**ТВЕРЖДАЮ:**
ОУ ВО «ГУМРФ имени
ла С.О. Макарова»
ческих наук, профессор

___ С.О. Барышников
«27» января 2020 г.

# ОТЗЫВ

## ведущей организации

на диссертационную работу Салахутдиновой Ксении Иркиновны «Методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность».

## 1. АКТУАЛЬНОСТЬ ТЕМЫ ДИССЕРТАЦИИ

В диссертационной работе решается научная задача разработки и обоснования научно-методического аппарата по идентификации версий исполняемых файлов, устанавливаемых на средства вычислительной техники. Применение комбинированного подхода на основе алгоритма машинного обучения и аддитивного критерия позволяет уменьшать число ошибок неверной классификации и максимизировать эмерджентность совокупности признакового пространства. Предлагаемый способ решения обеспечивает увеличение точности идентификации в условиях наличия различных версий, большого числа различных наименований программ и ограниченности числа объектов обучающей выборки.

Использование свободно распространяемого программного обеспечения в целях расширения возможностей по осуществлению процессов анализа, управления, принятия решений, является неотъемлемой частью современных информационных систем, эксплуатируемых в различных секторах экономики.

Однако наличие неконтролируемо распространяемых версий способно оказать деструктивное воздействие на автоматизированную систему и ее информационную безопасность в силу возможных дефектов программного обеспечения, наличия недекларированных возможностей, нелегального использования интеллектуальной собственности, применения специальных программ, направленных на преодоление установленной защиты.

В связи с этим возникает необходимость решения ряда задач идентификации, верификации и валидации программного обеспечения. Существующие решения ориентированы, в основном, на отслеживание

фиксированного состояния кода программ на носителях и в оперативной памяти, что не всегда позволяет оперативно определить санкционированные модификации, изменения версий.

В связи с вышеуказанным, тема диссертации «Методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ» является актуальной, а решаемая в диссертационной работе научная задача имеет как теоретическую, так и практическую значимость.

## 2. ЛИЧНОЕ УЧАСТИЕ АВТОРА И АПРОБАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Личное участие автора в проведенных исследованиях заключается в разработке методов, основывающихся на ряде отобранных наиболее информативных и устойчивых в проявлении признаков дизассемблированных кодов программ, обеспечивающих увеличение точности идентификации исполняемых файлов, объектов информационных процессов автоматизированной системы.

Автором разработаны метод формирования сигнатур исполняемых файлов, включающий ряд отобранных наиболее информативных и устойчивых в проявлении ассемблерных команд; метод сравнения сигнатур идентифицируемых исполняемых файлов с ранее сформированными эталонными сигнатурами программ, путем применения подхода на основе алгоритма машинного обучения и аддитивного критерия и методика идентификации программного обеспечения, основанная на комбинированном анализе характеристик дизассемблированного кода программ.

Положения, выносимые на защиту, и результаты работы базируются на теории информационной безопасности, методах математической статистики, теории предпочтений, методах машинного обучения, экспериментальных методах исследования.

Результаты работы докладывались и обсуждались на Всероссийских и международных научно-технических конференциях:

- 18th, 20th Conference of Open Innovations Association FRUCT and ISPIT 2017 seminar, 2016, 2017 г.г.
- IV, VI, VII, VIII Всероссийский конгресс молодых ученых, 2015, 2017, 2018, 2019 г.г.
- 11th IEEE International Conference on Application of Information and Communication Technologies, AICT 2017, 2017 г.
- IX, X Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России», 2015, 2017 г.г.
- Региональная информатика "РИ-2016", 2016 г.
- XLVII, XLVIII Научная и учебно-методическая конференция Университета ИТМО, 2018, 2019 г.г.
- International Conference on Next Generation Wired/Wireless Networking Conference on Internet of Things and Smart Spaces NEW2AN 2018, ruSMART,

2018 г.

– 28-я научно-техническая конференция. Методы и технические средства обеспечения безопасности информации, МиТСОБИ, 2019 г.

По результатам диссертационного исследования автором опубликовано 32 работы, из них статей в журналах, рекомендованных ВАК РФ – 8, входящих в базы цитирования Web of Science и Scopus – 8, в прочих изданиях – 10, свидетельств о государственной регистрации программы для ЭВМ – 6.

Результаты диссертационного исследования реализованы при выполнении НИР:

– Проект по программе Президиума РАН № 0073-2018-0008 «Теория и распределенные алгоритмы самоорганизации группового поведения агентов в автономной миссии», 2018, 2019 г.г.

– Проект по программе Президиума РАН № 0073-2018-0007 «Разработка масштабируемых устойчивых алгоритмов построения семантических моделей больших данных и их использование для решения прикладных задач кластеризации и машинного обучения», 2018, 2019 г.г.

– НИР-ФУНД «Разработка методов интеллектуального управления киберфизическими системами с использованием квантовых технологий» №617026 (2017-2018гг.).

– НИР-ФУНД «Разработка методов создания и внедрения киберфизических систем» № 619296 (2018-2019гг.).

## 3. НОВИЗНА ИССЛЕДОВАНИЯ И ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ, ВЫВОДОВ И РЕКОМЕНДАЦИЙ

В отличие от ранее проводимых исследований, для увеличения точности идентификации исполняемых файлов, автором предложено оригинальное идентификационное признаковое пространство, состоящее из характеристик дизассемблированного кода программ. Разработан метод формирования сигнатур исполняемых файлов, позволяющий произвести отбор наиболее информативных признаков, устойчиво проявляющихся в различных программах, вне зависимости от их функциональности, размера, области применения и т.д.

Особенностью предложенного подхода является дискретизация признаков, применение на полученном множестве сигнатур исполняемых файлов бустинга деревьев решений, что позволяет снизить время и повысить точность идентификации, по сравнению с существующими подходами, рассмотренными в работе.

В работе предложена методика идентификации исполняемых файлов, основанная на комбинированном анализе характеристик дизассемблированного кода программ, где в отличие от известных, для целей идентификации применяется бустинг деревьев решений, а также используются оригинальный метод формирования уникального признакового пространства, основанный на статическом анализе характеристик дизассемблированного кода программ, и теории полезности для принятия решения.

На основе такого подхода определены направления адаптации предложенных методов в средствах защиты информации, реализуемых в качестве

технических мер по аудиту программного обеспечения, обеспечивающих автоматизированную идентификацию исполняемых файлов в соответствии с формируемым архивом легитимных или нелегитимных программ.

В диссертационной работе Салахутдиновой Ксении Иркиновны получены следующие основные результаты, обладающие научной новизной:

1. Метод формирования сигнатур исполняемых файлов, основанный на построении частотного распределения каждой из градаций выделенной характеристики исполняемых файлов, отличающийся от существующих использованием ряда отобранных наиболее информативных и устойчивых в проявлении ассемблерных команд.

2. Метод сравнения сигнатур идентифицируемых исполняемых файлов с ранее сформированными эталонными сигнатурами программ, отличающийся от известных применением комбинированного подхода использования алгоритма машинного обучения и аддитивного критерия, способствующего снижению числа ошибочных результатов классификации и обеспечивающего увеличение точности от совокупного использования признакового пространства, а также учитывающий ряд изменений в коде исполняемых файлов, что позволяет идентифицировать не рассматриваемые на этапе обучения версии программ.

3. Методика идентификации ПО, основанная на комбинированном анализе характеристик дизассемблированного кода программ, отличающаяся от известных, применением уникального сформированного признакового пространства и теории полезности для принятия решения на основе аддитивного критерия, что позволяет распознавать версии программ, ранее не задействованных в создании эталонных сигнатур исполняемых файлов.

Особенностью результатов является использование комбинированного анализа характеристик дизассемблированного кода программ, для чего диссертантом впервые выполнено:

1. Обоснование информативной модели представления идентификатора исполняемого файла в виде математического кортежа, включающего характеристики дизассемблированного кода программ, получаемые в результате обработки исполняемых файлов, для решения задач идентификации программных объектов, устанавливаемых на компьютеры пользователей автоматизированных систем.

2. Исследование свойств дизассемблированного кода исполняемых файлов формата ELF Linux операционных систем с целью выявления характеристик, применимых в решении задач идентификации программных объектов, устанавливаемых на автоматизированные рабочие места.

3. Построение признакового пространства ассемблерных команд, дающих возможность решать задачи идентификации – классификации объектов автоматизированных систем. В методе предлагается использовать десять различных признаков, т.к. было установлено, что их различающая способность отличается на различных программах.

4. Построение эталонных сигнатур программ и сигнатур идентифицируемых исполняемых файлов, используемых в качестве идентификаторов, путем формирования частотных распределений встречаемости отобранных информативных признаков.

5. Обоснование выбора алгоритма бустинга деревьев решений в качестве

основы для разработанного метода идентификации исполняемых файлов.

6. Совершенствование методов идентификации программных объектов, устанавливаемых на компьютеры пользователей автоматизированных систем с целью: организации защищенного функционирования бизнес-процессов; противодействия возникновению новых уязвимостей и реализации угроз информационной безопасности, компьютерных преступлений; выявления недобросовестных пользователей, многократно устанавливающих запрещенное программное обеспечение, пытающихся обойти меры по обеспечению информационной безопасности; анализа конкретной программы на основе версий других, схожих по функционалу программ, для выявления степени принадлежности их области назначения.

7. Анализ применимости предлагаемых в исследовании методов и методики идентификации исполняемых файлов для обработки системами аудита электронных носителей информации.

Указанные научно обоснованные технические и технологические решения способствуют защищенности автоматизированной системы, так как, предлагаемая методика идентификации, в отличие от широко известных, эффективно показывает себя на ограниченном наборе данных и способна идентифицировать исполняемый файл, не задействованный при создании архива сигнатур, имеющий новую версию или внесенные в него изменения, а также исправленные, или отсутствующие, метаданные.

## 4. ТЕОРЕТИЧЕСКАЯ И ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Полученные в диссертации научные результаты являются еще одним существенным шагом на пути совершенствования технологий идентификации исполняемых файлов, позволяющим не основываясь на методах оценки целостности и анализе метаданных встроенными средствами операционных систем, решать задачи обнаружения несанкционированно установленных объектов.

Практическую ценность результатов диссертационной работы составляют методика и методы идентификации. Предложенная методика идентификации исполняемых файлов, позволяет достигать точности идентификации равной 99,19%, а показателя бикубической меры *F-measure* – 0,99.

Использование приведенных методов, на основе комбинированного анализа характеристик дизассемблированного кода программ, дает возможность проводить мульти-классификацию с числом классов равным числу известных программ и увеличить на 18% бикубическую меру по сравнению с другими методами идентификации. Это позволяет применить результаты в системе организации аудита электронных носителей информации.

Непосредственную практическую значимость имеют:

1. Методика идентификации исполняемых файлов как часть технических мер по поддержанию установленной политики безопасности в части запрета несанкционированной установки программного обеспечения.

2. Методы реализации специализированных программных систем выявления и обнаружения факта нарушения политики безопасности при попытке в преодоления установленных мер по обеспечению защиты информации.

3. Предложенная методика и полученный в результате диссертационного исследования архив эталонных сигнатур можно использовать для реализации предметно-ориентированных средств идентификации исполняемых файлов в таких областях как: обеспечение информационной безопасности, выявление схожих по функциональной направленности программ, противодействие нелегальному использованию интеллектуальной собственности, компьютерная криминалистика.

Совокупность предложенных в исследовании методов и методики позволяет осуществить решение задачи разработки и обоснования научно-методического аппарата по идентификации исполняемых файлов, устанавливаемых на средства вычислительной техники, уменьшить число ошибок неверной классификации и максимизировать эмерджентность совокупности признакового пространства, для увеличения точности идентификации в условиях наличия различных версий, большого числа различных наименований программ и ограниченности числа объектов обучающей выборки.

Практическая значимость подтверждается корректным функционированием программного обеспечения, использующего полученные результаты (получены свидетельства о регистрации программ для ЭВМ: «Программа мониторинга Google Apps for Business» №2016614206; «Программа для стеганографического сокрытия информации в медиа файлах» №2016614207; «Программа для поиска и анализа криптоконтейнеров с носителя информации» №2016611975; «Программа для криминалистического анализа IM ICQ и Jabber» №2016614048; «Программа для аудита событий информационной безопасности на основе модели MapReduce» №2016614208; «Сравнение сигнатур исполняемых файлов» №2019619363).

## 5. ДОСТОВЕРНОСТЬ И ОБОСНОВАННОСТЬ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Обоснованность и достоверность полученных научных результатов определяется:

- научной обоснованностью приводимых выкладок и математических преобразований;
- использованием методик и математического аппарата теории вероятности, теории информации, методов машинного обучения, проверенных экспериментами;
- системным анализом описания объекта исследования, учетом сложившихся практик и опыта в информационной безопасности;
- проведением сравнительного анализа предложенного метода с существующими решениями и результатами экспериментов;
- непротиворечивостью полученных результатов известным решениям.

Описанные в диссертации методы, апробированные в экспериментах, проводимых в НИР, показывают, что предложенная методика позволяет производить мульти-классификацию исполняемых файлов с точностью 99,19% в отличии от методов, способных достигать схожих величин лишь при использовании бинарной-классификации. Измеряемый показатель бикубической меры достигает значения 0,99, что значительно (в среднем на 18%) превосходит существующие методы идентификации, основанные на мульти-классификации.

## 6. РЕКОМЕНДАЦИИ ПО ИСПОЛЬЗОВАНИЮ РЕЗУЛЬТАТОВ И ВЫВОДОВ

Проведенные в диссертации исследования могут быть продолжены в направлении расширения практических возможностей разработанных методов, алгоритмов, интеллектуальной программной системы.

Полученные в диссертационной работе научные результаты, выводы и практические рекомендации могут найти применение в различных проектах при разработке систем защиты информации, аудита электронных носителей информации, функционирующих в организациях МО РФ, МВД РФ, РАН, СПбФ АО «НПК «ТРИСТАН», АО «Эврика» и промышленности, а также в действующих организациях, реализующих основные направления задач мониторинга информационной безопасности:

– обнаружение пользователей, нарушающих правила пользования программным обеспечением путем установки нелегитимных программ на компьютеры автоматизированной системы;

– регламентации политики использования только приобретенного или разработанного программного обеспечения внутри организации;

– поиск инцидентов информационной безопасности, связанных уязвимостями, возникшими по причине наличия дефектов программного обеспечения, недекларированных возможностей, нелегального использования интеллектуальной собственности, применения специальных программ, направленных на преодоление установленной защиты.

Полученные результаты используются в образовательном процессе факультета Безопасности информационных технологий Университета ИТМО по направлениям подготовки бакалавриата 10.03.01 и магистратуры 10.04.01 по дисциплинам «Организация и управление службой защиты информации», «Теория вероятностей», «Методы цифровой обработки видеоизображений», «Управление информационной безопасностью».

## 7. СООТВЕТСТВИЕ СОДЕРЖАНИЯ ДИССЕРТАЦИИ И АВТОРЕФЕРАТА

Автореферат отражает содержание диссертационной работы.

Содержание диссертации в полной мере соответствует паспорту специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность».

Автореферат и диссертация написаны в хорошем стиле и на понятном языке, удачно структурированы.

По предоставленному библиографическому списку и прилагаемому

перечню собственных публикаций автора можно сделать вывод о том, что основные положения диссертации достаточно полно изложены в печати и апробированы на конференциях.

## 8. ОСНОВНЫЕ ЗАМЕЧАНИЯ ПО ДИССЕРТАЦИИ

В качестве недостатков работы можно отметить следующие:

1. Исследование направлено на анализ исполняемых файлов формата ELF на языке ассемблера, однако недостаточно внимания уделено особенностям структуры данного формата, а также использованию различных трансляторов ассемблера, на основе которых реализовано признаковое пространство для решения задач идентификации исполняемых файлов.
2. Не вполне понятно, в соответствии с каким принципом было выбрано признаковое пространство, состоящее из ряда различных ассемблерных команд.
3. Недостаточно акцентировано внимание на преимуществах метода бустинга деревьев решений, повлекшего его выбор для решения обозначенных в диссертационном исследовании задач, мало внимания уделено ограничениям, возникающим в результате его использования.
4. Требуется анализ ресурсоемкости и вычислительной сложности предложенных решений, что позволило бы говорить о возможности применения в программном обеспечении, реализующем идентификацию исполняемых файлов.
5. В тексте диссертации недостаточно ясно отражена практическая реализация методики идентификации исполняемых файлов, ее место в задаче обеспечения информационной безопасности.

Перечисленные замечания и недостатки не снижают научный уровень проведенных исследований и не влияют на общий положительный вывод о качестве представленной к защите диссертации.

## 9. ЗАКЛЮЧЕНИЕ ПО ДИССЕРТАЦИОННОЙ РАБОТЕ

Представленная диссертация соответствует требованиям ВАК, предъявляемым к кандидатским диссертациям, обладает научной новизной и практической значимостью.

Диссертационное исследование Салахутдиновой Ксении Иркиновны является законченной самостоятельной научно-квалификационной работой, которая вносит значительный вклад в решение актуальной задачи разработки и обоснования научно-методического аппарата идентификации исполняемых файлов, устанавливаемых на средства вычислительной техники, на основе комбинированного подхода использования алгоритма машинного обучения и аддитивного критерия, в условиях наличия различных версий, большого числа различных наименований программ и ограниченности числа объектов обучающей выборки.

Содержание и основные научные результаты соответствуют паспорту специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность». Автореферат диссертации отражает основное содержание диссертационной работы. По оформлению работа соответствует требованиям, предъявляемым к диссертациям.

На основании изложенного можно сделать вывод, что диссертация «Методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ» соответствует критериям, изложенным в пунктах 9-14 Положения «О порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 года № 842 в редакции от 01.10.2018 года, предъявляемым к кандидатским диссертациям, а ее автор Салахутдинова Ксения Иркиновна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность».

Диссертационная работа Салахутдиновой Ксении Иркиновны обсуждена на заседании кафедры комплексного обеспечения информационной безопасности Федерального государственного бюджетного образовательного учреждения высшего образования «Государственный университет морского и речного флота имени адмирала С.О. Макарова», протокол № 5 от «24» января 2020 года.

Заведующий кафедрой
комплексного обес
информационной бе
д.т.н., доцент Соколов Сергей Сергеевич

Федеральное государств ое бюджетное образовательное учреждение высшего образования «Государственный университет морского и речного флота имени адмирала С.О. Макарова» (ФГБОУ ВО «ГУМРФ имени адмирала С.О. Макарова»).
адрес: Двинская ул., д. 5/7, г. Санкт-Петербург, 198035;
тел.: (812) 748-96-92;
e-mail: otd_o@gumrf.ru;
сайт: https://gumrf.ru/