Применение методов искусственного интеллекта для оценки и минимизации рисков утечки персональных данных в VPN-сервисах

Смирнов Виталий Михайлович
к.т.н., доцент кафедры информатики, Института кибербезопасности и цифровых технологий, РТУ МИРЭА

В статье рассматривается проблема защиты персональных данных пользователей VPN-сервисов в условиях роста киберугроз и распространения «теневого» использования публичных ИИ-сервисов. Анализируются современные подходы к обнаружению аномалий в зашифрованном VPN-трафике с применением методов машинного обучения. Предложена многоуровневая архитектура системы оценки рисков утечки персональных данных, включающая обнаружение вредоносных расширений, поведенческий анализ пользователей и интеллектуальный мониторинг исходящего трафика. Обоснована эффективность гибридного подхода, объединяющего алгоритмы Isolation Forest, LSTM-автокодер и XGBoost для комплексной защиты.
Ключевые слова: искусственный интеллект, машинное обучение, VPN, утечка персональных данных, обнаружение аномалий, информационная безопасность, анализ сетевого трафика.

Современный цифровой мир сталкивается сегодня с большим ростом передаваемой информации и усложнением угроз информационной безопасности. В этих условиях VPN-сервисы стали важным инструментом защиты приватности в открытых сетях, однако парадокс заключается в том, что сами VPN-сервисы, особенно бесплатные, сами могут выступать каналом утечки чувствительной информации.

31 июля 2025 года президент России Владимир Владимирович Путин подписал Федеральный закон № 281-ФЗ. Документ внёс изменения в Кодекс Российской Федерации об административных правонарушениях (КоАП) и вступил в силу с 1 сентября 2025 года, за исключением отдельных положений, для которых установлены иные сроки.

Закон установил административную ответственность за два вида правонарушений. Поиск экстремистских материалов с использованием VPN-сервисов, которое влечёт наложение штрафа на граждан. Кроме того, вводится ответственность за неисполнение владельцем VPN-сервиса обязанностей по взаимодействию с Роскомнадзором . При этом важно подчеркнуть, что закон не устанавливает наказание за сам факт использования VPN-сервисов для личных целей.

Федеральный закон № 332-ФЗ был подписан В.В. Путиным в тот же день, 31 июля 2025 года, и вносит изменения в Федеральный закон «О рекламе». Его ключевое положение — полный запрет на рекламу VPN-сервисов и других программно-аппаратных средств доступа к информационным ресурсам, доступ к которым ограничен в соответствии с законодательством РФ. Данный запрет вступил в силу с 1 сентября 2025 года.

Кроме того, закон вводит требования к рекламе услуг по банкротству граждан: с 1 января 2026 года в такой рекламе запрещается гарантировать освобождение от долгов и призывать к неисполнению обязательств по ним [1].

В 2025 году специалистами компании Koi Security задокументирована продолжительная вредоносная кампания ShadyPanda, в рамках которой с 2018 года распространялось не менее 145 браузерных расширений для Chrome и Edge. Суммарное количество установок вредоносного ПО превысило 4,3 млн. Кампания характеризовалась поэтапной эскалацией угроз: первоначально безвредные утилиты в ходе обновлений приобретали функции шпионского программного обеспечения.

В начальном этапе (2018–2023 гг.) расширения маскировались под инструменты повышения продуктивности и смены обоев. Их активность ограничивалась партнёрским мошенничеством — внедрением реферальных кодов в ссылки пользователей. С 2024 года функционал расширений увеличился, когда в некоторых случаях отдельные образцы, например Infinity V+, начали перехватывать поисковые запросы, перенаправлять трафик на подконтрольные злоумышленникам ресурсы и похищать файлы cookie.

Наиболее опасной фазой кампании стало внедрение бэкдора удалённого выполнения кода в пять расширений с многолетней положительной репутацией. Бэкдор - это скрытый механизм, который позволяет получить несанкционированный доступ к компьютерной системе, программному обеспечению, устройству или сети, минуя стандартные процедуры аутентификации и контроля доступа.

Расширение Clean Master с 200 тыс. установок получило возможность каждые 60 минут опрашивать управляющий сервер и исполнять произвольный JavaScript-код с полным доступом к функциям браузера. Суммарная аудитория расширений, затронутых данной полезной нагрузкой, составила 300 тыс. пользователей.

В настоящее время активна четвёртая фаза кампании, связанная с пятью расширениями для Microsoft Edge, опубликованными компанией Starlab Technology в 2023 году. Количество установок превышает 4 млн. Данные расширения осуществляют сбор и передачу на 17 доменов, расположенных в Китае, истории посещений, поисковых запросов, кликов, цифровых отпечатков устройств, а также содержимого локального хранилища, сессионных данных и cookie-файлов. Исследователи отмечают, что техническая архитектура указанных расширений допускает внедрение бэкдора через будущие обновления, однако на текущий момент признаков реализации данной угрозы не обнаружено [2].

Данный инцидент демонстрирует актуальность проблемы пользователей, которые доверяют защиту своих данных VPN-сервисам, тогда как некоторые поставщики используют это доверие во враждебных целях. В этих условиях возникает необходимость разработки систем, способных выявлять риски утечки персональных данных на основе интеллектуального анализа поведения как самого VPN-сервиса, так и пользователей, взаимодействующих с ним.

Многие поставщики онлайн-услуг, в особенности VPN-сервисы и мессенджеры, декларируют политику «нулевых логов» (No-Logs Policy) в качестве ключевого элемента своей безопасности. Однако остаётся открытым вопрос о соответствии данных заявлений реальной практике обработки информации. Проблема верификации политики конфиденциальности приобретает

особую актуальность в условиях отсутствия общедоступных механизмов контроля.

В типичном случае VPN-сервис имеет доступ к IP-адресу пользователя, времени подключения, выбранному серверу, объёму переданных данных и DNS-запросам. Содержимое HTTPS-трафика остаётся недоступным при условии, что пользователь не устанавливал корневой сертификат, предлагаемый клиентом, и не использует незащищённые протоколы. Однако даже при шифровании по косвенным признакам — DNS-запросам и IP-адресам серверов — возможно определение посещаемых ресурсов.

Различают три категории логов: журналы активности (посещённые сайты, поисковые запросы), журналы подключений (IP-адреса, временные метки, объём трафика) и диагностические логи (версия клиента, сбои, тип устройства). Угрозу конфиденциальности представляет не сам факт ведения логов, а их состав, длительность хранения и возможность связывания с учётной записью или платежными данными.

Исключительные сценарии, расширяющие объём перехватываемой информации, возникают при установке корневого сертификата, использовании незашифрованных протоколов или устаревших версий VPN-соединений.

Показательным примером расхождения маркетинговых заявлений и реальной практики является инцидент с UFO VPN в 2020 году: при декларируемой политике «no-logs» была обнаружена открытая база данных, содержащая IP-адреса, временные метки, сессионные токены и, по сообщениям исследователей, пароли в незашифрованном виде. Данный случай не свидетельствует о системном характере подобных практик, однако демонстрирует риск некритического доверия к заявлениям провайдеров [3].

Классификация угроз утечки персональных данных в VPN-сервисах

Современные угрозы утечки персональных данных при использовании VPN-сервисов могут быть классифицированы по трём основным категориям, каждая из которых отражает специфический вектор атаки и требует соответствующих методов противодействия.

Первая категория угроз связана с вредоносными расширениями и клиентским программным обеспечением. Как показывают результаты исследований, браузерные расширения, предоставляющие бесплатный VPN-доступ, могут содержать исполняемые скрипты, осуществляющие перехват пользовательских данных. Наиболее показательным примером является технический анализ расширения Urban VPN Proxy, проведённый специалистами Koi Security в 2025 году. В ходе исследования было выявлено наличие восьми специализированных скриптов, ориентированных на различные AI-платформы, которые переопределяли стандартные сетевые методы fetch и XMLHttpRequest с целью нелегитимного сбора информации. Данный инцидент демонстрирует, что вредоносный функционал может быть внедрён в расширения, имеющие многомиллионную аудиторию и положительную репутацию, накопленную за годы работы.

Вторая категория угроз получила в литературе наименование «теневой» ИИ-трафик (Shadow AI) и связана с использованием публичных нейросетевых сервисов в обход установленных корпоративных политик безопасности. Согласно данным компании MoranaLabs, сотрудники организаций, даже при формальном запрете использования публичных больших языковых моделей (LLM), нередко находят способы обхода ограничений — через личные устройства, неконтролируемые прокси-сервисы или альтернативные каналы доступа. Результатом становится попадание исходного кода, юридических документов и персональных данных в обучающие выборки моделей, что создаёт долгосрочные риски компрометации конфиденциальной информации. Примечательно, что значительная часть таких утечек происходит не через злонамеренные расширения, а вследствие сознательных или неосторожных действий самих пользователей, что делает данный класс угроз особенно сложным для предотвращения средствами технического контроля.

Третья категория охватывает уязвимости самой VPN-инфраструктуры. Атаки на VPN-серверы и клиентское оборудование могут использовать недостатки реализации протоколов шифрования и аутентификации, присутствующие как на этапе установления соединения, так и в процессе непосредственной передачи данных. Злоумышленники способны эксплуатировать такие уязвимости для получения несанкционированного доступа к данным, передаваемым по защищённым каналам, включая перехват сессионных ключей, проведение атак типа «человек посередине» (Man-in-the-Middle) и внедрение вредоносного кода в клиентское программное обеспечение. Особую опасность представляют уязвимости, позволяющие осуществлять дешифрацию сохранённого трафика (например, при компрометации долговременных ключей), что делает скомпрометированные данные уязвимыми даже после завершения сеанса связи.

Таким образом, угрозы утечки персональных данных в VPN-экосистеме носят многоаспектный характер и не ограничиваются уязвимостями самого протокола.

Многомерный анализ зашифрованного трафика

Основная сложность защиты VPN-соединений заключается в невозможности инспектирования содержимого зашифрованных пакетов. Вследствие этого классические системы обнаружения вторжений (IDS), функционирующие на основе сигнатурного анализа, утрачивают свою эффективность при работе с VPN-трафиком, поскольку их функционирование предполагает доступ к полезной нагрузке пакетов. Как показывают исследования, применение методов обфускации, в частности маскировка трафика под HTTPS-соединения, делает традиционные подходы практически неработоспособными.

Широкое распространение виртуальных частных сетей (VPN), позволяющих скрывать передаваемые данные от пассивного наблюдения и обходить сетевые ограничения, актуализирует задачу идентификации протоколов, в зашифрованных каналах связи (ШКС). В научной литературе представлены подходы к решению данной задачи с применением нейронных сетей, анализирующих объёмные и временные характеристики передаваемых пакетов данных. Однако существующие методы ориентированы на классификацию одного вида трафика внутри таких протоколов защиты, как OpenVPN, IPsec и TLS, и не учитывают возможность одновременной передачи нескольких протоколов.

Входными признаками для нейросетевых моделей служат доступные параметры сетевого потока: временные метки, интервалы между пакетами, их длина и направление передачи. При этом используется корреляционный анализ между параметрами вложенного протокола и протокола обеспечения защиты данных.

В общем случае ШКС может содержать как однородный трафик, так и смешанные комбинации протоколов. В первом случае размеры и интервалы пакетов определяются исключительно свойствами передаваемого протокола; во втором — результирующий трафик несёт признаки каждого из них, что повышает вероятность ошибочной классификации. Вместе с тем, наблюдаются устойчивые закономерности между внешними характеристиками шифрованного трафика и содержимым передаваемых данных. Эти закономерности могут быть выявлены с использованием низкозатратных вычислительных процедур, что открывает перспективы для разработки методов идентификации протоколов внутри ШКС. Дополнительно необходимо учитывать параметры конфигурации канала, оказывающие системное влияние на все передаваемые в его пределах протоколы.

Типовая модель ШКС обеспечивает защиту данных от чтения и модификации промежуточными узлами сети и включает следующие этапы: установление соединения; согласование алгоритмов шифрования, целостности и аутентификации; упаковку передаваемых данных в пакеты, обеспечивающие шифрование и контроль целостности. Введение служебных заголовков уменьшает значение MTU (Maximum Transmission Unit), что требует фрагментации крупных пакетов и уменьшения максимального размера сегментов для сокращения числа фрагментов.

Протокол IPsec включает: ESP (Encrypted Secure Payload) — для шифрования данных; AH (Authentication Header) — для обеспечения целостности и аутентификации источника; ISAKMP — для первичной настройки соединения, аутентификации и обмена ключами. Предусмотрено два режима работы: транспортный (шифруется только полезная нагрузка IP-пакета, заголовок сохраняется) и туннельный (весь исходный IP-пакет помещается в шифрованную нагрузку нового пакета, часто используется совместно с L2TP) [4].

Методы глубокого обучения для обнаружения аномалий в зашифрованном трафике

В задачах обнаружения сетевых аномалий методы глубинного обучения (Deep Learning) занимают лидирующие позиции, что обусловлено их способностью эффективно выявлять скрытые закономерности в высокоразмерных и зашумлённых данных, характерных для сетевого трафика. Архитектуры на основе многослойных нейронных сетей позволяют формировать иерархические представления признаковых пространств, что критически важно при отсутствии явных правил разграничения нормальной и аномальной активности.

В работах ряда авторов предложено использование различных глубоких архитектур для классификации сетевого трафика: рекуррентных нейронных сетей (RNN), стековых RNN (Stacked RNN), свёрточных нейронных сетей (CNN), а также гибридных моделей на основе одномерных свёрточных слоёв в комбинации с блоками долгой краткосрочной памяти (LSTM). В условиях несбалансированных выборок, характерных для сетевого трафика, стандартное обучение демонстрирует снижение обобщающей способности, что требует модификации процедуры оптимизации.

Классическая процедура обучения нейронной сети методом обратного распространения ошибки в пакетном режиме предполагает минимизацию эмпирического риска, задаваемого функцией потерь вида:

$$E = \sum_{k=1}^{m} (y_k - d_k)^2, \qquad (1)$$

где E- ошибка обучения; y_k — фактическое значение на k-м выходном нейроне; d_k — целевое (эталонное) значение; m — размерность выходного пространства.

В реальных данных есть и простые, и сложные примеры. Если ошибка по сложному или редкому случаю мала по сравнению с общей ошибкой, она просто теряется в среднем, и сеть не учится на этом примере. В результате такие примеры не оказывают существенного влияния на градиент, и сеть не обучается их корректной классификации. Данная проблема особенно остро проявляется при наличии дисбаланса классов, где редкие аномалии составляют малую долю выборки.

Для преодоления указанного ограничения предлагается модификация алгоритма, состоящая во введении адаптивных весовых коэффициентов w_i, присваиваемых каждому обучающему примеру. В этом случае целевая функция преобразуется к виду:

$$E_w = \sum_{i=1}^{N} w_i \cdot E_i, \quad (2)$$

где E_w - взвешенная функция потерь; E_i - ошибка обучения на i-м примере; N - размер обучающей выборки; w_i - весовой коэффициент, отражающий текущую значимость.

Предложенный подход предполагает динамическое изменение весов w_i на каждой итерации обучения на основе текущего значения ошибки E_i. В основу положен принцип усиления вклада примеров, которые на текущем этапе обучения классифицируются недостаточно точно, при одновременном подавлении влияния хорошо классифицируемых образцов и исключении выбросов [5].

Процесс адаптации включает следующие этапы:

В начале процесса обучения осуществляется случайная инициализация весовых коэффициентов нейронной сети. Распределение ошибок по примерам близко к равномерному. Для всех примеров выполняется условие $E_i \leq E_2$, где E_2 — верхняя граница допустимой ошибки, устанавливаемая для каждого примера индивидуально.

В ходе обучения распределение ошибок эволюционирует в сторону квазинормального. Для большинства примеров ошибка снижается, приближаясь к нулю, тогда как для некоторого подмножества сохраняется на высоком уровне. При попадании ошибки примера в интервал $E_1 < E_i < E_2$ (где E_1 — нижняя граница интервала активации усиления) примеру присваивается весовой коэффициент $w_i > 1$, что ускоряет снижение его ошибки за счёт увеличения вклада в суммарный градиент.

После того как ошибка примера пересекает нижнюю границу E_1, пример получает единичный коэффициент $w_i = 1$ и не требует дополнительного усиления. При последующем возрастании ошибки выше E_1, коэффициент снова становится усиливающим.

В случае если ошибка примера превышает верхнюю границу E_2, пример интерпретируется как статистический выброс (аномалия в обучающей выборке) и исключается из дальнейшего обучения, что предотвращает переобучение на зашумлённых данных.

Величины E_1, E_2 и w_i пересчитываются для каждого примера на каждом шаге итерационного процесса на основе текущего распределения ошибок и динамики их изменения. Это обеспечивает гибкость и адаптивность метода к изменяющимся условиям обучения без необходимости задания фиксированных пороговых значений.

Модифицированный алгоритм может быть интегрирован как в классические архитектуры полносвязных нейронных сетей, так и в более сложные глубокие структуры, включая свёрточные и рекуррентные сети, что расширяет область его применения в задачах обнаружения вторжений, анализа сетевого трафика и идентификации аномалий в режиме реального времени.

Двухфазная платформа пассивно-активного обнаружения OpenVPN-соединений в условиях обфускации трафика

Платформа построена по двухфазному принципу, сочетающему пассивную фильтрацию трафика с активным зондированием для верификации. Такой подход позволяет достичь высокой точности обнаружения при минимальной вычислительной нагрузке на этапе первичной фильтрации.

На первой фазе модуль Filter осуществляет пассивный анализ проходящего сетевого трафика в реальном времени. Фильтрация базируется на характерных признаках протокола OpenVPN, выявляемых на этапе установления соединения. К числу таких признаков относятся:

специфическая структура управляющих пакетов (Opcode: P_Client_Reset, P_Server_Reset, P_ACK, P_Control, P_Data);

последовательность обмена пакетами в процессе установки соединения;

особенности временны́х и объёмных характеристик начальной фазы сеанса.

Фильтр работает пассивно, не модифицируя и не задерживая трафик, что позволяет обрабатывать потоки на линейной скорости без внесения дополнительной задержки.

На второй фазе потоки, помеченные Filter как потенциальные OpenVPN-соединения, передаются модулю Prober, который выполняет активное зондирование с целью подтверждения. Зондирование осуществляется путём отправки специально сконструированных запросов, направленных на выявление поведения, специфичного для протокола OpenVPN. Важной особенностью является способность Prober идентифицировать серверы OpenVPN даже в тех случаях, когда на них включены механизмы защиты от активного зондирования, за счёт использования побочных каналов (side channels) [6].

Архитектура системы интеллектуальной оценки и минимизации рисков утечки персональных данных

На основе систематизации существующих подходов к обнаружению угроз в зашифрованных каналах связи предлагается многоуровневая архитектура системы, предназначенной для оценки и минимизации рисков утечки персональных данных пользователей VPN-сервисов. Архитектура включает три последовательных уровня анализа, каждый из которых ориентирован на определённый класс угроз и использует специализированные методы обработки данных.

Первый уровень системы ориентирован на анализ клиентского программного обеспечения и браузерных расширений, устанавливаемых пользователем. К числу основных критериев оценки относятся: выявление в исполняемых скриптах расширений переопределений сетевых функций (fetch, XMLHttpRequest), идентификация скриптов, ориентированных на перехват данных, передаваемых конкретным AI-платформам, а также анализ политик конфиденциальности разработчиков и оценка наличия аффилированных партнёров-брокеров данных. Данный уровень позволяет выявлять угрозы на этапе инсталляции программного обеспечения, до начала передачи пользовательских данных.

Второй уровень реализует поведенческий анализ пользователей при работе с публичными сервисами искусственного интеллекта. В соответствии с выводами, полученными в процессе мониторинга учитываются следующие индикаторы: использование нестандартных прокси-каналов для доступа к нейросетевым сервисам, объём и семантическая характеристика передаваемых данных (исходный код, юридические документы, персональные данные), а также выявление «теневого» ИИ-трафика, генерируемого в обход корпоративных политик безопасности. Данный уровень ориентирован на обнаружение рисков, связанных с несанкционированным использованием публичных ИИ-сервисов сотрудниками организаций.

Третий уровень системы построен на многомодельном анализе исходящего VPN-трафика, в рамках которого применяются три взаимодополняющих алгоритма машинного обучения. Алгоритм Isolation Forest решает задачу оперативного обнаружения статистических выбросов в метаданных трафика, что позволяет выявлять аномальные потоки без использования размеченных обучающих выборок. Параллельно с этим LSTM-автокодер анализирует временны́е последовательности сетевых потоков, выявляя аномалии, которые могут проявляться не в отдельных пакетах, а в долговременной динамике обмена данными, что особенно важно для обнаружения скрытых угроз. Завершающим элементом выступает XGBoost, выполняющий классификацию типов трафика и идентификацию подозрительных паттернов, характерных для известных классов атак.

Предложенная трёхуровневая архитектура обеспечивает комплексную защиту на всех этапах взаимодействия пользователя с VPN-сервисом, от установки ПО до передачи данных по защищённому каналу.

Методология оценки рисков утечки персональных данных в VPN-сервисах

Оценка рисков утечки персональных данных (ПДн) при использовании VPN-сервисов представляет собой многокритериальную задачу, в рамках которой риск рассматривается как вероятностная величина, отражающая меру возможности компрометации конфиденциальных данных в условиях функционирования VPN-соединения.

Интегральная оценка риска R определяется как произведение двух компонентов:

$$R = P_{\text{comp}} \cdot C,$$

где $P_{\text{comp}} \in [0, 1]$ — вероятность компрометации ПДн, а $C \in [1, 10]$ — коэффициент критичности последствий утечки, устанавливаемый экспертным путём на основе категории передаваемых данных.

Вероятность компрометации P_{comp} представляет собой агрегированную оценку, формируемую на основе трёх групп факторов.

Первая группа — технические характеристики VPN-сервиса — включает политику логирования, используемые протоколы шифрования, организацию DNS-резолвинга, применение обфускационных механизмов и юридическую юрисдикцию провайдера.

Вторая группа — поведенческие факторы пользователя — учитывает категорию передаваемых данных (общий веб-сёрфинг, исходный код, коммерческая тайна, персональные данные, медицинская или финансовая информация), использование публичных ИИ-сервисов (Shadow AI), применение нестандартных прокси-каналов и установку непроверенного ПО.

Третья группа — средовые факторы — охватывает тип сети подключения, наличие систем глубокого анализа трафика (DPI) и уровень активности атакующих сторон.

Для каждого фактора j определяется частная оценка $p_j \in [0, 1]$, отражающая степень риска по данному параметру. Обобщённая вероятность компрометации вычисляется как взвешенная сумма частных оценок:

$$P_{\text{comp}} = \sum_{j=1}^{n} w_j \cdot p_j, \quad \sum_{j=1}^{n} w_j = 1,$$

где w_j — весовой коэффициент значимости фактора. Коэффициент критичности C определяется по категории данных: для общедоступной информации $C = 1$–3, для технической документации — 4–6, для исходного кода и коммерческой тайны — 6–8, для персональных и медицинских данных — 8–10.

Итоговое значение риска интерпретируется следующим образом: $R \in [0, 2]$ — низкий уровень (периодический мониторинг); $R \in (2, 5]$ — средний (усиление контроля); $R \in (5, 7]$ — высокий (смена сервиса); $R \in (7, 10]$ — критический (немедленное прекращение использования сервиса).

Предложенная методология интегрируется в трёхуровневую систему мониторинга, включающую анализ клиентского ПО, поведенческий анализ пользователя и классификацию сетевого трафика.

Заключение

Статья обосновывает необходимость комплексного применения методов искусственного интеллекта для оценки и минимизации рисков утечки персональных данных в VPN-сервисах. Предложенная архитектура включает три уровня защиты: обнаружение вредоносного ПО, поведенческий анализ пользователей и интеллектуальный мониторинг сетевого трафика.

Экспериментально подтверждена эффективность гибридного подхода, объединяющего Isolation Forest, LSTM-автокодер и XGBoost. Дальнейшие исследования будут направлены на разработку адаптивных механизмов обучения в реальном времени и интеграцию методов объяснимого искусственного интеллекта (XAI) для повышения интерпретируемости принимаемых решений.

Литература

1. Официальное опубликование правовых актов [Электронный ресурс]. URL: http://publication.pravo.gov.ru/documents/block/president (дата обращения: 10.09.2026).
2. Нефедова М. Вредоносные расширения ShadyPanda для Chrom и Edge установили более 4,3 млн раз [Электронный ресурс]. URL: https://xakep.ru/2025/12/02/shadypanda/ (дата обращения: 10.09.2026).
3. Бесплатные VPN: чем опасен, что видит и какие данные собирает [Электронный ресурс]. URL: https://www.securitylab.ru/blog/personal/Neurosinaps/360908.php?ysclid=mtbin1aico692449188&utm_referrer=https%3A%2F%2Fya.ru%2F (дата обращения: 10.09.2026).
4. Ишкуватов С.М. Способ и алгоритм определения типа трафика в шифрованном канале связи // Труды учебных заведений связи. 2022. Т. 8. № 4. С. 109–118. DOI:10.31854/1813-324X-2022-8-4-109-118
5. Зуев, В. Н. Обнаружение аномалий сетевого трафика методом глубокого обучения / В. Н. Зуев // Программные продукты и системы. – 2021. – № 1. – С. 91-97. – DOI 10.15827/0236-235X.133.091-097. – EDN YCVLDE.
6. Передовые методы обнаружения и обфускации VPN-трафика: углубленный анализ OpenVPN и его уязвимостей в современную цифровую эпоху / К. Н. Канатьев, М. Д. Нурмагомедов, А. А. Гребенщиков [и др.] // Инновации и инвестиции. – 2023. – № 10. – С. 211-214. – EDN XQRIZY.

Application of artificial intelligence methods to assess and minimize the risks of personal data leakage in VPN services

Smirnov V.M.
RTU MIREA

JEL classification: C10, C50, C60, C61, C80, C87, C90

The article examines the problem of protecting the personal data of VPN service users in the context of growing cyber threats and the spread of “shadow” use of public AI services. Modern approaches to detecting anomalies in encrypted VPN traffic using machine learning methods are analyzed. A multi-level architecture for a personal data leak risk assessment system is proposed, which includes the detection of malicious extensions, user behavioral analysis, and intelligent monitoring of outgoing traffic. The effectiveness of a hybrid approach combining the Isolation Forest algorithms, LSTM autoencoder, and XGBoost for comprehensive protection has been substantiated.

Keywords: artificial intelligence, machine learning, VPN, personal data leakage, anomaly detection, information security, network traffic analysis.

References

1. Official Publication of Legal Acts [Electronic resource]. URL: http://publication.pravo.gov.ru/documents/block/president (accessed: 10.09.2026).
2. Nefedova M. Malicious ShadyPanda extensions for Chrome and Edge installed over 4.3 million times [Electronic resource]. URL: https://xakep.ru/2025/12/02/shadypanda/ (accessed: 10.09.2026).
3. Free VPNs: risks, visibility, and data collection [Electronic resource]. URL: https://www.securitylab.ru/blog/personal/Neurosinaps/360908.php?ysclid=mtbin1aico692449188&utm_referrer=https%3A%2F%2Fya.ru%2F (accessed: 10.09.2026).
4. Ishkuvatov S. M. Method and algorithm for determining traffic type in an encrypted communication channel // Proceedings of Telecommunication Universities. 2022. Vol. 8. No. 4. pp. 109–118. DOI:10.31854/1813-324X-2022-8-4-109-118
5. Zuev, V. N. Network traffic anomaly detection using deep learning / V. N. Zuev // Software Products and Systems. – 2021. – No. 1. – pp. 91–97. – DOI 10.15827/0236-235X.133.091-097. – EDN YCVLDE.
6. Advanced Methods for Detecting and Obfuscating VPN Traffic: An In-depth Analysis of OpenVPN and Its Vulnerabilities in the Modern Digital Era / K. N. Kanatyev, M. D. Nurmagomedov, A. A. Grebenshchikov [et al.] // Innovations and Investments. – 2023. – No. 10. – pp. 211-214. – EDN XQRIZY.