ОТЗЫВ

официального оппонента, доктора технических наук, профессора
Примакина Алексея Ивановича
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

Стремительный прогресс повсеместного внедрения информационных технологий во все сферы человеческой деятельности приводит к полной зависимости современной организации бизнес процессов от надлежащего функционирования информационных систем. При взаимодействии пользователя с информационной системой организации, ее ресурсами и обрабатываемой внутри информацией наступает потребность в регулировании такого взаимодействия, реализуемое за счет политики информационной безопасности, мер и средств защиты.

С каждым годом все более доступным становится доступ к различным страницам, сайтам сети Интернет, ориентированным на информирование интернет-пользователей о существующих уязвимостях операционных систем, программного обеспечения и эксплуатации нетривиальных способов обхода установленных на рабочих местах систем защиты информации.

Действия пользователей автоматизированных систем, направленные против установленной политики безопасности в организации, способны стать источником роста числа уязвимостей системы и повлиять на ее информационную безопасность. Таким образом, аудит электронных носителей информации на предмет выявления несанкционированно установленного программного обеспечения становится все более актуальной задачей.

Обеспечение противодействия угрозам нарушения информационной безопасности локальной сети, является важной задачей службы защиты информации, решение которой невозможно без создания эффективных методов идентификации, верификации и валидации программного обеспечения. Приводимые в исследовании решения, направленные на увеличение точности идентификации объектов автоматизированных систем, размещенных на электронных носителях информации, позволяют отслеживать несанкционированно установленные исполняемые файлы в локальной сети. Поэтому тема диссертационного исследования является весьма актуальной.

Степень обоснованности и достоверность научных положений, выводов и рекомендаций

В диссертационном исследовании Салахутдиновой К.И. для увеличения показателей точности процессов идентификации исполняемых файлов, предлагается использование ряда отобранных признаков, получаемых после приведения к дизассемблерному представлению программ.

Диссертация оформлена в соответствии с действующими требованиями ВАК и состоит из введения, четырех глав и заключения. В диссертационной работе автором приводятся выявленные отличительные черты дизассемблированного кода исполняемых файлов, связанные с наличием ряда структурных особенностей, заложенных на этапе создания программы, и направленностью их функционала.

Предлагается осуществлять итерационную и комбинированную обработку таких программ на основе признакового пространства и теории полезности. Особую ценность представляет решение, заключающееся в расчете и отборе информативных признаков для задачи идентификации (учет частотной составляющей присутствия признака в различных программах), что отличает данную работу от ряда других. Автором было обосновано применение данного подхода и экспериментально подтверждено, что для отобранного ряда ассемблерных команд различие выражается не только в форме получаемого распределения, но и порядке частоты встречаемости признака.

Выявленные в диссертационной работе закономерности дизассемблированных кодов исполняемых файлов позволяют создать профиль программы, содержащий информацию ряда ассемблерных команд. Используя методы машинного обучения, основанные на градиентном бустинге деревьев решений, диссертант предлагает методику идентификации исполняемых файлов на электронных носителях информации.

Обоснованность научных положений, рекомендаций и достоверность результатов исследования подтверждаются:

– результатами экспериментов и их сопоставлением с результатами, полученными другими исследователями изучаемой проблемы;

– согласованностью результатов, полученных на базе теоретических расчётов, с экспериментальными данными;

– корректностью постановки научной проблемы и принятых допущений и ограничений;

– использованием апробированного математического аппарата;

– практической апробацией в деятельности научно-производственных организаций и одобрением на научно-технических конференциях.

Значимость для науки и практики

Значимость для науки и практики результатов исследования заключается в следующем:

– полученные в процессе выполнения диссертационного исследования результаты направлены на решение задачи увеличения точности идентификации устанавливаемого программного обеспечения на основе методов статистического анализа, в условиях наличия различных версий, большого числа наименований программ и ограниченности числа объектов обучающей выборки;

– результаты проведенного диссертационного исследования доведены до уровня, обеспечивающего возможность их непосредственного практического использования в системах, обеспечивающих идентификацию исполняемых файлов, аудита программного обеспечения и системах обнаружения потенциальных источников уязвимостей.

Результаты диссертационной работы реализованы в процессе проведения ряда НИОКР, по результатам которых автором получены акты о реализации проведенных исследований.

Научная новизна

Новизна исследования и полученных результатов заключается в том, что лично автором впервые:

– обоснована структура отображения исполняемого файла в виде сигнатуры, содержащей статистические значения ряда идентификационных признаков;

– произведена оценка использования эталонных сигнатур программ, формируемых на основании расчета различающей способности признаков на различных видах программ и их версий;

– разработан метод идентификации, основанный на градиентном бустинге деревьев решений и последующем применении аддитивного критерия Фишберна, позволяющего достигать наименьшего числа ошибок неверной классификации и максимальной эмерджентности совокупности признакового пространства;

– разработана методика идентификации исполняемых файлов на электронных носителях информации, основанная на статическом характере анализа дизассемблированного кода программ.

Кроме того, необходимо отметить, что автором диссертационного исследования произведен анализ характеристических особенностей ассемблерных команд и дана оценка их влияния на точность идентификации исполняемых файлов, в условиях малого количества версий программ и их большого разнообразия.

Модификация и теоретическое обобщение предложенных методов предоставляют возможность совершенствования и адаптации технологий идентификации программного обеспечения.

Новыми научными результатами, полученными лично автором, являются:

1. Метод формирования сигнатур исполняемых файлов, основанный на построении частотного распределения каждой из градаций выделенной характеристики исполняемых файлов.

Отличается 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.

Отличается 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 и позволяющий идентифицировать не рассматриваемые на этапе обучения версии программ.

3. Методика идентификации программного обеспечения, основанная на комбинированном анализе характеристик дизассемблированного кода программ.

Отличается 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.

Каждый из предложенных методов имеет собственный набор отличий от аналогов.

Общая оценка диссертационной работы

В рассматриваемой диссертационной работе предложены методы и методика, большинство из которых может быть адаптировано для любой из задач идентификации, а также реализовано как часть систем мониторинга средств вычислительной техники автоматизированной системы предприятия и предотвращения возникновения сторонних уязвимостей.

Полученные автором научные результаты позволили увеличить точность идентификации исполняемых файлов.

Автором диссертационной работы в процессе решения поставленных задач частично разработано программное обеспечение, вошедшее в состав реально работающих систем, что подтверждается полученными свидетельствами о регистрации программ для ЭВМ.

Результаты диссертационной работы Салахутдиновой Ксении Иркиновны доложены на 14 научно-технических конференциях, опубликованы в 32 научных работах (в том числе ВАК РФ – 8, Scopus/WoS – 8) и в шести свидетельствах о регистрации программ.

Однако работа не лишена ряда недостатков, основными из которых, являются следующие:

1. Слабо определены границы применимости и ограничения методик и алгоритмов, предлагаемых в диссертационном исследовании.

2. Возникает вопрос о влиянии обучающей выборки, связанном с изменением функциональности, изменением порядка следования отдельных программных блоков ассемблерного кода и других характеристик исполняемых файлов, на полученные результаты.

3. Отсутствует рассмотрение динамической составляющей при анализе

исполняемых файлов, при каком проценте внесенных изменений в код программы, ее распознавание, приведенной методикой, станет невозможно.

4. Встречаются отклонения от стандартов при оформлении диссертационного исследования, орфографических, стилистических ошибок.

Отмеченные недостатки не являются доминирующими и существенно не снижают общего научного уровня диссертации Салахутдиновой К.И. Они могут послужить основой для дальнейших исследований в рамках данного научного направления.

Работа базируется на достоверной и достаточно полной статистической информации, полученной автором. Достоверность и обоснованность результатов исследования определяются применением апробированных средств и методов исследования, корректностью принятых допущений и ограничений, достоверностью исходных данных, оказывающих существенное влияние на анализ предметной области, серией расширенных экспериментов, непротиворечивостью полученных результатов и их согласованностью с результатами исследований, проведенных другими авторами по тематике, близкой к теме диссертационной работы, а также актами внедрения и публикациями в рецензируемых изданиях.

Диссертация написана хорошим литературным языком, грамотно и аккуратно оформлена. По каждой главе и работе в целом сделаны четкие выводы.

Диссертационная работа, выполненная Салахутдиновой Ксенией Иркиновной, способствует повышению показателей защищенности ресурсов автоматизированной системы за счет увеличения качественных характеристик (точности) процессов идентификации исполняемых файлов, путем использования структурных особенностей программ, их выделения, обработки и постобработки.

В диссертации изложены научно обоснованные технические и технологические решения, обеспечивающие решение задачи повышения точности идентификации исполняемых файлов, устанавливаемых на средства вычислительной техники, обеспечивающего увеличение точности идентификации в условиях наличия различных версий, большого числа различных наименований программ и ограниченности числа объектов обучающей выборки, имеющей важное значение в области информационной безопасности.

Содержание представленной диссертации соответствует специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность».

Автореферат полностью соответствует основному содержанию диссертации и составлен в соответствии с требованиями «Положения о присуждении ученых степеней».

Заключение

Диссертация Салахутдиновой Ксении Иркиновны является законченной научно-квалификационной работой, выполненной самостоятельно на хорошем научном уровне.

В диссертационной работе разработаны теоретические положения,

направленные на увеличение точности идентификации исполняемых файлов, устанавливаемых на средства вычислительной техники. Заявленные в работе цели достигнуты.

Диссертация 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, 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».

Официальный оппонент:
доктор технических наук, профессор
начальник кафедры специальных информационных технологий
Федерального государственного казенного образовательного учреждения
высшего образования «Санкт-Петербургский [illegible]истерства
внутренних дел Российской Федерации»

«22» января 2020 г. Примакин Алексей Иванович

Сведения о составителе отзыва:
ФИО: Примакин Алексей Иванович
Ученая степень: доктор технических наук
Ученое звание: профессор
Место работы: Федеральное государственное казенное
образовательное учреждение высшего образования
«Санкт-Петербургский университет Министерства
внутренних дел Российской Федерации»
Должность: начальник кафедры специальных информационных технологий
Почтовый адрес: 198206, Санкт-Петербург, ул. Летчика Пилютова, д. 1
Телефон: (812) 744-70-00
Эл. почта: a.primakin@mail.ru