В Диссертационный совет 24.2.385.09
при федеральном государственном
бюджетном образовательном учреждении
высшего образования
«Санкт-Петербургский
государственный университет
промышленных технологий и дизайна»

ОТЗЫВ
Официального оппонента д.т.н., доцента Лавровой Дарьи Сергеевны на диссертацию Баркова Вячеслава Валерьевича «Классификация противоправных и нежелательных мобильных приложений методами машинного обучения в потоковом режиме», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.6. Методы и системы защиты информации, информационная безопасность.

1. Актуальность темы исследования

Задача выявления мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, приобретает особую актуальность в связи с активным развитием мобильных устройств. 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, посвященных проблеме классификации приложений на основе анализа сетевого трафика, слабо учитывается требование выявления неизвестного сетевого трафика. При проектировании моделей классификации приложений оно полностью исключается в предположении наличия только известных классов, обучение осуществляется на данных из ограниченного числа классов приложений, а тестирование – 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.

Обобщая вышеизложенное, диссертационная работа Баркова Вячеслава Валерьевича, направленная на повышение эффективности классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, на основе анализа сетевого трафика методами машинного обучения в потоковом режиме в условиях априорной неопределённости является *актуальной*.

Целью 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. Для достижения поставленной цели диссертантом предложено использовать искусственные нейронные сети (ИНС), в частности, автокодировщики (АК).

Объектом исследования в диссертационной работе является сетевой трафик, генерируемый мобильными приложениями.

Предметом исследования являются методы ML для классификация мобильных приложений на основе анализа сетевого трафика в условиях априорной неопределённости числа приложений.

2. Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Обоснованность первого положения, выносимого диссертантом на защиту, подтверждается обоснованием применения методики отбора атрибутов на основе анализа их информативности при фиксировании допустимой вероятности ложной классификации. При этом определены ограничения как по числу потоков, так и по числу пакетов в потоке, в то время как общепринятые подходы предполагают расчет характеристик на основе данных всего потока.

Обоснованность второго положения подтверждается новизной предложенного модифицированного алгоритма классификации мобильных приложений в условиях неконтролируемого фонового трафика, обладающего более высокой эффективностью и отличающегося от известных алгоритмов каскадным включением нейронной сети с архитектурой АК.

Обоснованность третьего положения подтверждается использованием хорошо апробированных моделей обнаружения смены концепта при классификации мобильных приложений на основе анализа сетевого трафика, отличающаяся от известных включением АК в качестве базовой модели обнаружения смены концепта.

Обоснованность четвёртого положения подтверждается новизной предложенного алгоритма обнаружения смены концепта мобильных приложений в потоковом режиме с обработкой в скользящем окне в режиме накопления с «конечной памятью». Алгоритм отличается от известных учетом «старения» данных в окне обработки и негауссовским характером изменяющихся параметров классифицируемых приложений.

Обоснованность пятого положения подтверждается новизной предложенного модифицированного алгоритма Adaptive Random Forest (MARF) со встроенной моделью обнаружения смены концепта, что позволяет осуществлять классификацию анализируемых приложений быстрее, чем известные алгоритмы.

3. Достоверность и новизна научных положений, выводов и рекомендаций

Достоверность результатов первого и второго научных положений подтверждается строгостью применяемого математического аппарата и результатами имитационного моделирования.

Достоверность результатов третьего научного положения подтверждается результатами компьютерного моделирования в среде Python при оценке эффективности предложенных моделей.

Достоверность результатов четвёртого научного положения подтверждается строгостью применяемого математического аппарата, результатами компьютерного моделирования в среде Python при оценке эффективности алгоритма.

Достоверность результатов пятого научного положения подтверждается строгостью применяемого математического аппарата, результатами компьютерного моделирования в среде Python при оценке эффективности алгоритма MARF.

Научная новизна состоит в следующем:

1) Предложена методика отбора значимых атрибутов классификации мобильных приложений на основе анализа сетевого трафика, обеспечивающая высокую достоверность классификации приложений. Отличительным признаком новизны методики является её инвариантность по отношению к разным типам сетевого трафика.

2) Предложен алгоритм классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного или вредоносного контента, состоящий из последовательно включённых автокодировщиков (для каждого приложения) и типовой модели классификации. Отличительным признаком новизны алгоритма является то, что он не требует разметки фоновых приложений в случае их внезапного появления. Алгоритм обеспечивает повышение достоверности классификации приложений по сравнению с известными алгоритмами в условиях априорной неопределенности и неконтролируемого фонового трафика.

3) Предложена модель обнаружения смены концепта классифицируемых мобильных приложений. Отличительный признак новизны модели состоит в использовании АК, что позволяет повысить точность обнаружения смены концепта в потоковом режиме.

4) Предложен 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. Отличительным признаком новизны алгоритма является учёт «старения» данных в окне обработки и негауссовским характером изменяющихся параметров классифицируемых приложений.

5) Предложен алгоритм MARF, обеспечивающий обнаружение смены концепта на этапе предсказания. Отличительный признак новизны алгоритма состоит в использовании предложенной модели обнаружения смены концепта как на этапе обучения, так и на этапе предсказания.

Перечисленные результаты диссертационного исследований являются новыми и достоверными.

4. Теоретическая и практическая значимость результатов работы

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.

Практическая значимость результатов работы заключается в разработке алгоритмов и реализации программного комплекса для классификации мобильных приложений на основе анализа сетевого трафика в потоковом режиме в условиях смены концепта для предотвращения распространения противоправного, вредоносного и нежелательного контента. Сформированная экспериментальная база данных сетевого трафика мобильных приложений 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.

Практические результаты диссертационных исследований внедрены в АО «Лаборатория Касперского» при разработке межсетевых экранов, а также в учебный процесс МТУСИ, что подтверждается соответствующими актами внедрения.

5. Апробация работы и публикации по диссертации

Основные результаты диссертационных исследований обсуждены и одобрены на 7 конференциях и опубликованы в 18 научных печатных работах, в том числе: 5 – в рецензируемых научных изданиях, рекомендованных ВАК Минобрнауки России; 1 – в научных рецензируемых изданиях по базе Scopus; 11 – в материалах конференций и других изданиях. Получено свидетельство о Государственной регистрации программы для ЭВМ. Среди опубликованных работ две написаны лично диссертантом. Исходя из апробации результатов исследований, можно сделать вывод об их качественном обсуждении и апробации.

6. Соответствие работы паспорту научной специальности

Диссертация соответствует двум пунктам паспорта специальности 2.3.6 «Методы и системы защиты информации, информационная безопасность (технические науки)» ВАК Министерства науки и высшего образования РФ:

п. 13. «Методы и модели выявления и противодействия распространению ложной и вредоносной информации»;

п. 15. «Принципы и решения (технические, математические, организационные и др.) по созданию новых и совершенствованию существующих средств защиты информации и обеспечения информационной безопасности».

7. Оценка содержания автореферата и текста диссертации

Диссертация содержит введение, четыре главы, заключение, список использованных источников, содержащий 118 наименования, и пять приложений. Основной текст работы изложен на 107 страницах, проиллюстрирован 25 рисунками и 26 таблицами.

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 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), *F*1-мера.

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. Предложена методика отбора рационального числа атрибутов по максимуму их информативности, фиксируя допустимую вероятность ложной классификации. Для повышения эффективности классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента в условиях фонового трафика предложено использовать ИНС – АК, выполняющий роль предварительного фильтра. 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, а также отсутствие дополнительных механизмов для определения ФТ и повторного обучения при возрастании числа фоновых приложений.

Третья глава посвящена разработке модели обнаружения смены концепта (МОСК) с применением АК, который обучается на статистике данных нормального сетевого трафика. 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. Для учета эффекта «старения» данных в потоковом режиме предложена МОСК, в которой изменение текущих статистических характеристик атрибутов сетевого трафика мобильных приложений определяется с помощью двух перемещающихся во времени окон. 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. 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. 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. Указанный подход был реализован путем модификации известного алгоритма ARF. Для обнаружения смены концепта в MARF используется MOCK 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 программный комплекс «Система анализа трафика» (ПК САТ), позволяет автоматизировать процесс сбора сетевого трафика и проведение исследований эффективности моделей классификации сетевого трафика.

Структура диссертации логична и последовательна. В работе в достаточной степени раскрыты вопросы актуальности и новизны темы, теоретической и практической значимости исследования. Чётко сформулированы цели и задачи работы, а также обоснованы основные положения, выносимые на защиту. В рамках исследования решена научная задача классификации мобильных приложений, осуществляющих распространение противоправного, вредоносного и нежелательного контента.

Автореферат диссертации полностью отражает основные положения диссертационной работы

8. Замечания по диссертации

1) В главе 2 при описании условий эксперимента и полученных экспериментальных данных не указано, как обрабатываются TCP-сессии, для которых потеряна фаза установления соединения.

2) В главе 2 в таблице 2.3 на странице 44 приведено описание 23-х атрибутов, описывающих поведение анализируемых приложений. Однако в дальнейшем в тексте написано, что количество атрибутов уменьшилось и стало равным 21. Осталось неясным, какие атрибуты и по какой причине были исключены из дальнейшего рассмотрения.

3) В главе 3 в пункте 3.1.4 на странице 82 при описании полученного автором алгоритма указаны пороги экземпляра, группы и подсчёта, однако не представлены формулы, по которым эти пороги могут быть рассчитаны.

4) В пункте 3.2 на странице 84 при описании решающего правила для обнаружения смены концепта с учетом «старения» обрабатываемых данных приведены численные значения коэффициентов затухания α_1 и α_2, но отсутствует обоснование их выбора.

Отмеченные недостатки носят частный характер и, в целом, не влияют на высокое качество представленной на отзыв диссертационной работы.

9. Заключение о соответствии диссертации критериям, установленным Положением о присуждении учёных степеней

Диссертационная работа Баркова Вячеслава Валерьевича «Классификация противоправных и нежелательных мобильных приложений методами машинного обучения в потоковом режиме» по актуальности, научной новизне, теоретической и практической значимости соответствует требованиям п. 9-14 «Положения о присуждении ученых степеней» ВАК Минобрнауки России, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук, так как является научно-квалификационной работой, в которой изложены научно обоснованные технические и программные решения в области классификации мобильных приложений, осуществляющих распространение противоправного, нежелательного и

вредоносного контента, имеющие существенное значение для развития систем обеспечения информационной безопасности страны, использующих методы интеллектуального анализа данных.

Тема и содержание диссертации «Классификация противоправных и нежелательных мобильных приложений методами машинного обучения в потоковом режиме» полностью соответствует выбранной специальности 2.3.6. «Методы и системы защиты информации, информационная безопасность (технические науки)».

Барков Вячеслав Валерьевич заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.6. Методы и системы защиты информации, информационная безопасность (технические науки).

Официальный оппонент:
Лаврова Дарья Сергеевна,
доктор технических наук (специальность 05.13.19. Методы и системы защиты информации, информационная безопасность), доцент,
профессор Высшей школы кибербезопасности
Института компьютерных наук и кибербезопасности,
Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

195251, г. Санкт-Петербург,
ул. Политехническая, д. 29, литера Б
Тел.: +7 (950) 034-95-78
E-mail: lavrova@ibks.spbstu.ru

15 ноября 2024 г.