МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ / METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY

DOI: https://doi.org/10.23670/IRJ.2023.132.39

МЕТОД АНАЛИЗА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ ЧАСТОТНОЙ ХАРАКТЕРИСТИКИ СИМВОЛОВ В КОДЕ ДЛЯ ОБНАРУЖЕНИЯ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Научная статья

Васенин Р.С.[1], Таров Е.В.[2,] *

[2] ORCID : 0009-0003-9631-3609;

[1, 2] Санкт-Петербургский государственный университет Телекоммуникаций им. профессора М. А. Бонч-Бруевича, Санкт-Петербург, Российская Федерация

* Корреспондирующий автор (tarov25[at]mail.ru)

Аннотация

В данной статье представляется новый подход к оценке кода на основании частотной характеристики символов, который позволяет определять наличие вредоносных частей. Для этого строится таблица частотности символов на основе некоторого количества примеров вредоносных программ, и используется для анализа исследуемых кодов. Таким образом, предлагается новая методика оценки кода на наличие вредоносных частей, которая может быть эффективна в случае, если вредоносный код является новым и неизвестным. Данная работа вносит научную новизну в область информационной безопасности и может быть полезна для специалистов в области информационной безопасности, а также для разработки новых методов защиты от вредоносного программного обеспечения.

Ключевые слова: информационная безопасность, вредоносный код, анализ программ, защита информации, частотный анализ.

A SOFTWARE ANALYSIS METHOD BASED ON FREQUENCY CHARACTERIZATION OF CHARACTERS IN CODE FOR INFORMATION SECURITY THREAT DETECTION

Research article

Vasenin R.S.[1], Tarov E.V.[2,] *

[2] ORCID : 0009-0003-9631-3609;

[1, 2] The Bonch-Bruevich Saint Petersburg State University of Telecommunications, Saint-Petersburg, Russian Federation

* Corresponding author (tarov25[at]mail.ru)

Abstract

This article presents a new approach to code evaluation based on character frequency response, which allows to determine the presence of malicious parts. For this aim, a character frequency table is constructed based on a number of malware samples and used to analyse the studied codes. Thus, a new methodology is proposed to evaluate the code for the presence of malicious parts, which can be effective if the malicious code is new and unknown. This work brings scientific novelty to the field of information security and can be useful for information security professionals and for the development of new methods of protection against malicious software.

Keywords: information security, malware, software analysis, information security, frequency analysis.

Введение

Информационная безопасность является одной из наиболее важных задач в настоящее время. Вместе с ростом объемов информации, передаваемой и хранимой в электронной форме, возрастает и угроза ее утечки или несанкционированного доступа. Каждый день мы сталкиваемся с новыми видами кибератак, которые могут привести к серьезным последствиям, таким как потеря конфиденциальной информации, финансовые потери, и нарушение работоспособности систем [1], [2], [3], [4].

В данной работе предлагается метод оценки кода на основе частотности символов. Для этого создается таблица частотности для вредоносного кода и на ее основе проводится анализ исследуемых кодов, чтобы определить, содержат ли они в себе вредоносные элементы.

Научная новизна данного исследования заключается в разработке нового метода анализа вредоносного ПО. Использование частотной характеристики символов в коде является нестандартным подходом к оценке кода и может помочь в улучшении эффективности методов обнаружения и защиты от вредоносного ПО. Таким образом, данное исследование имеет высокую актуальность в сфере информационной безопасности и может быть использовано для развития более эффективных методов обнаружения и защиты от вредоносного ПО.

Современные методы анализа ПО

Современные методы анализа ПО широко применяются для обеспечения безопасности и качества ПО. Некоторые из современных методов анализа ПО включают в себя [5], [6], [7], [8]:

1. Статический анализ кода. Это метод анализа, при котором программный код анализируется без его выполнения. В этом методе анализаторы проверяют код на наличие ошибок, уязвимостей и других потенциальных проблем.

2. Динамический анализ кода. Этот метод включает выполнение кода в контролируемой среде и анализ его поведения. Динамический анализ используется для обнаружения уязвимостей, скрытого поведения и других проблем, которые могут быть пропущены статическим анализом.

3. Fuzz-тестирование. Этот метод включает создание случайных входных данных для программы с целью проверки ее устойчивости и безопасности. Fuzz-тестирование используется для обнаружения уязвимостей, которые могут быть использованы злоумышленниками для атак на ПО.

4. Обратная разработка. Этот метод включает анализ скомпилированного кода с целью получения информации о его структуре и функциональности. Обратная разработка используется для анализа вредоносных программ и уязвимостей в ПО.

5. Анализ бинарного кода. Этот метод включает анализ скомпилированного кода без доступа к исходному коду. Анализ бинарного кода используется для обнаружения уязвимостей, анализа вредоносных программ и оптимизации кода.

6. Исследование угроз. Этот метод включает анализ потенциальных угроз и рисков для ПО. Исследование угроз используется для разработки стратегий защиты ПО и уменьшения рисков для организации.

Проблемы современных методов анализа ПО

Исследователи в области обнаружения вредоносного ПО сталкиваются с множеством проблем, которые могут затруднять точное определение наличия вредоносной программы в системе. Некоторые из этих проблем включают в себя:

- изменение кода вредоносной программы: злоумышленники могут изменять код вредоносных программ, чтобы обойти существующие системы обнаружения вредоносного ПО (это делает обнаружение вредоносных программ более сложным);
- использование новых методов атаки: злоумышленники постоянно разрабатывают новые методы атаки, которые могут быть трудными для обнаружения существующими системами безопасности;
- высокое количество ложных срабатываний: традиционные системы обнаружения вредоносного ПО могут давать высокий процент ложных срабатываний, что означает, что система может считать некоторые безвредные программы вредоносными;
- сложность обработки большого объема данных: системы обнаружения вредоносного ПО могут столкнуться с проблемой обработки большого объема данных, что может привести к недостаточной скорости обнаружения вредоносных программ.

Метод анализа ПО на основе частотной характеристики символов в коде может решить некоторые из этих проблем. Данный подход позволяет создавать таблицу частотности для заранее известных вредоносных кодов и анализировать коды исследуемых программ на основе этой таблицы, что может помочь точно определить наличие вредоносной программы в системе. Этот подход может также помочь в решении проблемы высокого процента ложных срабатываний, так как он анализирует код, а не поведение программы, что может дать более точные результаты. Кроме того, этот подход может обрабатывать большие объемы данных быстро, что помогает в улучшении скорости обнаружения вредоносных программ.

Суть предлагаемого метода

Предлагаемый метод основан на том, что в различных типах программ (включая вредоносный код) используются различные символы с разной частотой. Например, во вредоносном коде могут быть часто встречающиеся символы «x90» (опкод для инструкции NOP в ассемблере). Подобный метод анализа можно наблюдать в криптографии – частотный анализ [9], [10].

Для начала из базы данных вирусов необходимо взять большое количество примеров вредоносного ПО конкретного семейства. Затем для удобства каждый символ преобразовать в соответствующее число по таблице символов ASCII (т.е. от 0 до 255). Далее для каждого из 256 чисел найти процент его содержания относительно длинны всех анализируемых кодов. Данное значение можно назвать весом числа, которое рассчитывается по следующему правилу:

$$\omega_i = \frac{N_i}{N}, \quad (1)$$

где N – суммарный листинг всех анализируемых кодов, N_i – количество чисел i во всем коде N, ω_i – вес числа i, где i = 0, 1, 2, …, 255.

После того как для каждого числа 0-255 были определены соответствующие веса, необходимо построить частотную таблицу. Приведем пример части такой таблицы – таблица 1.

Таблица 1 - Веса некоторых чисел вредоносных кодов типа троянов

DOI: https://doi.org/10.23670/IRJ.2023.132.39.1

i	Значение вредоносного ПО, %
6	0,41
13	0,99
19	0,34
139	0,95
192	0,49

224	0,29
238	0,25
255	2,85

Таблица 1 показывает процент содержания каждого значения символа в коде, преобразованного в десятичную систему счисления. Например, символ с кодом 255 встречается в данном коде в 2,85% случаев, а символ с кодом 238 практически ни разу (0,25%).

Для того чтобы иметь возможность оценить вредоносность анализируемого ПО, введем следующую метрику:

$$P = \sum_{i=0}^{225} x, \tag{2}$$

где значение x рассчитывается для каждого значения i по следующему правилу:

$$x = \begin{cases} \max(\omega_i, \omega_{iT}) / \min(\omega_i, \omega_{iT}), \min(\omega_i, \omega_{iT}) \neq 0 \\ \max(\omega_i, \omega_{iT}) / 0.0001, \min(\omega_i, \omega_{iT}) = 0 \end{cases}, \tag{3}$$

здесь ω_i – вес числа i из таблицы 1, ω_{iT} – вес числа i относительно анализируемого кода, $max(\omega_i, \omega_{iT})$ – максимальное значение между ω_i и ω_{iT}, $min(\omega_i, \omega_{iT})$ – минимальное значение между ω_i и ω_{iT}, 0,0001 – самое минимальное значение весов чисел, полученное в результате большого количества опытов (введение данного числа обусловлено тем, что для некоторых кодов значение $min(\omega_i, \omega_{iT})$ может быть ноль).

Далее возьмем большое количество тестовых кодов – обычных и вредоносных (семейства троянов). Для каждого из данных кодов вычислим значения метрики *P* и возьмем среднее значение. В результате чего можно сформировать таблицу 2.

Таблица 2 - Значение метрики P в зависимости от типа и от размерности ПО

DOI: https://doi.org/10.23670/IRJ.2023.132.39.2

Тип ПО	Размерность, КБ	Значения метрики *P*
Вредоносное	4000	369
	2000	365
	500	444
	100	1212
Обычное	4000	118397
	2000	128691
	500	146748
	100	36355

Как видно из таблицы 2, значения метрики *P* для вредоносного и обычного ПО имеют значительные расхождения, например, значение *P* для обычного кода в 320 раз больше значения *P* для вредоносного ПО. Однако данная разница уменьшается с уменьшением размера анализируемых кодов, но даже так разница остается ощутимой.

Заключение

Предложенный метод имеет несколько преимуществ по сравнению со стандартными методами анализа вредоносного ПО:

Во-первых, он основан на частотной характеристике символов в коде, что позволяет определить характерные особенности вредоносного ПО, которые могут быть пропущены другими методами анализа.

Во-вторых, данный метод позволяет проводить анализ большого количества кода в автоматическом режиме, что ускоряет процесс обнаружения вредоносных программ.

В-третьих, данный метод может быть полезен при анализе обфусцированного вредоносного ПО, поскольку обфускация кода часто основывается на замене символов и перестановке строк, что может затруднять традиционный анализ кода. Однако, даже после обфускации, частота использования различных символов в коде остается примерно одинаковой. Поэтому, если анализатор безопасности строит таблицу частотности символов для нескольких образцов вредоносных программ, он может использовать эту таблицу для идентификации общих паттернов, которые могут присутствовать даже после обфускации. Кроме того, вредоносные программы, созданные одним автором или группой авторов, могут иметь сходство в частоте использования символов. С помощью метода анализа на основе частотной характеристики символов в коде можно идентифицировать эти сходства и использовать их для поиска других образцов вредоносных программ, созданных тем же автором или группой авторов

Наконец, этот метод может быть использован в сочетании с другими методами анализа для более эффективного обнаружения вредоносных программ. В целом, данный метод представляет собой важный инструмент для обнаружения и борьбы с вредоносным ПО.

Существует несколько направлений, в которых можно проводить будущие исследования:

Использование методов машинного обучения: можно использовать алгоритмы машинного обучения, такие как нейронные сети, для автоматического обнаружения вредоносного кода на основе частотной характеристики символов.

Анализ других аспектов кода: помимо частотной характеристики символов, можно проводить исследования, направленные на анализ других аспектов кода, таких как синтаксическая структура, использование API и т.д.

Исследования, связанные с защитой: можно исследовать методы защиты от вредоносного кода на основе частотной характеристики символов, такие как обнаружение и удаление вредоносных элементов, защита от атак, связанных с изменением частотной характеристики символов и т.д.

Расширение области исследований: можно расширить область исследований, включив в них не только вредоносный код, но и другие виды кода, например, код, используемый в различных областях, таких как банковское дело, медицина и т.д.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Пат. 2020617705 Russian Federation, МПК2020616731 .. Программная реализация средств предотвращений вторжений и аномалий сетевой инфраструктуры / Красов А.В., Гельфанд А.М., Фадеев И.И. и др.; заявитель и патентообладатель Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича. — № 2020616731; заявл. 2020-06-29; опубл. 2020-07-10, — 1 с.

2. Орлов Г. А. Применение Big Data при анализе больших данных в компьютерных сетях / Г. А. Орлов, А. В. Красов, А. М. Гельфанд // Наукоемкие технологии в космических исследованиях Земли. — 2020. — Т. 12. — № 4. — С. 76-84. — DOI: 10.36724/2409-5419-2020-12-4-76-84.

3. Гельфанд А.М. Исследование распределенного механизма безопасности для устройств интернета вещей с ограниченными ресурсами / А. М. Гельфанд, А. А. Казанцев, А. В. Красов [и др.] // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2020). — Санкт-Петербург: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2020. — С. 321-326.

4. Гельфанд А. М. Разработка модели распространения самомодифицирующегося кода в защищаемой информационной системе / А. М. Гельфанд, И. Е. Пестов, А. И. Катасонов [и др.] // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2018. — № 8. — С. 91-97.

5. Буйневич М. В. Сравнительный анализ подходов к поиску уязвимостей в программном коде / М. В. Буйневич, К. Е. Израилов, Д. И. Мостович // Актуальные проблемы инфотелекоммуникаций в науке и образовании. — Санкт-Петербург: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2016. — С. 256-260.

6. Израилов К. Е. Метод анализа вредоносного программного обеспечения на базе fuzzy hash / К. Е. Израилов, Н. В. Гололобов, Г. А. Краскин // Информатизация и связь. — 2019. — № 2. — С. 36-44. — DOI: 10.34219/2078-8320-2019-10-2-36-44.

7. Буйневич М. В. Основы кибербезопасности: способы анализа программ / М. В. Буйневич, К. Е. Израилов. — Санкт-Петербург: Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, 2022. — 92 с.

8. Буйневич М.В. Основы кибербезопасности: способы защиты от анализа программ / М.В. Буйневич, К.Е. Израилов. — Санкт-Петербург: Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, 2022. — 76 с.

9. Коржик В. И. Основы криптографии / В. И. Коржик, В. А. Яковлев. — Санкт-Петербург: Интермедия, 2017. — 312 с.

10. Коржик В. И. Основы криптографии / В. И. Коржик, В. П. Просихин. — Санкт-Петербург: Линк, 2008. — 249 с.

Список литературы на английском языке / References in English

1. Pat. 2020617705 Russian Federation, MPK2020616731 .. Programmnaya realizaciya sredstv predotvrashhenij vtorzhenij i anomalij setevoj infrastruktury' [Software Implementation of Intrusion and Anomaly Prevention Tools for Network Infrastructure] / Красов А.В., Гельфанд А.М., Фадеев И.И. и др.; the applicant and the patentee St. Petersburg State University of Telecommunications named after prof. M.A. Bonch-Bruevich. — № 2020616731; appl. 2020-06-29; publ. 2020-07-10, — 1 p. [in Russian]

2. Orlov G. A. Primenenie Big Data pri analize bol'shih dannyh v komp'juternyh setjah [Application of Big Data in the Analysis of Big Data in Computer Networks] / G. A. Orlov, A. V. Krasov, A. M. Gel'fand // Naukoemkie tehnologii v

kosmicheskih issledovanijah Zemli [Science-Intensive Technologies in Space Exploration of the Earth]. — 2020. — Vol. 12. — № 4. — P. 76-84. — DOI: 10.36724/2409-5419-2020-12-4-76-84 [in Russian]

3. Gel'fand A. M. Issledovanie raspredelennogo mehanizma bezopasnosti dlja ustrojstv interneta veshhej s ogranichennymi resursami [Exploring a Distributed Security Mechanism for IoT Devices with Limited Resources] / A. M. Gel'fand, A. A. Kazancev, A. V. Krasov [et al.] // Aktual'nye problemy infotelekommunikacij v nauke i obrazovanii (APINO 2020) [Actual Problems of Infotelecommunications in Science and Education (APISE 2020)] — Saint-Petersburg: St. Petersburg State University of Telecommunications named after Prof. M.A. Bonch-Bruevich, 2020. — P. 321-326. [in Russian]

4. Gel'fand A. M. Razrabotka modeli rasprostranenija samomodificirujushhegosja koda v zashhishhaemoj informacionnoj sisteme [Development of a Self-Modifying Code Distribution Model in a Protected Information System] / A. M. Gel'fand, I. E. Pestov, A. I. Katasonov [et al.] // Sovremennaja nauka: aktual'nye problemy teorii i praktiki. Serija: Estestvennye i tehnicheskie nauki [Modern Science: Actual Problems of Theory and Practice. Series: Natural and Technical Sciences]. — 2018. — № 8. — P. 91-97. [in Russian]

5. Bujnevich M. V. Sravnitel'nyj analiz podhodov k poisku ujazvimostej v programmnom kode [A Comparative Analysis of Approaches to the Search for Vulnerabilities in the Program Code] / M. V. Bujnevich, K. E. Izrailov, D. I. Mostovich // Aktual'nye problemy infotelekommunikacij v nauke i obrazovanii [Actual Problems of Infotelecommunications in Science and Education]. — Sankt-Peterburg: St. Petersburg State University of Telecommunications named after Prof. M.A. Bonch-Bruevich, 2016. — P. 256-260. [in Russian]

6. Izrailov K. E. Metod analiza vredonosnogo programmnogo obespechenija na baze fuzzy hash [Fuzzy Hash-Based Malware Analysis Method] / K. E. Izrailov, N. V. Gololobov, G. A. Kraskin // Informatizacija i svjaz' [Informatization and Communication]. — 2019. — № 2. — P. 36-44. — DOI: 10.34219/2078-8320-2019-10-2-36-44 [in Russian]

7. Bujnevich M. V. Osnovy kiberbezopasnosti: sposoby analiza programm [Fundamentals of Cybersecurity: Methods of Program Analysis] / M. V. Bujnevich, K. E. Izrailov. — Sankt-Peterburg: St. Petersburg University of the State Fire Service of the Ministry of the Russian Federation for Civil Defense, Emergencies and Disaster Relief named after the Hero of the Russian Federation, General of the Army E.N. Zinichev, 2022. — 92 p. [in Russian]

8. Bujnevich M. V. Osnovy kiberbezopasnosti: sposoby zashhity ot analiza programm [Fundamentals of Cybersecurity: Ways to Protect against Program Analysis] / M. V. Bujnevich, K. E. Izrailov. — Sankt-Peterburg: St. Petersburg University of the State Fire Service of the Ministry of the Russian Federation for Civil Defense, Emergencies and Disaster Relief named after the Hero of the Russian Federation, General of the Army E.N. Zinichev, 2022. — 76 p. [in Russian]

9. Korzhik V. I. Osnovy kriptografii [Fundamentals of Cryptography] / V. I. Korzhik, V. A. Jakovlev. — St. Petersburg: Intermedija, 2017. — 312 p. [in Russian]

10. Korzhik V. I. Osnovy kriptografii [Fundamentals of Cryptography] / V. I. Korzhik, V. P. Prosihin. — St. Petersburg: Link, 2008. — 249 p. [in Russian]