На правах рукописи

БАРКОВ ВЯЧЕСЛАВ ВАЛЕРЬЕВИЧ

КЛАССИФИКАЦИЯ ПРОТИВОПРАВНЫХ И НЕЖЕЛАТЕЛЬНЫХ МОБИЛЬНЫХ ПРИЛОЖЕНИЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ В ПОТОКОВОМ РЕЖИМЕ

Специальность 2.3.6. Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2024

Работа выполнена в ордена Трудового Красного Знамени федеральном государственном бюджетном образовательном учреждении высшего образования «Московский технический университет связи и информатики» (МТУСИ) на кафедре «Информационная безопасность»

Научный руководитель	Шелухин Олег Иванович – Заслуженный деятель науки РФ, доктор технических наук, профессор, ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики», заведующий кафедрой «Информационная безопасность»
Официальные оппоненты:	Лаврова Дарья Сергеевна – доктор технических наук, доцент, федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого», Высшая школа кибербезопасности, Институт компьютерных наук и кибербезопасности, профессор
	Красов Андрей Владимирович – кандидат технических наук, доцент, федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича», кафедра «Защищённые системы связи», заведующий кафедрой
Ведущая организация	Федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет», г. Ростов-на-Дону

Защита диссертации состоится 10 декабря 2024 г. в 11 часов на заседании диссертационного совета 24.2.385.09, созданного на базе федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет промышленных технологий и дизайна» по адресу: 191186, Санкт-Петербург, ул. Большая Морская, д. 18, 437 аудитория.

С диссертацией можно ознакомиться на сайте федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет промышленных технологий и дизайна» и в библиотеке по адресу: 190068, Санкт-Петербург, Вознесенский пр., д. 46, https://sutd.ru/nauka/dissertacii/.

Автореферат разослан «____» _____ 2024 г.

Ученый секретарь
диссертационного совета
24.2.385.09
кандидат экономических наук, доцент Климова Наталья Сергеевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Задача выявления мобильных приложений, осуществляющих распространение противоправного, нежелательного или ***вредоносного контента***, приобретает особую актуальность в связи с активным развитием мобильных устройств. Под противоправным контентом понимается информация, содержание которой противоречит законодательству Российской Федерации. В числе типовых нарушений, которые должны выявляться в противоправном контенте — призывы к массовым беспорядкам, оскорбление общества, государственной власти, официальных государственных символов, конституции или исполнительной власти, призывы к суицидам, информация об изготовлении и приобретении наркотиков и др.

Федеральное законодательство предусматривает порядок ограничения к таким ресурсам в сети Интернет и к соответствующим программным приложениям. Ограничение доступа к приложениям предусматривается в случаях распространения информации с нарушением авторских и/или смежных прав, а также в случае установления факта неисполнения организатором распространения информации в сети Интернет обязанностей, предусмотренных законодательством. Для операторов мобильной связи информация об использовании пользователями тех или иных приложений необходима для получения статистики по наиболее востребованным, в том числе противоправным. При необходимости мониторинг приложений в потоковом режиме обеспечивает ограничение доступа к подобным сетевым ресурсам.

Для решения подобных задач широкое распространение получили методы интеллектуального анализа данных (Data Mining, DM) и машинного обучения (Machine Learning, ML), позволяющие адаптироваться к непрерывно изменяющейся структуре Интернет-ресурсов и учитывающие специфику сетевого трафика. Внедрение таких методов позволяет с достаточно высокой эффективностью производить классификацию, анализ и фильтрацию сетевого трафика мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента.

Вместе с тем в известных работах, посвященные проблеме классификации приложений на основе анализа сетевого трафика, слабо учитывается требование выявления неизвестного сетевого трафика. При проектировании моделей классификации приложений он полностью исключается в предположении наличия только известных классов, обучение осуществляется на данных из ограниченного числа классов приложений и тестируется с помощью других данных из тех же известных классов.

Отсутствие полной и достоверной информации о структуре фонового трафика значительно снижает качество классификации интересуемых мобильных приложений. Известным методом повышения качества классификации является использование архитектуры искусственных нейронных сетей (ИНС) автокодировщик (АК).

Степень разработанности темы. Теоретическую базу диссертации в области методов ML в области информационной безопасности составляют работы таких ученых, как *Зегжда П. Д., Зегжда Д. П., Лаврова Д. С., Козачок А. В., Марков А. С., Молдовян Н. А., Крундышева В. М., Калинин М. О.*, Котенко И. В., Шелухин О. И., M. Pietrzyk, Z. Chen, B. Yang, J. Erman, K. Balachandran, J. H. Broberg, T. Bujlow, V. Carela-Español, C. C. Aggarwal, Y. Wang, G. S. o Han, J. Erman, M. Arlitt, A. Mahanti и др. Однако задаче классификации

мобильных приложений на основе анализа сетевого трафика было уделено недостаточно внимания.

Вышесказанное обусловливает актуальность настоящего исследования, направленного на повышение эффективности классификации мобильных приложений на основе анализа сетевого трафика методами машинного обучения в потоковом режиме.

Целью диссертационного исследования является повышение эффективности классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, на основе анализа сетевого трафика методами машинного обучения в потоковом режиме.

Достижение поставленной цели предусматривает решение **частных задач:**

1) Сравнительный анализ известных алгоритмов классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, на основе анализа сетевого трафика в условиях априорной неопределенности в режиме offline.
2) Разработка нового алгоритма на основе использования ИНС в виде автокодировщика.
3) Разработка модели обнаружения смены концепта в наблюдаемых атрибутах при классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, на основе анализа сетевого трафика.
4) Разработка алгоритмов классификации мобильных приложений на основе анализа сетевого трафика в потоковом режиме с «конечной» и «бесконечной» памятью на базе созданных репрезентативных выборок.
5) Разработка программного комплекса (ПК) «Система анализа трафика» (САТ) для автоматизации процесса классификации мобильных приложений с помощью анализа сетевого трафика.

Объект исследования: сетевой трафик, генерируемый мобильными приложениями.

Предмет исследования: методы ML для классификация мобильных приложений на основе анализа сетевого трафика в условиях априорной неопределенности числа приложений.

Методы исследования: методы математического моделирования, теория вероятности и математической статистики, методы машинного обучения и интеллектуального анализа (обработки) данных. Методологической основой исследования является системный подход.

Основные положения, выносимые на защиту:

1) Методика отбора рационального числа атрибутов на основе анализа их информативности при фиксировании допустимой вероятности ложной классификации. При этом определены ограничения на структуру анализируемых данных по числу пакетов и потоков, в то время как ***общепринятые*** подходы предполагают расчет характеристик на основе данных всего потока.
2) Модифицированный алгоритм классификации мобильных приложений в условиях неконтролируемого фонового трафика, ***отличающийся от известных*** алгоритмов каскадным включением нейронной сети с архитектурой АК, выполняющей предварительную фильтрацию, и вторичной модели классификации.
3) Статистическая ***модель обнаружения смены концепта*** при классификации мобильных приложений на основе анализа сетевого трафика, ***отличающаяся от известных включением АК*** в качестве базовой модели обнаружения смены концепта, в котором момент наступления смены концепта определяется посредством оценок

ошибок восстановления анализируемых приложений и превышения пороговых значений.

4) ***Новый алгоритм обнаружения смены концепта*** мобильных приложений в потоковом режиме с обработкой в скользящем окне в режиме накопления с «конечной памятью», как с равномерной, так и неравномерной интенсивностью поступления данных, ***отличающийся от известных учетом «старения» данных*** в окне обработки и негауссовским характером изменяющихся параметров классифицируемых приложений.
5) Модифицированный алгоритм Adaptive Random Forest (MARF) со встроенной моделью обнаружения смены концепта, позволяющей обнаруживать смену концепта не только во время обучения, но и во время предсказания, т.к. не использует истинные метки, осуществляет классификацию быстрее чем алгоритмы Random Forest (RF), Hoeffding Adaptive Tree (HAT), K nearest neighbors (KNN), Oza Bagging (OB).

Теоретическая значимость исследования состоит в разработке и совершенствовании математических моделей и алгоритмов, позволяющих путём применения методов машинного обучения осуществлять в потоковом режиме классификацию мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, в условиях априорной неопределенности относительно состава и числа классифицируемых приложений и возможной смены концепта.

Научная новизна состоит в следующем:

1) Методика отбора значимых атрибутов классификации, обеспечивающая повышение качества классификации, высокую достоверность классификации мобильных приложений, осуществляющих шифрование сетевого трафика, более 90% при ограниченном размере обучающей выборки (300 потоков с 16-58 пакетами в каждом, в зависимости от приложения), в то время как ***общепринятые подходы*** предполагают расчет характеристик по данным всего потока. Предложенная методика является инвариантной по отношению к разным типам сетевого трафика.
2) ***Алгоритм классификации*** мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, состоящий из последовательно включенных АК и типовой модели классификации, ***обеспечивает в условиях*** априорной неопределенности и неконтролируемого ***фонового трафика повышение достоверности*** (accuracy) ***классификации приложений на 7% по сравнению с известными алгоритмами***, ***не требуя разметки фоновых приложений*** в случае их внезапного появления.
3) ***Модель обнаружения смены концепта*** классифицируемых мобильных приложений, ***отличающаяся от известных включением АК*** (для каждого приложения), в которой момент смены концепта определяется по ошибкам восстановления анализируемых мобильных приложений и превышению пороговых значений, ***что повышает точность обнаружения смены концепта*** в потоковом режиме.
4) ***Алгоритм обнаружения смены концепта*** и классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента в потоковом режиме с накоплением и обработкой в скользящем окне в условиях ограниченной памяти для равномерной и неравномерной интенсивности поступления данных, ***отличающийся от известных алгоритмов учетом «старения» данных***.

5) ***Модифицированный адаптивный MARF в отличие от стандартного алгоритма ARF***, использующего модель обнаружения смены концепта только на этапе обучения и использующего истинные метки класса, ***позволяет обнаруживать смену концепта на этапе предсказания***.

Практическая ценность работы заключается в разработке алгоритмов и реализации программного комплекса для классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, на основе анализа сетевого трафика в потоковом режиме в условиях априорной неопределенности состава и числа классифицируемых приложений, в условиях смены концепта.

Сформирована экспериментальная база данных сетевого трафика мобильных приложений, которая может быть использована в системах обнаружения вторжений, для блокировки мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, в том числе приложений, использующих шифрование сетевого трафика.

Достоверность результатов диссертационной работы подтверждается сходимостью результатов имитационного моделирования с результатами экспериментальных данных, корректным использованием современного математического аппарата, а также достаточно широким рядом публикаций, обсуждением основных положений со специалистами на научных конференциях и семинарах.

Внедрение результатов работы.

Результаты диссертационных исследований, подтвержденные соответствующими актами внедрения, используются в АО «Лаборатория Касперского» при разработке межсетевых экранов, а также в учебном процессе МТУСИ.

Апробация результатов.

Основные результаты работы обсуждались и получили одобрение на конференциях: Международная научно-техническая конференция «Телекоммуникационные и вычислительные системы – 2018».; Международная конференция «Technology & entrepreneurship in digital society»; Научно-техническая конференция РОСИНФОКОМ-2018 «Беспроводная связь и информационная безопасность интернета»; Международная научно-техническая конференция «Фундаментальные проблемы радиоэлектронного приборостроения «INTERMATIC-2018»; IX Всероссийская научно-техническая конференция «Безопасные информационные технологи»; Международная научно-техническая конференция «Systems of signals generating and processing in the field of on board communications - 2019»; II Всероссийская научная школа-семинар «Современные тенденции развития методов и технологии защиты информации».

Публикации.

Основные положения диссертации опубликованы в 18 научных печатных работах, в том числе: 5 – в научных журналах перечня ВАК; 1 – в научных рецензируемых изданиях по базе Scopus; 11 – в материалах конференций и других изданиях. Получено свидетельство о Государственной регистрации программы для ЭВМ.

Соответствие паспорту специальности. Диссертация соответствует п.13. «Методы и модели выявления и противодействия распространению ложной и вредоносной информации» и п.15. «Принципы и решения (технические, математические, организационные и др.) по созданию новых и совершенствованию существующих средств защиты информации и обеспечения информационной безопасности» паспорта специальности 2.3.6.

Личный вклад автора. Основные научные результаты, в том числе разработка алгоритмов обнаружения смены концепта, методических рекомендаций по подбору параметров алгоритмов, получены автором лично. Вклад соавторов ограничивался постановкой задач на исследования и обсуждением полученных результатов.

Связь работы с научными программами, темами, грантами. Исследования выполнялись в инициативном порядке, в рамках работы по гранту аспирантам, соискателям и молодым ученым на исследования, направленные на обеспечение информационной безопасности для задач цифровой экономики и при государственной поддержке ведущих научных школ Российской Федерации в области информационной безопасности.

Объём и структура диссертации.

Диссертация состоит из введения, четырёх глав, заключения, списка используемых источников (118 наименований) и пяти приложений. Основное содержимое диссертации (без списка используемых источников и приложений) изложено на 107 страницах машинописного текста, иллюстрировано 25 рисунками и содержит 26 таблиц.

Основное содержание работы

Во введении обоснованы актуальность и степень разработанности темы исследования, на основе которых сформулированы цель работы и решаемые задачи для ее достижения.

В первой главе представлен анализ основных подходов к классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, методами машинного обучения в условиях априорной неопределенности и изменения характеристик сетевого трафика. Для этого на мобильных устройствах под управлением ОС Android с помощью Android VPN API осуществлялся сбор необработанных данных сетевого трафика в виде IP-пакетов.

Для классификации приложений в работе использовались известные алгоритмы машинного обучения: Logistic Regression (LR), KNN, Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), Naive Bayes (NB), C4.5, AdaBoost, SVM, ИНН. В результате тестирования определялись основные метрики модели классификации: Accuracy (достоверность), $Precision$ (точность), Recall (полнота), $F1$-мера.

Отличительной особенностью задач, решаемых в исследовании, являлись условия априорной неопределенности о составе, характеристиках, количестве классифицируемых приложений, а также наличие смены концепта, при котором статистические свойства класса (целевой переменной) изменяются со временем случайным образом.

На основании проведённого анализа сформулированы частные задачи, решением которых достигается цель исследования.

Во второй главе рассмотрены вопросы построения алгоритмов классификации нежелательных или вредоносных мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, на основе анализа сетевого трафика в режиме offline как в условиях априорной неопределенности, под которой понимается наличие фонового трафика (ФТ), не предусмотренного в классических задачах ML, так и при полной информации о числе и характеристиках классифицируемых приложений.

Выявлен полный список из 23 атрибутов приложений, формируемый серверным ПО. С использованием алгоритмов Principal Components Analysis (PCA), InfoGain, Correlation-based Feature Selector (CFS), оберточного алгоритма (Wrapper) определены состав и число атрибутов, в которых сосредоточена основная информация о мобильных приложениях.

Предложена методика отбора рационального числа атрибутов по максимуму их информативности, фиксируя допустимую вероятность ложной классификации.

Рассмотрены нежелательные и вредоносные мобильные приложения, использующие шифрование сетевого трафика при передаче по сети. Классификация таких приложений не предусматривала дешифрование сетевого трафика, так что данные, содержащиеся внутри пакетов, оставались конфиденциальными. При классификации мобильных приложений, осуществляющих шифрование трафика, рассматривались приложения, использующие протоколы SSL/TLS (SB, MI_RU) и собственные протоколы (такие как SP).

На первом этапе при классификации возможное присутствие неизвестного типа сетевого трафика не учитывалось. При обучении и тестировании моделей классификации он полностью исключался из рассмотрения, и предполагалось наличие только известных приложений. В условиях априорной определенности результаты классификации на основе алгоритмов LR, KNN, DT, RF и GB подтвердили правильность предложенного отбора атрибутов. Показано, что при использовании 20 из 23 атрибутов вероятность ложной классификации снижается в 2,5 раза (с 0,007 до 0,003) при уменьшении вероятности правильной классификации с 0,998 до 0,995.

Для приложений, использующих шифрование, проведен сравнительный анализ наиболее распространенных алгоритмов классификации: NB, C4.5, AdaBoost, SVM, RF. Наилучшие результаты обеспечивает алгоритм RF, при этом наименьшее время на этапе обучения показывают NB, C4.5 и AdaBoost, а на этапе тестирования – C4.5, RF, AdaBoost и SVM. Установлено, что для обеспечения высокого качества классификации мобильных приложений, осуществляющих шифрование сетевого трафика, достаточно ограничиться 13 наиболее информативными атрибутами. Для достижения достоверности более 90% размер обучающей выборки алгоритма RF не превышает 300 потоков, а число анализируемых сетевых пакетов в потоке составляет от 16 до 58 (в зависимости от приложения). Дальнейшее увеличение числа пакетов в потоке не приводит к заметному улучшению показателей эффективности.

На втором этапе учитывалось присутствие неизвестного типа нежелательного или вредоносного трафика. Исследования подтвердили, что отсутствие полной информации о структуре фонового трафика (ФТ) существенно снижает качество классификации интересующих приложений. Эти результаты иллюстрируются на рисунке 1а.

Рассмотрены два варианта решения этой проблемы: кластеризация сетевого трафика и введение в рассмотрение виртуального класса «Неизвестное приложение». Второй вариант является более простым и легко реализуемым. Из рисунка 1б видно, что введение класса «Неизвестное приложение» существенно улучшает точность, F-меру и достоверность классификации при наличии ФТ, однако при этом несколько возрастает интенсивность ошибок 1 рода: достоверность возрастает в среднем на 20% число ложноположительных решений увеличивается до 5%.

Недостатком данного подхода является то, что класс «Неизвестное приложение» тоже необходимо размечать и использовать при обучении моделей классификации, что не всегда возможно, например, при появлении ранее неизвестного приложения.

Для повышения эффективности классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента в условиях фонового трафика предложено использовать ИНС – автокодировщик, выполняющий роль предварительного фильтра. Автокодировщик – это модель глубокого обучения, обладающая способностью изучать закодированное представление на входном слое

ИНС, а затем воспроизводить входные данные на выходе. Эффективность использования АК вместе с моделью классификации обеспечивается предварительной фильтрации признаков перед классификацией.

Были выбраны 2 мобильных приложения, осуществляющие шифрование сетевого трафика, у которых анализировались 5 наиболее значимых атрибутов (x_4, x_3, x_2, x_6, x_8) без учета IP-адресов. Эффективность АК оценивалась по величине уменьшения среднеквадратической ошибки (MSE) выходных данных выбранных атрибутов АК относительно входных для каждого приложения.

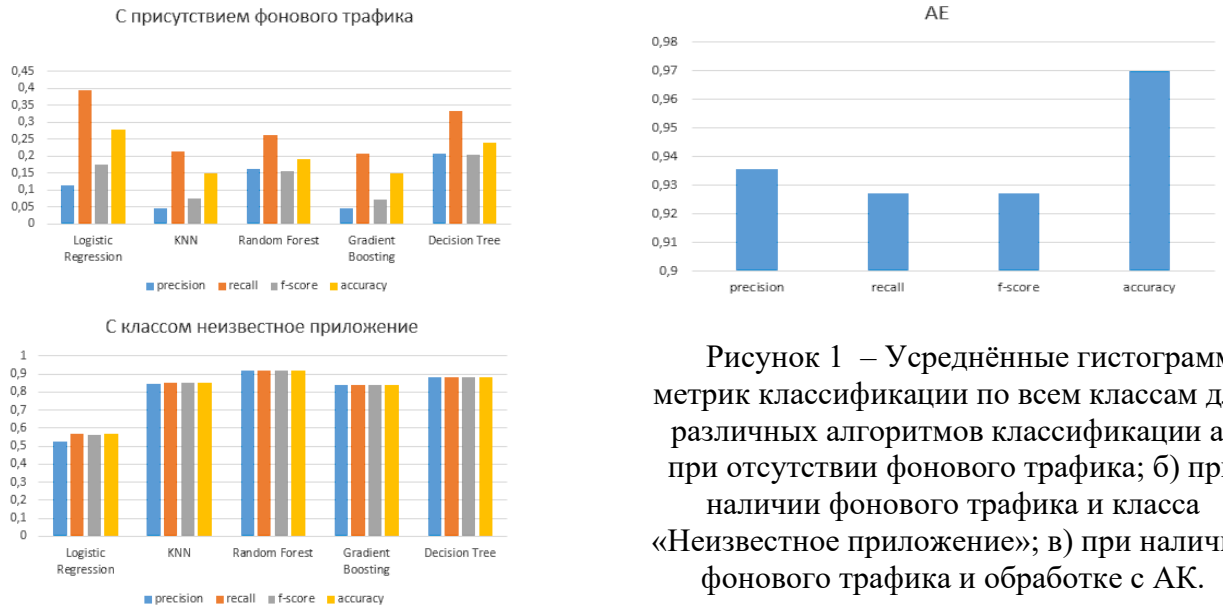

Рисунок 1 – Усреднённые гистограммы метрик классификации по всем классам для различных алгоритмов классификации а) при отсутствии фонового трафика; б) при наличии фонового трафика и класса «Неизвестное приложение»; в) при наличии фонового трафика и обработке с АК.

Показано, что диапазоны значений наиболее значимых атрибутов мобильных приложений уменьшаются на 11-50% за счет использования АК, используемого как часть модели бинарной классификации. В режиме тестирования АК восстанавливает атрибуты приложения, на которых он обучался с минимальной ошибкой, в то время как при восстановлении атрибутов других приложений ошибка будет значительно больше.

Для определения уровня допустимой ошибки часть обучающей выборки использовалась для подбора порогов. Эта часть выборки разделялась на группы по 32 экземпляра, и в каждой из них рассчитывалось пороговое значение. Если при восстановлении текущего вектора атрибутов величина MSE превышает порог экземпляра, то принимается решение, что экземпляр не принадлежит к рассматриваемому классу; в противном случае считается, что экземпляр принадлежит.

Структура модели классификации на основе АК показана на рисунке 2.

В процессе её обучения для каждого класса в обучающей выборке обучается свой АК и подбирается порог экземпляра. В процессе предсказания атрибуты экземпляра подаются на вход всех АК, вычисляется ошибка восстановления экземпляра и сравнивается с порогом экземпляра. Предсказываемый класс выбирается среди тех АК, MSE которых не превысила порог и является минимальной. Если пороги превышены для всех АК, приложение считается фоновым или неизвестным.

Проведенные исследования показали, что качество классификации с использованием АК в условиях фонового трафика значительно лучше по сравнению с лучшим алгоритмом классификации RF. Выигрыш по достоверности достигает 5%, для метрик Precision, Recall F1-score составляет около 2%. Дополнительным преимуществом использования АК является отсутствие необходимости разметки фоновых приложений (особенно в случае их внезапного

появления), а также отсутствие дополнительных механизмов для определения ФТ и повторного обучения при возрастании числа фоновых приложений.

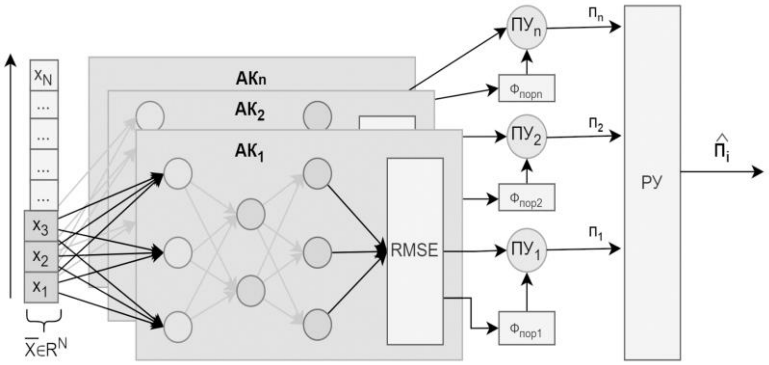

Рисунок 2 – Процедура классификации с помощью АК

Важным практическим аспектом классификации приложений в условиях априорной неопределенности является работа в потоковом (online) режиме. Отличительной особенностью потокового режима является смена концепта, которая происходит, когда исследуемый сетевой трафик является нестационарным. В таком случае классификация мобильных приложений должна осуществляться в совокупности с моделью обнаружения смены концепта (МОСК).

Третья глава посвящена разработке МОСК с применением АК, который обучается на статистике данных нормального сетевого трафика. Обнаружение смены концепта в потоке представляет собой задачу многозначной классификации, т.к. возможны случаи, когда статистические характеристики изменяются сразу для нескольких приложений. Обнаружение осуществляется моделью D отдельно для каждой метки класса $y_k = \psi(a_k); k = \overline{1,K} | a_k \in A$. Для этого строятся МОСК $D^{(y_k)}$ для потока с заданной меткой класса $y_k \in Y$. Смена концепта считается обнаруженной в приложении, если хотя бы одна модель $D^{(y_k)}$ её обнаружит.

Разделение размеченного множества $X^{(Y)}$ на подмножества, содержащие только экземпляры с одинаковыми метками классов, осуществляется с помощью функции β. Получая на вход некоторое подмножество $X^{(Y)\prime} \subset X^{(Y)}$, она разделяет его на не пересекающиеся подмножества $X^{(Y)\prime\prime}$:

$$\beta: \{X^{(Y)\prime} \subset X^{(Y)}\} \to \{\{X^{(y_k)\prime} \subset X^{(Y)\prime}\} | X^{(Y)\prime} \subset X^{(Y)}\};$$
$$\beta(X^{(Y)\prime}) = \{X^{(y_k)\prime} = \{\vec{x}_m; m = \overline{1, |X^{(Y)\prime}|} | (\vec{x}_m, y_k) \in X^{(Y)\prime}\}; k = \overline{1,K}\}. \quad (3)$$

В качестве меток классов при разделении на этапах обучения и тестирования используются истинные, а на этапе предсказания – предсказанные моделью классификации. Смена концепта приложения обнаруживается моделью $D^{(y_k)}$ с помощью модели $D^{(y_k)}_{\text{атриб}}$, которая основана на механизме обнаружения смены концепта в группах экземпляров $G^{(r)}$. На первом этапе АК использует набор весовых коэффициентов для отображения входных данных в кодирующий вектор на скрытом слое. На втором этапе АК использует набор генеративных весовых коэффициентов для восстановления закодированного вектора в исходный входной сигнал на выходном слое.

Если АК обучен только на “доброкачественных” экземплярах, он реконструирует нормальные наблюдения, но не может восстанавливать неизвестные аномальные наблюдения. В результате, когда АК фиксирует существенную ошибку восстановления, происходит классификация этого наблюдения как аномального. Момент наступления смены концепта

оценивается по превышению порогов: экземпляра k-ого приложения $T_{\text{экз}}^{(k)}$, группы k-ого приложения $T_{\text{гр}}^{(k)}$, подсчета k-ого приложения $T_{\text{п}}^{(k)}$.

Метрика качества $Q_{\text{в}}^{(y_k)} = Q(X_{AE_{\text{в}}}^{(y_k)}, \hat{X}_{AE_{\text{в}}}^{(y_k)})$ обученного автокодировщика $AE^{(y_k)}$ необходима для последующего действия. При низком качестве хотя бы одного АК требуется его повторное обучение. Смена концепта в атрибуте n считается обнаруженной, если в множестве экземпляров G обнаруживается w подряд идущих групп, для которых модель $D_{\text{атриб}}^{\text{гр}}$ обнаруживает смену концепта. Модель фиксирует номер группы, с которой началась смена концепта. В противном случае устанавливает нулевое значение.

МОСК приложения в потоке $D_{\text{прил}}$ определяется через МОСК атрибута $D_{\text{атриб}}$:

$$D_{\text{прил}}\left(G, \hat{G}, T_{\text{гр}}^{(k)}, T_{\text{экз}}^{(k)}, T_{\text{п}}^{(k)}, w\right) =$$
$$\{(n, r) | r = D_{\text{атриб}}\left(G, \hat{G}, n, T_{\text{гр}}^{(k)}, T_{\text{экз}}^{(k)}, T_{\text{п}}^{(k)}, w\right), r > 0, n = \overline{1, N}\}$$

где G – исходное множество групп экземпляров; $\hat{G}$ – множество реконструированных групп экземпляров; $T_{\text{гр}}^{(k)}$ – множество порогов группы для всех атрибутов; $T_{\text{экз}}^{(k)}$ – множество порогов экземпляров для всех атрибутов; $T_{\text{п}}^{(k)}$ – множество порогов подсчёта для всех атрибутов; w – количество подряд идущих групп, в которых определяется смена концепта.

С помощью модели для каждого атрибута определяется наличие смены концепта. Смена концепта приложения обнаруживается, если она обнаруживается хотя бы в одном атрибуте этого приложения. Модель $D_{\text{прил}}$ определяет множество пар: номер атрибута n, в котором обнаружена смена концепта, и номер группы r, с которой она началась. Если в n-ом атрибуте смена концепта не обнаружена ($r = 0$), то он не попадает в результирующее множество. Если смена концепта не обнаружена ни в одном атрибуте, результирующее множество будет пустым. Итоговая МОСК имеет вид: $D\left(X^{(Y)\prime}, w\right) = \{(y_k, n, r) | (n, r) \in D_{\text{прил}}\left(G^{(y_k)}, AE^{(y_k)}\left(G^{(y_k)}\right), T_{\text{гр}}^{(k)}, T_{\text{экз}}^{(\text{k})}, T_{\text{п}}^{(\text{k})}, w\right), G^{(y_k)} \in g\left(X^{(y_k)}, V\right), X^{(y_k)} \in \beta(X^{(Y)\prime})\}$, где $X^{(Y)\prime}$ – размеченное множество TCP сеансов в виде множества пары вектора атрибутов $\vec{x}_m \in X$ и метки класса $y'_m \in Y$; w – количество подряд идущих групп, в которых должна быть выявлена смена концепта; $X^{(y_k)}$ – размеченное подмножество TCP сеансов, сформированное приложением с меткой класса y_k.

Обнаружение смены концепта может осуществляться не только на этапах обучения и тестирования, но и на этапе предсказания, т.к. не предполагает наличия истинных меток и заключается либо в использовании меток модели классификации, либо в попытке восстановления данных с использованием всех АК путём подсчёта ошибок восстановления. Структурная схема предлагаемой МОСК, основанной на механизме обнаружения смены концепта в группах экземпляров, приведена на рисунке 3. При обучении множество АК используют обычные, размеченные данные.

Обнаружение смены концепта заключается в определении приложений, статистические характеристики которых значительно изменяются во времени, приводя к снижению качества классификации. В режиме предсказания АК восстанавливает атрибуты приложения, на котором обучался с минимальной ошибкой, в то время как при восстановлении атрибутов других приложений ошибка будет значительной.

Обучающее множество $X_{\text{об}}^{(Y)}$ разделяется на k подмножеств, равное числу идентифицируемых приложений.

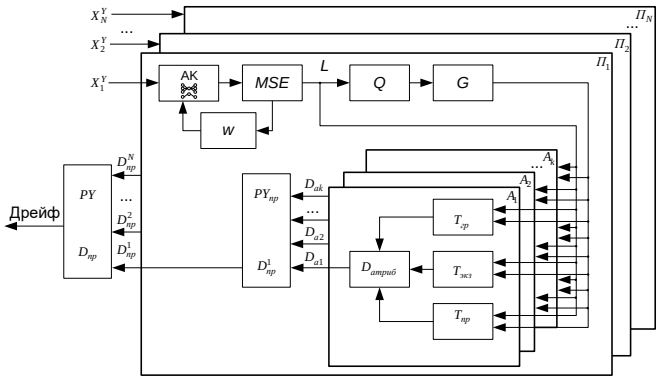

Рисунок 3 – Структурная схема МОСК

Каждое подмножество $X_{\text{об}}^{(y_k)}$, в свою очередь, разделяется на обучающее $X_{AE_{\text{об}}}^{(y_k)}$, и валидационное $X_{AE_{\text{в}}}^{(y_k)}$ множества, а также множество для настройки порогов приложения $X_{AE_{\text{п}}}^{(y_k)}$. Множество $X_{AE_{\text{об}}}^{(y_k)}$ используется для обучения АК, а $X_{AE_{\text{в}}}^{(y_k)}$ – для его оценки. Множество $X_{AE_{\text{п}}}^{(y_k)}$ используется для вычисления порогов экземпляра $T_{\text{экз}}^{(k)}$, группы $T_{\text{гр}}^{(k)}$ и подсчета $T_{\text{п}}^{(k)}$. Потери атрибутов в каждой группе усредняются. Полученные значения усредняются по всем группам. В результате для каждого атрибута формируются пороги группы $T_{\text{гр}}^{(k)} = \left\{T_{\text{гр}}^{(kn)};\ n = \overline{1,N}\right\}$.

Каждый экземпляр в группе реконструируется (восстанавливается) с помощью обученного АК. Для каждого экземпляра в группах рассчитывается функция потерь – среднеквадратичная ошибка $\text{MSE} = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \hat{x}_i)^2$. Если при восстановлении векторов признаков MSE фиксируется частое превышение порога, то атрибут приложения считается дрейфующим.

Для каждого класса модели обучается отдельный АК и подбирается порог экземпляра. В процессе предсказания атрибуты экземпляра подаются на вход всех АК, вычисляется ошибка восстановления экземпляра и сравнивается с порогом.

Реализация алгоритма обнаружения смены концепта реализована на языке программирования Python. Количество обучаемых АК выбиралось равным количеству анализируемых приложений.

Для учета эффекта «старения» данных в потоковом режиме изменение текущих статистических характеристик атрибутов сетевого трафика мобильных приложений определяется с помощью двух перемещающихся во времени окон W_1 и W_2. Решающее правило для обнаружения смены концепта с учетом «старения» обрабатываемых данных имеет вид: $R_t^J = \frac{M_{W_2}^J(t)}{M_{W_1}^J(t)} > \lambda$, где $M_{W_1}^J(t) = \frac{\frac{1}{K}\sum_{i=N2}^{N1+N2-1}\sum_{j=1}^{K}\alpha_1^i y_{t-i,j}^J}{\sum_{i=0}^{N1-1}\alpha_1^i}$; $M_{W_2}^J(t) = \frac{\frac{1}{K}\sum_{i=0}^{N2-1}\sum_{j=1}^{K}\alpha_2^i y_{t-i,j}^J}{\sum_{i=0}^{N2-1}\alpha_2^i}$;

$y_{t,j}^J$ – значение элементов наблюдаемого потока $Y_j^J = \{y_{0,j}^J, y_{1,j}^J, \dots, y_{N-1,j}^J\}$ (j атрибута J приложения), измеренных в момент времени $t \in T = \{0,1,\dots,N-1\}$; N – размер множества Y; N_1 – размер окна W_1; N_2 – размер окна W_2; K – количество атрибутов; α_1, α_2 – коэффициенты затухания, характеризующие «память» измеренных значений соответственно в первом W_1 и втором W_2 окнах ($0 < \alpha_2 < \alpha_1 < 1$).

Использование МОСК позволяют снизить вероятность ошибки классификации примерно на 5%. Показано, что модели классификации, использующие МОСК на основе

коэффициентов затухания, начинают хуже работать сразу после смены концепта. Модели классификации, не использующие МОСК, допускают на 5–7% больше ошибок.

В четвертой главе представлены результаты исследования потоковой классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, на основе анализа сетевого трафика. Сетевой трафик получен с помощью разработанного программного комплекса CAT, включающего сервер баз данных, сервер приложений, Web-приложение и клиентское программное обеспечение (ПО) для мобильных устройств под управлением ОС Android.

Проведен сравнительный анализ известных алгоритмов классификации мобильных приложений в режиме online: ARF, HAT, KNN, OB в режимах с «конечной» и «бесконечной» памятью на примере следующих мобильных приложений: ISG, SP, MI_RU, SB, HSN, PKB. Для каждого приложения исследовалось 5000 TCP сеансов, часть из которых была захвачена с момента начала «рукопожатия», другая – уже во время передачи информации.

Анализировались два сценария распределения интенсивности трафика. Первый, когда нежелательные или вредоносные приложения поступали равномерно и непрерывно. Второй, когда анализируемые приложения появлялись со случайной интенсивностью и длительностью. Из экспериментальной последовательности потоковых данных формировались 100 периодов длительностью $T = T_{\text{об}} + T_{\text{тест}}$, где $T_{\text{об}}$ – длительность интервала обучения, $T_{\text{тест}}$ – длительность интервала тестирования модели классификации; при этом параметр $\text{К} = T_{\text{об}}/\text{Т}$ характеризует скважность периодов обучения. Формирование показателей качества осуществлялось на основе обработки последовательностей длиной n периодов T в окне размера $W = n \cdot T$.

Для первого сценария получено, что наилучшие результаты качества классификации показывает алгоритм ARF. В работе приведены результаты исследования показателя $Precision$, характеризующего качество модели классификации ARF в первом режиме с накоплением данных для двух приложений. Для случая неравномерного поступления данных получены оценки эффективности алгоритма ARF на всем потоковом трафике.

Исследования показали, что алгоритм ARF «хорошо» справляется с задачами классификации как при равномерном, так и при неравномерном поступлении сетевого трафика нежелательных или вредоносных мобильных приложений. В случае равномерной интенсивности поступления данных рекомендуемое значение параметра $\text{К} \approx 0{,}5$. При неравномерной интенсивности наилучшие результаты обеспечиваются при $\text{К} \approx 0{,}8$ в режиме накопления с фиксированным размером окна порядка $W \approx 20\text{Т}$. Указанный подход был реализован путем модификации известного алгоритма ARF. Для обнаружения смены концепта в MARF используется МОСК на основе автокодировщиков, не требующая истинных меток.

Использование MARF позволяет не только осуществлять классификацию в 2-3 раза быстрее, чем базовый алгоритм RF и алгоритмы HAT, KNN и OB, но и обнаруживать смену концепта при использовании модели. Это делает MARF предпочтительным для решения задач классификации мобильных приложений на основе анализа сетевого трафика в реальном масштабе времени.

Разработанный ПК CAT, позволяет в автоматическом режиме собирать с мобильных устройств пакеты сетевого трафика и сохранять их в БД, группировать пакеты сетевого трафика в потоки. По запросу пользователя формировать наборы данных с заданными характеристиками (число потоков конкретного приложения, контроль наличия фонового трафика).

В состав разработанного ПК входят: серверное ПО, реализующее алгоритмы классификации сетевого трафика; сервер БД для управления БД потоков сетевого трафика; клиентское ПО для мобильных устройств, реализующее сбор пакетов трафика, отправку их на сервер и управление исследованиями; клиентское ПО для Web-браузера для управления исследованиями и визуализации результатов.

В качестве аппаратного обеспечения использовались: сервер IMB под управлением ОС Microsoft Windows Server 2016 Standard; мобильные устройства под управлением ОС Android 4.0 и выше. Для серверного ПО использовалась платформа Java Enterprise Edition 8, библиотеки WEKA и MOA. В качестве сервера приложений выбран Oracle Glassfish 5.0, для сервера БД использована бесплатная СУБД MySQL 5.7. При разработке мобильного приложения использовались языки программирования Java и C++. Для клиентского программного обеспечения Web-браузера выбраны JavaScript, HTML 5.0, CSS 3.0 и библиотеки Bootstrap и jQuery.

Обучение и тестирование моделей классификации осуществлялось с использованием библиотек WEKA и MOA. Разработанная абстракция моделей классификации позволяют использовать другие библиотеки, в том числе написанные на других языках программирования, а также реализовывать собственные алгоритмы классификации.

Разработанный ПК показал высокие эксплуатационные возможности не только при проведении исследований, но и использован в учебном процессе МТУСИ.

Заключение

В диссертационном исследовании решена актуальная научная задача разработки программно-методического инструментария для обнаружения и классификации в потоковом режиме нежелательных или вредоносных мобильных приложений путём анализа сетевого трафика. Получены следующие результаты.

1. Предложена методика отбора значимых атрибутов нежелательных или вредоносных мобильных приложений, обеспечивающая высокую достоверность и снижение вероятности ложной классификации до 2,5 раз. Это позволило применить классификацию мобильных приложений в потоковом режиме, которая является инвариантной по отношению к разным типам сетевого трафика.
2. Для обеспечения достаточно высокой достоверности классификации нежелательных или вредоносных мобильных приложений, осуществляющих шифрование сетевого трафика, размер обучающей выборки достаточно ограничить 300 потоками с 16-58 пакетами в каждом и числом информативных атрибутов не более 13 (в зависимости от приложения), в то время как общепринятые подходы предполагают использование данных всего потока.
3. Разработан новый алгоритм классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента, в условиях априорной неопределенности и неконтролируемого фонового трафика посредством последовательного включения АК и типовой модели классификации, что позволило повысить достоверность классификации приложений на 5-7% по сравнению с типовыми алгоритмами, не требуя разметки фоновых приложений.
4. Разработана статистическая модель обнаружения смены концепта при классификации нежелательных или вредоносных мобильных приложений путём анализа сетевого трафика, отличающаяся от известных тем, что АК включён в качестве базовой модели обнаружения, а момент наступления смены концепта определяется по превышению

порогов ошибок восстановления атрибутов мобильных приложений, что позволяет повысить точность обнаружения смены концепта в реальном времени до 10 %.

5. Разработан алгоритм MARF (модификация алгоритма ARF) со встроенной моделью обнаружения смены концепта, позволяющей обнаруживать смену концепта не только во время обучения, но и во время предсказания, т.к. не использует истинные метки и осуществляющий классификацию в 2–3 раза быстрее, чем известные алгоритмы (RF, HAT, KNN, OB) при выборе показателя скважности $K \approx 0{,}5$; для неравномерного сетевого трафика лучшие результаты обеспечивает $K \approx 0.8$, при котором достоверность классификации приближается к 99%.
6. Разработан и реализован ПК, включающий сервер баз данных, сервер приложений, Web-приложение и ПО для мобильных устройств под управлением ОС Android. Он позволяет: собирать с мобильных устройств пакеты сетевого трафика и сохранять их в БД; группировать пакеты сетевого трафика в потоки; по запросу пользователя формировать наборы данных с заданными характеристиками, автоматизировать процесс классификации мобильных приложений путём анализа сетевого трафика и др.

Полученные результаты позволяют заключить, что все поставленные задачи решены и цель диссертационного исследования достигнута.

Список публикаций

Статьи в научных журналах, входящих в перечень ВАК по специальности 2.3.6

1. **Барков, В. В.** Классификации противоправных и нежелательных мобильных приложений с помощью модифицированного алгоритма Adaptive Random Forest в условиях смены концепта // Вестник Санкт-Петербургского государственного университета технологии и дизайна: Серия 1. Естественные и технические науки. 2024. № 2. С. 73–80.
2. **Барков, В. В.** Повышение эффективности классификации противоправных и нежелательных мобильных приложений с использованием автокодировщиков // Вестник Санкт-Петербургского государственного университета технологии и дизайна: Серия 1. Естественные и технические науки. 2024. № 1. С. 95–99.
3. Шелухин, О. И., **Барков, В. В**., Маторин, Ф. А. Повышение эффективности классификации противоправных и нежелательных приложений в условиях фонового трафика с помощью автокодировщиков // Вестник Санкт-Петербургского государственного университета технологии и дизайна: Серия 1. Естественные и технические науки. 2023. № 3. С. 159–165.
4. Шелухин, О. И., **Барков, В. В.**, Симонян, А. Г. Обнаружение дрейфа концепта при классификации мобильных приложений с использованием автокодировщиков // Наукоемкие технологии в космических исследованиях Земли. 2023. Т. 15. № 3. С. 20–29.
5. Шелухин, О. И., **Барков, В. В**., Полковников, М. В. Сравнительный анализ алгоритмов оценки количества и структуры атрибутов в задачах классификации мобильных приложений // Наукоемкие технологии в космических исследованиях Земли. 2019. Т. 11. № 2. С. 90–100.

Статьи в научных журналах, входящих в Scopus

6. Sheluhin, O. I., Erokhin, S. D., Osin, A. V., **Barkov**, **V. V.** Experimental Studies of Network Traffic of Mobile Devices with Android OS // Systems of Signals Generating and Processing in the Field of on Board Communications. 2019.

Свидетельства о государственной регистрации программы для ЭВМ

7. Шелухин, О. И., Ерохин, С. Д., **Барков, В. В**. Программный комплекс для онлайн классификации сетевого трафика // Свидетельство о государственной регистрации программы для ЭВМ № 2019615330 от 24 апреля 2019 г.

Публикации в научно-рецензируемых журналах и материалах конференций

8. Sheluhin, O. I., **Barkov, V. V.**, Sekretarev, S. A. The Online Classification of the Mobile Applications Traffic Using Data Mining Techniques. // T-Comm. 2019. vol. 13. no.10. pp.60-67.
9. Шелухин, О. И., Ерохин, С. Д., **Барков, В. В.** Создание базы данных сетевого трафика для автоматизации классификации мобильных приложений под управлением операционной системы Android // Нейрокомпьютеры: разработка, применение. 2019. №1. С.40-51.
10. Шелухин, О. И., **Барков, В. В.**, Полковников, М. В. Классификация зашифрованного трафика мобильных приложений методом машинного обучения // Вопросы кибербезопасности. 2018. № 4(28). С.21-28.
11. Sheluhin, O. I., **Barkov, V. V.** Influence of Background Traffic on the Effectiveness of Mobile Applications Traffic Classification Using Data Mining Techniques // T-Comm. 2018. vol.12, no.10. pp.52-57.
12. Шелухин, О. И., **Барков, В. В.**, Секретарев, С. А. Алгоритмы обнаружения дрейфа концепта при потоковой классификации трафика мобильных приложений // REDS: Телекоммуникационные устройства и системы 2020. №3. С.19-27.
13. Sheluhin, O. I., **Barkov, V. V.**, Polkovnikov, M. V. Classification of encrypted Applications of Traffic Mobile Devices using the Data Mining // Proceedings of the International Conference Technology & Entrepreneurship in Digital Society (TEDS): Proceedings of the International Conference, Moscow, 07 ноября 2018 года. Moscow: Издательский дом "Реальная экономика", 2019. p.98-101
14. **Барков, В. В.**, Секретарев, С. А. Классификация мобильных приложений в реальном масштабе времени методом машинного обучения // Технологии информационного общества: Материалы XIII Международной отраслевой научно-технической конференции, Москва, 20–21 марта 2019 года. Том 1. Москва: ООО "Издательский дом Медиа паблишер", 2019. С.323-325
15. Шелухин, О. И., **Барков, В. В.** Экспериментальные исследования и создание базы данных сетевого трафика мобильных устройств под управлением операционной системы Android // Фундаментальные проблемы радиоэлектронного приборостроения: «INTERMATIC-2018». М.: МИРЭА. 2018. Т.18. №4. С.1011-1017.
16. Шелухин, О. И., **Барков, В. В.** Методы сбора сетевого трафика с мобильных устройств под управлением операционной системы Android с целью классификации по типам приложений // Сб. тр. XII Межд. отраслевой науч.-тех. конф. «Технологии информационного общества» (14-15 марта 2018 г., Москва). М.: МТУСИ. Т.2. С.20-21.
17. **Барков, В. В.** Проектирование и разработка экспертно-аналитической системы "Система анализа трафика" для исследования алгоритмов классификации трафика мобильных устройств под управлением операционной системы Android // Безопасные информационные технологии: Сб. тр. 9-й всерос. науч.-тех. конф. М.: МГТУ им. Н.Э. Баумана, 2018. С.2-13.
18. Шелухин, О. И., **Барков, В. В.** Разработка инфраструктуры для классификации сетевого трафика мобильных приложений с применением алгоритмов машинного обучения // Телекоммуникационные и вычислительные системы - 2017. Тр. межд. научно-тех. конф. М.: МТУСИ, 2017. С.180-181.