ОТЗЫВ

официального оппонента, доктора технических наук, профессора
Бурлова Вячеслава Георгиевича
на диссертационную работу Салахутдиновой Ксении Иркиновны на тему «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»

Актуальность темы

Применение открытого программного обеспечения обуславливает необходимость разработки дополнительных методов, систем и средств защиты информации. Возможные дефекты программного обеспечения, наличие не декларированных возможностей, нелегальное использование интеллектуальной собственности, применение специальных программ, направленных на преодоление установленной защиты, могут привести к росту числа уязвимостей и повлиять на информационную безопасность систем.

Качество идентификации исполняемых файлов существующими подходами по таким характеристикам как контрольная сумма, процесс выполнения программного кода, характерные последовательности байтов, характеристики байтовых и ассемблерных кодов и другие, является достаточно низким. В особенности это проявляется в случаях, когда известная нам программа подвергается какой-либо модификации со стороны разработчика при выпуске новой версии или же со стороны пользователя, как с целью придания ей новой функциональности, так и с целью обхода методов идентификации ориентированных, на отслеживание фиксированного состояния кода программ на носителях и в оперативной памяти.

Для качественного улучшения решения данной задачи могут применяться методы из области статического анализа характеристик дизассемблированного кода программ, разработанные для идентификации исполняемых файлов.

Задача идентификации программного обеспечения имеет широкую известность, однако, применение существующих решений для не вредоносных программ сталкивается с рядом трудностей различного характера.

Существующее противоречие между научно-методическим уровнем технологий идентификации исполняемых файлов и требуемым состоянием особенно остро проявляется в связи с постоянным ростом производительности и вычислительной мощности, который позволяет существенно улучшить качество идентификации за счет использования методов идентификации по статистическим характеристикам распределения ассемблерных команд.

В связи с этим актуальными являются работы, направленные на изучение и разработку методов идентификации не вредоносных исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ, позволяющих увеличить точность идентификации.

Степень обоснованности и достоверность научных положений, выводов и рекомендаций

Обоснованность и достоверность полученных научных результатов определяется:

– корректностью постановки задачи исследования и введенными ограничениями;

– использованием апробированного математического аппарата;

– серией практических экспериментов и сравнительным анализом с существующими методами идентификации;

– обоснованностью применения апробированных методов исследования, согласованностью результатов, полученных при теоретическом и исследовании, с результатами экспериментов;

– практической апробацией при реализации научно-исследовательских работ, а также одобрением на научно-технических конференциях.

Достоверность результатов подтверждается также корректным функционированием программного обеспечения (получены свидетельства о регистрации программ для ЭВМ), использующего полученные результаты, разработанного в ходе проведения научно-исследовательской работы «Теория и распределенные алгоритмы самоорганизации группового поведения агентов в автономной миссии» для проекта по программе Президиума РАН № 0073-2018-0008.

Оценка сущности и содержания диссертации

В настоящее время известны различные подходы к идентификации программ, основанные на анализе динамики исполнения программного кода в изолированной среде, а также выделении отличительных последовательностей, присущих конкретному виду программного обеспечения. Сегодня идентификация на основе статичного характера анализа программного кода исполняемых файлов остается достаточно перспективной. В этом направлении наиболее распространенными являются выделение признаков из кода программ, представленных на языках высокого и низкого уровней программирования, а также в бинарном представлении. В данной работе предложено использование уникальной комбинации ассемблерных характеристик не вредоносных исполняемых файлов. В основе решения задачи идентификации лежат методы математической статистики, машинного обучения и теория полезности.

Подход автора к изучению рассматриваемой в основании проблемы отличается новизной и продуктивностью. Структуру представленной диссертации отличает продуманность и логичность изложения. Текст диссертации отражает ход и результаты исследования, состоит из введения, четырех глав, заключения, списка литературы, включающего 101 источник.

Во введении обоснован выбор темы исследования и ее актуальность, представлена степень разработанности темы, определены объект, предмет и цель исследования; описаны частные задачи, обоснована теоретическая и практическая значимость получаемых результатов; раскрыты принципы используемых подходов и разработанной методики; сформулированы положения, выносимые на защиту и проведена апробация результатов исследования.

В первой главе представлен анализ существующей проблематики современного подхода к идентификации программного обеспечения. На основе выделения объекта защиты, цели реализации угроз, потенциальных уязвимостей и видов ущерба разработаны модель угроз и нарушителя информационной безопасности при обработке информации конфиденциального характера в информационной системе.

Во второй главе произведен обзор современных решений в области идентификации исполняемых файлов, представленных в отечественной и зарубежной литературе. Проанализированы различные подходы к сбору характеристик файлов и методы сравнения наборов данных характеристик для нескольких файлов. Выявлены достоинства и недостатки данных методов. Сформулирована постановка задачи обеспечения безопасности автоматизированных систем от потенциальных угроз со стороны нелегитимно установленного программного обеспечения.

В третьей главе произведен анализ структуры и характеристик ELF-файла, его дизассемблирование и представление исходного кода на низкоуровневом языке ассемблера. Описаны подход к представлению программного обеспечения и модель представления исполняемого файла в настоящем исследовании. Произведено выделение признакового пространства и описано его дальнейшее использование в различных методах формирования сигнатур.

Представлено подробное описание разработанных методов формирования сигнатур. Сформулированы методы сравнения сигнатур. Описана методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ, включающая в себя также и этап постобработки результатов сравнения сигнатур для увеличения точности идентификации исполняемых файлов. Представлены ограничения, накладываемые на методику и условия ее использования.

В четвертой главе с целью проверки точности идентификации исполняемых файлов при помощи разработанной методики была проведена серия экспериментов, направленных на каждый разработанный метод сравнения сигнатур файлов в отдельности. Полученные результаты прошли сравнение с результатами других исследователей и на данном основании сделан вывод о том, что в условиях наличия множества классов и ограниченности объема обучающей выборки, разработанная методика обеспечивает более высокую

точность идентификации. Определено, что на практике методика идентификации программного обеспечения может быть применена для повышения безопасности информационной системы организации, путем внедрения ее в качестве технической меры по аудиту программного обеспечения на электронных носителях информации, она позволяет осуществлять автоматизированную идентификацию ELF-файлов в соответствии с формируемыми архивами легитимных или нелегитимных программ. Выделен ряд смежных задач, решаемых разработанной методикой.

В заключении приведены основные результаты и выводы по разработанной методике идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ.

Оценка новизны

Научная новизна рассматриваемой диссертационной работы заключается в следующем:

– разработан метод формирования сигнатур исполняемых файлов, характеризующийся использованием ряда отобранных наиболее информативных и устойчивых в проявлении ассемблерных команд;

– представлен метод сравнения сигнатур идентифицируемых исполняемых файлов с ранее сформированными эталонными сигнатурами программ, использующий комбинированный подход соединяющий алгоритм машинного обучения и аддитивного критерия;

– разработана методика идентификации исполняемых файлов, сочетающая в себе и проводящая итерационный процесс описанных выше методов формирования и сравнения сигнатур, увеличивающая точность идентификации программ, ранее не задействованных в создании эталонных сигнатур исполняемых файлов.

Предложенные методы и методика отличаются от известных, использованием оригинального признакового пространства, оригинального метода формирования сигнатур исполняемых файлов, построения модели классификации на основе бустинга деревьев решений и постобработки результатов при помощи аддитивного критерия Фишберна, что позволяет увеличить точность идентификации исполняемых файлов.

Автореферат правильно отражает содержание диссертации. В нем отражены основные научные положения, результаты, выводы и рекомендации, полученные в рамках диссертационного исследования.

Теоретическая и практическая значимость

Теоретическая значимость результатов научного исследования заключается в возможности дальнейшего развития разработанного научно-методического аппарата в области обеспечения информационной безопасности автоматизированной сети, а также повышении гибкости и эффективности

систем аудита и обнаружения несанкционированно устанавливаемого программного обеспечения.

Практическая значимость работы состоит в широком использовании результатов диссертационной работы в различных научно-исследовательских работах, а также дисциплинах по соответствующим направлениям подготовки бакалавриата и магистратуры.

Теоретическая и практическая значимость полученных результатов подтверждается рядом публикаций в рецензируемых изданиях, свидетельствами о регистрации программного обеспечения для ЭВМ, актами о внедрении и реализации результатов диссертационного исследования в различных образовательных, научных учреждениях и коммерческих организациях.

Замечания

Отмечая достаточно высокий научный и прикладной уровень рецензируемой работы, можно выделить следующие замечания и рекомендации:

1. Недостаточно внимания уделено описанию рассматриваемых в работе информационных процессов.

2. Требует пояснения какой критерий идентификации исполняемых файлов используется, возможно ли установить некоторый пороговые значения.

3. Недостаточно внимания уделено разработке архива эталонной информации об исполняемых файлов (сигнатур) и организации взаимодействия с ней.

4. При рассмотрении предложений, связанных с использованием полученных результатов в системах мониторинга средств вычислительной техники, не рассмотрен вопрос подделки формируемой сигнатуры исполняемого файла злоумышленником, хотя это является важным для корректного функционирования средств мониторинга.

5. В исследовании слабо освещена оценка затрат вычислительных ресурсов при использовании предлагаемых подходов. Повышение ряда показателей точности идентификации может приводить к лавинообразному росту объема баз данных шаблонной информации, что вызывает необходимость описания ограничений, накладываемых на предлагаемые решения.

Отмеченные недостатки не меняют положительной оценки диссертационной работы в целом.

Заключение

Диссертационная работа, выполненная автором самостоятельно на высоком научно-техническом уровне, представляет законченный научно-исследовательский труд, содержащий решение актуальной научной задачи, отличается научной новизной и практической значимостью.

Диссертационная работа Салахутдиновой Ксении Иркиновны «Методика идентификации исполняемых файлов на основе статического анализа

характеристик дизассемблированного кода программ» полностью соответствует требованиям п.9-14 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24.09.2013 года № 842, предъявляемым к кандидатским диссертациям.

Считаю, что Салахутдинова Ксения Иркиновна заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность».

Официальный оппонент:
доктор технических наук, профессор кафедры информационных технологий и систем безопасности Федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный гидрометеорологический университет»

«27» _января_2020 г. Бурлов Вячеслав Георгиевич

Сведения о составителе отзыва.
ФИО: Бурлов Вячеслав Георгиевич
Ученая степень: доктор технических наук
Место работы: Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский государственный гидрометеорологический университет»
Должность: профессор кафедры информационных технологий и систем безопасности
Почтовый адрес: 192007, г. Санкт-Петербург, Воронежская улица, 79,
Телефон: (812) 633-01-86, +7 911 100 41 01.
Эл.почта: burlovvg@mail.ru