DOI 10.31891/2307-5732-2024-333-2-60
УДК 004.056.57

ПЕДЯШ ВОЛОДИМИР
Міжнародний гуманітарний університет
https://orcid.org/0000-0002-4071-357X
e-mail: vpedyash@gmail.com
ЛЕДОВСЬКИЙ ЄВГЕН
Міжнародний гуманітарний університет
https://orcid.org/0009-0003-2176-9265
e-mail: ledovskyi_yevhen@ukr.net
ТКАЧ ВОЛОДИМИР
Міжнародний гуманітарний університет
https://orcid.org/0009-0003-0517-6355
e-mail: tkach_volodym@ukr.net

СУЧАСНІ МЕТОДИ ВИЯВЛЕННЯ ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

В роботі розглянуто сучасні методи виявлення шкідливого програмного забезпечення (ПЗ) та їх аналіз у контексті кібербезпеки. Розглядаються різні підходи, включно з сигнатурним аналізом, евристичними методами, статичним і динамічним аналізом. Також розглянуто інноваційні технології останніх років із використанням нейронних мереж та перетворенням послідовності байт програмного коду на графічні зображення. Обговорено переваги та обмеження кожного методу з урахуванням їхньої ефективності, швидкості роботи і можливості виявлення невідомих загроз.

Ключові слова: шкідливе програмне забезпечення, кібербезпека, інформаційна безпека, сигнатурний аналіз, евристичні методи, нейронні мережі

PEDYASH VOLODYMYR, LEDOVSKYI YEVHEN, TKACH VOLODYMYR,
International Humanitarian University

MODERN METHODS OF MALWARE DETECTION

Modern information society faces an ever-increasing threat of malware, which requires continuous improvement of detection and control methods. This paper presents an overview and analyses different approaches to malware detection. Traditional methods such as signature analysis and euristic methods continue to be important tools in cybersecurity. However, with the emergence of new threats and technologies, there is a need to develop and apply modern approaches to malware detection. Static and dynamic analyses of program code provide additional opportunities to detect hidden and evolving threats. Innovative techniques, such as neural networks and code-to-image, offer new perspectives on malware detection. Neural networks capable of learning from large amounts of data can effectively detect complex patterns in malware behaviour. Converting a sequence of bytes of code into images allows information about the code to be presented in a new context. Using deep learning to analyze images opens up new opportunities in malware detection. However, these innovative methods also face challenges such as complexity of model training and computational resource requirements. Continuous improvement and development of detection methods is necessary for effective malware detection. An important area of research is to develop methods that can detect new and unknown threats. The combination of different approaches and technologies can provide better protection of information systems from cyberattacks. The development of advanced malware detection techniques is critical to security in the digital environment. Further research and development in this area is key to ensuring that information systems are protected from ever-evolving cyber threats.

Keywords: malware, cybersecurity, information security, signature analysis, heuristic methods, neural networks.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

У наш час, коли цифрові технології проникають в усі сфери нашого життя, захист інформації стає невід'ємною складовою забезпечення безпеки як на рівні індивідуального користувача, так і на рівні великих корпорацій і державних організацій. Стрімке зростання кількості інформації, переданої та збереженої в комп'ютерних системах, супроводжується збільшенням загроз кібербезпеці [1]. Кібератаки стають дедалі складнішими, постійно пристосовуючись до нових методів захисту, що підкреслює актуальність проблеми захисту інформації. Щодня кіберзлочинці масштабують свої атаки, використовуючи найсучасніші методи злому та експлуатації вразливостей, що призводить до величезних матеріальних і репутаційних збитків. Економічні втрати від кіберзлочинності оцінюються в мільярди доларів щорічно і ця цифра продовжує зростати, що актуалізує нагальну потребу з розробки нових методів захисту даних.

Однією з головних проблем у сфері кібербезпеки є нездатність традиційних засобів захисту, таких як антивірусні програми та брандмауери ефективно протистояти новим атакам. Існуючі системи виявлення часто не можуть виявити приховані або замасковані атаки, що робить їх вразливими перед сучасними загрозами. Це призводить до необхідності пошуку нових підходів до захисту інформації, які можуть забезпечити вищий рівень безпеки. Іншою проблемою є необхідність оперативної реакції на кібератаки. Існуючі методи захисту інформації можуть вимагати певного часу для виявлення атаки та вжиття заходів щодо її запобігання, що може призвести до серйозних наслідків. Розвиток машинного навчання і штучного інтелекту надає нові можливості для створення більш інтелектуальних систем захисту, здатних адаптуватися до мінливої загрози і виявляти приховані атаки.

Аналіз останніх досліджень

Завдання по аналізу методів виявлення шкідливого ПЗ наведено в роботах вітчизняних та зарубіжних дослідників. Наприклад, в монографіях [2–4] та статтях [5, 6] систематизовано матеріал по існуючим методам виявлення шкідливого ПЗ та запропоновано ряд супутніх алгоритмів. Узагальнивши викладену по даному напряму інформацію, можна зробити висновок про доцільність застосування методів статичного та динамічного аналізу програмного коду з метою пошуку шкідливих фрагментів. Однак завжди існує низка складних і непередбачуваних загроз, які можуть залишатися непоміченими після аналізу вказаними методами. У зв'язку з цим дедалі більшу увагу привертає практика застосування нейронних мереж на базі машинного навчання для вирішення поставленої проблеми [7]. Нейронні мережі здатні обробляти великі обсяги даних і виявляти складні патерни, що робить їх ефективними в боротьбі з комплексними загрозами.

Метою цієї статті є аналіз останніх досліджень з розробки методів виявлення шкідливого програмного забезпечення та визначити перспективні напрями його розвитку.

Виклад основного матеріалу

У сфері кібербезпеки важливими інструментами для виявлення шкідливого програмного забезпечення є групи статичних, динамічних та комплексних методів. Вони засновані на різних принципах аналізу і можуть ефективно доповнювати один одного. Статичні методи виявлення шкідливого ПЗ аналізують його структуру і код без його фактичного виконання. Цей підхід включає в себе статичний аналіз коду виконуваних файлів та їх метаданих. Методи статичного аналізу можуть включати: сканування антивірусними програмами, використання сигнатурних баз даних і евристичний аналіз, заснований на відомих патернах шкідливого ПЗ. Динамічні методи виявлення шкідливого ПЗ аналізують його поведінку під час виконання в режимі реального часу. Цей підхід охоплює запуск програм у контрольованому середовищі та моніторинг їхньої активності, мережевої взаємодії, системних викликів і змін у файловій системі. Динамічні методи можуть включати використання віртуальних машин, емуляторів і налагоджувачів для аналізу шкідливого ПЗ у реальному часі. Хоча статичні та динамічні методи виявлення шкідливого ПЗ мають свої переваги та обмеження, найефективнішим підходом є використання комплексного методу, що об'єднує обидва підходи.

Тепер розглянемо основні методи, що входять до вищевказаних груп. Першим, найбільш фундаментальним методом виявлення шкідливого програмного забезпечення є сигнатурний аналіз. Він заснований на пошуку унікальних сигнатур або характеристик, специфічних для кожної конкретної загрози. Ці сигнатури зазвичай являють собою рядки коду, хеш-суми файлів, певні послідовності байтів або інші унікальні ознаки, які допомагають ідентифікувати шкідливе ПЗ. Сигнатурний аналіз починається зі створення бази даних сигнатур, яка містить інформацію про існуючі шкідливі програми та їхні унікальні характеристики. Потім антивірусне програмне забезпечення використовує цю базу даних для сканування файлів або системи на наявність збігів з відомими загрозами. Якщо виявляється схожа сигнатура, антивірусна програма сигналізує про загрозу і вживає відповідних заходів для її нейтралізації. Основною перевагою сигнатурного аналізу є його висока точність при виявленні вже відомих загроз. Цей метод має низький рівень помилкових спрацьовувань і забезпечує швидку реакцію на відомі загрози. Однак, існують і обмеження для цього методу. Він нездатний виявляти нові або раніше невідомі загрози, оскільки не може розпізнати їх без відповідної сигнатури. Крім того, сигнатурний аналіз може бути вразливим до обходу систем захисту, оскільки шкідливі програми можуть змінювати свої сигнатури або використовувати методи шифрування для приховування своєї ідентифікації. Сигнатурний аналіз, як і раніше, залишається важливим компонентом у системах безпеки, проте сучасні тенденції в кібербезпеці наголошують на необхідності комбінування цього методу з іншими техніками виявлення.

Другим способом дослідження ПЗ є евристичний аналіз, що ґрунтується на здатності системи навчитися розпізнавати незвичайні та підозрілі патерни поведінки програм. Він є надзвичайно важливим інструментом в арсеналі засобів безпеки, оскільки дає змогу виявляти нові загрози, які могли б оминути традиційні методи, такі як сигнатурний аналіз. Використовуючи набір евристичних правил або алгоритмів, метод визначає потенційно небезпечні сценарії або поведінку. Ці правила постійно оновлюються для відображення змін у тактиках розробників шкідливого ПЗ. Метод аналізує поведінку програм у системі, включаючи їх взаємодію з файловою системою, реєстром, мережею та іншими компонентами операційної системи. Однак, також існує ризик помилкових спрацьовувань, коли звичайне (доброякісне) програмне забезпечення може бути неправильно класифіковане через схожість у поведінці.

Третім поширеним методом є поведінковий аналіз для виявлення шкідливого програмного забезпечення, який зосереджений на вивченні дій програм у реальному часі. Він базується на ідеї, що шкідливе ПЗ часто проявляє певні аномальні або підозрілі патерни поведінки, які можуть відрізнятися від типової поведінки звичайних програм. Під час використання поведінкового аналізу система аналізує дії програми в реальному часі та ідентифікує аномалії або незвичні шаблони поведінки, які можуть вказувати на наявність шкідливої активності. Наприклад, такі дії можуть включати в себе спроби зміни системних файлів, несанкціоновані мережеві запити, запуск прихованих процесів і багато іншого. Для проведення поведінкового аналізу можуть використовуватися різні техніки та методи: моніторинг системних викликів (API-викликів), аналіз мережевого трафіку, відстеження змін у файловій системі, контроль активності процесів і багато іншого. Комбінація цих технік дає змогу системі ефективно виявляти аномалії, що свідчать про наявність

шкідливої активності.

Однією з головних переваг поведінкового аналізу є його здатність виявляти нові та раніше невідомі загрози, які можуть обійти традиційні методи виявлення, такі як сигнатурний аналіз. Оскільки поведінковий аналіз аналізує безпосередні дії програм у реальному часі, він може виявляти навіть приховані або замасковані загрози. Однак, поведінковий аналіз також має свої обмеження. Деякі нешкідливі програми можуть проявляти аномальну поведінку, яка може бути ідентифікована як шкідливе ПЗ, що може призвести до хибних спрацьовувань. Крім того, складність аналізу великих обсягів даних і високий ступінь імовірності помилкових спрацьовувань є викликами для розробників і аналітиків. Здатність методу виявляти нові та невідомі загрози робить його важливим компонентом у стратегіях кібербезпеки.

Оскільки традиційні методи виявлення шкідливого ПЗ стикаються з проблемами, не встигаючи за витонченістю сучасних атак, то необхідні інноваційні підходи для аналізу ПЗ. Один з останніх трендів в цьому напрямку є візуальній метод пошуку шкідливого ПЗ. Основна ідея цього методу полягає в тому, що кожен виконуваний файл програми може бути представлений у вигляді графічного зображення, де кожен байт коду відображається як піксель або набір пікселів [8].

На даний момент запропоновано кілька варіантів формування зображень. В роботі [9] запропонована реалізація даного методу для класифікації набору даних, що складається з 9458 шкідливих програм, що згруповані в 25 сімейств. Задача класифікації вирішувалася шляхом застосування вейвлет перетворення до сформованих на основі двійкового коду зображень в градаціях сірого кольору. Точність класифікації запропонованим способом становила близько 97%. В подальших роботах інші дослідники зосередили увагу на підвищенні продуктивності алгоритмів розпізнавання образів. Досить ефективно ця задача вирішена в роботі [10] шляхом використання машин опорних векторів (SVM). Для набору даних, що містить 25000 шкідливих програм і 12000 безпечних зразків була отримана точність 95%. Більш пізніші дослідження використовують підхід, що передбачає розпізнавання шкідливого ПЗ на основі марківських зображень. Так в роботі [11] досліджено метод формування марківських зображень та подальшого їх розпізнавання нейронною мережею. Для набору даних з 12971 доброякісних зразків та 9339 зразків шкідливого програмного забезпечення отримана точність 99,16%.

Використання даного методу передбачає наступні стандартні кроки: попередня обробка даних, навчання моделі, оцінювання ефективності її роботи та по необхідності доопрацювання. Перший крок передбачає збір різноманітного набору даних зі зразків шкідливих програм, що представляють різні сімейства та категорії. Цей набір повинен охоплювати широкий спектр шкідливої поведінки та характеристик, щоб забезпечити надійність та узагальнюючу здатність класифікаційної моделі. Далі двійкові файли шкідливого ПЗ проходять попередню обробку, де вони перетворюються на візуальні образи за допомогою відповідних методів кодування. Ця операція передбачає перетворення двійкового коду в структуроване візуальне представлення, що зберігає важливі ознаки та закономірності.

Наступним важливим кроком цього методу є вибір відповідної архітектури глибокого навчання для задачі класифікації. Зазвичай для задач класифікації зображень обирають згорткові штучні нейронні мережі (ШНП) завдяки їхній здатності знаходити певні індивідуальні ознаки безпосередньо з необроблених піксельних даних. Однак для досягнення максимальної продуктивності повинні бути ретельно оптимізовані архітектура та параметри ШНМ. Після визначення архітектури моделі починається процес її навчання з використанням попередньо сформованого набору зображень. Під час навчання ШНМ вчиться виокремлювати значущі ознаки з візуальних образів і співвідносити їх з відповідними категоріями шкідливих програм. Це досягається за допомогою ітеративного процесу оптимізації, в якому ваги ШНМ коригуються на основі помилок прогнозування, розрахованих за допомогою алгоритму оптимізації, наприклад, стохастичного градієнтного спуску або методу Адама. Після навчання оцінюється продуктивність моделі за допомогою окремого тестового набору даних, щоб оцінити її здатність визначати невідомі зразки шкідливого програмного забезпечення. Для кількісної оцінки ефективності по класифікації різних категорій шкідливого ПЗ визначаються такі показники методу, як точність та достовірність.

Для подальшого покращення продуктивності класифікаційної моделі можна застосувати методи її точного налаштування та оптимізації. Вони включають такі параметри ШНМ як швидкість навчання, розміри навчального, тестового набору зображень та ряд інших. Після того як модель класифікації навчена і перевірена, її можна розгортати в реальних системах кібербезпеки для подальшого тестування. Модель також може бути інтегрована в існуючі системи безпеки, такі як системи виявлення вторгнень (Intrusion Detection System, IDS) або платформи захисту кінцевих точок (Endpoint Protection Platform, EPP) для розширення їх можливостей і забезпечення проактивного захисту від нових загроз.

Порівняльний аналіз розглянутих методів виявлення шкідливого ПЗ представлений у табл. 1, де вони оцінюються за кількома критеріями, такими як ефективність виявлення, переваги та недоліки. Також важливо зазначити, що кожен метод має свої особливості та може бути застосований залежно від конкретних потреб і умов середовища.

Таблиця 1

Порівняльна характеристика методів пошуку шкідливого ПЗ

Метод	Опис	Переваги	Недоліки
Сигнатурний аналіз	Зіставлення сигнатур шкідливого ПЗ з базою даних відомих загроз	Висока ефективність під час виявлення відомих загроз	Неефективний проти нових і змінених варіантів ПЗ
Евристичний аналіз	Виявлення підозрілої поведінки, що не відповідає нормальному зразку	Можливість виявлення нових і невідомих загроз	Може давати хибні спрацьовування, вимагає налаштування
Статичний аналіз	Аналіз програмного коду без його виконання	Ефективний під час виявлення вразливостей і неявних загроз	Вимагає доступу до вихідного коду, не враховує динаміку виконання
Динамічний аналіз	Дослідження поведінки програми в реальному часі під час її виконання	Враховує динаміку роботи ПЗ, здатний виявити приховані загрози	Може пропустити загрози, вимагає високих обчислювальних ресурсів
Нейромережевий аналіз	Застосування нейронних мереж для виявлення шкідливого ПЗ	Висока точність виявлення, здатність до навчання на великих об'ємах даних	Вимагає великих об'ємів даних для навчання, складність інтерпретації результатів

Висновки

В даній роботі було виконано аналіз методів виявлення шкідливого програмного забезпечення. Відмічено, що традиційні підходи (сигнатурний аналіз і евристичні методи), залишаються важливими інструментами в галузі кібербезпеки. Інноваційні методи, такі як використання нейронних мереж і перетворення коду на зображення, являють собою перспективні підходи до виявлення загроз. Перетворення послідовності байт коду на зображення дає змогу представити інформацію про код у новому контексті та полегшує його аналіз. Нейронні мережі на основі навчання на великих обсягах даних, мають високу ефективність розпізнавання складних патернів шкідливого ПЗ. Однак використання інноваційних методів також пов'язане з низкою викликів і обмежень. Основними перешкодами на шляху застосування даного методу в системах захисту інформації є складність навчання моделей і високі вимоги до обчислювальних ресурсів.

Подальше дослідження і розвиток у цій галузі слід проводити у напрямку створення нових алгоритмів для перетворення двійкового коду програм в цифрове зображення, що дозволить підвищити ступінь виявлення шкідливого ПЗ і покращити боротьбу з кіберзагрозами в майбутньому.

References

1. Iona L.G. Banking security systems: training. Manual. Odesa: DUITZ, 2022. 192 p.

2. Abhijit M., Anoop S. Malware Analysis And Detection Engineering: A Comprehensive Approach To Detect And Analyze Modern Malware. Berkeley: Apress, 2020. 928 p.

3. Molinari S. Malware Science: A comprehensive guide to detection, analysis, and compliance. Birmingham: Packt Publishing, 2023. 230 p.

4. Saxe J., Sanders H. Malware Data Science. San Francisco: No Starch Press, 2018. 276 p.

5. Lysenko S.., Schuka R.V. Analysis of malware detection methods in computer systems. Herald of Khmelnytskyi National University. 2020. Issue 2. P. 101-107.

6. Souri A., Hosseini R. A state-of-the-art survey of malware detection approaches using data mining techniques. Human-centric Computing and Information Sciences. 2018. Volume 8, Article number 3. P. 1-22.

7. Stamp M., Alazab M., Shalaginov A. Malware Analysis Using Artificial Intelligence and Deep Learning. Cham: Springer, 2021. 655 p.

8. KyoungSoo H., Hyun Lim J., Eul Gyu I. Malware analysis method using visualization of binary files. RACS '13: Proceedings of the 2013 Research in Adaptive and Convergent Systems, October 2013. P. 317–321.

9. Nataraj L., Karthikeyan S., Jacob G., Manjunath B.S. Malware images: visualization and automatic classification. VizSec '11: Proceedings of the 8th International Symposium on Visualization for Cyber SecurityJuly, Pittsburgh, Pennsylvania, USA, July 2011. P. 1-7.

10. Kesav Kancherla, Kesav Kancherla Image visualization based malware detection. IEEE Symposium on Computational Intelligence in Cyber Security (CICS), Singapore, 16-19 April 2013. P. 40-44.

11. Pinhero A., Anupama M.L., Vinod P., Visaggio C.A., Aneesh N., Abhijith S., AnanthaKrishnan S. Malware detection employed by visualization and deep neural network. Computers & Security. 2021. № 105(12). P. 1-30.