УДК 004.942:519.23/.25 Вестник УрФО № 3(53) / 2024, с. 44–53

Медведев М. А.

DOI: 10.14529/secur240305

ИССЛЕДОВАНИЕ И ОЦЕНКА МЕТОДОВ АВТОМАТИЗАЦИИ СБОРА И ОБРАБОТКИ СЕТЕВОГО ТРАФИКА ДЛЯ ЗАДАЧ КОНТЕНТ-ФИЛЬТРАЦИИ

Цель данной статьи – проведение оценки эффективности методов автоматизации сбора и обработки сетевого трафика. В условиях постоянно растущих объемов данных и увеличивающихся требований к их обработке предложен алгоритм, способный автоматически собирать и обрабатывать сетевой трафик с удаленных узлов, минимизируя задержки и повышая производительность системы. Основное внимание уделено трем ключевым аспектам: исследованию способов преобразования сетевого трафика, определению критериев для создания датасета в области семантической контент-фильтрации и оценке эффективности предложенных методов автоматизации. В рамках исследования выполнено преобразование сетевого трафика с использованием методов нормализации, фильтрации нерелевантных данных и разделения на сессии. Для оценки эффективности методов были предложены и использованы четыре критерия: масштабируемость, точность и качество данных, скорость обработки, а также управление потоками данных.

Новизна работы заключается в разработке стандартизированного подхода для сбора и обработки сетевого трафика, что способствует повышению точности моделей машинного обучения в задачах семантической контент-фильтрации. Предлагаемые результаты указывают возможные пути улучшения в рамках автоматизации обработки больших объемов сетевого трафика с высокой производительностью.

Ключевые слова: *автоматизация сбора сетевого трафика, преобразование трафика, нормализация данных, семантическая контент-фильтрация, масштабируемость, точность данных, фильтрация трафика, управление данными.*

RESEARCH AND EVALUATION OF METHODS FOR AUTOMATING NETWORK TRAFFIC COLLECTION AND PROCESSING FOR CONTENT FILTERING TASKS

The aim of this paper is to assess the effectiveness of methods for automating network traffic collection and processing. In the context of continuously increasing data volumes and growing demands for processing, an algorithm capable of automatically collecting and processing network traffic from remote nodes is proposed, minimizing delays and enhancing system performance. The focus is on three key aspects: investigating methods for transforming network traffic, defining criteria for creating datasets in semantic content filtering, and evaluating the effectiveness of the proposed automation methods. The study involved transforming network traffic using methods of normalization, irrelevant data filtering, and session separation. Four criteria were proposed and used to assess the effectiveness of the methods: scalability, data accuracy and quality, processing speed, and data flow management.

The novelty of the work lies in the development of a standardized approach for collecting and processing network traffic, which contributes to improving the accuracy of machine learning models in semantic content filtering tasks. The proposed results indicate possible improvements in automating the processing of large volumes of network traffic with high performance.

Keywords: *automation of network traffic collection, traffic transformation, data normalization, semantic content filtering, scalability, data accuracy, traffic filtering, data management.*

Введение

С каждым годом объем сетевого трафика неуклонно растет, что ставит перед специалистами в области сетевой безопасности и управления данными задачу разработки эффективных методов автоматизации сбора и обработки сетевого трафика. Актуальность данного исследования обусловлена необходимостью повышения точности и скорости обработки данных, что напрямую влияет на своевременное обнаружение угроз и аномалий в сетях. Автоматизация этих процессов позволяет не только улучшить производительность систем, но и позволит существенно снизить нагрузку на вычислительные мощности, что особенно важно для современных крупных сетей и облачных инфраструктур.

Цель данного исследования – провести оценку эффективности методов автоматизации сбора и обработки сетевого трафика для повышения точности, масштабируемости и скорости обработки данных.

Для достижения цели были поставлены следующие задачи:

1) исследовать и проанализировать существующие способы преобразования сетевого трафика;
2) определить критерии для создания датасета, используемого в задачах семантической контент-фильтрации;
3) оценить эффективность методов автоматизации сбора и обработки сетевого трафика, опираясь на предложенные критерии: масштабируемость, точность и качество данных, скорость и управление потоками данных.

Ряд исследований в области сетевой безопасности и обработки трафика демонстрируют необходимость применения методов машинного обучения и алгоритмов фильтра-

ции для повышения точности анализа трафика. Исследования Чжао и Ву [1] предложили подход на основе подпространств для обнаружения аномалий в кампусных сетях, который показал высокую эффективность при работе с большими объемами данных. В то же время, работы Вэйбао Чжана и Жоан П. Ласаро [2] подробно анализируют методы обнаружения аномалий, используя статистические методы и машинное обучение. Однако большинство решений все еще сталкиваются с проблемами масштабируемости и управлением потоками данных в реальном времени, что делает задачу автоматизации сбора и обработки трафика актуальной и незавершенной. В исследовании Гебрей, Ванг и Ли [3] особое внимание уделяется критическим аспектам извлечения и маркировки данных для машинного обучения в системах обнаружения атак в сетях IoT. Авторы подробно описывают методологии, направленные на повышение точности и эффективности моделей машинного обучения при работе с изменчивыми и большими объемами данных IoT. В работе Адила и Альфоуди [4] исследуется применение алгоритмов машинного обучения для классификации трафика в сетях с программно-конфигурируемой инфраструктурой (SDN). Исследование Фана и коллег [5] предлагает новый подход к анализу трафика IoT с помощью «отпечатков трафика», основанных на анализе переходов в пакетных данных. Авторы [5] рассматривают уникальные паттерны трафика различных устройств IoT, что может помочь в оптимизации и улучшении сетевой безопасности.

Основная проблема, которую решает данное исследование, заключается в разработке методов автоматизации сбора и обработки сетевого трафика с учетом масштабируемости, точности данных и скорости их обработки. Предполагается, что применение программного зеркалирования и методов нормализации данных на уровне удаленных узлов позволит существенно улучшить производительность системы и снизить время отклика на потенциальные угрозы.

Для решения поставленных задач был разработан алгоритм автоматизации сбора сетевого трафика, основанный на использовании программного зеркалирования, фильтрации нерелевантных данных и нормализации трафика. Применялись методы фильтрации трафика, разделения данных на сессии и создания стандартизированных CSV-файлов для повышения точности классификации данных. Оценка эффективности предложенных методов была проведена на основе четырех ключевых критериев: масштабируемость, точность и качество данных, скорость обработки и управление потоками данных.

Статья состоит из нескольких разделов. В первой части исследуются существующие методы преобразования сетевого трафика и проводится их анализ с точки зрения масштабируемости и точности. Во второй части описываются критерии, которые были использованы для создания датасета для задач семантической контент-фильтрации. В третьей части подробно рассматриваются разработанные методы автоматизации и их оценка по каждому из критериев. В заключении представлены выводы о проделанной работе.

Исследование способов преобразования сетевого трафика

Сложность и динамика современного сетевого трафика требуют от исследователей не только глубокого понимания существующих подходов и технологий, но и разработки новых алгоритмов, способных адекватно реагировать на постоянные изменения в сетевой активности. Исследование нацелено на решение проблемы эффективного сбора и обработки данных, что позволит формировать более качественные наборы данных на основании, которых можно проводить дальнейшее исследование сетевого трафика. Перейдем к рассмотрению основных научных статей, которые затрагивают аспекты автоматизации сбора и обработки данных.

Статья Сяофэн Чжао и Цюйбин Ву [1] представляет собой значимый вклад в область обнаружения аномалий в сетевом трафике, используя методы анализа подпространства. Этот подход основывается на выделении основных компонент трафика, которые могут указывать на нестандартное или аномальное поведение. Метод анализа подпространства может быть формализован через применение техник линейной алгебры для сокращения размерности данных [1]. Основное предположение состоит в том, что большинство нормальных данных сетевого трафика находятся в низкоразмерном подпространстве, в то время как аномалии будут проявляться в виде отклонений от этого подпространства [1]. Математически это можно описать следующей моделью (формула 1):

$$X = T\sum V^T + E, \quad (1)$$

где X – исходные данные, T и V – матрицы, описывающие подпространства, Σ – диагональная матрица сингулярных значений, указывающая на важность каждого подпространства, и E – матрица ошибок или шума.

Применение данной методики в контексте мониторинга сетевого трафика позволяет обнаруживать аномалии путем анализа резидуальных компонент E, которые не укладываются в основное подпространство [1]. Эта методика позволяет не только обнаруживать существующие угрозы, но и прогнозировать потенциальные проблемы, обеспечивая тем самым более высокий уровень предсказуемости и контроля за сетевой средой.

В статье Вэйбао Чжана и Жоан П. Ласаро [2] представлен широкий обзор методов, применяемых в современной сетевой безопасности для анализа трафика и обнаружения аномалий. Авторы детально изучают и сравнивают различные подходы – от традиционных статистических методов до современных алгоритмов машинного обучения. Это исследование помогает определить, какие методы наиболее эффективны в различных операционных условиях, а какие могут быть менее результативными.

Статистические методы, как первые инструменты анализа сетевого трафика, включают вычисление таких метрик, как среднее значение, стандартное отклонение и коэффициенты корреляции между типами трафика. Эти простые, но важные показатели позволяют выявлять отклонения в сетевом поведении. Например, стандартное отклонение, рассчитываемое по следующей формуле (формула 2), помогает оценивать степень вариации данных:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \mu\right)^2}, \quad (2)$$

где σ – стандартное отклонение, N – количество измерений, x_i – значение каждого измерения, μ – среднее значение измерений.

Авторы [2] акцентируют внимание на понимании преимуществ и недостатков каждого подхода, что позволяет более точно настраивать системы защиты, улучшать процессы обработки данных и обеспечивать высокую безопасность сетевых ресурсов.

Исследование Хайелома Гебрея, Юна Ванга и Фагена Ли [3] посвящено важнейшим аспектам извлечения и маркировки данных в сетях Интернета вещей (IoT), что играет ключевую роль в разработке и внедрении систем обнаружения атак на основе машинного обучения. Главная проблема заключается в обеспечении высокой точности и надежности данных, которые поступают на вход алгоритмов машинного обучения, что важно для своевременного и точного обнаружения аномалий и угроз в условиях динамически изменяющегося и разнообразного трафика IoT [3].

Исследователи [3] акцентируют внимание на том, что для успешной работы алгоритмов необходимо обеспечить точный захват данных, соответствующих специфическим задачам. Одним из подходов, предложенных авторами [3], является фильтрация трафика на основе заранее определенных критериев, что позволяет существенно снизить объем данных для обработки и, как следствие, повысить скорость и точность последующего анализа [3]. Например, один из применяемых методов может быть представлен в виде алгоритма классификации (формула 3):

$$y = f(x \mid \theta), \quad (3)$$

где y – метка класса, x – входные данные, θ– параметры модели.

Методики, разработанные для извлечения и маркировки данных в сетях IoT, позволят улучшить качество данных, используемых для обучения моделей машинного обучения, что напрямую скажется на эффективности систем обнаружения атак [3].

Статья Самаха Адила и Али Саида Альфоуди [4] посвящена исследованию применения различных алгоритмов машинного обучения для классификации сетевого трафика в средах с программно-конфигурируемой инфраструктурой (SDN). Одним из центральных аспектов анализа работы авторов [4] является вычисление метрик качества классификации, таких как точность (Accuracy), точность по классу (Precision), полнота (Recall) и F1-мера. Эти метрики рассчитываются по следующим формулам (формулы 4, 5, 6, 7):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

$$Precision = \frac{TP}{TP + FP}, \quad (5)$$

$$Recall = \frac{TP}{TP + FN}, \quad (6)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (7)$$

где TP – истинно положительные, TN – истинно отрицательные, FP – ложно положительные, FN – ложно отрицательные результаты.

Результаты исследования демонстрируют, что различные алгоритмы машинного обучения могут эффективно применяться для улучшения управления сетевым трафиком и обеспечения безопасности в SDN. Это исследование предоставляет важные сведения для разработки интеллектуальных и автоматизированных систем управления трафиком, способствующих повышению общей производительности и безопасности сетевой инфраструктуры [4].

В статье Мингруй Фана [5] представлен инновационный метод анализа трафика, основанный на изучении переходных моделей в пакетных данных, с целью создания уникальных «отпечатков трафика» для устройств Интернета вещей (IoT). Эти отпечатки позволяют идентифицировать и классифицировать различные типы устройств IoT на основе их характерных паттернов трафика [5]. Метод основан на анализе изменений в последовательностях, передаваемых и получаемых данных, с использованием статистических и машинно-обучаемых моделей. Это позволяет генерировать детализированные профили каждого типа устройства, где «отпечатки» формируются на основе частоты и последовательности передач пакетов данных. Например, вероятность перехода между состояниями вычисляется по формуле (формула 8):

$$p_{ij} = \frac{n_{ij}}{\sum_k n_{ik}}, \quad (8)$$

где P_{ij} – вероятность перехода от состояния i к состоянию j, n_{ij} – количество переходов от i к j, а $\sum_k n_{ik}$ – общее количество переходов из состояния i.

Исследование подчеркивает необходимость детализированного анализа трафика в условиях постоянно растущего числа подключенных устройств в современных сетях [5].

В данной главе подробно рассмотрены различные подходы и методы преобразования сетевого трафика, направленные на повышение эффективности мониторинга, анализа и управления сетевыми ресурсами. Авторы предложили методы, способствующие улучшению сетевой безопасности и более рациональному использованию ресурсов. Проведенный анализ алгоритмов позволил выявить ключевые критерии, которые могут служить основой для разработки алгоритма автоматизации сбора и обработки сетевого трафика. Эти критерии будут использоваться для дальнейшего совершенствования методов анализа и обработки данных, что позволит повысить точность и оперативность обнаружения угроз в сети.

Формирование критериев для создания датасета

Следующий этап настоящего исследования будет включать сравнительный анализ подходов определенных в исследованиях авторов с целью формирования критериев для создания датасета в области семантической контент-фильтрации с последующей разработкой алгоритма автоматизации сбора и обработки трафика. Проведем сравнительный анализ описанных выше исследований (таблица 1), для выделения основных характеристик, целей и задач.

Таблица 1 представляет методы, цели и решаемые задачи, описанные в каждой статье, демонстрируя разнообразие методов, направленных на улучшение сбора и обработки сетевого трафика. В первой, четвертой и пятой статьях представлены специализированные алгоритмы для решения задач, как обнаружение аномалий, классификация трафика и создание «отпечатков» трафика. Вторая статья представляет обзор существующих методов, предоставляя сравнительный анализ различных подходов. В третьей статье представлены методы совершенствования процессов извлечения и маркировки данных, а также оценка методов аутентификации. Вторая и третья статьи – вносят предложения для обучения и оценки технологий, что повышает их эффективность и углубляет понимание методов.

На основе проведенного анализа можно выделить следующие критерии, необходимые для автоматизации сбора и обработки сетевого трафика:

1) критерий масштабируемости – способность системы эффективно обрабатывать увеличивающиеся объемы данных без снижения производительности;

Таблица 1

Основные характеристики и методы

п/п	Название статьи	Методы и подходы	Цели исследования	Основные задачи
1	Subspace-Based Anomaly Detection for Large-Scale Campus Network Traffic	Математический анализ подпространств	Обнаружение аномалий в кампусных сетях	Выявление аномалий в сетевом трафике
2	A Survey on Network Security Traffic Analysis and Anomaly Detection Techniques	Обзор и сравнение существующих методов	Обзор текущих и перспективных технологий анализа трафика и обнаружения аномалий	Сравнительный анализ методов обнаружения аномалий
3	Traffic Data Extraction and Labeling for Machine Learning Based Attack Detection in IoT Networks	Методы извлечения и маркировки данных	Улучшение машинного обучения для обнаружения атак в IoT сетях	Извлечение и маркировка данных для обучения моделей
4	Traffic Data Classification in SDN Network Based on Machine Learning Algorithms	Алгоритмы машинного обучения для классификации трафика	Оптимизация безопасности и управления в SDN через классификацию трафика	Классификация сетевого трафика в SDN
5	Traffic Fingerprints for Homogeneous IoT Traffic Based on Packet Payload Transition Patterns	Анализ паттернов переходов в пакетных данных	Создание уникальных «отпечатков» трафика для идентификации устройств IoT	Формирование отпечатков IoT трафика на основе анализа пакетов

2) критерий точности и качества данных – необходимость точного захвата, классификации и маркировки данных для их последующего анализа и использования в моделях машинного обучения;
3) критерий скорости – возможность системы обрабатывать и классифицировать данные в режиме реального времени;
4) критерий управления и обработки данных – процессы фильтрации, агрегации и хранения, упрощающие дальнейший анализ и снижающие нагрузку на системы хранения.

Разработка алгоритма автоматизации сбора и обработки сетевого трафика требует тщательного понимания характеристик современных сетевых инфраструктур, которые постоянно сталкиваются с увеличением объемов данных. Выделенные критерии позволят эффективно структурировать процесс создания алгоритмов, ориентированных на решение реальных технических задач. Анализ рисков и проблем, основанный на этих критериях, поможет глубже оценить потенциальные угрозы и разработать оптимальные стратегии для их минимизации.

Оценка эффективности методов автоматизации сбора и обработки сетевого трафика

Разработка методов автоматизации сбора и обработки сетевого трафика является важной задачей в условиях роста объемов данных и повышенных требований к их обработке. В работе оцениваются предложенные методы на основе четырех критериев. Рассматриваются такие подходы, как программное зеркалирование и обработка данных с удаленных узлов.

Программное зеркалирование является инструментом для захвата трафика, позволяя дублировать данные с удаленных узлов и передавать их на центральный сервер для последующей обработки. Это особенно эффективно для распределенных сетей, где центральный сервер не имеет прямого доступа ко всем узлам. Такой подход снижает нагрузку на удаленные узлы и обеспечивает централизованное управление трафиком. Схема подключения удаленных узлов к серверу показана на рисунке 1.

Данный подход повышает точность сбора данных и ускоряет их обработку благодаря первичной фильтрации и нормализации на

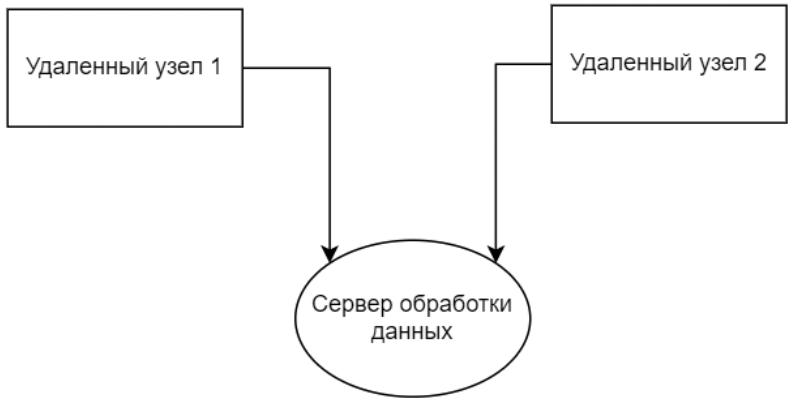

Рис. 1. Схема подключений

удаленных узлах. Распределение вычислительных ресурсов между узлами и сервером оптимизирует процесс обработки трафика, что особенно важно при работе с большими объемами данных (формула 10) [6]:

$$T_{total} = \sum_{i=1}^{n} T_i, \quad (10)$$

где T_{total} – общий объем трафика, который обрабатывается сервером, T_i – объем трафика, поступающий с каждого удаленного узла i, а n – количество удаленных узлов.

Масштабируемость системы является способность системы адаптироваться к увеличению объемов трафика без потери производительности. Применение программного зеркалирования распределяет нагрузку между удаленными узлами и центральным сервером, позволяя системе оперативно обрабатывать данные даже при росте объема трафика, без существенных задержек.

После передачи данных на основной сервер с помощью программного зеркалирования применяется метод обработки трафика с удаленных узлов, представленный на рисунке 2.

Процесс включает в себя:

1) извлечение пакетов из протокола TZSP, используемого для передачи захваченного трафика, который после служит основой для дальнейшей обработки;
2) удаление локальных и нерелевантных адресов, что сокращает объем данных для анализа и оптимизирует обработку;
3) нормализация пакетов, стандартизирующая формат данных и преобразующая их в pcap для совместимости с аналитическими инструментами;
4) разделение трафика на сессии, упрощающее классификацию и анализ сетевых данных;
5) создание CSV-формата, который структурирует данные с разделением на категории True и False, улучшая их пригодность для обучения моделей машинного обучения.

Процесс нормализации пакетов и их разделения можно представить по следующей формуле (11) [7]:

$$S = \{p_1, p_2, ..., p_n\}, \quad T_{true}, T_{false} \subset S, \quad (11)$$

где S – множество пакетов в сессии, T_{true} – подмножество нормальных пакетов, и T_{false} – подмножество аномальных пакетов.

Точность захвата и качество данных являются ключевыми факторами успешного анализа сетевого трафика. Основные методы – это нормализация, фильтрация нерелевантной информации и разделение трафика на сессии (рисунок 2) – минимизируют ошибки и упрощают анализ.

Пусть D – исходный объем данных, поступающих в систему. Фильтрация и нормализация уменьшают объем данных до D', где (формула 12):

$$D' = D - F(D), \quad (12)$$

где F(D) – функция, удаляющая нерелевантные или дублирующиеся данные.

Для упрощения обработки данные разделяются на сессии S_i, где i – номер сессии, а n – общее количество сессий (формула 13):

$$S = \{S_1, S_2, ..., S_n\}, \quad (13)$$

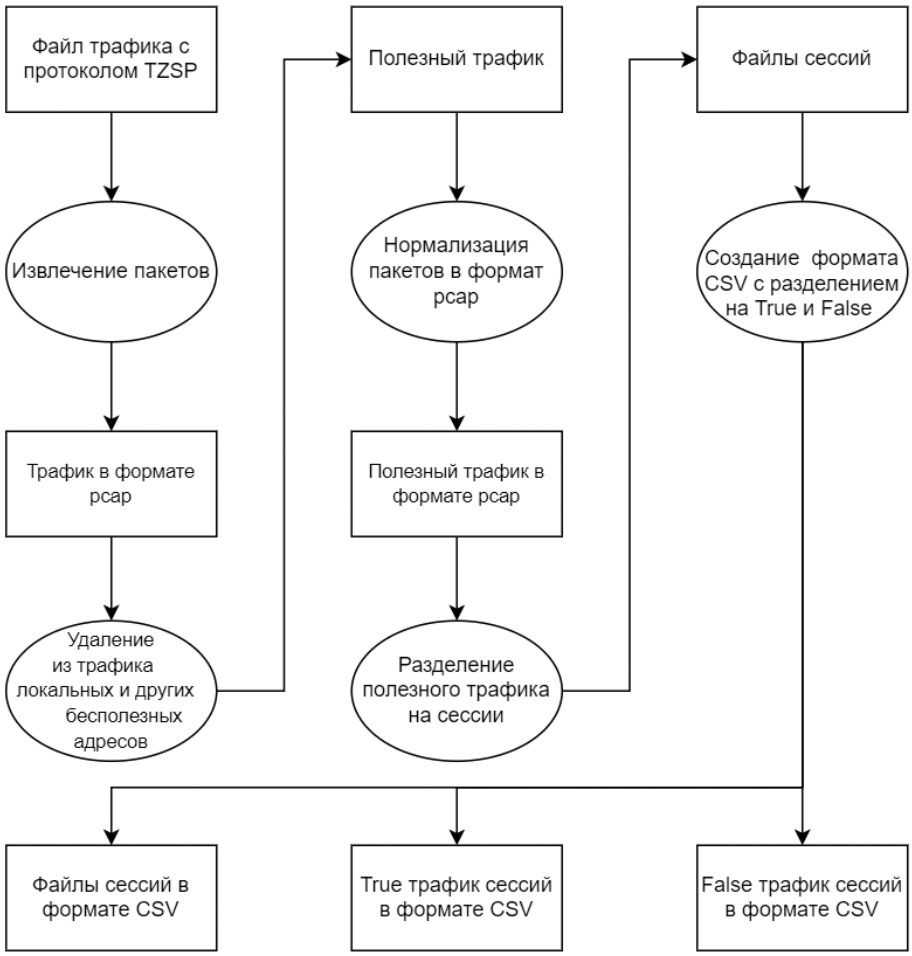

Рис. 2. Алгоритм сбора и обработки трафика

Каждая сессия нормализуется и преобразуется в формат CSV, что упрощает дальнейшую обработку.

Эффективность фильтрации оценивается через временную сложность $O_n(f(D))$, где D – исходный объем данных. Общая сложность нормализации и фильтрации может быть выражена как (формула 14):

$$O_n(f(D)) = O_n(D \log D), \qquad (14)$$

Это предполагает использование эффективных алгоритмов, что существенно сокращает объем данных и время на их обработку.

Создание стандартизированного формата CSV с разделением на True и False трафик улучшает качество данных для машинного обучения. Пусть X – множество всех пакетов данных, тогда каждый пакет $x \in X$ классифицируется как True или False, что формально выражается так (формула 15):

$$\begin{aligned} x \in T \quad & \text{если } x \text{ соответствует нормальному трафику} \\ x \in F \quad & \text{если } x \text{ соответствует аномальному трафику} \end{aligned} \qquad (15)$$

Точность классификации моделей машинного обучения измеряется метрикой точности (формула 16) [4]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \qquad (16)$$

где TP – истинно положительные, TN – истинно отрицательные, FP – ложно положительные, FN – ложно отрицательные результаты. Чем выше точность захвата и фильтрации данных, тем больше TP и TN, что повышает общую точность модели.

Скорость обработки, управление потоками и эффективная обработка трафика играют

ключевую роль в обеспечении надежной и производительной системы автоматизации сбора трафика. Эти аспекты работают совместно, позволяя системе быстро реагировать на угрозы, минимизировать задержки и снижать нагрузку на ресурсы. Внедрение методов программного зеркалирования и обработки трафика на удаленных узлах позволит сократить задержки при передаче данных на сервер и их последующей обработке [6].

Управление потоками данных и их эффективная обработка дополняют скорость системы, уменьшая нагрузку на сервер за счет фильтрации нерелевантных данных на этапе сбора [8]. Фильтрация удаляет локальные и незначимые адреса, сокращая объем данных и снижая требования к системам хранения и анализа [8]. Нормализация пакетов позволит ускорить процесс обработки и стандартизировать данные для дальнейшего анализа и классификации.

Процесс обработки и управления потоками можно описать математической моделью (формула 17) [8]:

$$T_{proc} = \frac{\sum_{i=1}^{n} P_i - F}{C_s}, \qquad (17)$$

где T_{proc} – общее время обработки, P_i – объем трафика с узла i, F – объем отфильтрованных данных, C_s – вычислительная мощность системы. Фильтрация на удаленных узлах снижает объем данных, передаваемых на сервер, что уменьшает время обработки и увеличивает скорость системы [8].

Объединение методов управления и обработки данных с высокой скоростью позволит системе эффективно справляться с увеличением объемов трафика. Ускоренная обработка и управление потоками не только снизят нагрузку на сервер, но и обеспечат быструю реакцию на потенциальные угрозы, что особенно важно при работе с большими объемами данных.

На основании проведенного исследования автоматизации сбора и обработки сетевого трафика сделаны выводы по четырем ключевым критериям: масштабируемость, точность, скорость и управление данными. Программное зеркалирование для распределенной обработки на удаленных узлах демонстрирует высокую масштабируемость, позволяя системе эффективно справляться с ростом трафика, поддерживая производительность и снижая нагрузку на центральные ресурсы. Точность и качество данных обеспечиваются нормализацией, фильтрацией и разделением трафика на сессии, с последующей стандартизацией в формате CSV для улучшения анализа. Скорость обработки возрастет благодаря фильтрации на удаленных узлах, что сократит объем передаваемых данных и уменьшит задержки. Эффективное управление потоками будет достигаться за счет ранней фильтрации и нормализации, что снизит нагрузку на системы хранения и вычислительные ресурсы. Внедрение данных критериев позволяет разработать надежный и эффективный алгоритм автоматизации сбора и обработки трафика.

Заключение

Целью данного исследования являлась оценка эффективности методов автоматизации сбора и обработки сетевого трафика для повышения точности, масштабируемости и скорости обработки данных. В ходе работы были решены следующие задачи: проведён анализ существующих методов преобразования сетевого трафика, определены критерии для создания датасета, используемого в задачах семантической контент-фильтрации, и выполнена оценка эффективности методов автоматизации на основе предложенных критериев – масштабируемости, точности и качества данных, скорости, управления потоками и обработкой данных.

Анализ методов программного зеркалирования, нормализации и фильтрации трафика на удалённых узлах показал их высокую эффективность при увеличении объёмов данных. Эти подходы обеспечат точный захват, минимизируют нагрузку на центральный сервер и поспособствуют улучшению масштабируемости системы.

Поставленные задачи исследования были достигнуты: предложенный алгоритм улучшил точность обработки данных, снизил задержки и повысил масштабируемость системы. Кроме того, выполнена задача по формированию критериев оценки методов для задач семантической контент-фильтрации. Перспективы дальнейшего развития системы связаны с внедрением потоковой обработки данных в реальном времени, что позволит ещё более эффективно управлять сетевыми потоками и своевременно реагировать на угрозы.

Литература

1. Zhao X., Wu Q. Subspace-Based Anomaly Detection for Large-Scale Campus Network Traffic, Journal of Applied Mathematics, Volume 2023, No. 1, 2023, pp. 1–12.

2. Zhang W., Lazaro J.P. A Survey on Network Security Traffic Analysis and Anomaly Detection Techniques, International Journal of Emerging Technology and Advanced Applications, Volume 1, No. 4, 2024, pp 1–9.

3. Gebrye H., Wang Y., Li F. Traffic Data Extraction and Labeling for Machine Learning Based Attack Detection in IoT Networks, International Journal of Machine Learning and Cybernetics, Volume 14, No. 7, 2023, pp. 1–16.

4. Adil S., Alfoudi A.S. Traffic Data Classification in SDN Network Based on Machine Learning Algorithms, Wasit Journal of Pure Sciences, Volume 3, No. 2, 2024, pp. 161–171.

5. Fan M., Gao J., He Y., Lu Y. Traffic Fingerprints for Homogeneous IoT Traffic Based on Packet Payload Transition Patterns, Electronics, Volume 13, No. 5, 2024, p. 930.

6. Salau A., Beyene M.M. Software Defined Networking Based Network Traffic Classification Using Machine Learning Techniques, Scientific Reports, Volume 14, No. 1, 2024, pp. 1–17.

7. Rekhi A., Saxena A., Sharma C., Easwar S. A Python-Based Packet Sniffer Simulation for Network Traffic Analysis, Preprint, 2023, pp. 1–9.

8. Vorabbi L., Maltoni D., Santi S. Optimizing Data Flow in Binary Neural Networks, Sensors, Volume 24, No. 15, 2024, pp. 1–15.

References

1. Zhao X., Wu Q. Subspace-Based Anomaly Detection for Large-Scale Campus Network Traffic, Journal of Applied Mathematics, Volume 2023, No. 1, 2023, pp. 1–12.

2. Zhang W., Lazaro J.P. A Survey on Network Security Traffic Analysis and Anomaly Detection Techniques, International Journal of Emerging Technology and Advanced Applications, Volume 1, No. 4, 2024, pp 1–9.

3. Gebrye H., Wang Y., Li F. Traffic Data Extraction and Labeling for Machine Learning Based Attack Detection in IoT Networks, International Journal of Machine Learning and Cybernetics, Volume 14, No. 7, 2023, pp. 1–16.

4. Adil S., Alfoudi A.S. Traffic Data Classification in SDN Network Based on Machine Learning Algorithms, Wasit Journal of Pure Sciences, Volume 3, No. 2, 2024, pp. 161–171.

5. Fan M., Gao J., He Y., Lu Y. Traffic Fingerprints for Homogeneous IoT Traffic Based on Packet Payload Transition Patterns, Electronics, Volume 13, No. 5, 2024, p. 930.

6. Salau A., Beyene M.M. Software Defined Networking Based Network Traffic Classification Using Machine Learning Techniques, Scientific Reports, Volume 14, No. 1, 2024, pp. 1–17.

7. Rekhi A., Saxena A., Sharma C., Easwar S. A Python-Based Packet Sniffer Simulation for Network Traffic Analysis, Preprint, 2023, pp. 1–9.

8. Vorabbi L., Maltoni D., Santi S. Optimizing Data Flow in Binary Neural Networks, Sensors, Volume 24, No. 15, 2024, pp. 1–15.

МЕДВЕДЕВ Михаил Александрович, старший преподаватель кафедры защиты информации, Новосибирский государственный технический университет. 630073, г. Новосибирск, проспект К. Маркса, 20. E-mail: m.medvedev@corp.nstu.ru

MEDVEDEV Mikhail Alexandrovich, Senior lecturer of the Department of Information Security, Novosibirsk State Technical University. 630073, Novosibirsk, K. Marksa Ave., 20. E-mail: m.medvedev@corp.nstu.ru