БЕЗОПАСНОСТЬ ЦИФРОВЫХ ТЕХНОЛОГИЙ. – 2022. – № 3 (106). – 49–61

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.056 DOI: 10.17212/2782-2230-2022-3-49-61

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ СЕТЕВОГО ТРАФИКА В ТЕХНОЛОГИЯХ ДОВЕРЕННОГО ВЗАИМОДЕЙСТВИЯ*

А.Б. АРХИПОВА[1], М.А. МЕДВЕДЕВ[2], В.В. РЕУТОВ[3], И.В. КОРОТКИХ[4]

[1] *630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доцент кафедры защиты информации. E-mail: arhipova@corp.nstu.ru*
[2] *630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, ассистент кафедры защиты информации. E-mail: m.medvedev@corp.nstu.ru*
[3] *630008, РФ, г. Новосибирск, ул. Бориса Богаткова, 63/1, ООО СИБ, начальник коммерческого отдела ООО СИБ. E-mail: rvv@sib-nsk.ru*
[4] *109129, РФ, г. Москва, ул. 8-я Текстильщиков, 11, АРСИБ, руководитель регионального отделения. E-mail: nsk@aciso.ru*

Информационная сфера оказывает значительное влияние на современного человека и непосредственно зависит от технологий и ресурсов с информацией, их безопасностью и качеством. Значительное влияние на человечество в целом оказала компьютеризация, современные общественные отношения невозможно представить без участия, применения современных IT-технологий. Соответственно у пользователей появляется необходимость в создании надежной комплексной безопасности разрабатываемых информационных автоматизированных систем.

В настоящей статье дано понятие сетевого трафика, рассмотрена классификация сетевого трафика, в которой были выделены классификация по номерам портов, глубокий анализ пакетов, стохастический анализ пакетов, использование машинного обучения. Определены методы защиты информации с использованием доверенных технологий, рассмотрено общее представление технологий доверия. Сделаны основные выводы по перспективности использования машинного обучения для классификации сетевого трафика в доверенных технологиях.

Ключевые слова: доверие, управление доверием, доверенные системы, информационная безопасность, защита информации, машинное обучение.

* Статья получена 05 августа 2022 г.

ВВЕДЕНИЕ

В настоящее время классификация сетевого трафика крайне необходима ввиду того, что полученные результаты могут быть применимы в различного рода приложениях, которые оказывают существенное значение как для администрирования сети, так и для конечного пользователя.

Понятие протоколов, приложений и типов приложений по потокам данных в сети, с точки зрения провайдера, используется [1]:

– для контроля сети и трафика в ней;

– для обеспечения высокого качества обслуживания клиентов посредством эффективного выделения наиболее приоритетных потоков и регулирования скорости передачи отдельных пакетов;

– для планирования размещения и использования ресурсов;

– для оптимизации предоставляемых сервисов и алгоритмов маршрутизации [1].

Пользователи активно используют Сеть, поэтому у них постоянно возникает и изменяется всё больше потребностей к ней. Для этого необходимо отслеживать изменения потребностей, анализировать запросы пользователей и вовремя модифицировать сеть. Соответственно нужно уметь моделировать устройство сети в режиме здесь и сейчас и необходимо понимать, куда направлено движение изменений и ее развитие.

Классификацию интернет-потоков можно использовать с точки зрения информационной безопасности как особый признак при выявлении кибератак, неправомерных действий пользователя и т. д. Всё это дает большую возможность повысить безопасность сети в целом.

Методы, используемые при классификации интернет-трафика, также постоянно подвергаются изменениям согласно текущим потребностям пользователей, с глобальными непостоянными изменениями в устройстве самого интернет-трафика. Это влечет за собой далеко не положительное влияние на качество работы ранее применяемых способов, так как внедряются адаптированные новые технологии. Это приводит к созданию и развитию всё большего количества различных новых подходов.

1. СЕТЕВОЙ ТРАФИК

В статье рассматривается классификация трафика, но перед тем как перейдем к его рассмотрению, следует определить, что такой сам сетевой трафик [2].

Сетевой трафик рассматриваем как некоторое количество информации, передаваемой через компьютерную сеть за выбранный временной промежуток, измеряющийся в битах, байтах, гигабайтах и т. д. [2].

Сетевой трафик включает в себя множество приложений, сервисов, протоколов, т. е. сетевой трафик неоднороден. И большинство приложений уникальны по требованиям к сети, в связи с этим возникает необходимость в выполнении требований, чтобы все условия для работы приложений были удовлетворены (к примеру, скорость, джиттер и пр.). В первую очередь пользователи должны быть удовлетворены качеством поставленных в задачу условий к работе сети.

Сетевой трафик существует нескольких видов, а именно исходящий и входящий, внутренний и внешний [2]. Исходящий вид трафика отличается от входящего тем, что исходящий поступает во внешнюю сеть, а входящий, наоборот, из внешней сети. Внутренний определяется в пределах определенной сети/локальной сети, внешний же определяется за ее пределами, т. е. интернет-трафик [2].

Мы изучили статистику на основании данных британской компании SimilarWeb и приходим к выводу, что в современном обществе большую часть объема информации интернет-трафика сформировали переходы на сайты из браузеров [3].

Эти переходы возникают в связи с использованием пользователями всевозможных поисковых систем, переходов по ссылкам с других ресурсов [3]. Чуть меньшую долю объемов информации интернет-трафика формируют переходы из социальных сетей на сайты, такая же история возникает с переходами на сайты из Email и разного рода рекламы [3]. К примеру, Российская Федерация вошла в топ-10 стран по объему информации интернет-трафика, а также в топ входят Великобритания, США и т. д.

После того как мы дали понятие сетевого трафика и изучили статистику на основании данных британской компании SimilarWeb, во втором разделе рассмотрим подробно классификацию сетевого трафика.

2. КЛАССИФИКАЦИЯ СЕТЕВОГО ТРАФИКА

Доступ к сети Интернет ориентирован на пользователей, которые задают тренды к качеству работы Интернета [4]. Например, одни из ключевых требований – это высокая скорость Интернета, отсутствие «торможения», приложения пользователей должны быть с удовлетворительной скоростью скачивания, необходимость в быстром отклике приложений, четкая и плавная звуковая передача при звонках и видеозвонках. Информация о протоколах и приложениях позволяет IT-администратору реализовать требуемую политику безопасности для защиты пользователей сети [4].

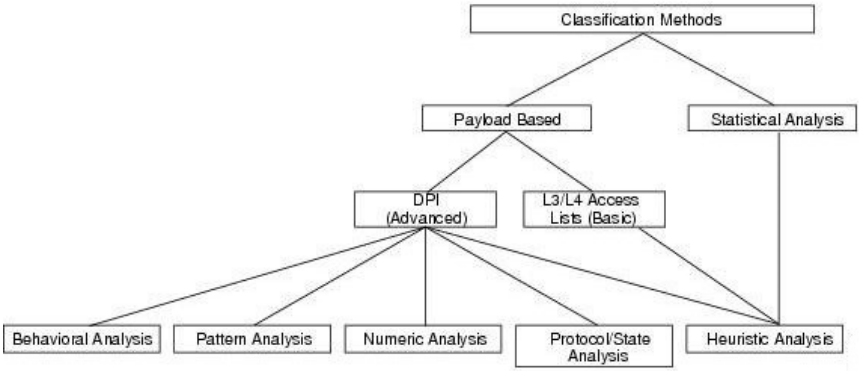

Рис. 1. Классификация трафика

Fig. 1. Traffic classification

Для идентификации приложений и протоколов, которые передаются по сети, осуществляется классификация трафика, позволяющая оптимизировать возможность управлять им [5]. После классификации все пакеты становятся отмеченными по принадлежности к определенному протоколу или приложению, что позволяет сетевым устройствам применять политики обслуживания (QoS), опираясь на эти метки и флаги. QoS (Quality of Service) – технология предоставления различным классам трафика различных приоритетов в обслуживании [5].

2.1. КЛАССИФИКАЦИЯ ПО НОМЕРАМ ПОРТОВ

Первоначальные системы классификации трафика зарождались на возможности извлекать из пакетов номера портов, а также осуществлять их сопоставление со списком Internet Assigned Numbers Authority, «Администрация адресного пространства Интернет» (IANA) [6]. Этот список регистрирует и выделяет номера портов, которые использовались для четких специфичных целей [6]. Для определения примерного типа деятельности пользователя IT-администраторы используют информацию о протоколе. Этот метод работает достаточно быстро, и благодаря ему не требуется хранения данных о потоке и вычисления элементарны. Это позволяет, например, удобно использовать его в межсетевых экранах для фильтрации трафика. Однако он обладает рядом существенных недостатков, которые по мере эволюции устройства Сети негативно влияют на результаты его работы [6].

Номер порта определен не для всех протоколов. В списке IANA уже содержатся категории «известно несколько применений наряду с зарегистрированным» и «порт не зарегистрирован IANA». Какая-то небольшая доля протоколов случайной выборки выбирает себе порты для обмена данными, такие как FTP. Часть протоколов маскируются под известными номерами совершенно других портов, но в том случае, если другой протокол, с точки зрения интернет-провайдера, становится более предпочтительным [7].

Например, протокол BitTorrent может таким образом маскироваться под HTTP, чтобы избегать блокировок или ограничений на скорость передачи данных. Появляющиеся в последнее время протоколы также могут не успевать получить зарезервированный за собой порт [7].

Этот метод хорошо подходит для определения протоколов, однако не способен хорошо различать приложения. Например, серфинг веб-страниц, VoIP и просмотр видео будут использовать 80-й (HTTP) или 443-й (HTTPS) порты для своей работы. Но у этих приложений абсолютно разные сценарии использования, поэтому на практике нам хотелось бы их различать.

Широкое распространение технологий туннелирования, инкапсулирующих протоколы, шифрование на уровне IP, использование Network Address Translation (NAT), или преобразование сетевых адресов, и Network Address and Port Translation, или трансляция сетевых адресов и портов (NAPT), – всё это влияет на применимость данного метода. Поэтому точность систем, основанных на определении номеров портов, невысока (по разным оценкам, от 30 до 70 %) и продолжает ухудшаться. В настоящее время этот признак может служить лишь одним из многих, выступая как источник дополнительной информации при принятии решения, основанного на других критериях.

2.2. ГЛУБОКИЙ АНАЛИЗ ПАКЕТОВ

Deep Packet Inspection, глубокий анализ пакетов, или DPI-технология, является следующим этапом развития классификации интернет-трафика [8]. Согласно технологии фильтрация сетевых трафиков проводится по содержимому, а именно производится анализ полностью интернет-трафика на уровнях модели OSI со второго уровня и выше, а не только заголовков [8]. DPI имеет высокоточность выполняемой работы, и при получении разметки достаточно часто ее принимают за совершенный пример для данных с неизвестными классами [8]. Для классификации с помощью DPI создается библиотека сигнатур и шаблонов пакетов, и для каждого пакета производится поиск соответствий в этой библиотеке [8].

Было замечено, что некоторые проприетарные протоколы передают информацию на уровне битов, что привело к созданию инструментов, работающих и на этом уровне [9]. Генерируемые маски содержат значения «0», «1» и «*» или вероятность единицы в данном бите [9].

Метод DPI имеет множество достоинств, но всё же имеются существенные недостатки в работе. Основной недостаток – это отсутствие возможности взаимодействовать с зашифрованным трафиком и большие требования к ресурсам [10]. Всё больше пользователей применяют в своих требованиях зашифрованный трафик, но данный метод не может отвечать их запросам, так как требуется огромный объем памяти, чтобы хранить библиотеку сигнатур и сами пакеты данных, что приводит к увеличению числа известных классов, после чего еще дополнительно увеличивается размер библиотеки и мы получаем продолжительный промежуток времени на поиск соответствия в библиотеке [10]. Становится очевидным, что этот метод не подходит для работы в высокоскоростных сетях в режиме реального времени.

Отдельно стоит вопрос защиты приватности пользователей Сети – проблема, которая актуальна для всех систем, использующих в своей работе полезную нагрузку пакетов. Законодательную сторону этого аспекта нужно учитывать при создании, обучении и работе систем глубокого анализа пакетов. Некоторые методы пытаются ограничить количество используемых данных пакета, например, первыми 40 битами, но полностью проблему это не решает.

2.3. СТОХАСТИЧЕСКИЙ АНАЛИЗ ПАКЕТОВ

Стохастический анализ пакетов (SPI, Stochastic packet inspection) используется для их классификации [11]. Например, в сетевом трафике используется критерий Хи-квадрат Пирсона для изучения случайности распределения первых байтов полезной нагрузки пакета [11]. Таким образом, строится модель синтаксиса протокола, используемого приложением. В потоке определяются как зашифрованные, так и незашифрованные пакеты на основании энтропии первого пакета. Вычисление энтропии первых байтов полезной нагрузки идентифицирует тип содержимого как текст, бинарный файл или зашифрованный файл, что позволяет приоритизировать передачу некоторых файлов [11].

Однако такую классификацию сложно назвать точной или детализированной, так как для одного и того же приложения возможно использование всех видов содержимого. Кроме того, хотя стохастический анализ и использует более простые операции, чем глубокий анализ пакетов, он всё равно использует большой объем памяти для анализа. В связи с этим данный метод не получил широкого распространения.

2.4. ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ТРАФИКА

В современном обществе происходит тенденция изменения сетевого трафика. Тренды к сетевому трафику задают пользователи. К примеру, к современным требованиям можно отнести всеобщее распространение шифрования, высокоскоростную передачу и обработку данных, в связи с этим растет количество новых способов классификации трафика для выполнения современных требований [12]. Подытоживая ранее сказанное, мы подходим к заключению, что в современном мире не обойтись без машинного обучения. Машинное обучение дает возможность упростить работу с созданием наборов различающих характеристик классов, автоматизируя этот процесс на основе анализа большого количества примеров этих классов [12].

Большинство описанных методов не работает с полезной нагрузкой пакетов, а взаимодействует только по одинаковым, общим признакам пакетов. Но благодаря этим методам решаются проблемы с шифрованием и защитой данных активных пользователей. В конечном счете у этих методов есть огромное преимущество в уменьшении объема памяти для принятия решения, что способствует высокой скорости классификации трафика [12].

3. МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ ДОВЕРЕННЫХ ТЕХНОЛОГИЙ

В рамках статьи доверие – это субъективное ожидание агентом *A* будущего поведения агента *B* на основе истории их взаимодействий.

Доверенные технологии/платформы не являются чем-то новым. Они подразумевают полное отсутствие несанкционированных действий в момент вычислений, т. е. в момент эксплуатации вычислительной системы [13].

При работе данных технологий необходимо предлагать алгоритмы, которые основаны на логистической регрессии, обеспечивающей надежную статистическую основу для предсказания доверия [13]. И тогда получаем новую вариацию определения доверия как качественное или количественное свойство доверенного лица, оцениваемое доверителем как измеримое убеждение субъективным или объективным образом, для данной задачи в определенном контексте в течение заданного периода времени [13].

Описав всё вышеперечисленное, рассмотрим машинное обучение в классическом контексте [14]. Само по себе машинное обучение имеет определенные сложности в использовании, так как включает ряд этапов реализации.

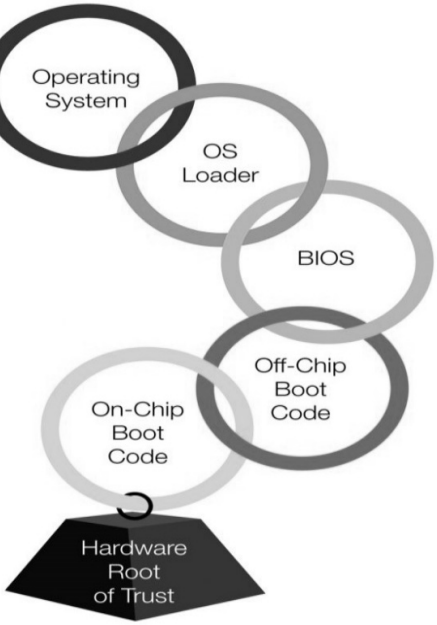

Рис. 2. Общее представление технологий доверия

Fig. 2. Introduction of trusted technologies

Сначала мы обучаем систему на тренировочном наборе данных, а затем предлагаем ей для работы неизвестную совокупность данных (генеральную совокупность) [14]. Генеральная совокупность имеет характеристики, отличающиеся от тренировочных. Так как происходит постоянное изменение характеристик данных, то положительные результаты обучения, полученные на этапах тренировки, валидации и тестирования обучаемой системы, могут быть бесполезными при самостоятельной работе системы. В связи с этим есть вероятность того, что можно самостоятельно изменить данные на любых этапах машинного обучения. Весь этот процесс можно определить как атаки на системы машинного обучения. Однозначно нельзя определить, какие использовать средства для кибербезопасности при атаках на системы машинного обучения, эти атаки выходят за рамки ранее известных стандартов. В связи с этим у пользователей возникает необходимость в исследовании и устранении проблем кибербезопасности систем искусственного интеллекта.

В стандартном варианте проблемы с моделями машинного обучения рассматриваются через призму сохранения устойчивости к вариабельности данных в генеральной выборке как отсутствие устойчивости. Алгоритм, в котором погрешность, допущенная в начальных данных или допускаемая при вычислениях, с каждым шагом не увеличивается или увеличивается незначительно, называется устойчивым [15]. Но если у погрешности значительный рост значений от шага к шагу – алгоритм неустойчивый. Устойчивым алгоритм называется тогда, когда есть мера его чувствительности ко всем изменениям в исходных данных. Устойчивость к машинному обучению означает сохранение показателей, достигнутых на всех этапах (тренировка, тестирование, валидация) в момент эксплуатации системы.

ЗАКЛЮЧЕНИЕ

В этой статье авторы привели понятие сетевого трафика и статистику по его использованию в мире, разобрали основные способы его классификации. Высока вероятность, что в ближайшее время IT-специалисты найдут еще один или более способов классификации, так как скорость развития технологий очень высока, люди нетерпеливы, им нужны новшества для комфортной жизни.

На сегодняшний день больше половины населения планеты использует Интернет. И чтобы чувствовать себя комфортно в Сети, нужно повышать скорость, эргономичность, комфортность и, самое главное, конфиденциальность Сети. И в области машинного обучения существуют определенные перспективы использования. Основная проблема внедрения системы ИИ в критических приложениях – это доверие, так как без вмешательства в работу вычислительной системы не всегда можно обойтись.

СПИСОК ЛИТЕРАТУРЫ

1. *Гетьман А.И.*, *Евстропов Е.В., Маркин Ю.В.* Анализ сетевого трафика в режиме реального времени: обзор прикладных задач, подходов и решений. – М.: ИСП РАН, 2015. – 52 с. – (Препринт / ИСП РАН; 28).

2. *Гетьман А.И.*, *Иконникова М.К.* Обзор методов классификации сетевого трафика с использованием машинного обучения // Труды Института системного программирования РАН. – 2020. – Т. 32, № 6. – С. 137–154.

3. *Шелухин О.И.* Классификация IP-трафика методами машинного обучения. – М.: Горячая линия-Телеком, 2021. – 284 с.

4. *Луганин И.В., Никонова М.К.* Классификация сетевого трафика и его использование. – М.: ИСП РАН, 2021. – 124 с.

5. *Галаминов Л.И.* Анализ IP-трафика и его классификация. – М.: Горячая линия-Телеком, 2021. – 285 с.

6. *Гришина Н.В.* Интернет-трафик: защита сетевых ресурсов организации // Прикладная информатика. – 2006. – № 1. – С. 72–76.

7. *Калишенко Е.Л., Кринкин К.В.* Подходы к прогнозированию трафика в беспроводных mesh-сетях // Прикладная информатика. – 2010. – № 6. – С. 62–68.

8. Классификация трафика и Deep Packet Inspection / VAS Experts. – URL: https://vasexperts.ru/blog/klassifikatsiya-trafika-i-deep-packet-inspection/ (дата обращения: 29.08.2022).

9. SimularWeb: web-сайт. – URL: https://www.similarweb.com/ru/ (дата обращения: 29.08.2022).

10. Internet traffic classification demystified: myths, caveats, and the best practices / H. Kim, M. Fomenkov, D. Barman, M. Faloutsos, K. Lee // Proceedings of the 4th Conference on Emerging Network Experiment and Technology, December 09–12, 2008. – Madrid, Spain, 2008. – P. 112–124.

11. *Sen S., Spatscheck O., Wang D.* Accurate, scalable In-network identification of P2P traffic using application signatures // Proceedings of the 13th International conference on World Wide Web, May 17–20, 2004. – New York, USA, 2004. – P. 512–521.

12. *Zhanh L., Tang J.* Characterization and performance study of IP traffic in WDM networks // Computer Communications. – 2001. – N 24. – P. 1702–1713.

13. *Jiang D., Tang C., Zhang A.* Cluster analysis for gene expression data: a survey // IEEE Transactions on Knowledge and Data Engineering. – 2004. – Vol. 16, N 12. – P. 1370–1386.

14. *Hubballi N., Swarnkar M., Conti M.* BitProb: probabilistic bit signatures for accurate application identification // IEEE Transactions on Network and Service Management. – 2020. – Vol. 17, N 3. – P. 1730-1741.

15. *Dorfinger P., Panholzer G., John W*. Entropy estimation for real-time encrypted traffic identification // Traffic Monitoring and Analysis. TMA 2011. – Berlin; Heidelberg: Springer, 2011. – P. 164–171.

Архипова Анастасия Борисовна, канд. техн. наук, доцент кафедры защиты информации Новосибирского государственного технического университета. В настоящее время специализируется в области программирования и информационной безопасности. E-mail: arhipova@corp.nstu.ru

Медведев Михаил Александрович, ассистент кафедры защиты информации Новосибирского государственного технического университета. В настоя-

щее время специализируется в области машинного обучения и информационной безопасности. E-mail: M.medvedev@corp.nstu.ru

Реутов Владимир Владимирович, начальник коммерческого отдела ООО СИБ. В настоящее время занимается вопросами информационной безопасности. E-mail: rvv@sib-nsk.ru

Коротких Игорь Валерьевич, руководитель регионального отделения АРСИБ. В настоящее время занимается вопросами информационной безопасности. E-mail: nsk@aciso.ru

DOI: 10.17212/2782-2230-2022-3-49-61

The perspective of using machine learning to classify network traffic in trusted interaction technologies*

A.B. Arkhipova[1], M.A. Medvedev[2], V.V. Reutov[3], I.V. Korotkih[4]

[1] *Novosibirsk State Technical University, 20 Karl Marx Prospekt, Novosibirsk, 630073, Russian Federation, associate professor of Information Security Department. E-mail: arhipova@corp.nstu.ru*
[2] *Novosibirsk State Technical University, 20 Karl Marx Prospekt, Novosibirsk, 630073, Russian Federation, assistant of the Information Security Department. E-mail: m.medvedev@corp.nstu.ru*
[3] *SIB LLC, 63/1 st. Borisa Bogatkova, Novosibirsk, 630008, Russian Federation, head of the commercial department of SIB LLC. E-mail: E-mail: rvv@sib-nsk.ru*
[4] *ARSIB, 11 8th Tekstilshchikov Street, Moscow, 109129, Russian Federation, ARSIB, head of the regional branch. E-mail: nsk@aciso.ru*

The current stage of development of the world community is characterized by an ever-increasing role of the information sphere and is completely dependent on information resources and technologies, their quality and security. Computerization of all aspects of life has become the main reason that a significant part of the elements of social relations cannot be implemented without the use of new IT in various subject areas, and hence without the implementation of a reliable system of integrated security of the developed information automated systems. This article gives the concept of network traffic, considers the classification of network traffic, in which classification by port numbers, deep packet analysis, stochastic packet analysis, and the use of machine learning were identified. Methods for protecting information using trusted technologies were defined, where a general presentation of trust technologies was considered. The main conclusions are drawn on the prospects of using machine learning to classify network traffic in trusted technologies.

Keywords: trust, trust management, trusted systems, information security, information security, machine learning

* *Received 05 August 2022.*

REFERENCES

1. Get'man A.I., Evstropov E.V., Markin Yu.V. *Analiz setevogo trafika v rezhime real'nogo vremeni: obzor prikladnykh zadach, podkhodov i reshenii* [Wirespeed network traffic analysis: survey of applied problems, approaches and solutions]. Preprint no. 28. Institute for System Programming of the RAS. Moscow, 2019. 52 p.

2. Getman A.I., Ikonnikova M.K. Obzor metodov klassifikatsii setevogo trafika s ispol'zovaniem mashinnogo obucheniya [A survey of network traffic classification methods using machine learning]. *Trudy Instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute of System Programming of the RAS*, 2020, vol. 32, no. 6, pp. 137–154.

3. Shelukhin O.I. *Klassifikatsiya IP-trafika metodami mashinnogo obucheniya* [IP traffic classification using machine learning methods]. Moscow, Goryachaya liniya-Telekom Publ., 2021. 284 p.

4. Luganin I.V., Nikonova M.K. *Klassifikatsiya setevogo trafika i ego ispol'zovanie* [Classification of network traffic and its use]. Moscow, Institute for System Programming of the RAS Publ., 2021. 124 p.

5. Galaminov L.I. *Analiz IP-trafika i ego klassifikatsiya* [IP traffic analysis and classification]. Moscow, Goryachaya liniya-Telekom Publ., 2021. 285 p.

6. Grishina N.V. Internet-trafik: zashchita setevykh resursov organizatsii [Internet traffic: protecting your organization's network resources]. *Prikladnaya informatika = Journal of Applied Informatics*, 2006, no. 1, pp. 72–76.

7. Kalishenko E.L., Krinkin K.V. Podkhody k prognozirovaniyu trafika v besprovodnykh mesh-setyakh [Traffic forecasting approaches in wireless mesh-networks]. *Prikladnaya informatika = Journal of Applied Informatics*, 2010, no. 6, pp. 62–68.

8. VAS Experts. *Klassifikatsiya trafika i Deep Packet Inspection* [Traffic classification and Deep Packet Inspection]. Available at: https://vasexperts.ru/blog/klassifikatsiya-trafika-i-deep-packet-inspection/ (accessed 29.08.2022).

9. SimularWeb: website. Available at: https://www.similarweb.com/ru/ (accessed 29.08.2022).

10. Kim H., Fomenkov M., Barman D., Faloutsos M., Lee K. Internet traffic classification demystified: myths, caveats, and the best practices. *Proceedings of the 4th Conference on Emerging Network Experiment and Technology*, December 09–12, Madrid, Spain, 2008, pp. 112–124.

11. Sen S., Spatscheck O., Wang D. Accurate, scalable In-network identification of P2P traffic using application signatures. *Proceedings of the 13th International conference on World Wide Web*, New York, USA, May 17–20, 2004, pp. 512–521.

12. Zhanh L., Tang J. Characterization and performance study of IP traffic in WDM networks. *Computer Communications*, 2001, no. 24, pp. 1702–1713.

13. Jiang D., Tang C., Zhang A. Cluster analysis for gene expression data: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 2004, vol. 16, no. 12, pp. 1370–1386.

14. Hubballi N., Swarnkar M., Conti M. BitProb: probabilistic bit signatures for accurate application identification. *IEEE Transactions on Network and Service Management*, 2020, vol. 17, no. 3, pp. 1730-1741.

15. Dorfinger P., Panholzer G., John W. Entropy estimation for real-time encrypted traffic identification. *Traffic Monitoring and Analysis. TMA 2011*. Berlin, Heidelberg, Springer, 2011, pp. 164–171.

Для цитирования:

Перспективность использования машинного обучения для классификации сетевого трафика в технологиях доверенного взаимодействия / А.Б. Архипова, М.А. Медведев, В.В. Реутов, И.В. Коротких // Безопасность цифровых технологий. – 2022. – № 3 (106). – С. 49–61. – DOI: 10.17212/2782-2230-2022-3-49-61.

For citation:

Arkhipova A.B., Medvedev M.A., Reutov V.V., Korotkih I.V. Perspektivnost' ispol'zovaniya mashinnogo obucheniya dlya klassifikatsii setevogo trafika v tekhnologiyakh doverennogo vzaimodeistviya [The perspective of using machine learning to classify network traffic in trusted interaction technologies]. *Bezopasnost' tsifrovykh tekhnologii = Digital Technology Security*, 2022, no. 3 (106), pp. 49–61. DOI: 10.17212/2782-2230-2022-3-49-61.