ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
САНКТ-ПЕТЕРБУРГСКИЙ ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(СПИИРАН)

УТВЕРЖДАЮ
Директор СПИИРАН
профессор РАН

_ А.Л. Ронжин

«25» ноября 2019 г.

ЗАКЛЮЧЕНИЕ
Федерального государственного бюджетного учреждения науки
Санкт-Петербургского института информатики и автоматизации
Российской академии наук

Диссертационная работа «Методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ» выполнена в лаборатории интеллектуальных систем Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН).

В 2017 году Салахутдинова Ксения Иркиновна окончила Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики» (Университет ИТМО) по направлению подготовки 10.04.01 – «Информационная безопасность». Диплом об окончании магистратуры № 107824 1782536 выдан в 2017г. Университетом ИТМО.

В период подготовки диссертационной работы соискатель Салахутдинова К.И. являлась аспирантом очной формы обучения СПИИРАН.

В период подготовки диссертационной работы соискатель Салахутдинова К.И. работала в качестве тьютора в Университете ИТМО, а также в качестве младшего научного сотрудника лаборатории интеллектуальных систем в СПИИРАН по совместительству.

Научный руководитель – Лебедев Илья Сергеевич, доктор технических наук, профессор, главный научный сотрудник – руководитель лаборатории интеллектуальных систем СПИИРАН.

По результатам рассмотрения диссертации «Методика идентификации исполняемых файлов на основе статического анализа характеристик дизассемблированного кода программ» принято следующее заключение.

Актуальность работы

Свободно распространяемое программное обеспечение (ПО) является неотъемлемой частью современных информационных систем, эксплуатируемых в различных секторах экономики, оно позволяет расширить возможности по осуществлению процессов анализа, управления, принятия решений.

Применение открытого ПО 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.

В связи с этим возникает необходимость решения ряда задач идентификации, верификации и валидации программного обеспечения. Предлагаемые 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.

Данная работа направлена на решение задачи идентификации программного обеспечения, в условии изменения версий распространяемого ПО, на основе процедуры построения информативной модели в виде математического кортежа по выбранному признаковому пространству, характеристики которой позволяют найти однозначное соответствие между анализируемой последовательностью и хранящимся эталоном исполняемого файла.

Объект, предмет, цель диссертационной работы

Объектом исследования является разнообразие версий программного обеспечения и процесс выявления их несанкционированной установки. Предметом исследования методы идентификации программного обеспечения на основе статического сбора характеристик дизассемблированного кода программ. Целью работы является увеличение точности идентификации установленного программного обеспечения за счет разработки и обосновании научно-методического аппарата по 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.

Теоретическая и практическая значимость результатов

Теоретическая значимость работы состоит в разработке методов формирования и сравнения сигнатур исполняемых файлов ELF формата Linux операционных систем, а также в разработке методики их идентификации, позволяющей производить распознавание исполняемого файла как той или иной программы при существующих ограничениях на наличие различных версий программного обеспечения, большого числа различных наименований программ и ограниченности числа объектов обучающей выборки, обусловленное реальным состоянием выпускаемых версий того или иного программного обеспечения.

Практическая значимость работы состоит в максимизации точности идентификации программного обеспечения, позволяющими обнаруживать нарушения информационной безопасности при обработке конфиденциальной информации, вызванные несанкционированной установкой программ. Проведенные вычислительные эксперименты подтверждают результативность предложенной методики на реальных данных. Результаты работы могут быть использованы специалистами по информационной безопасности для проведения аудита электронных носителей информации.

Научная новизна работы

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 и позволяющий идентифицировать не рассматриваемые на этапе обучения версии программ.

3. Разработанная 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.

Степень обоснованности результатов проведенных исследований

Обоснованность и достоверность полученных результатов подтверждается использованием апробированного математического аппарата и подтверждается проведением сравнительного анализа с существующими методами; серией практических экспериментов по идентификации исполняемых файлов; проверкой адекватности положений и выводов; согласованностью результатов, полученных при теоретическом исследовании с результатами проведенных экспериментов; практической апробацией результатов исследования в докладах и публикациях на отечественных и зарубежных научных конференциях.

Основные результаты работы были представлены на следующих конференциях:

1. С 07 по 10 апреля 2015г., Санкт-Петербург, IV Всероссийский конгресс молодых ученых, 2015г. (доклад).
2. С 28 по 30 октября 2015г., Санкт-Петербург, Информационная безопасность регионов России (ИБРР-2015), 2015г. (доклад).
3. С 26 по 28 октября 2016г., Санкт-Петербург, Региональная информатика "РИ-2016", 2016г. (доклад).
4. С 18 по 22 апреля 2016г., Санкт-Петербург, 18th Conference of Open Innovations Association FRUCT and ISPIT 2016 seminar, 2016г. (доклад).
5. С 03 по 07 апреля 2017г., Санкт-Петербург, 20th Conference of Open Innovations Association FRUCT and ISPIT 2017 seminar, 2017г. (доклад).
6. С 18 по 21 апреля 2017г., Санкт-Петербург, VI Всероссийский конгресс молодых ученых, 2017г. (доклад).
7. С 17 по 20 апреля 2018г., Санкт-Петербург, VII Всероссийский конгресс молодых ученых, 2018г. (доклад).
8. С 15 по 19 апреля 2019г., Санкт-Петербург, VIII Всероссийский конгресс молодых ученых, 2019г. (доклад).
9. С 20 по 22 сентября 2017г., Москва, 11th IEEE International Conference on Application of Information and Communication Technologies, AICT 2017, 2017г. (доклад).
10. С 01 по 04 ноября 2017г., Санкт-Петербург, Юбилейная X Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России (ИБРР-2017)», 2017г. (доклад).

11. С 30 января по 02 февраля 2018г., Санкт-Петербург, XLVII Научная и учебно-методическая конференция Университета ИТМО, 2018г. (доклад).

12. С 29 января по 02 февраля 2019г., Санкт-Петербург, XLVIII Научная и учебно-методическая конференция Университета ИТМО, 2019г. (доклад).

13. С 29 по 31 августа 2018г., Санкт-Петербург, International Conference on Next Generation Wired/Wireless Networking Conference on Internet of Things and Smart Spaces NEW2AN 2018, ruSMART, 2018г. (доклад).

14. С 24 по 27 июня 2019г., Санкт-Петербург, 28-я научно-техническая конференция. Методы и технические средства обеспечения безопасности информации, МиТСОБИ, 2019г. (доклад).

Результаты, полученные в диссертации, были реализованы в рамках выполнения следующих НИР: Проект по программе Президиума РАН № 0073-2018-0008 «Теория и распределенные алгоритмы самоорганизации группового поведения агентов в автономной миссии», 2018,2019гг.; 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.). Результаты работы использовались при разработке системы мониторинга состояния внутренних сетей компании АО «НПК «ТРИСТАН». Полученные результаты используются в образовательном процессе факультета БИТ 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».

Личное участие соискателя в получении результатов, изложенных в диссертационной работе

Результаты диссертационной работы получены автором самостоятельно. Автором проведен анализ существующих методов идентификации программного обеспечения. Проанализированы условия и ограничения применения каждого из методов. Научное исследование было проведено при общем руководстве д.т.н. Лебедева И.С.

Полнота изложения материалов диссертации в работах, опубликованных соискателем

Основные результаты диссертационной работы изложены в печатных трудах в необходимой полноте: опубликовано 32 печатных работы, среди них статей в журналах, рекомендованных ВАК РФ – 8, входящих в базы цитирования Web of Science и Scopus – 8, свидетельств о государственной регистрации программы для ЭВМ – 6, в прочих изданиях – 10.

Публикации в рецензируемых изданиях ВАК

1. Бажаев Н., Давыдов А.Е., Кривцова И.Е., Лебедев И.С., Салахутдинова К.И. Подход к анализу состояния информационной безопасности беспроводной сети // Прикладная информатика - 2016. - Т. 11. - № 6(66). - С. 121-128. 0,63 п.л. / 0,13 п.л.

2. Кривцова И.Е., Салахутдинова К.И., Юрин И.В. Метод идентификации исполняемых файлов по их сигнатурам // Вестник Государственного университета морского и речного флота имени адмирала С.О. Макарова - 2016. - № 1(35). - С. 215-224. 0,71 п.л. / 0,24 п.л.

3. Салахутдинова К.И., Лебедев И.С., Кривцова И.Е. Подход к выбору информативного признака в задаче идентификации программного обеспечения // Научно-технический вестник информационных технологий, механики и оптики. 2018. - Т. 18. - № 2(114). - С. 278–285. 0,77 п.л. / 0,26 п.л.

4. Салахутдинова, К.И. Лебедев И.С., Кривцова И.Е., Сухопаров М.Е. Исследование влияния выбора признака и коэффициента (ratio) при формировании сигнатуры в задаче по идентификации программ // Проблемы информационной безопасности. Компьютерные системы. 2018. № 1. С. 136–141. 0,33 п.л. / 0,1 п.л.

5. Салахутдинова, К.И. Лебедев И.С., Кривцова И.Е. Алгоритм градиентного бустинга деревьев решений в задаче идентификации программного обеспечения // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18, № 6. - С. 1016-1022. 0,7 п.л. / 0,23 п.л.

6. Салахутдинова К.И., Малков В.В., Кривцова И.Е. Сравнительный анализ подходов к идентификации программного обеспечения // Безопасность информационных технологий. 2019. - Т. 26. - № 2. - С. 58-66. 0,4 п.л. / 0,13 п.л.

7. Салахутдинова К.И., Лебедев И.С., Кривцова И.Е., Анисимов А.С. Идентификации программного обеспечения в задаче аудита электронных носителей информации // Авиакосмическое приборостроение. 2019. - № 9. - С. 20-28. 0,54 п.л. / 0,14 п.л.

8. Салахутдинова К.И. Повышение точности идентификации программного обеспечения путем использования аддитивного критерия Фишберна // Информационные технологии. 2019. - Т. 25. -№ 10. - С. 609-614. 0,46 п.л. / 0,46 п.л.

Публикации тезисов и докладов в трудах конференций

9. Салахутдинова К.И., Овсяникова В.В., Трофимов А.А., Бессонова Е.Е., Ефремов А.А., Настека А.В. Анализ защищенности систем "Умный дом"//Региональная информатика (РИ-2014). XIV Санкт-Петербургская международная конференция «Региональная информатика (РИ-2014)». Санкт-Петербург, 29-31 октября 2014 г.: Материалы конференции. \ СПОИСУ. – СПб, 2014. - 2014. - С. 124

10. Ефремов А.А., Трофимов А.А., Настека А.В., Салахутдинова К.И., Овсяникова В.В. Защита управляющих сигналов в системе «Умный дом»//Сборник тезисов докладов конгресса молодых ученых. Электронное издание. – СПб: Университет ИТМО, 2015. – 2015

11. Овсяникова В.В., Настека А.В., Ефремов А.А., Салахутдинова К.И., Трофимов А.А. Защита системы «Умный дом» от программных сбоев//Сборник тезисов докладов конгресса молодых ученых. Электронное издание. – СПб: Университет ИТМО, 2015. – 2015

12. Druzhinin N.K., Salakhutdinova K.I. Identification of executable file by dint of individual feature//ISPIT-2015. International Conference on Information Security and Protection of Information Technology. St. Petersburg, Russia, November 5-6, 2015, IET - 2015, pp. 45-47

13. Кривцова И.Е., Салахутдинова К.И. Применение x2-критерия для идентификации elf-файлов. V Всероссийский конгресс молодых ученых//Сборник ВКМУ – 2016

14. Салахутдинова К.И., Кривцова И.Е. Условия применения критерия Пирсона для идентификации исполняемых файлов. Региональная информатика "РИ-2016" - 2016

15. Салахутдинова К.И., Дружинин Н.К. Идентификация исполняемых файлов по их ассемблерным командам // Научные работы участников конкурса «Молодые ученые Университета ИТМО» 2016 года. 2017.

16. Салахутдинова К.И., Лебедев И.С. Применение методов статистического анализа для идентификаций версий программного обеспечения удаленных автономных объектов транспортных систем //Информационная безопасность регионов России (ИБРР-2017). 2017.

17. Салахутдинова К.И., Лебедев И.С. Использование градиентного бустинга в задаче сравнения сигнатур программ //Сборник тезисов докладов конгресса молодых ученых. 2018. 136-141 с.

18. Салахутдинова, К.И. Обзор существующих подходов по аудиту электронных носителей информации // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. 2019.

Публикации в изданиях, индексируемых в реферативных базах Scopus

19. Lebedev I.S., Korzhuk V., Krivtsova I., Salakhutdinova K., Sukhoparov M.E., Tikhonov D. Using Preventive Measures for the Purpose of Assuring Information Security of Wireless Communication Channels // Proceedings of the 18th Conference of Open Innovations Association FRUCT - 2016, pp. 167-173. 0,7 п.л. / 0,12 п.л.

20. Bazhayev N., Lebedev I.S., Krivtsova I.E., Sukhoparov M.E., Salakhutdinova K., Davydov A.E., Shaparenko I.M. Evaluation of the available wireless remote devices subject to the information impact // 10th IEEE International Conference on Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings (Azerbaijan, Baku, 12-14 October 2016) - 2016, pp. 1-6. 0,65 п.л. / 0,1 п.л.

21. Lebedev I.S., Krivtsova I.E., Korzhuk V., Bazhayev N., Sukhoparov M.E., Pecherkin S., Salakhutdinova K. The Analysis of Abnormal Behavior of the System Local Segment on the Basis of Statistical Data Obtained from the Network Infrastructure Monitoring // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 2016, Vol. 9870, pp. 503-511. 0,36 п.л. / 0,1 п.л.

22. Krivtsova I.E., Lebedev I.S., Salakhutdinova K.I. Identification of Executable Files on the basis of Statistical Criteria//Proceedings of the 20th Conference of Open Innovations Association FRUCT, IET. 2017, pp. 202-208. 0,72 п.л. / 0,24 п.л.

23. Salakhutdinova K., Lebedev I.S., Krivtsova I.E., Bazhayev N., Sukhoparov M.E., Smirnov P.I., Markelov D.V., Davydov A.E., Pecherkin S., Kolcherin D.V., Shaparenko I.M., Iskanderov Y. A Frequency Approach to Creation of Executable File Signatures for their Identification//11th IEEE International Conference on Application of Information and Communication Technologies, AICT 2017 - Conference Proceedings (Moscow, 20-22 september 2017), IET. 2017, pp. 261-267. 0,69 п.л. / 0,23 п.л.

24. Salakhutdinova, K.I. Krivtsova I.E., Lebedev I.S., Sukhoparov M.E. An Approach to Selecting an Informative Feature in Software Identification // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. С. 318–327. 0,34 п.л. / 0,1 п.л.

25. Salakhutdinova K.I., Lebedev I.S., Krivtsova I.E., Sukhoparov M.E. Studying the Effect of Selection of the Sign and Ratio in the Formation of a Signature in a Program Identification Problem // Automatic Control and Computer Sciences. 2018, Vol. 52, No. 8, pp. 1101–1104. 0,35 п.л. / 0,1 п.л.

26. Semenov, Viktor V., Ilya S. Lebedev, Mikhail E. Sukhoparov and Kseniya I. Salakhutdinova. Application of an Autonomous Object Behavior Model to Classify the Cybersecurity State. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2019, ruSMART 2019. Lecture Notes in Computer Science - 2019, Vol. 11660, pp. 104-112. 0,57 п.л. / 0,14 п.л.

Зарегистрированное программное обеспечение

27. Ильин А.Г., Салахутдинова К.И., Юшковский А.В., Катаева В.А., Первушин А.О., Павлов К.С. Программа мониторинга Google Apps for Business. №2016614206, 18.04.2016.

28. Ильин А.Г., Салахутдинова К.И., Юшковский А.В., Катаева В.А., Первушин А.О., Павлов К.С. Программа для стеганографического сокрытия информации в медиа файлах. №2016614207,18.04.2016.

29. Ильин А.Г., Салахутдинова К.И., Юшковский А.В., Катаева В.А., Первушин А.О., Павлов К.С. Программа для поиска и анализа криптоконтейнеров с носителя информации. №2016611975, 15.02.2016.

30. Ильин А.Г., Салахутдинова К.И., Юшковский А.В., Катаева В.А., Первушин А.О., Павлов К.С. Программа для криминалистического анализа IM ICQ и Jabber. №2016614048, 13.04.2016.

31. Ильин А.Г., Салахутдинова К.И., Юшковский А.В., Катаева В.А., Первушин А.О., Павлов К.С. Программа для аудита событий информационной безопасности на основе модели MapReduce. №2016614208, 18.04.2016.

32. Салахутдинова К.И. Сравнение сигнатур исполняемых файлов. Свидетельство о государственной регистрации программы для ЭВМ №2019619363, 16.07.2019.

Диссертационная работа соответствует требованиям п. 9-14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 №842 (ред. от 01.10.2018) и пунктам 6, 13 и 14 Паспорта специальности ВАК (технические науки) по специальности 05.13.19 – Методы и системы защиты информации, информационная безопасность.

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

Заключение принято на расширенном семинаре лабораторий интеллектуальных систем, информационно-вычислительных систем и технологий программирования, проблем компьютерной безопасности, кибербезопасности и постквантовых криптосистем. В составе 15 чел. Результаты голосования: «за» – 15 чел., «против» – 0 чел., «воздержалось» – 0 чел., протокол № 3 от «11» октября 2019г.

Председатель расширенного семинара
доктор технических наук, профессор
главный научный сотрудник лаборатории
кибербезопасности и постквантовых
криптосистем А.А. Молдовян

Секретарь семинара
кандидат технических наук
старший научный сотрудник Л.Н. Федорченко