ОТЗЫВ

официального оппонента Аникина Игоря Вячеславовича на диссертационную работу Куртуковой Анны Владимировны «Методика и программный комплекс для идентификации автора программного кода» на соискание ученой степени кандидата технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность

Актуальность диссертационного исследования обусловлена сложностью решения множества задач, возникающих при идентификации авторства программного кода. Существующие подходы демонстрируют недостаточную устойчивость к атакам на методы идентификации (обфускация, оптимизация, имитация стиля). Большинство методов разработано для закрытого множества авторов, что не соответствует реальным условиям анализа инцидентов с участием неизвестных субъектов. Отсутствуют унифицированные подходы к обработке различных представлений кода (исходный код, дизассемблированный код, абстрактное синтаксическое дерево, граф потока управления), что ограничивает применимость методов в различных сценариях.

Разработка комплексного решения, обеспечивающего высокую точность идентификации автора программного кода в условиях современной программной разработки, необходима для повышения эффективности реализации мер по обеспечению кибербезопасности, в том числе противодействия внедрению недекларированных возможностей, защите от вредоносного ПО, несанкционированным модификациям и целевым искажениям процессов сборки и распространения ПО. Данное решение также позволит повысить эффективность реализации процессов в области цифровой криминалистики и реагирования на инциденты информационной безопасности.

Основное содержание диссертационной работы

Диссертационная работа объемом 178 страниц включает введение, 5 глав, заключение, список литературы из 114 наименований и 2 приложения. Структура работы соответствует поставленной цели и задачам, материал изложен целостно и аргументированно.

Первая глава представляет систематизированный обзор методов идентификации авторства программного кода, рассматриваемых как для исходных кодов, так и для бинарных артефактов. Прослеживается эволюция подходов от статистических метрик через лексико-синтаксические методы к нейросетевым решениям. Выделены недостатки существующих подходов, среди которых чувствительность к обфускации, проблемы масштабируемости, слабая калибровка уверенности моделей, а также атаки на методы идентификации авторства.

Вторая глава описывает формирование репрезентативного набора данных. На основе исходных кодов построены альтернативные представления: абстрактные синтаксические деревья, дизассемблированные коды, графы потока управления. При создании набора учтены особенности современных практик разработки ПО и атаки на методы идентификации авторства программного кода. Процесс сбора данных хорошо документирован с использованием UML-диаграмм и четких критериев отбора.

В третьей главе представлен основной научный результат – методика идентификации авторства на основе multi-view подхода. Архитектура основана на

специализированных энкодерах для каждого представления кода. Векторные представления используются для решения трех задач: идентификация в закрытом множестве, в открытом множестве и верификация автора. При верификации применена калибровка вероятностей посредством изотонической регрессии. Предложен иерархический подход с трехступенчатой схемой идентификации для масштабирования на большое число авторов.

Четвертая глава содержит экспериментальную оценку методики. Выполнены эксперименты и получены оценки метрик для различных сочетаний представлений ПО.

Пятая глава описывает практическую реализацию программного комплекса, реализующего полученные результаты.

В заключении подведены итоги проведенного исследования, сформулированы общие выводы и рекомендации.

Каждая глава диссертации завершается краткими выводами, отражающими основные результаты соответствующего этапа работы. Содержание диссертации логично структурировано, материал изложен последовательно.

Научная новизна

К основным научным результатам, обладающим новизной, относятся:

1. Методика идентификации автора программного кода, отличающаяся использованием multi-view подхода к получению векторных представлений, позволяющая определять автора в случаях открытого и закрытого множества авторов-кандидатов.

2. Метод верификации авторства программного кода, отличающийся применением сиамской нейросетевой архитектуры с контрастивным обучением и адаптивной калибровкой прогностических вероятностей, позволяющий определять принадлежность программного кода конкретному автору.

3. Алгоритмическое обеспечение для экстракции и кодирования четырех видов представлений программного кода (исходный код, дизассемблированный код, абстрактное синтаксическое дерево, граф потока управления), позволяющее верифицировать и идентифицировать его автора.

Теоретическая и практическая значимость

Теоретическая значимость работы заключается в том, что автором предложена новая методика идентификации авторства программного кода, основанная на едином multi-view энкодере для различных представлений кода. Интеграция различных представлений программного кода позволяет унифицировать решение задач в сценариях закрытого и открытого множества, а также верификации авторства, что расширяет теоретическую базу анализа программных артефактов и способствует выявлению общих закономерностей между признаками на различных уровнях анализа кода. Разработан иерархический подход к идентификации автора программного кода, обеспечивающий высокую достоверность идентификации среди большего числа возможных авторов.

Практическая значимость результатов диссертации обусловлена тем, что

1. Сформирован репрезентативный набор данных, включающий четыре вида представлений программного кода и содержащий 255 тыс. образцов кода 3811 авторов, а также учитывающий случаи командной разработки, следования стандартам кодирования, искусственной-генерации, имитации стиля и обфускации.

2. Предложен метод идентификации автора программного кода в закрытом множестве, достигающий точности в среднем 0,97 на 50 авторах для 13 различных языков программирования (на 0,04 выше аналогов) в простых случаях и до 0,93 в случаях атак на методы идентификации автора программного кода.

3. Разработан метод идентификации автора программного кода в открытом множестве, достигающий метрики 0,68 по F1-макро и 0,85 по F1@10 на 500 авторах-кандидатах (на 0,12-0,18 выше в сравнении с одиночными представлениями).

4. Предложен метод верификации автора программного кода, применяющий калибровку прогностических вероятностей, достигающий точности 0,98 (на 0,02 выше аналогов).

5. Разработано программное обеспечение, позволяющее извлекать и кодировать четыре вида представлений программного кода, идентифицировать и верифицировать автора программного кода в простых и сложных случаях, включая обфускацию, оптимизацию и командную разработку.

Результаты диссертационной работы использованы в рамках выполнения 8 проектов, где соискатель участвовал как исполнитель либо руководитель. Результаты диссертационной работы внедрены в процессы работы компаний ООО «ТомскСофт» и ООО «НТР Томск» в г. Томске, а также в учебный процесс ТУСУРа, что подтверждено тремя актами о внедрении. Положительный эффект от внедрения подтверждает практическую значимость полученных результатов.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность

Полученные в диссертации научные результаты и выводы являются обоснованными и достоверными. Это обеспечено корректным применением соискателем современных методов исследований - машинного и глубокого обучения, математической статистики, вычислительного эксперимента, объектно-ориентированного программирования. Автор провел обширные экспериментальные исследования, что позволило всесторонне проверить работоспособность предложенных методов. Использованы стандартные метрики оценки качества классификации, выполнено их сравнение с результатами, полученными другими авторами. В диссертации прослеживается согласованность полученных результатов с выдвинутыми гипотезами и целями исследования. Работоспособность предложенных решений подтверждена на масштабном репрезентативном наборе данных.

Полнота опубликования результатов работы

Основные результаты диссертационного исследования опубликованы в достаточном объеме и соответствуют установленным требованиям ВАК.

Основные положения опубликованы в 7 статьях в журналах, рекомендованных ВАК РФ, в 6 публикациях, включенных в реферативные базы данных WOS и Scopus. Соискателем получено 1 свидетельство о регистрации программы для ЭВМ и 2 свидетельства о регистрации баз данных.

Соответствие содержания диссертации паспортам научной специальности

Содержание диссертации соответствует следующим пунктам паспорта специальности 2.3.6:

- п. 13: Методы и модели выявления и противодействия распространению ложной и вредоносной информации.

- п. 5: Методы, модели и средства (комплексы средств) противодействия угрозам нарушения информационной безопасности в открытых компьютерных сетях, включая Интернет.

Замечания

1. В диссертационной работе представлен трехступенчатый иерархический подход к идентификации автора программного кода, обладающий значительной научной новизной и множеством преимуществ. Однако, в явном виде данный метод не заявлен ни в научной новизне диссертационной работы, ни в положениях, выносимых на защиту. Включение данного результата в научную новизну и в положения, выносимые на защиту, только усилило бы полученные результаты.

2. В разделе 3.2 диссертационной работы описываются гиперпараметры обучения. Возникает вопрос - как выбирались конкретные значения гипермараметров? Проводился ли поиск лучших значений? Насколько чувствительны к ним полученные результаты?

3. В п.3. научной новизны и п.3. положений, выносимых на защиту, заявлено алгоритмическое обеспечение, позволяющее верифицировать и идентифицировать автора программного кода. Несомненно, в диссертационной работе данное алгоритмическое обеспечение присутствует. Однако, разработанные алгоритмы следовало бы четко перечислить и представить в более формализованном виде. Например, следовало представить блок-схемы алгоритмов, реализующих отдельные этапы декомпозиций, изображенных на рис. 3.4. – 3.8. Также, в разделе 5.2.2. многие алгоритмы описаны словесно, что ухудшает их восприятие.

4. В главе 3 следовало представить в графическом виде архитектуры применяемых нейронных сетей.

5. Раздел 4.1. следовало перенести в Главу 3 с целью обоснования выбора конкретных энкодеров в разработанной методике идентификации программного кода.

Отмеченные недостатки не снижают ценность диссертационной работы и не ставят под сомнение основные научные и практические результаты диссертационного исследования.

Заключение

Диссертационная работа А.В. Куртуковой соответствует паспорту специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Диссертация написана самостоятельно, содержит новые научные результаты и положения. Структура и содержание диссертации соответствуют поставленной цели и задачам исследования. Текст диссертации изложен грамотно, цель исследования достигнута, поставленные задачи решены. Автореферат диссертации достаточно полно раскрывает ее основное содержание и положения, выносимые на защиту. Содержание автореферата соответствует основным положениям диссертационной работы. В нём изложены все основные результаты, выносимые на защиту, дано достаточно подробное представление о научно-практической значимости работы.

Соискателем решена важная научно-техническая задача разработки методики и программного комплекса для идентификации автора программного кода, обеспечивающих высокую в сравнении с аналогами точность как в простых, так и в сложных случаях. Решение данной задачи вносит существенный вклад в развитие основ для создания безопасных программных систем и информационных технологий.

Диссертационная работа соответствует требованиям пп. 9–14 «Положения о присуждении ученых степеней» (Постановление Правительства РФ от 24.09.2013 г. № 842), предъявляемым к кандидатским диссертациям. Автор, Куртукова Анна Владимировна, заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Я, Аникин Игорь Вячеславович, даю свое согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

Официальный оппонент,
доктор технических наук, профессор, заведующий кафедрой систем информационной безопасности Казанского национального исследовательского технического университета им. А.Н. Туполева-КАИ

«28» 05 2026 Аникин Игорь Вячеславович

Диссертация на соискание ученой степени доктора технических наук защищена по специальности 05.13.19 – Методы и системы защиты информации, информационная безопасность (2018 г).

ФГБОУ ВО Казанский национальный исследовательский
технический университет им. А.Н. Туполева-КАИ (КНИТУ-КАИ)
420111, Республика Татарстан, город Казань, ул. Карла Маркса, д. 10.
Тел: +7 (843) 231-97-34
Email: anikinigor777@mail.ru

Подпись Аникина Игоря Вячеславовича заверяю

Подпись
заверяю. Начальник управления
делопроизводства и контроля