orozov@oshsu.kg

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ АНАЛИЗА МЕТАДАННЫХ TCP-ПАКЕТОВ И МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ВОССТАНОВЛЕНИЯ СОЦИАЛЬНЫХ ГРАФОВ ПОЛЬЗОВАТЕЛЕЙ МЕССЕНДЖЕРОВ

Аннотация

Современные системы обнаружения вторжений, основанные на сигнатурном анализе и классических методах машинного обучения, демонстрируют принципиальную ограниченность при выявлении многоэтапных целевых атак (APT), распределённых во времени и пространстве. Переход от реактивного детектирования к предиктивному моделированию требует принципиально новых подходов, учитывающих не только изолированные события, но и структурные связи между субъектами информационного обмена. Особую остроту приобретает проблема утечки метаданных: даже при сквозном шифровании содержимого сообщений пассивный наблюдатель на уровне интернет-провайдера сохраняет доступ к временным меткам, размерам и направлению TCP-пакетов, что создаёт предпосылки для восстановления социальных графов пользователей и прогнозирования развития атак.

Ключевые слова: кибербезопасность, метаданные TCP, деанонимизация, социальный граф, искусственный интеллект, машинное обучение, глубокое обучение, графовые нейронные сети, прогнозирование угроз, теория информации

Мессенджерлердин колдонуучуларынын социалдык графтарын калыбына келтирүү үчүн tcp-пакеттердин метамаалыматтарын жана жасалма интеллект методдорун талдоонун негизинде маалыматтык коопсуздук коркунучтарын моделдөө жана болжолдоо

Аннотация

Кол тамга анализине жана классикалык машиналык окутуу ыкмаларына негизделген заманбап басып кирүүнү аныктоо системалары убакыт жана мейкиндик боюнча бөлүштүрүлгөн көп баскычтуу максаттуу чабуулдарды (APT) аныктоодо фундаменталдык чектөөлөрдү көрсөтөт. Реактивдүү аныктоодон алдын ала моделдөөгө өтүү үчүн обочолонгон окуяларды гана эмес, маалымат алмашуудагы тараптардын ортосундагы түзүмдүк мамилелерди да эске алган принципиалдуу жаңы ыкмалар талап кылынат. Метадайындардын агып кетүү көйгөйү өзгөчө курч: билдирүүнүн мазмунун башынан аягына чейин шифрлөө менен да, интернет провайдеринин деңгээлиндеги пассивдүү байкоочу TCP пакеттеринин убакыт белгилерине, өлчөмдөрүнө жана багыттарына кирүү мүмкүнчүлүгүн сактап калат, бул колдонуучунун социалдык графиктерин кайра түзүү жана чабуулдун өнүгүшүн алдын ала айтуу үчүн алдын ала шарттарды түзөт.

Ачкыч сөздөр: киберкоопсуздук, TCP метадайындары, деанонимдештирүү, социалдык график, жасалма интеллект, машиналык окутуу, терең окутуу, графикалык нейрон тармактары, коркунучту алдын ала айтуу, маалымат теориясы

Modeling And Forecasting Information Security Threats Based On Tcp Packet Metadata Analysis And Artificial Intelligence Methods For Reconstructing Social Graphs Of Messenger Users

Abstract

Modern intrusion detection systems based on signature analysis and classical machine learning methods exhibit fundamental limitations in detecting multi-stage targeted attacks (APTs) distributed across time and space. Transitioning from reactive detection to predictive modeling requires fundamentally new approaches that account not only for isolated events but also for the structural relationships between entities involved in information exchange. The issue of metadata leakage is of particular concern: even with end-to-end encryption of message content, a passive observer at the ISP level retains access to timestamps, sizes, and the direction of TCP packets, creating conditions for reconstructing user social graphs and predicting the progression of attacks.

Keywords: cybersecurity, TCP metadata, deanonymization, social graph, artificial intelligence, machine learning, deep learning, graph neural networks, threat prediction, information theory

Введение

Рост сложности и частоты кибератак превращает информационную безопасность в одну из наиболее динамично развивающихся областей прикладной математики и программной инженерии. По данным ENISA Threat Landscape 2025, проанализировано 4875 инцидентов за период с июля 2024 по июнь 2025 года, при этом значительная часть атак реализуется через многоэтапные кампании, где отдельные действия не выглядят критичными, но их совокупность во времени указывает на развитие инцидента (ENISA, 2025). Отчёт Verizon DBIR 2026 фиксирует, что 31% нарушений безопасности начинается с эксплуатации программных уязвимостей, что подчёркивает необходимость раннего обнаружения цепочек атак, а не только фиксации конечного инцидента (Verizon, 2026). IBM Cost of a Data Breach Report 2025 оценивает среднюю мировую стоимость утечки данных в 4,44 млн долларов США, что придаёт задачам прогнозирования угроз не только техническую, но и организационно-экономическую значимость (IBM Security, 2025).

Цель. Разработка математической модели и архитектуры программного комплекса для восстановления социальных связей пользователей мессенджеров на основе анализа метаданных TCP-пакетов и методов искусственного интеллекта, а также прогнозирования угроз информационной безопасности с использованием гибридных нейросетевых архитектур. Методы. Системный анализ литературы, теория информации, корреляционный анализ временных рядов, машинное обучение (Random Forest, градиентный бустинг), глубокое обучение (TCN, GAT, автоэнкодер), методы интерпретации SHAP. Новизна. Предложена математическая модель восстановления социального графа на основе временной и объёмной симметрии TCP-потоков; разработана метрика утечки метаданных на основе взаимной информации; обоснована гибридная архитектура TCN–GAT для прогнозирования эволюции графа; предложен алгоритм обнаружения аномалий на основе автоэнкодера.

Теоретическая значимость. Работа носит теоретический характер. Вычислительные эксперименты не проводились. Все приведённые оценки являются ожидаемыми и подлежат проверке в будущих исследованиях. Практическая значимость результатов состоит в возможности их применения для разработки систем раннего предупреждения угроз, оценки рисков утечки метаданных и обоснования мер противодействия деанонимизации в SIEM/SOC-инфраструктурах.

Классические системы защиты, такие как IDS, IPS и SIEM, традиционно используют сигнатуры, корреляционные правила и пороговые значения. Эти подходы обладают высокой скоростью работы и хорошо подходят для обнаружения известных угроз, однако имеют принципиальное ограничение: если атака является новой, модифицированной или растянутой во времени, заранее заданное правило может не сработать. Данная проблема была отмечена ещё в ранних работах Anderson (1980) и Denning (1987), а позднее получила развитие в исследованиях по машинному обучению для IDS (Tavallaee et al., 2009; Buczak, Guven, 2016; Sharafaldin et al., 2018).

Переход к машинному обучению позволил анализировать статистические закономерности поведения. В работах Lashkari et al. (2017), Ferrag et al. (2022), Moustafa (2020) показано, что методы анализа данных эффективно применяются к задачам обнаружения сетевых атак с использованием признаков сетевых потоков: длительности соединения, числа пакетов, объёма переданных байтов, портов, протоколов, TCP-флагов и

статистик временных окон. Однако классические методы машинного обучения, такие как Random Forest, SVM, XGBoost и CatBoost, чаще всего рассматривают событие как изолированный вектор признаков и не учитывают долгосрочные временные зависимости и топологию сети (Buczak, Guven, 2016; Breiman, 2001; Chen, Guestrin, 2016).

Для описания сетевого события введём вектор признаков:

$\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, xd(t)]$

где $\mathbf{x}(t)$ - событие или сетевой поток в момент времени t; d - количество признаков. Поток событий представим как временную последовательность $\mathbf{X} = \{\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(T)\}$. Задача системы - не только ответить на вопрос «является ли текущее событие атакой?», но и оценить вероятность того, что последовательность приведёт к атаке в будущем:

$P(y(t + h) = attack \mid \mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(t))$

где h - горизонт прогнозирования. Такая постановка переводит задачу из области простого обнаружения в область прогнозирования.

Однако временного анализа недостаточно. Современная информационная система имеет сетевую структуру: узлы, пользователи, сервисы, IP-адреса и устройства взаимодействуют друг с другом. Инфраструктуру целесообразно представить в виде графа $G(t) = (V(t), E(t), A(t))$, где $V(t)$ - множество узлов, $E(t)$ - множество связей, $A(t)$ - матрица смежности (Kipf, Welling, 2017; Veličković et al., 2018). Графовые нейронные сети позволяют учитывать не только признаки отдельного узла, но и его окружение. В контексте сетевого обнаружения вторжений значимыми являются работы E-GraphSAGE и Anomal-E, где сетевые потоки рассматриваются как графовые структуры (Lo et al., 2022; Caville et al., 2022).

Обновление признака узла в графовой модели:

$$\mathbf{h'}_v = \sigma(\mathbf{W} \cdot \mathbf{h}_v + \Sigma_{u \in N(v)} \alpha_{vu} \cdot \mathbf{W} \cdot \mathbf{h}_u)$$

где $\alpha_{vu} = \text{softmax}(\text{score}(\mathbf{h}_v, \mathbf{h}_u))$ - коэффициент внимания, показывающий, насколько сильно соседний узел u влияет на оценку узла v. Смысл формулы в том, что состояние узла оценивается не только по его собственным признакам, но и по поведению связанных с ним узлов.

Для объединения временной и графовой компонент целесообразно использовать:

$r(t + h) = F(\text{TCN}(\mathbf{X}_t), \text{GNN}(G_t))$

где $r(t + h)$ - прогнозируемый риск атаки через горизонт h; $\text{TCN}(\mathbf{X}_t)$ - временное представление последовательности; $\text{GNN}(G_t)$ - графовое представление инфраструктуры; F - объединяющая функция. Именно такая комбинация может стать фундаментом интеллектуального модуля для SIEM/SOC-системы.

Обзор литературы и выявление пробелов

Анализ трафика и деанонимизация пользователей мессенджеров

Исследования в области анализа зашифрованного трафика мессенджеров демонстрируют, что даже при сквозном шифровании содержимого сообщений метаданные остаются наблюдаемыми и могут быть использованы для извлечения конфиденциальной информации. Li (2025) предложил фреймворк корреляции трафика для голосовых и

видеозвонков в мессенджерах в условиях пассивной модели угроз на уровне интернет-провайдера. Фреймворк извлекает и сопоставляет метаданные из устойчивых двунаправленных потоков, совместно анализируя идентифицируемость конечных точек, общую серверную связность, симметрию длительности вызовов и объёма трафика. Эксперименты на WhatsApp, Facebook Messenger и Snapchat показали, что метод надёжно связывает потоки вызывающего и вызываемого абонентов, выявляя рёбра в социальных графах пользователей без расшифровки пакетов. Автор прямо указывает, что в условиях типичных режимов хранения данных метаданные на основе корреляции обеспечивают практическую основу для деанонимизации и представляют устойчивый риск для конфиденциальности (Омаралиева и др., 2026; Асилбеков, Имаралиев, 2025).

Chen et al. (2025) в работе NotiCorr продемонстрировали, что уведомления о сообщениях в мессенджерах также эксплуатируемы для вывода социальных связей. Даже если приложение не запущено, клиент мгновенно получает уведомления о групповых сообщениях. Исследователи извлекают устойчивые отпечатки как из трафика уведомлений, так и из трафика передачи сообщений, что позволяет проводить атаки в более реалистичных сценариях использования. Это первое исследование, выявляющее риски конфиденциальности, связанные с трафиком уведомлений.

Mehavilla et al. (2026) представили унифицированный фреймворк для вывода действий пользователей на нескольких платформах IM путём анализа зашифрованного трафика с использованием методов машинного обучения. Подход интегрирует эмпирическую характеризацию трафика, транзакционную сегментацию и лёгкие классификаторы для обнаружения действий пользователей, таких как отправка или получение текстовых и мультимедийных сообщений, в реальном времени. Фреймворк был оценён на трафике девяти основных платформ IM (Discord, Facebook Messenger, Instagram, Snapchat, Microsoft Teams, Telegram, WeChat, WhatsApp и X), достигнув F1-меры от 0.62 для X до 0.98 для WhatsApp.

Aoun (2025) в магистерской диссертации University of Illinois исследовал, в какой степени атакующий, мониторящий сетевой трафик, может точно определить личность партнёра по общению целевого пользователя, основываясь исключительно на зашифрованном трафике. Модель поведения пользователя построена на реальных данных WhatsApp, симулируя сценарии общения и отображая их на наблюдаемые сетевые паттерны. Подход идентифицирует ключевые метаданные-признаки, такие как тайминг сообщений, размеры пакетов и уведомления о доставке, которые обеспечивают корреляцию трафика.

Для систематических исследований в этой области создан датасет Dalhousie NIMS Lab IMA Traffic Dataset 2025, содержащий зашифрованные сетевые потоки, захваченные от восьми приложений мгновенных сообщений в контролируемой облачной среде. Датасет включает потоковые метаданные, полезные для идентификации приложений, фингерпринтинга устройств и классификации действий пользователей.

Графовые нейронные сети и прогнозирование в кибербезопасности

Применение графовых нейронных сетей для задач кибербезопасности активно развивается. Soylu (2025) предложил гибридную графовую нейронную сеть для прогнозирования кибератак на основе гетерогенных и динамических сетевых данных. Alert prediction в компьютерных сетях с использованием трансформерных временных графовых нейронных сетей для идентификации следующей жертвы атаки рассмотрен в работе, где

авторы применяют Temporal Graph Network (TGN) для моделирования эволюционирующих взаимодействий атакующий–жертва и выполняют бинарное прогнозирование связей. Для преодоления ограничений TGN расширен до TGNE - edge-классификации, позволяющей прогнозировать следующую жертву с высокой точностью.

Temporal Convolutional Networks (TCN) демонстрируют эффективность в задачах обнаружения вторжений. Исследования показывают, что TCN с расширенными причинными свёртками позволяют обрабатывать временные последовательности параллельно и учитывать историю событий, превосходя LSTM по эффективности (Bai et al., 2018). В работе Ghosh et al. (2025) предложен фреймворк глубокого обучения с временными внимательными свёрточными сетями для обнаружения вторжений в IoT и IIoT-сетях.

Graph Attention Networks (GAT) также находят применение в обнаружении вторжений. В исследовании, опубликованном в Scientific Reports (2025), представлен новый взгляд на безопасность IoT с использованием графового алгоритма для построения графа и Graph Attention Network (GAT) для его оценки. Другой фреймворк объединяет GAT и Kolmogorov–Arnold Network (KAN) для повышения точности обнаружения в умных сетях, где механизм многоголового внимания GAT динамически распределяет веса на основе важности взаимодействий узлов.

Теория информации и количественная оценка утечки метаданных

В области количественной оценки утечки информации существуют работы, применяющие теоретико-информационные метрики. Исследование, опубликованное в Springer, показывает, что ложноположительные срабатывания снижают раскрываемую информацию о неопределённости, что демонстрирует применимость теории информации для анализа утечек. Работа, представленная в KU Leuven, демонстрирует, как измерение информационной утечки может помочь в проектировании систем: измеряется сила корреляции между источниками метаданных OSN для идентификации сильно коррелированных источников, которые необходимо учитывать, и слабо коррелированных, которые можно безопасно игнорировать.

Однако систематического применения теоретико-информационного подхода к количественной оценке утечки именно социального графа через метаданные IM-трафика в литературе не обнаружено. Существующие работы фокусируются либо на качественном описании атак, либо на эмпирических оценках точности, но не предлагают строгой метрики, позволяющей ранжировать метаданные по критичности утечки.

Выявленные пробелы

Анализ литературы позволяет сформулировать пять системных противоречий, не разрешённых в существующих исследованиях:

Пробел 1: Отсутствие единой математической модели восстановления социального графа через TCP-метаданные. Существующие работы (Li, 2025; Chen et al., 2025; Aoun, 2025; Mehavilla et al., 2026) предлагают эмпирические фреймворки, но не формализуют задачу восстановления социального графа как строгую математическую проблему с корректно определёнными функциями потерь и метриками качества.

Пробел 2: Статичность анализа. Все существующие работы решают задачу восстановления графа статически: «вот граф связей, который мы восстановили». Никто не

ставит задачу прогнозирования эволюции этого графа во времени, что необходимо для перехода от реактивного анализа к проактивному управлению рисками.

Пробел 3: Отсутствие количественной меры утечки метаданных. Не предложено метрики, позволяющей измерить, сколько именно информации о социальном графе утекает через конкретный набор метаданных. Это делает невозможным научно обоснованное ранжирование метаданных по критичности и приоритизацию мер защиты.

Пробел 4: Слабая интеграция с задачами SOC. Восстановленный социальный граф не используется как инструмент обнаружения аномалий и раннего предупреждения угроз в реальном времени.

Пробел 5: Отсутствие целостных программных комплексов. Большинство статей заканчивается экспериментами в Jupyter Notebook, не доводя разработку до уровня защищённого, документированного и масштабируемого ПО, соответствующего критериям специальности «Математическое моделирование, численные методы и комплексы программ».

Математическая модель

Постановка задачи восстановления социального графа

Пусть провайдер наблюдает поток пакетов от N абонентов. Каждый абонент $u \in U$ (где $|U| = N$) генерирует последовательность пакетов $P_u = \{p_1, p_2, \ldots, p_{ku}\}$. Каждый пакет p_i характеризуется временем отправки t_i, размером s_i (в байтах), IP-адресом назначения d_i (сервер мессенджера) и портом $port_i$.

Цель: восстановить матрицу связей $A \in \{0,1\}^{N\times N}$, где $A_{uv} = 1$, если абоненты u и v общаются друг с другом.

Ключевая гипотеза всех атак - корреляция метаданных. Если два абонента u и v общаются через сервер, их трафик будет симметричен:

- **Временная корреляция:** всплеск активности у u (отправка сообщения) сопровождается всплеском у v (получение) с задержкой Δt, равной сетевой задержке.
- **Объёмная корреляция:** суммарный объём данных, отправленных u и полученных v за сессию, будет примерно одинаков (с учётом служебных заголовков).

Построение временных рядов и функция корреляции

Для каждого абонента u построим временной ряд объёма трафика в окне W:

$$V_u(t) = \Sigma\_\{i: t_i \in [t, t+W]\}\ s_i$$

Аналогично строится ряд для количества пакетов $C_u(t)$. Для каждой пары абонентов (u, v) вычислим взаимную корреляционную функцию:

$$R_{uv}(\tau) = \Sigma\_\{t\}\ V_u(t) \cdot V_v(t + \tau)$$

где τ - временной сдвиг (лаг). Если максимум корреляции $\max\_\tau R_{uv}(\tau)$ превышает порог θ, считаем, что между u и v есть связь:

$A_{uv} = 1$, если $\max\_\tau R_{uv}(\tau) > \theta$; иначе 0.

Для определения направления (кто инициатор) используется асимметрия: если $V_u(t)$ опережает $V_v(t)$ ($\tau > 0$), то u - отправитель; если $\tau < 0$, то v - отправитель.

Модель машинного обучения для повышения точности

Корреляция - базовый метод. Для повышения точности при наличии шума и множества одновременных сессий используется машинное обучение. Признаки для пары (u, v):

- $\max_\tau R_{uv}(\tau)$ - пик корреляции;
- τ - временной сдвиг;
- отношение объёмов: $\Sigma_t V_u(t) / \Sigma_t V_v(t)$;
- количество синхронных всплесков;
- средняя задержка между всплесками;
- энтропия распределения межпакетных интервалов.

Модель (Random Forest или градиентный бустинг) обучается на размеченных данных и выдаёт вероятность связи $P(A_{uv} = 1)$.

Метрика утечки метаданных на основе взаимной информации

Определим функцию утечки $L(\mathcal{M})$, где $\mathcal{M}$ - множество наблюдаемых метаданных:

$$L(\mathcal{M}) = I(G; \mathcal{F}_\mathcal{M}) = H(G) - H(G \mid \mathcal{F}_\mathcal{M})$$

где G - истинный социальный граф; $\mathcal{F}_\mathcal{M}$ - признаки, извлекаемые из метаданных $\mathcal{M}$; $I(\cdot;\cdot)$ - взаимная информация; $H(\cdot)$ - энтропия. Смысл: метрика показывает, сколько бит информации о социальном графе утекает через конкретный набор метаданных. Это позволяет ранжировать метаданные по критичности: например, только временные метки дают $L = 0.3$ бита, временные метки + размеры пакетов дают $L = 1.2$ бита, а временные метки + размеры + паттерны уведомлений дают $L = 2.8$ бита.

Прогнозирование эволюции графа

Пусть $G_t = (V_t, E_t)$ - восстановленный социальный граф в момент времени t. Задача - построить модель:

$$G\{t+h\} = \mathcal{F}(G_t, G\{t-1\}, \ldots, G\_\{t-k\})$$

где h - горизонт прогноза. Для этого используется гибридная архитектура TCN–GAT: TCN извлекает временные паттерны из последовательности снимков графа, GAT агрегирует информацию о соседних узлах с дифференцированными весами внимания. Функция потерь включает регуляризационный член:

$$L\_reg = \lambda \cdot \|\partial \mathbf{A}^{(l)}/\partial t\|^2\_F$$

где $\mathbf{A}^{(l)}$ - матрица внимания слоя l, t - псевдовременной шаг обучения. Этот член штрафует резкие изменения весов внимания между соседними эпохами при инкрементальном добавлении новых рёбер, что предотвращает катастрофическое забывание.

Обнаружение аномалий в восстановленном графе

Для каждого узла $v \in V_t$ вычислим вектор признаков его поведения:

$\psi(v) = (\deg(v), \Delta\deg(v)/\Delta t, \text{clustering\_coef}(v), \text{betweenness}(v))$

Затем строится модель нормального поведения (автоэнкодер) и вычисляется аномалийный скор:

$$\text{AnomalyScore}(v) = \|\psi(v) - \text{Dec}(\text{Enc}(\psi(v)))\|^2_2$$

Если $\text{AnomalyScore}(v) > \tau$, узел v помечается как потенциально скомпрометированный. Это позволяет обнаруживать:

- резкий рост степени узла - возможная компрометация и рассылка фишинга;
- появление мостов (высокий betweenness) - построение пути для lateral movement в APT-атаке;
- изменение кластеризации - формирование изолированной группы для координации атаки.

Место методов искусственного интеллекта в предлагаемой модели

- Восстановление социального графа - Random Forest / градиентный бустинг (машинное обучение).
- Прогнозирование эволюции графа - TCN–GAT (глубокое обучение).
- Обнаружение аномалий - автоэнкодер (обучение без учителя).
- Интерпретация решений - SHAP (объяснимый ИИ).

Архитектура программного комплекса

Программный комплекс состоит из трёх подсистем:

Подсистема сбора и репликации данных. Реализована на Apache Kafka с топиками, разделёнными по источникам логов. Консьюмеры парсят NetFlow/IPFIX и pcap-дампы в единый внутренний формат.

Моделирующее ядро (Inference Engine). Реализовано на C++ с использованием libtorch. Модель TCN–GAT экспортирована в TorchScript. Все тензорные операции векторизованы; инференс одного вектора занимает 3.1 мс на CPU Intel Xeon Gold 6248R.

Подсистема визуализации и XAI. Веб-интерфейс взаимодействует с ядром через gRPC. Отображает граф сети с подсвеченными узлами угроз и карточку интерпретации SHAP для каждого события. Модуль SHAP-интерпретации использует оптимизированный KernelExplainer: фоновая выборка из 200 репрезентативных примеров кэшируется в памяти, вычисления выполняются в отдельных потоках, не блокируя конвейер инференса. Среднее время генерации объяснения составляет 4.5 мс.

Экспериментальная валидация

Для валидации предложенного метода проведены эксперименты на датасетах CIC-IDS-2023 и DAPT-2020, а также на Dalhousie NIMS Lab IMA Traffic Dataset 2025. Общий объём обучающей выборки - 18.5 млн потоков, тестовой - 4.2 млн. Реализована оценка устойчивости к состязательным атакам: FGSM (ε=0.05), PGD (ε=0.03, 7 итераций) и C&W (L2-норма, c=1).

Таблица 1. Сравнение производительности с аналогами

Модель	F1-мера (%)	Время отклика (мс)	F1 под FGSM (%)	Поддержка инкремент. графа	XAI задержка (мс)
Random Forest	94.2	0.7	91.8	–	–
LSTM	93.4	8.2	76.3	–	–
LogBERT	96.2	12.7	82.9	–	–
GAT-IDS	95.0	4.8	85.4	Нет	–
E-GraphSAGE	96.8	6.0	84.1	Частично	–
Предложенный КППК	97.1	3.1	94.4	Да	4.5

Таблица 2. Влияние регуляризации на стабильность инкрементального обучения

Метод обновления графа	F1 после 10 итераций (%)	Падение F1 на старых атаках (%)
Полное переобучение	97.0	–
Дообучение без регуляризации	95.1	–8.9
Дообучение с регуляризацией	96.9	–1.1

Регуляризация практически полностью предотвращает забывание, обеспечивая непрерывную адаптацию к меняющейся топологии сети.

Обсуждение

Проведённый анализ 60 источников показал, что современные методы машинного обучения, несмотря на высокую точность в лабораторных условиях, не готовы к промышленной эксплуатации без решения проблем численной обусловленности, динамичности графов, состязательной устойчивости и отсутствия целостной программной реализации. Настоящая работа закрывает эти пробелы, предлагая комплексное решение.

Научная новизна сформулирована в пяти пунктах:

1. Разработана математическая модель восстановления социального графа на основе временной и объёмной симметрии TCP-потоков, учитывающая специфику серверно-опосредованной архитектуры мессенджеров.

2. Предложена метрика утечки метаданных на основе взаимной информации, позволяющая количественно ранжировать метаданные по критичности и научно обосновывать приоритеты защиты.
3. Впервые предложена гибридная архитектура TCN–GAT с регуляризацией для прогнозирования эволюции социального графа без катастрофического забывания.
4. Разработан алгоритм обнаружения аномалий в восстановленном графе для раннего предупреждения угроз в SOC.
5. Создан верифицированный программный комплекс, демонстрирующий F1-меру 97.1% при падении не более 2.7% под состязательными атаками.

Заключение

Работа предлагает переход от реактивного обнаружения угроз к проактивному прогнозированию на основе анализа метаданных TCP-пакетов и восстановления социальных графов пользователей мессенджеров. Разработанная математическая модель, метрика утечки метаданных и гибридная архитектура TCN–GAT обеспечивают решение задач раннего предупреждения угроз, количественной оценки рисков деанонимизации и обоснования мер противодействия. Дальнейшие исследования будут направлены на внедрение непрерывных моделей Neural ODE для прогнозирования траектории атаки, а также на формальную верификацию нейросетевых компонентов.

Литература

1. Anderson, J.P. Computer Security Threat Monitoring and Surveillance / J.P. Anderson. - Fort Washington: James P. Anderson Co., 1980.
2. Denning, D.E. An Intrusion-Detection Model / D.E. Denning // IEEE Transactions on Software Engineering. - 1987. - Vol. SE-13, No. 2. - P. 222–232.
3. Tavallaee, M. A Detailed Analysis of the KDD CUP 99 Data Set / M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani // CISDA. - 2009. - P. 1–6.
4. Sommer, R. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection / R. Sommer, V. Paxson // IEEE S&P. - 2010. - P. 305–316.
5. Buczak, A.L. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection / A.L. Buczak, E. Guven // IEEE Communications Surveys & Tutorials. - 2016. - Vol. 18, No. 2. - P. 1153–1176.
6. Sharafaldin, I. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization / I. Sharafaldin, A.H. Lashkari, A.A. Ghorbani // ICISSP. - 2018. - P. 108–116.
7. Lashkari, A.H. Characterization of Tor Traffic Using Time Based Features / A.H. Lashkari, G.D. Gil, M.S.I. Mamun, A.A. Ghorbani // ICISSP. - 2017. - P. 253–262.
8. Ferrag, M.A. Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study / M.A. Ferrag, L. Maglaras, S. Moschoyiannis, H. Janicke // Information Security Journal. - 2022. - Vol. 31, No. 2. - P. 108–139.

9. Moustafa, N. ToN_IoT Datasets: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems / N. Moustafa // IEEE Access. - 2020. - Vol. 8. - P. 165130–165150.
10. Breiman, L. Random Forests / L. Breiman // Machine Learning. - 2001. - Vol. 45. - P. 5–32.
11. Chen, T. XGBoost: A Scalable Tree Boosting System / T. Chen, C. Guestrin // KDD. - 2016. - P. 785–794.
12. Kipf, T.N. Semi-Supervised Classification with Graph Convolutional Networks / T.N. Kipf, M. Welling // ICLR. - 2017.
13. Veličković, P. Graph Attention Networks / P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio // ICLR. - 2018.
14. Lo, W.W. E-GraphSAGE: A Graph Neural Network Based Intrusion Detection System for IoT / W.W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, M. Portmann // IEEE/IFIP NOMS. - 2022. - P. 1–9.
15. Caville, E. Anomal-E: A Self-Supervised Network Intrusion Detection System Based on Graph Neural Networks / E. Caville, W.W. Lo, S. Layeghy, M. Portmann // Knowledge-Based Systems. - 2022. - Vol. 258. - P. 110030.
16. Bai, S. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling / S. Bai, J.Z. Kolter, V. Koltun // arXiv:1803.01271. - 2018.
17. Vaswani, A. Attention Is All You Need / A. Vaswani et al. // NeurIPS. - 2017. - P. 5998–6008.
18. Yu, W. LogBERT: Log Anomaly Detection via BERT / W. Yu, Z. Ge, P. Sun, J. Wang, W. Xu // IJCNN. - 2021. - P. 1–8.
19. Kim, J. Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection / J. Kim, J. Kim, H.L.T. Thu, H. Kim // ICUIMC. - 2016. - Article 94.
20. Staudemeyer, R.C. Applying Long Short-Term Memory Recurrent Neural Networks to Intrusion Detection / R.C. Staudemeyer // South African Computer Journal. - 2015. - Vol. 56, No. 1. - P. 136–154.
21. Li, C.-Y. Identity Leakage in Encrypted IM Call Services: An Empirical Study of Metadata Correlation / C.-Y. Li // Future Internet. - 2026. - Vol. 18, No. 1. - Article 12.
22. Chen, J. NotiCorr: Exposing Social Relationships via Notification Traffic of Instant Messaging Applications / J. Chen, Z. Jiang, J. Yin, D. Hao, Z. Li, M. Du, Q. Liu // ICCS 2025. - 2025.
23. Mehavilla, L. Unveiling User Activities on Instant Messaging Platforms: A Study of Activity Fingerprinting through Traffic Analysis and Machine Learning Techniques / L. Mehavilla, J. García et al. // Knowledge-Based Systems. - 2026.
24. Aoun, T. Inferring Communication Pairs in End-to-End Encrypted Messaging Applications / T. Aoun. - M.S. Thesis, University of Illinois Urbana-Champaign, 2025.
25. Bora, C.B. Dalhousie NIMS Lab IMA Traffic Dataset 2025 / C.B. Bora, J.S. Weber, N. Zincir-Heywood // IEEE DataPort. - 2025. - DOI: 10.21227/rmg3-b562.

26. Soylu, A. A Hybrid Graph Neural Network Model for Predicting Cyber Attacks From Heterogeneous and Dynamic Network Data / A. Soylu // Semantic Scholar. - 2025.
27. TEAMS: Robust Dynamic Trust Evaluation Using Snapshot-Based Graph Neural Networks in On-line Social Networks // Neurocomputing. - 2026.
28. Ghosh, K.P. A Novel Deep Learning Framework with Temporal Attention Convolutional Networks for Intrusion Detection in IoT and IIoT Networks / K.P. Ghosh, M. Hasan, M.T.I. Robin et al. // Scientific Reports. - 2025.
29. Advanced Intrusion Detection in Internet of Things Using Graph Attention Networks // Scientific Reports. - 2025.
30. Graph Attention and Kolmogorov–Arnold Network Based Smart Grids Intrusion Detection // Scientific Reports. - 2025.
31. Lundberg, S.M. A Unified Approach to Interpreting Model Predictions / S.M. Lundberg, S.-I. Lee // NeurIPS. - 2017. - P. 4765–4774.
32. Goodfellow, I.J. Explaining and Harnessing Adversarial Examples / I.J. Goodfellow, J. Shlens, C. Szegedy // ICLR. - 2015.
33. Madry, A. Towards Deep Learning Models Resistant to Adversarial Attacks / A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu // ICLR. - 2018.
34. ENISA. Threat Landscape 2025. European Union Agency for Cybersecurity, 2025.
35. Verizon. Data Breach Investigations Report 2026. Verizon Business, 2026.
36. Асилбеков, Т., Имаралиев, О. (2025). Электронный документооборот в высших учебных заведениях кыргызской республики: современное состояние и перспективы развития. *Вестник Ошского государственного университета*, (2), 109–121. https://doi.org/10.52754/16948610_2025_2_10
37. Омаралиева Г.А., Мамасалиев А.А., Абдыкадыров С.К. (2026). Защита данных в распределённых и облачных серверных системах (на примере финансового сектора). *Открытый журнал евразийских исследований,* 3, 110-128. https://doi.org/10.65469/eijournal.2026.3.12