УДК 004.057.4

В. Ю. Сапегин[1]✉

Метод обнаружения маскировки сетевого трафика

[1]Национальный исследовательский ядерный университет «МИФИ»,
г. Москва, Российская Федерация
e-mail: gvlad224@gmail.com

Аннотация. В работе исследуются современные методы и принципы, используемые для сокрытия сетевого трафика в корпоративных вычислительных сетях, функционирующих на основе стека протоколов TCP/IP версий 4 и 6, а также техники их идентификации в реальном времени. В рамках исследования разработан новый подход к детектированию зашифрованных и замаскированных VPN-каналов, основанный на комбинации анализа независимых компонент (ICA) и ансамблевого машинного обучения, включающего алгоритмы градиентного бустинга и случайного леса. В ходе исследования был разработан экспериментальный стенд, развернутый с использованием системы автоматизированного управления конфигурацией Ansible, что обеспечило высокую воспроизводимость тестовых испытаний. На базе созданного стенда сформирован специализированный датасет, включающий как валидный сетевой трафик, так и трафик виртуальных защищенных каналов в условиях их маскировки, что позволило провести комплексную валидацию предложенного метода обнаружения. Проведенные тесты подтвердили высокую результативность предложенного решения, показав его преимущества в точности и скорости обнаружения скрытых соединений, что способствует оперативному реагированию на угрозы в корпоративных сетях. Ключевым фактором повышения эффективности стало применение метода ICA и отбор признаков объектов классификации, обеспечивающей более надежное выявление аномалий в реальных условиях. Полученные результаты могут быть использованы при создании и оптимизации систем защиты данных, а также при разработке механизмов фильтрации и мониторинга сетевой активности.

Ключевые слова: технологии маскировки, виртуальные защищенные каналы, VPN, стек протоколов TCP/IP, классификация информационных потоков, методы обнаружения VPN

V. Yu. Sapegin[1]✉

Method for detecting network traffic obfuscation

[1]National Research Nuclear University «MEPhI» (Moscow Engineering Physics Institute),
Moscow, Russian Federation
e-mail: gvlad224@gmail.com

Abstract. This study examines modern methods and principles used to conceal network traffic in corporate computer networks operating on TCP/IP protocol stacks (versions 4 and 6), as well as techniques for their real-time identification. As part of the research, a novel approach for detecting encrypted and obfuscated VPN channels was developed, based on a combination of independent component analysis (ICA) and ensemble machine learning, including gradient boosting and random forest algorithms. An experimental testbed was designed and deployed using the Ansible automated configuration management system, ensuring high reproducibility of test trials. A specialized dataset was compiled using this testbed, containing both legitimate network traffic and traffic from obfuscated secure virtual channels, enabling comprehensive validation of the proposed detection method. The conducted tests confirmed the high efficacy of the proposed solution, demonstrating its advantages in accuracy and speed of detecting covert connections, thereby facilitating prompt threat response in

corporate networks. A key factor in improving detection efficiency was the application of ICA and feature selection for classification objects, ensuring more reliable anomaly detection in real-world conditions. The obtained results can be utilized in the development and optimization of data protection systems, as well as in designing network activity filtering and monitoring mechanisms.

Keywords: obfuscation techniques, virtual secure channels, VPN, TCP/IP protocol stack, traffic flow classification, VPN detection methods

Введение

В эпоху цифровой трансформации вопросы обеспечения информационной безопасности корпоративных сетей выходят на первый план. Парадоксальным образом, технологии сокрытия сетевой активности, изначально разработанные для защиты конфиденциальности, одновременно создают существенные риски для информационной инфраструктуры. Их способность эффективно маскировать передаваемые данные делает их одинаково востребованными как в легитимных целях защиты информации, так и для организации скрытых каналов передачи данных в кибератаках.

Данное противоречие особенно актуально в свете современных тенденций к усилению защиты персональных данных и одновременного ужесточения требований к мониторингу сетевой активности. Представленное исследование направлено на комплексный анализ современных механизмов маскировки трафика в IPv4/IPv6-сетях с разработкой инновационного подхода к идентификации скрытых Virtual Private Network (VPN) соединений в корпоративных сетевых инфраструктурах. Так, согласно данным исследования F.A.C.C.T. [1], в 2023 году наблюдалась устойчивая тенденция к использованию VPN-технологий при организации целевых атак на корпоративные сети Российской Федерации. Данный факт объясняется фундаментальной способностью виртуальных частных сетей к имитации легитимного сетевого трафика, что существенно осложняет процесс своевременного обнаружения вредоносной активности. Таким образом особое внимание уделяется методам, позволяющим эффективно дифференцировать легитимное использование технологий защиты трафика от их злонамеренного применения.

Современные научные группы предлагают различные подходы к решению указанной проблемы. В работах Wu H. и соавторов [2, 3] рассматриваются перспективные методы нейросетевой классификации трафика, демонстрирующие высокую эффективность при детекции сложных протоколов, включая V2Ray и Shadowsocks. Однако практическая реализация данных методов сталкивается с существенными вычислительными затратами, что ограничивает их применение в реальных корпоративных сетях.

В исследованиях Канатьева К. Н. и соавторов [4, 5] проводится системный анализ современных методов обнаружения виртуальных защищенных каналов связи. Авторы акцентируют внимание на принципиальных ограничениях существующих методов детекции, обусловленных постоянным совершенствованием технологий маскировки трафика. В своих работах они обосновывают необходимость разработки новых аналитических подходов, способных адаптироваться к

эволюции методов обхода защиты. Особое внимание уделяется перспективам создания комплексных систем мониторинга, сочетающих различные методы анализа сетевой активности.

Особое место в исследованиях занимают работы научной школы «Сетевая информационная безопасность» под руководством Басыни Е. А. [6–8], посвященные анализу оверлейных сетей типа Tor и I2P. В работах детально исследуются механизмы обхода фильтрации, применяемые в рассматриваемых протоколах, а также разрабатываются методы их детекции на основе анализа параметров и характеристик сетевого трафика. Эти исследования имеют особую ценность в контексте постоянного совершенствования технологий анонимизации и соответствующих методов противодействия.

Эволюция технологий сокрытия сетевой активности оказывает существенное влияние на формирование правового поля в сфере информационной безопасности. Ярким примером такого регулирования является изданный Роскомнадзором Приказ № 168 от 08.11.2023 Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций «О внесении изменений в Критерии оценки материалов и (или) информации, необходимых для принятия Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций решений, являющихся основаниями для включения доменных имен и (или) указателей страниц сайтов в информационно-телекоммуникационной сети «Интернет», а также сетевых адресов, позволяющих идентифицировать сайты в информационно-телекоммуникационной сети «Интернет», в единую автоматизированную информационную систему «Единый реестр доменных имен, указателей страниц сайтов в информационно-телекоммуникационной сети «Интернет» и сетевых адресов, позволяющих идентифицировать сайты в информационно-телекоммуникационной сети «Интернет», содержащие информацию, распространение которой в Российской Федерации запрещено», утвержденные приказом Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций от 27 февраля 2023 г. N 25» от 08.11.2023 № 168, который предписывает создание Единого реестра VPN- и прокси-сервисов. Данная мера обусловлена способностью этих технологий обходить установленные ограничения доступа к ресурсам с запрещенным контентом. Принятие подобных нормативных актов подчеркивает острую необходимость в разработке эффективных методов обнаружения маскированного трафика и предотвращения его использования для обхода систем фильтрации. Особую значимость приобретают решения, позволяющие осуществлять мониторинг VPN-соединений в корпоративных сетях без нарушения требований к защите персональных данных. Реализация таких мер требует комплексного подхода, сочетающего технические решения с учетом постоянно меняющегося законодательного ландшафта в области информационной безопасности.

Постановка задачи

Целью настоящей работы является разработка нового метода выявления замаскированных сетевых потоков виртуальных защищенных каналов связи в

корпоративной вычислительной сети, функционирующей на базе стека протоколов TCP/IP версии 4 и версии 6.

Результаты работы

В рамках исследования создана комплексная тестовая инфраструктура (рис. 1) для анализа методов обнаружения виртуальных защищенных каналов связи. Архитектура стенда включает: клиентские устройства на платформах iOS (FoXray), Android (V2RayNG), Windows и Linux (Xray-core [9]); VPN-серверы в Японии и Польше для моделирования реальных сетевых задержек; а также коммутационное оборудование и межсетевой экран для захвата трафика. В качестве тестового окружения использовалось решение Xray с поддержкой протоколов VLESS-reality, VLESS-vision и VLESS-mkcp, обеспечивающих продвинутую маскировку трафика под стандартные TLS/DTLS-соединения.

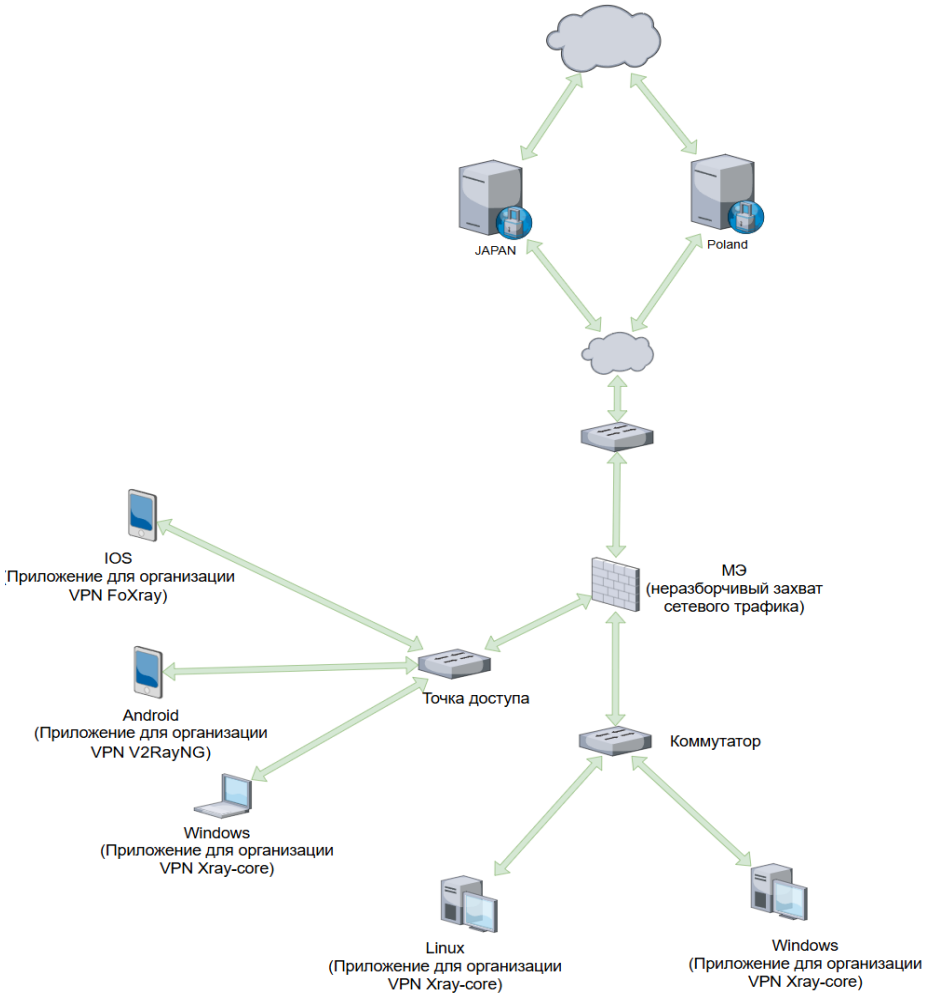

Рис. 1. Топология сети виртуального лабораторного стенда

Стенд обеспечивает генерацию зашифрованного трафика с различных платформ, его маршрутизацию через коммутатор и захват пакетов межсетевым экраном для последующего анализа. Автоматизированное развертывание стенда, при

помощи системы управления конфигурацией Ansible, гарантирует воспроизводимость результатов. Конфигурация оборудования позволяет оценивать эффективность методов обнаружения ВЗКС в условиях, приближенных к реальным, включая анализ устойчивости к современным технологиям маскировки трафика и влияние сетевых задержек на качество детектирования.

В ходе работы был разработан метод обнаружения (рис. 2). Предложенный метод реализует трехуровневый конвейер обработки данных, начиная со сбора сетевого трафика (DataCollectionTask). Первый этап (DataProcessing) включает параллельное выполнение двух ключевых операций: фильтрацию данных (DataFilteringTask) для устранения шума и нормализации, и извлечение признаков (FeatureExtractionTask) таких как временные метки и размеры пакетов.

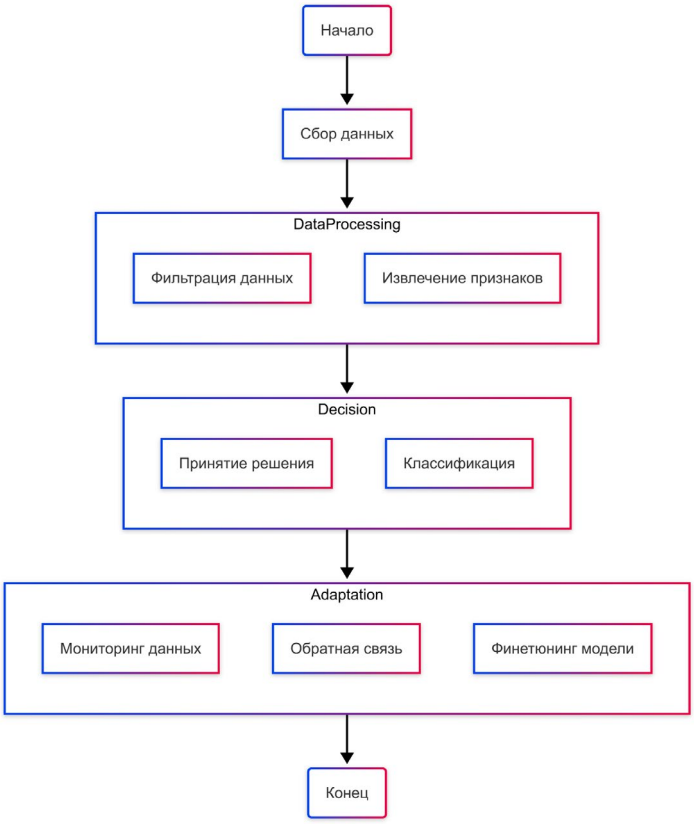

Рис. 2. Высокоуровневый метод обнаружения виртуальных защищенных каналов связи в условиях маскировки

Обработанные данные поступают на этап принятия решений (Decision), где сначала выполняется классификация трафика (ClassifierTask) с использованием современных ML-алгоритмов, а затем формируется итоговое решение (DecisionMakingTask) о характере сетевой активности. Финальный адаптационный этап (Adaptation) обеспечивает непрерывное совершенствование системы через мониторинг данных (DataMonitoringTask), анализ ошибок (FeedbackLoopTask) и тонкую настройку модели (ModelFinetuningTask), создавая замкнутый цикл самообучения системы.

На основе разработанного метода и экспериментального стенда был произведен сбор репрезентативного набора данных, включающего как обычный легитимный сетевой трафик, так и замаскированные VPN-соединения. Полученная выборка охватывает широкий спектр сценариев передачи данных, что создает надежную основу для валидации алгоритмов детектирования.

Для анализа сетевых потоков были выделены ключевые временные и пространственные характеристики сетевого трафика:

- interpacket_interval – временные интервалы между последовательными пакетами;
- packet_length - размеры отдельных пакетов в порядке их передачи;
- packet_number_per_<time> – количество пакетов в заданных временных интервалах (характеризуют плотность сетевых потоков);
- <aggregate_func>interpacket_interval_per<time> – агрегированные значения временных интервалов (среднее, медиана и др.);
- <aggregate_func>packet_length_per<time> – статистические показатели размеров пакетов в временных окнах.

Метод извлечения признаков реализует принцип скользящего анализа, аналогичный алгоритму скользящего среднего. Такой подход обеспечивает:

- эффективное подавление сетевого шума;
- точное описание мультиплексированных соединений;
- выявление паттернов маскировки в зашифрованных потоках.

Методология извлечения признаков позволяет детектировать даже сложные случаи обфускации трафика, сохраняя высокую чувствительность к аномальным сетевым активностям.

В процессе разработки модели был выполнен комплексный feature engineering, позволивший идентифицировать наиболее информативные признаки для классификации сетевых потоков (рис. 3).

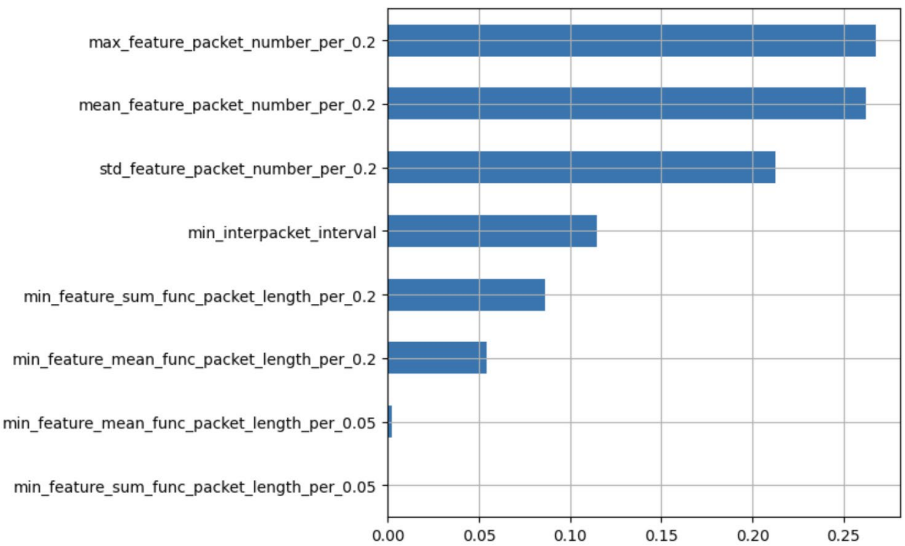

Рис. 3. Сравнительная диаграмма значимости сетевых признаков для задачи классификации VPN и Non-VPN сетевых потоков

Анализ значимости признаков выявил, что ключевую роль в детекции играют агрегированные статистические характеристики, в частности <aggregate_func>packet_length_per<time> и packet_number_per_<time>.

Эти признаки демонстрируют наибольшую дискриминативную способность, что объясняется их устойчивостью к шуму и способностью отражать фундаментальные различия в паттернах легитимного и замаскированного трафика. Их высокая значимость подтверждает гипотезу о том, что временные и размерные характеристики потока содержат наиболее релевантную информацию для задач идентификации VPN-туннелей.

Отбор этих признаков позволил оптимизировать модели классификации сетевых потоков, при этом сохранив высокую точность классификации при сокращении размерности feature space.

Выявленные высокозначимые признаки были использованы в комбинации с методом анализа независимых компонент (ICA), что позволило существенно улучшить качество классификации (рис. 4). ICA продемонстрировал исключительную эффективность в выделении статистически независимых компонент, которые максимально разделяют VPN и Non-VPN трафик в признаковом пространстве. Это связано со способностью ICA:

1) выделять скрытые факторы, лежащие в основе наблюдаемых параметров трафика;
2) устранять корреляции между признаками;
3) усиливать различия в фундаментальных характеристиках потоков.

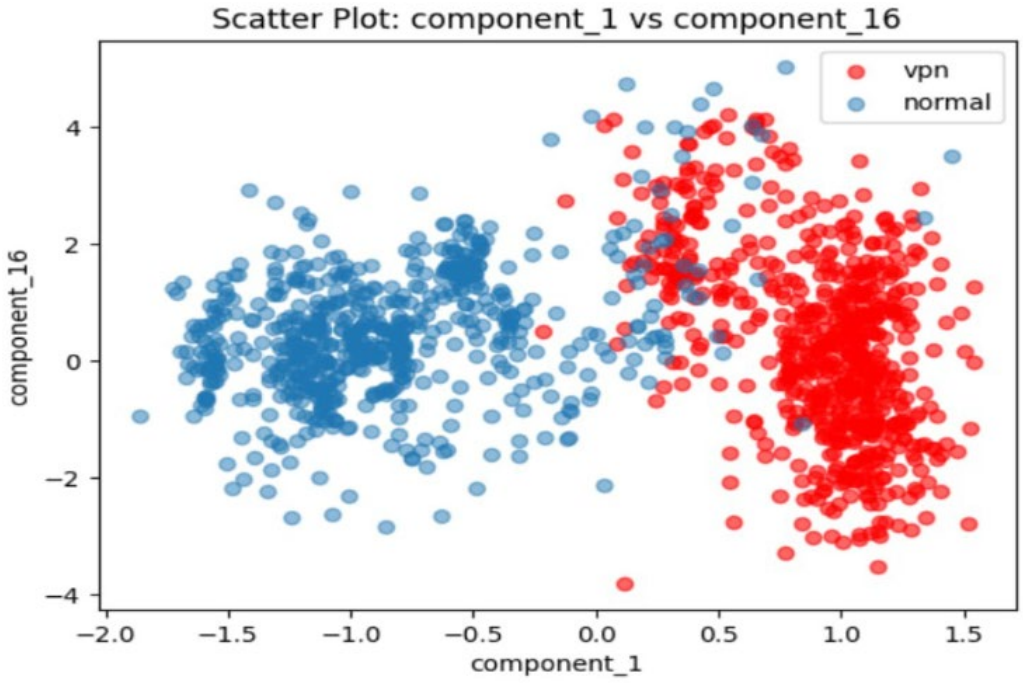

Рис. 4. Диаграмма рассеивания сетевых потоков
относительно двух выделенных

Полученные независимые компоненты были использованы в качестве входных данных для ансамблевых алгоритмов (XGBoost и Random Forest), что привело к значительному улучшению показателей классификации (рис. 5).

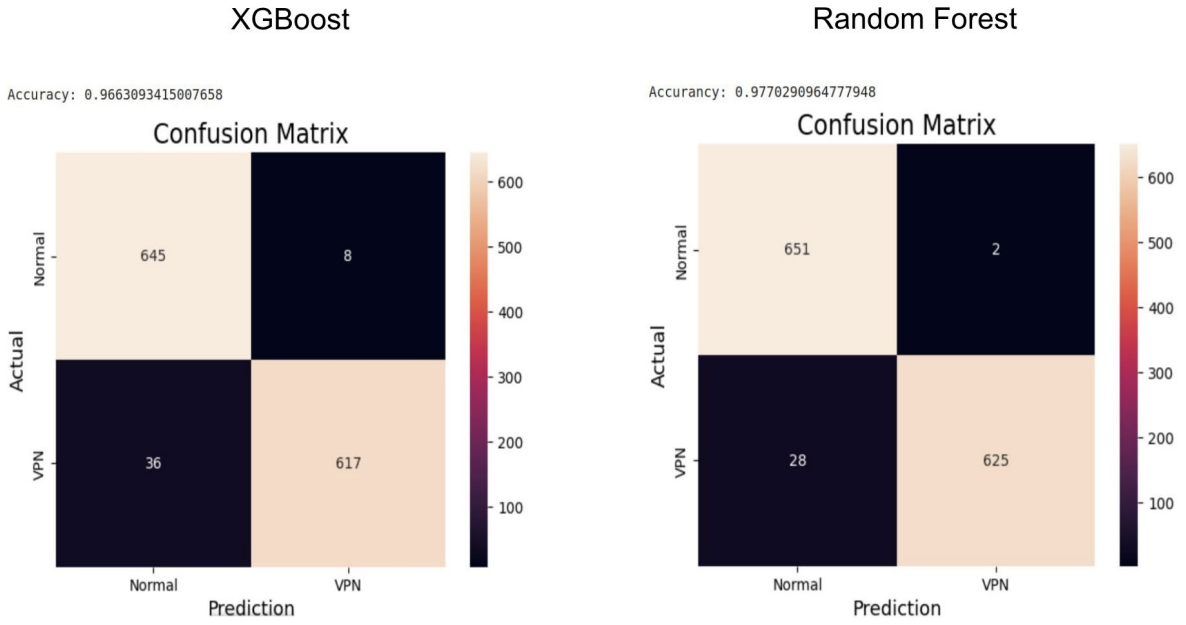

Рис. 5. Матрица ошибок классификации XGBoost и RandomForest в связке с методом ICA

Результаты подтверждают, что комбинация ICA с современными методами машинного обучения образует мощный инструмент для анализа сетевого трафика, особенно эффективный против современных технологий обфускации.

Заключение

В данной работе представлен комплексный анализ современных технологий сокрытия сетевой активности в инфраструктурах IPv4 и IPv6. Исследование включает детальное изучение принципов работы различных методов маскировки трафика, их классификацию по степени эффективности и способам противодействия системам обнаружения. Для экспериментальной проверки был развернут виртуальный лабораторный стенд, позволяющий моделировать реальные условия работы корпоративных сетей.

В ходе работы сформирован уникальный датасет, содержащий как легитимный, так и замаскированный сетевой трафик. Применение методов feature engineering позволило выделить наиболее значимые характеристики потоков данных, среди которых особую информативность показали временные и объемные параметры передачи пакетов. Использование анализа независимых компонент (ICA) продемонстрировало высокую эффективность в разделении VPN и Non-VPN трафика, что подтверждается результатами классификации с точностью 0,966 при использовании ансамблевых методов машинного обучения.

Полученные результаты имеют существенное значение для развития систем информационной безопасности. Разработанный метод может служить основой для создания эффективных решений по обнаружению и блокировке скрытых каналов передачи данных в корпоративных сетях. Особую ценность представляет адаптивность предложенного подхода, позволяющая противодействовать постоянно эволюционирующим технологиям маскировки трафика в современных сетевых инфраструктурах.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Universal Encryptor: The number of cyberattacks in Russia increased by 140% in 2023 // F.A.C.C.T. URL: https://www.facct.ru/media-center/press-releases/cyberattacks-2023/ (accessed: 21.04.2024).

2. Wu H. et al. Real-time identification of VPN traffic based on counting Bloom filter and chained hash table from sampled data in high-speed networks // ICC 2022-IEEE International Conference on Communications. – IEEE, 2022. – P. 5070-5075.

3. Wu H. et al. RT-CBCH: Real-time VPN Traffic Service Identification based on Sampled Data in High-speed Networks // IEEE Transactions on Network and Service Management. – 2023.

4. Обзор методов обнаружения VPN и DNS-анализ для классификации на примере Hotspot Shield Free / К. Н. Канатьев, С. Р. Шишкин, И. И. Басыров [и др.] // Инновации и инвестиции. – 2023. – № 10. – С. 215-219.

5. 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.

6. Басыня Е. А. Система самоорганизующегося виртуального защищенного канала связи / Е. А. Басыня // Защита информации. Инсайд. — 2018. — № 5 (83). — С. 10–15.

7. System of a self-organizing virtual secure communication channel based on stochastic multilayer encryption and overlay technologies / E. A. Basinya, Z. B. Akhayeva, D. H. Omarkhanova, G. B. Tolegenova [et al.]. — Text : direct // Journal of Theoretical and Applied Information Technology. — 2022. — Vol. 100, iss. 16. — P. 4918-4927.

8. Басыня Е. А. Программная реализация и исследование системы интеллектуально-адаптивного управления информационной инфраструктурой предприятия / Е. А. Басыня // Вестник Самарского государственного технического университета. Серия: Технические науки. — 2020. — Т. 28, № 1 (65). — С. 6-21.

9. Xray [Electronic resource]. — URL: https://github.com/XTLS/Xray-core (accessed: 21.04.2024)