https://doi.org/10.31891/2219-9365-2024-78-5
УДК 004

ЧАЙКОВСЬКИЙ Максим
Хмельницький національний університет
https://orcid.org/0000-0002-9596-6697
e-mail: max.chaikovskyi@gmail.com

КОМПЛЕКСНИЙ ПІДХІД ДО ВИЯВЛЕННЯ ТА АНАЛІЗУ ПОЛІМОРФНОГО ЗЛОВМИСНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

У статті досліджено особливості сучасного поліморфного зловмисного програмного забезпечення та його вплив на функціонування комп'ютерних систем. Розглянуто існуючі підходи та методи його виявлення та аналізу, такі як: алгоритм пошуку рядка, інтелектуальний аналіз даних, аналіз в пісочниці, машинне навчання, метод розробки структурних функцій. Визначено їх переваги та недоліки. Аргументовано необхідність застосування нового підходу, а саме виявлення зловмисного програмного забезпечення за допомогою імовірнісних логічних мереж. Визначено його переваги та перспективи розвитку. Запропоновано комплексний підхід до виявлення та аналізу поліморфного зловмисного програмного забезпечення, який полягає в системному поетапному поєднанні розглянутих методів з метою мінімізації їх недоліків. Даний підхід дозволить максимізувати ймовірність успішного виявлення поліморфного зловмисного програмного забезпечення.

Ключові слова: зловмисне програмне забезпечення, алгоритм пошуку рядка, інтелектуальний аналіз даних, аналіз в пісочниці, машинне навчання, метод розробки структурних функцій, імовірнісні логічні мережі, комплексний підхід.

CHAIKOVSKYI Maksym
Khmelnytskyi National University

COMPREHENSIVE APPROACH TO THE DETECTION AND ANALYSIS OF POLYMORPHIC MALWARE

The article examines the features of modern polymorphic malware and its impact on the functioning of computer systems. Existing approaches and methods of its detection and analysis are considered, such as: string search algorithm, intelligent data analysis, sandbox analysis, machine learning, method of developing structural functions. Their advantages and disadvantages are determined. The necessity of using a new approach, namely the detection of malicious software using probabilistic logical networks, is argued. Its advantages and development prospects are determined. In the study, a comprehensive approach consisting of 3 stages is proposed for the detection of polymorphic malware. The first one uses string search algorithms. The second is a complex of methods, including intelligent data analysis, sandbox analysis, machine learning, and the method of developing structural functions. In the third step, the use of probabilistic logical networks is proposed, which will allow establishing the probability that the software belongs to polymorphic malware. The use of the proposed integrated approach will also allow to determine the necessary methods for neutralization of detected malicious software. This approach will maximize the probability of detecting polymorphic malware.

Keywords: malicious software, string search algorithm, intelligent data analysis, sandbox analysis, machine learning, structural function development method, probabilistic logic networks, complex approach.

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Пошук та знешкодження комп'ютерних вірусів з кожним роком стає все більш актуальною та складною проблемою, адже вони несуть загрозу безперешкодному функціонуванню комп'ютерних систем, які використовуються у все більш критичних сферах діяльності людства. Тому розробка методів та засобів знешкодження зловмисного програмного забезпечення (ЗПЗ) є одним із перспективних та пріоритетних завдань досліджень у сфері комп'ютерних наук. Незважаючи на постійне вдосконалення антивірусного програмного забезпечення, генерація та поширення ЗПЗ збільшується з року в рік. Одна з найсерйозніших проблем, з якою стикається розробники антивірусного програмного забезпечення - це автоматична мутація коду зловмисної програми. Механізм мутації та перестановки коду зловмисної програми називається поліморфізмом. Поліморфне ЗПЗ неможливо ідентифікувати сигнатурним аналізом. Тому для цього необхідно використовувати нові, удосконалені методи аналізу сучасного ЗПЗ, а також комплексне поєднання існуючих методів та підходів.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Серед науковців, які досліджували питання виявлення та аналізу зловмисного програмного забезпечення, можна виділити наступних: Алазаб М. [1], Чень Х. [3], Рад Б. [4], Андерсон Б. [5], Більге Л. [6], Дауд Е. [7], Христодореску М. [10] та ін.

Серед новітніх методів аналізу сучасного ЗПЗ є деякі алгоритми штучного інтелекту (машинного навчання) які аналізують шкідливу програму у віртуальній машині (VM). Віртуальна машина може запускати запакований потенційно небезпечний файл і динамічно його аналізувати, автоматично тестуючи код та поведінку. Крім того, виглядають перспективними останні дослідження, де антивірусне програмне

забезпечення (ПЗ) використовує сучасні методи машинного навчання та аналіз поведінки у режимі реального часу у комплексі зі статичними методами для ідентифікації підозрілої активності та запобігання загрозам [1]. Такий підхід до виявлення ЗПЗ називається гібридним. Про важливість та актуальність теми захисту від ЗПЗ свідчать і статистичні дані. Так, за інформацією статистичної компанії Statistica, кількість кібератак на комп'ютерні системи невпинно зростає з року в рік, що відображено на (рис.1), а кількість атак на комп'ютерні системи за видами ЗПЗ на рис. 2.

Поліморфне ЗПЗ — це тип вірусів, які можуть змінювати свій код, зберігаючи основну функціональність. Ці віруси зазвичай мають механізм мутації на основі методів обфускації коду, упаковки та метаморфізму, який може шифрувати або розшифровувати код вірусу, щоразу створюючи унікальний програмний код [3]. Ця адаптивна поведінка робить статичні методи виявлення на основі сигнатур неефективними, оскільки код ЗПЗ відрізняється з кожною ітерацією інфікування. Таким чином, потреба в динамічних і проактивних методах виявлення та знешкодження для протидії поліморфному ЗПЗ стала більш важливою, ніж будь-коли. Поліморфні віруси використовують кілька адаптивних стратегій, щоб гарантувати, що вони не будуть виявлені та знешкоджені. Однією з найпоширеніших стратегій є шифрування коду за допомогою унікальних алгоритмів шифрування [4]. Це шифрування ускладнює виявлення вірусу антивірусним ПЗ, оскільки він виглядає як нешкідливий файл. Загальновідома блок-схема виявлення поліморфного ЗПЗ показана на рис 3.

Крім того, вірус може використовувати програму розпакування, яка запускається лише під час відкриття файлу, що ускладнює його виявлення. Нарешті, поліморфне ЗПЗ часто використовує методи антианалізу, щоб перешкодити спробам зворотного проєктування. Це може включати такі методи, як обфускація коду, процедури запобіганню зворотної розробки та інші [5]. Завдяки застосуванню цих методів, поліморфне ЗПЗ стає ще більш невловимим, що робить виявлення та аналіз досить складним завданням.

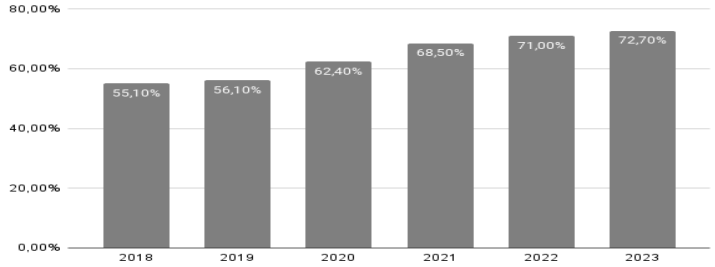

Рис. 1. Зростання кількості кібератак по рокам в млн [2]

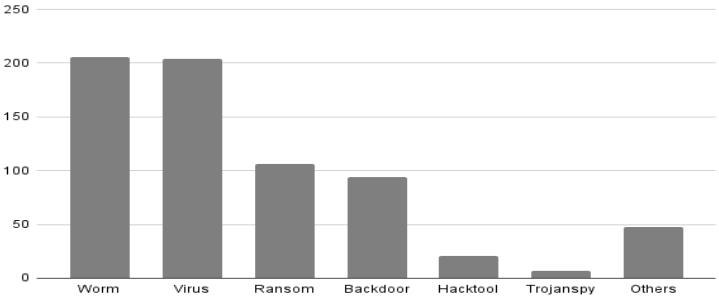

Рис. 2. Статистика кількості кібератак за видами ЗПЗ [2]

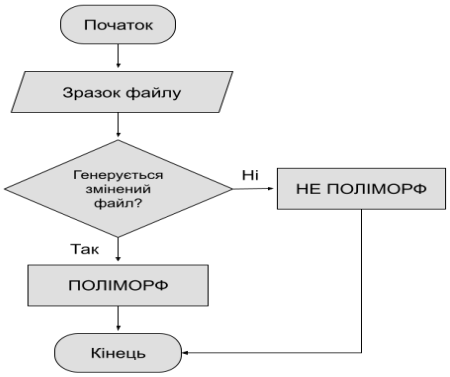

Рис. 3. Блок-схема виявлення поліморфного ЗПЗ

Виявлення поліморфного ЗПЗ вимагає використання комбінації методів статичного та динамічного аналізу [6]. Хоча статичний аналіз може дати початкове уявлення про поведінку ЗПЗ, він часто є неефективним через швидку зміну коду поліморфного ЗПЗ. Тому методи динамічного аналізу є важливими для ефективного виявлення та знешкодження загрози. Динамічний аналіз передбачає запуск ЗПЗ в контрольованому середовищі, наприклад віртуальній машині або пісочниці, для спостереження за його поведінкою [7]. Відстежуючи системні дії, модифікації файлів, мережеві підключення та інші показники, аналітики безпеки можуть ідентифікувати підозрілу поведінку та відповідним чином класифікувати ЗПЗ [8]. Для покращення можливостей виявлення часто використовуються методи аналізу поведінки. Ці методи включають моніторинг поведінки файлу під час виконання, аналіз його дій та оцінку ризиків, який він становить. Порівнюючи поведінку потенційно зловмисного виконуваного файлу з відомими шаблонами та евристиками, засоби безпеки можуть швидко ідентифікувати екземпляри поліморфного ЗПЗ. Крім того, алгоритми машинного навчання відіграють важливу роль у виявленні поліморфного ЗПЗ. Навчаючись на моделях та великих наборах даних відомого ЗПЗ, ці алгоритми навчаються ідентифікувати шкідливі файли та відрізняти поліморфне ЗПЗ від законного ПЗ. Цей підхід забезпечує ефективне та масштабоване рішення для боротьби зі все більш зростаючою загрозою поліморфного ЗПЗ. Оскільки поліморфне ЗПЗ продовжує розвиватися та ухилятися від традиційних методів виявлення та знешкодження, то реалізація ефективних контрзаходів стає все більш нагальною потребою. Нездатність зменшити загрозу від поліморфного ЗПЗ може призвести до катастрофічних наслідків, таких як витік даних, фінансові втрати та репутаційна шкода.

Метод пошуку ЗПЗ — це ефективний метод, який використовується в кібербезпеці для виявлення потенційного ЗПЗ в системі [9]. Він передбачає сканування двійкового коду або коду програми для пошуку певних рядків даних, які зазвичай асоціюються зі ЗПЗ.

Одним із найпоширеніших інструментів для пошуку рядків є команда strings у системах на базі Unix. Ця команда сканує файл і виводить будь-які послідовності друкованих символів, які часто вказують на зрозумілі людині рядки в коді програми.

У контексті виявлення шкідливих програм ці рядки можуть надати цінну інформацію про потенційну поведінку підозрілого файлу. Наприклад, вони можуть виявити наявність підозрілих викликів API, шляхів до файлів, URL-адрес або ключів реєстру, які часто пов’язані зі зловмисною діяльністю.

Однак пошук рядків не є надійним методом. Досвідчені автори ЗПЗ часто використовують методи обфускації, щоб приховати свої рядки, або вони можуть взагалі уникати використання підозрілих рядків. Крім того, легальні програми також можуть містити підозрілі на вигляд рядки випадково.

Таким чином, хоча пошук рядків може бути корисним першим кроком у аналізі ЗПЗ, важливо підтвердити результати за допомогою інших методів. Це може включати динамічний аналіз (спостереження за поведінкою програми під час виконання), статичний аналіз (перевірка коду програми без її запуску) або евристичний аналіз (порівняння поведінки програми або шаблонів коду з відомими сигнатурами ЗПЗ).

Таким чином, метод пошуку рядків ЗПЗ є цінним інструментом в арсеналі аналітиків з кібербезпеки, але його слід використовувати як частину ширшого, комплексного підходу до виявлення та аналізу ЗПЗ.

Одним із найбільш перспективних способів виявлення ЗПЗ є використання **методів аналізу даних**. Ці методи включають аналіз великих наборів даних для виявлення шаблонів, асоціацій або аномалій, які можуть вказувати на зловмисну діяльність [10].

Першим кроком у методі виявлення інтелектуального аналізу даних є збір даних. Це передбачає збір широкого діапазону даних, таких як журнали мережевого трафіку, відстеження системних викликів і дії користувачів. Дані можуть бути зібрані з однієї машини або мережі комп’ютерів для більш широкого аналізу.

Коли дані зібрані, їх часто попередньо обробляють, щоб перетворити їх у відповідний формат для аналізу даних. Наприклад, необроблені дані, можливо, потрібно буде перетворити в числовий формат або відфільтрувати нерелевантні дані.

Потім, попередньо оброблені дані піддаються алгоритмам інтелектуального аналізу даних. Існує кілька типів методів інтелектуального аналізу даних, які можна використовувати, включаючи класифікацію, кластеризацію, регресію та виявлення аномалій. Ці методи можуть допомогти виявити шаблони або аномалії, які можуть свідчити про наявність ЗПЗ.

Нарешті, результати можуть бути представлені у форматі, який аналітики з комп'ютерної безпеки легко інтерпретують, наприклад, візуальна панель або система сповіщень.

Класифікація, наприклад, передбачає навчання моделі розпізнаванню характеристик відомого ЗПЗ, а потім використання цієї моделі для класифікації нових даних як безпечних або шкідливих. З іншого боку, кластеризація групує подібні дані разом, що може допомогти виявити шаблони в даних, які можуть вказувати на атаку.

Після процесу інтелектуального аналізу даних результати часто піддаються постобробці, щоб видалити будь-які хибні позитивні чи негативні результати. Це може включати перехресну перевірку результатів за допомогою інших методів виявлення або ручну перевірку виявлення ЗПЗ.

Варто зазначити, що хоча видобуток даних може бути потужним інструментом для виявлення ЗПЗ, він не безпомилковий. Іноді він може видавати хибні спрацювання чи давати негативну відповідь. Тому він може виявити не всі типи ЗПЗ. Однак у поєднанні з іншими методами виявлення інтелектуальний аналіз даних може значно підвищити здатність системи виявляти загрози ЗПЗ та реагувати на них.

Аналіз ЗПЗ в пісочниці – це техніка, яка використовується фахівцями з кібербезпеки для аналізу та розуміння поведінки ЗПЗ в контрольованому середовищі [11]. Він передбачає запуск ЗПЗ у віртуальному або ізольованому середовищі, відомому як пісочниця, для спостереження за його діями та збору цінної інформації.

Метою аналізу ЗПЗ є розкриття можливостей ЗПЗ, ідентифікація потенційних загроз і розробка ефективних заходів протидії. Виконуючи ЗПЗ в контрольованому середовищі, аналітики можуть вивчати його взаємодію з операційною системою, мережею та іншими компонентами ПЗ.

Під час аналізу використовуються різні динамічні та статичні методики. Динамічний аналіз включає моніторинг поведінки ЗПЗ під час виконання, наприклад, модифікації файлової системи, мережевий зв'язок і системні виклики. З іншого боку, статичний аналіз зосереджується на дослідженні коду та структури ЗПЗ без виконання.

Інформація, зібрана в результаті аналізу поведінки зловмисної програми в ізольованому програмному середовищі, допомагає визначити вектори зараження, інфраструктуру та методи керування, механізми доставки корисного навантаження та потенційні методи викрадення даних. Ці знання мають вирішальне значення для розробки ефективних методів виявлення, оновлення елементів керування безпекою та пом'якшення впливу атак шкідливих програм.

Підсумовуючи, аналіз в ізольованому програмному середовищі є важливою складовою сучасних практик кібербезпеки. Він надає цінну інформацію про поведінку та характеристики ЗПЗ, дозволяючи організаціям в сфері кібербезпеки покращувати свої механізми захисту та розробляти перспективні методи протидії новим загрозам.

Традиційним методам виявлення ЗПЗ часто важко йти в ногу зі стрімко розвиваючим ландшафтом атак ЗПЗ. Методи машинного навчання стали потужним інструментом для покращення виявлення ЗПЗ та боротьби з цими загрозами.

Алгоритми машинного навчання можуть аналізувати великі обсяги даних і витягувати шаблони та функції, які можна використовувати для виявлення зловмисної поведінки. Навчаючи моделі на відомих зразках ЗПЗ та законному ПЗ, алгоритми машинного навчання можуть навчитися розрізняти їх і точно класифікувати нові та невідомі файли.

Однією з ключових переваг використання машинного навчання для виявлення ЗПЗ є його здатність адаптуватися та вчитися на нових загрозах. У міру появи нових типів ЗПЗ моделі машинного навчання можна оновлювати та перенавчати для ефективного виявлення цих нових загроз.

Існує кілька підходів до виявлення ЗПЗ за допомогою машинного навчання, включаючи статичний аналіз і динамічний аналіз. Статичний аналіз передбачає перевірку коду та структури файлу без його виконання, тоді як динамічний аналіз передбачає запуск файлу в контрольованому середовищі для спостереження за його поведінкою. Обидва підходи можуть надати цінну інформацію для виявлення ЗПЗ.

Чезаре та Сян запропонували метод класифікації поліморфного ЗПЗ під назвою Malwise, який використовує емуляцію на рівні програми для розпакування коду ЗПЗ [12].

Однак, важливо зазначити, що виявлення ЗПЗ за допомогою машинного навчання не без труднощів. Змагальні атаки, коли зловмисники маніпулюють ЗПЗ, щоб уникнути виявлення, можуть становити значну проблему. Крім того, великий обсяг даних і необхідність постійного оновлення та перенавчання моделей вимагають значних обчислювальних ресурсів.

Підсумовуючи, машинне навчання пропонує багатообіцяючі рішення для виявлення ЗПЗ, використовуючи свою здатність аналізувати величезні обсяги даних і виявляти шаблони. Постійно вдосконалюючи та оновлюючи моделі, машинне навчання може підвищити безпеку комп'ютерних систем і мереж від нових загроз ЗПЗ.

Розробка структурних функцій є ключовим аспектом розробки ефективних моделей виявлення ЗПЗ [13]. Витягаючи значущі функції зі структурованих даних, аналітики та дослідники даних можуть підвищити точність і надійність своїх систем виявлення ЗПЗ. У наступних кроках описано метод розробки структурних функцій, спеціально розроблений для виявлення ЗПЗ:

1. Розуміння даних: отримання повного розуміння структури та характеристик даних ЗПЗ. Визначення відповідних змінних, їх типів та будь-які шаблони чи зв'язки, присутні в наборі даних.
2. Ідентифікація функцій: визначення функцій, які можуть бути інформативними для виявлення ЗПЗ. Цього можна досягти за допомогою знання домену, дослідницького аналізу даних або статистичних методів, спеціально розроблених для виявлення ЗПЗ.
3. Вилучення функцій: витягнення вибраних функцій з необроблених даних ЗПЗ та перетворення їх у відповідний формат для аналізу. Застосування методів математичних перетворень, масштабування, нормалізації або кодування для попередньої обробки функцій.

4. Побудова функцій: створення нових функцій, поєднуючи або змінюючи існуючі функції таким чином, щоб охопити важливі аспекти поведінки ЗПЗ. Це може включати агрегації, математичні операції або взаємодію між змінними.

5. Вибір функцій: обрання найбільш релевантних функцій, які значною мірою сприяють виявленню ЗПЗ. Це допомагає зменшити розмірність і підвищити ефективність і точність моделі виявлення.

6. Кодування ознак: кодування категоріальних ознак в числові представлення, які можуть бути оброблені алгоритмами машинного навчання. Використання таких методів, як одноразове кодування, кодування міток або цільове кодування для ефективного представлення категоріальних змінних.

7. Масштабування функцій: масштабувати функції до загального діапазону, щоб переконатися, що вони мають порівнювані величини. Для цього можна використовувати методи стандартизації та нормалізації.

8. Перевірка функцій: перевірка розроблених функцій, оцінивши їх продуктивність у моделі виявлення ЗПЗ. Використання таких методів, як перехресна перевірка та показники оцінки моделі, щоб виміряти ефективність розроблених функцій і за необхідності їх ітеративно вдосконалювати.

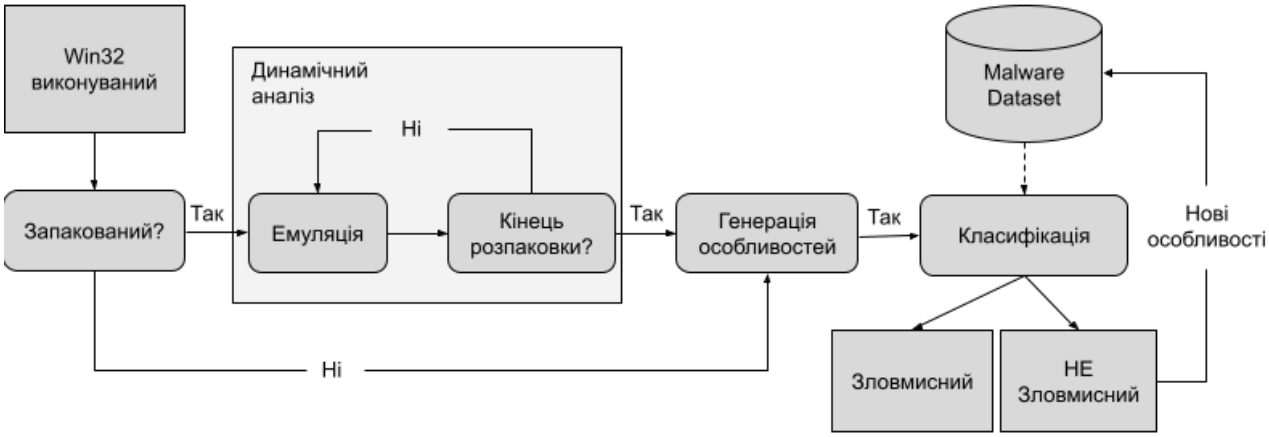

Рис.4. Блок-схема системи класифікації ЗПЗ [12]

Дотримуючись цього методу розробки структурних особливостей, аналітики та дослідники даних можуть підвищити точність і надійність своїх систем виявлення ЗПЗ, що призведе до покращених заходів кібербезпеки та захисту від ЗПЗ.

Недоліки розглянутих методів вимагають нових підходів до виявлення та аналізу ЗПЗ. Серед них - виявлення ЗПЗ за допомогою імовірнісних логічних мереж (PLN).

ВИДІЛЕННЯ НЕВИРІШЕНИХ РАНІШЕ ЧАСТИН ЗАГАЛЬНОЇ ПРОБЛЕМИ, КОТРИМ ПРИСВЯЧУЄТЬСЯ СТАТТЯ

Зважаючи на значну кількість досліджень у даному напрямку [1-15], питання комплексного підходу до виявлення та аналізу поліморфного ЗПЗ вивчено недостатньо та вимагає подальших досліджень.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою статті є розробка комплексного підходу до виявлення та аналізу поліморфного ЗПЗ.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Виявлення ЗПЗ є критично важливим аспектом кібербезпеки. PLN [14] пропонують потужний підхід до виявлення та пом’якшення загроз ЗПЗ. PLN поєднують імовірнісне міркування з логічним висновком для моделювання складних взаємозв’язків і залежностей у виявленні ЗПЗ.

PLN — це гібридна структура, яка об’єднує ймовірнісні графічні моделі з логікою першого порядку. Вони забезпечують гнучке та виразне представлення для фіксації невизначеності та міркування про складні сфери. PLN використовують сильні сторони як імовірнісного міркування, так і логічного висновку, що робить їх придатними для виявлення ЗПЗ.

Однією з ключових переваг PLN у виявленні ЗПЗ є їх здатність обробляти невизначену та неповну інформацію. Призначаючи ймовірності різним гіпотезам, PLN можуть оцінювати ймовірність присутності ЗПЗ та приймати обґрунтовані рішення. Це ймовірнісне міркування дозволяє використовувати більш точні та адаптивні механізми виявлення.

PLN чудово вловлюють складну поведінку та моделі шкідливих програм. Вони можуть представляти як статичні, так і динамічні характеристики ЗПЗ, включаючи структуру коду, взаємодію

системи та механізми розповсюдження. Моделюючи таку поведінку, PLN можуть ефективно відрізняти законне ПЗ від ЗПЗ.

Щоб навчити PLN виявляти ЗПЗ, потрібен великий набір даних відомих зразків ЗПЗ та доброякісного ПЗ. Для вивчення параметрів і структури PLN з цих даних можна використовувати методи машинного навчання. Ітеративно удосконалюючи PLN за допомогою навчальних прикладів, його можна налаштувати для точного виявлення та класифікації ЗПЗ.

Переваги PLN для виявлення ЗПЗ:

– гнучкість: PLN забезпечують гнучку структуру для моделювання та обґрунтування поведінки ЗПЗ, дозволяючи адаптуватися до нових загроз;

– обробка невизначеності: імовірнісний характер PLN дає змогу обробляти невизначену та неповну інформацію, підвищуючи точність виявлення ЗПЗ;

– виразність: PLN можуть фіксувати складні зв’язки та залежності, виявлені у ЗПЗ, забезпечуючи більш комплексні можливості виявлення;

– навчання з даних: PLN можна навчити за допомогою методів машинного навчання, що дозволяє постійно вдосконалюватись на основі нових зразків ЗПЗ.

Проблеми в PLN для виявлення ЗПЗ:

– масштабованість: зі збільшенням складності ЗПЗ і розміру наборів даних масштабування PLN для обробки великомасштабного виявлення стає проблемою;

– розробка знань: створення бази знань і визначення логічних правил для виявлення ЗПЗ вимагає досвіду та знань у галузі;

– обчислювальна складність: виконання висновків і навчання в PLN може бути вимогливим до обчислень, вимагаючи ефективних алгоритмів і систем.

Отже, у дослідженні для виявлення поліморфного ЗПЗ [15] пропонується комплексний підхід (рис.5), який складається з 3-х етапів. На першому використовуються алгоритми пошуку рядка. На другому – комплекс методів, серед яких інтелектуальний аналіз даних, аналіз в пісочниці, машинне навчання, метод розробки структурних функцій. На третьому кроці пропонується використання PLN, які дозволять встановити ймовірність належності ПЗ до поліморфного ЗПЗ. Використання запропонованого комплексного підходу також дозволить визначити необхідні методи для знешкодження виявленого ЗПЗ.

Експерименти

Для визначення ефективності запропонованої методики була проведена серія експериментів. Для отримання модифікованих поліморфних версій вірусів, взятих з [16], використовували різні типи поліморфних генераторів. Усі поліморфні версії, створені ними генератори були скомпільовані з опціями антиналагодження та антиемуляції. Для проведення першого експерименту було згенеровано 100 вірусів. Щоб оцінити ефективність запропонованої методики, визначалося відсоткове співвідношення виявлених вірусів на кожному кроці запропонованого у дослідженні комплексного підходу.

Результати проведеного експерименту відображені у таблиці 1.

Отже, на кроці 1 було виявлено лише 12 % вірусів, на кроці 2 – 61 %, а на кроці 3 із використанням PLN – 89 %. Ефективність запропонованої методики згідно проведеного експерименту становить 28 % завдяки використанню PLN. Також з 89 % виявлених вірусів за допомогою PLN 9 % було віднесено у діапазон ймовірності належності до ЗПЗ на рівні 0-25 % (низький рівень), на рівні 25-75 % (середній рівень) - 19 %, на рівні 75-100 % - 72 % (високий рівень). Використання PLN дозволило не лише збільшити ефективність виявлення ЗПЗ, але й класифікувати за рівнем ймовірності належності до ЗПЗ.

Таблиця 1.

Відсоткове співвідношення виявлених вірусів на кожному кроці запропонованого комплексного підходу

Кількість згенерованих вірусів	Відсоток виявлених вірусів алгоритмами пошуку рядка (крок 1)	Відсоток виявлених вірусів методами кроку 2	Відсоток виявлених вірусів із використанням PLN (крок 3)	Діапазон ймовірність належності вірусів до поліморфного ЗПЗ	Кількість вірусів у діапазоні ймовірності належності до ЗПЗ
100	12 %	61 %	89 %	0-25 % (low)	9 %
				25-75 % (medium)	19 %
				75-100 % (high)	72 %

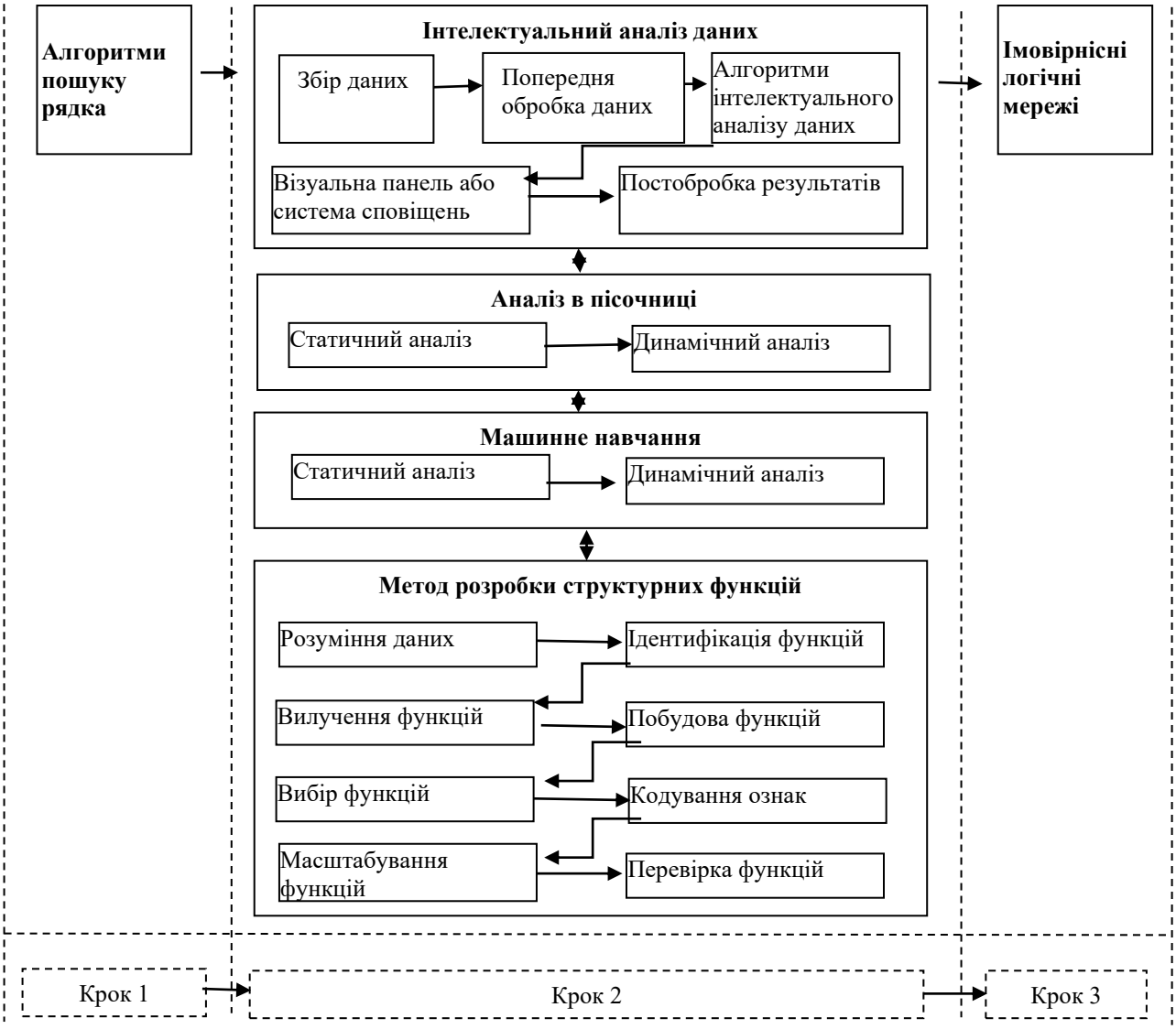

Рис.5. Комплексний підхід до виявлення та аналізу поліморфного ЗПЗ

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У дослідженні запропоновано комплексний підхід до виявлення та аналізу поліморфного ЗПЗ. Даний підхід складається з трьох етапів. На першому використовуються алгоритми пошуку рядка. На другому – комплекс методів, серед яких інтелектуальний аналіз даних, аналіз в пісочниці, машинне навчання, метод розробки структурних функцій. На третьому кроці пропонується використання PLN, які дозволять встановити ймовірність належності ПЗ до поліморфного ЗПЗ. Ефективність запропонованої методики згідно проведеного експерименту становить 28 % завдяки використанню PLN. Використання PLN дозволило не лише збільшити ефективність виявлення ЗПЗ, але й класифікувати за рівнем ймовірності належності до ЗПЗ.

Література

1. Alazab M. Malware Detection Based on Structural and Behavioral Features of API Calls / M. Alazab, R. Layton, S. Venkataraman, P. Watters Proceedings of the 1-st International Cyber Resilience Conference. – Perth Weste (Australia), August 23, 2010. – Pp. 1–8.
2. https://www.statista.com/
3. Chen X. Towards an Understanding of Anti-virtualization and Antidebugging Behavior in Modern Malware / X. Chen, J. Andersen, Z. M. Mao, M. Bailey, J. Nazario // Proceedings of the 5th IEEE International Conference on Malicious and Unwanted Software. – Anchorage (USA), June 24-27, 2008. – Pp. 177-186.
4. Rad B.B. Camouflage in Malware: From Encryption to Metamorphism / B.B. Rad, M. Masrom, S. Ibrahim // International Journal of Computer Science and Network Security. – 2012. – Vol. 12. – Pp. 74-83.

5. Anderson B. OS fingerprinting: New techniques and a study of information gain and obfuscation / B. Anderson, D. McGrew // 2017 IEEE Conference on Communications and Network Security (CNS). – 9.10.2017. – Pp. 1–9.

6. Bilge L. Riskteller: Predicting the risk of cyber incidents / L. Bilge, Y. Han, M. Dell'Amico // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (Dallas, Texas, USA). — October 30, 2017. – Pp. 1299-1311.

7. Daoud E. A. Computer Virus Strategies and Detection Methods / E.A. Daoud, I.H. Jebril, B. Zaqaibeh // International Journal of Open Problems in Computer Science and Mathematics. – 2008. – Vol. 1. – No. 2. – Pp. 29-36.

8. Pomorova O. A Technique for detection of bots which are using polymorphic code / O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, A. Nicheporuk // Communications in Computer and Information Science. – 2014. – Vol. 431. – Pp. 265-276.

9. Christodorescu M. Semantics-aware malware detection / M. Christodorescu, S. Jha, S. A. Seshia, D. Song, R. E. Bryant // Proceedings of the 15-th IEEE Symposium on Security and Privacy. – Oakland (USA), May 08–11, 2005. – Pp. 32-46.

10. Santos I. Opcode sequences as representation of executables for data mining-based unknown malware detection / I. Santos, F. Brezo, X. Ugarte-Pedrero, P.G. Bringas // Information Sciences. – 2013. – Vol. 231. – Pp. 64-82.

11. Bhatia J.S. Botnet Command Detection using Virtual Honeynet / J.S. Bhatia, R.K. Sehgal, S. Kumar // International Journal of Network Security & Its Applications. – 2011. – Vol. 3. – No. 5. – Pp. 177-189.

12. Cesare S. (2010). Classification of malware using structured control flow / S. Cesare, Y. Xiang // Proc. 8th Australasian Symposium on Parallel and Distributed Computing (AusPDC 2010), Brisbane, Australia. – 2010. URL: https://dl.acm.org/doi/pdf/10.5555/1862294.1862301

13. Structural Feature Engineering approach for detecting polymorphic malware / E. Masabo, K.S. Kaawaase, J. Sansa-Otim, D. Hanyurwimfura // IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing. – 2017. URL: https://www.academia.edu/76286606/Structural_Feature_Engineering_Approach_for_Detecting_Polymorphic_Malware?auto=download

14. Qu M. Probabilistic Logic Neural Networks for Reasoning / M. Qu, J. Tang // 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada. – 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/13e5ebb0fa112fe1b31a1067962d74a7-Paper.pdf

15. Savenko O. Approach for the Unknown Metamorphic Virus Detection / O. Savenko, S. Lysenko, A. Nicheporuk, B. Savenko // Proceedings of the 8-th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. – Bucharest (Romania), September 21-23, 2017. – Pp. 71-76.

16. VX Heavens [online] URL: http://vxheaven.org/

References

1. Alazab M. Malware Detection Based on Structural and Behavioral Features of API Calls / M. Alazab, R. Layton, S. Venkataraman, P. Watters Proceedings of the 1-st International Cyber Resilience Conference. – Perth Weste (Australia), August 23, 2010. – Pp. 1–8.

2. https://www.statista.com/

3. Chen X. Towards an Understanding of Anti-virtualization and Antidebugging Behavior in Modern Malware / X. Chen, J. Andersen, Z. M. Mao, M. Bailey, J. Nazario // Proceedings of the 5th IEEE International Conference on Malicious and Unwanted Software. – Anchorage (USA), June 24-27, 2008. – Pp. 177-186.

4. Rad B.B. Camouflage in Malware: From Encryption to Metamorphism / B.B. Rad, M. Masrom, S. Ibrahim // International Journal of Computer Science and Network Security. – 2012. – Vol. 12. – Pp. 74-83.

5. Anderson B. OS fingerprinting: New techniques and a study of information gain and obfuscation / B. Anderson, D. McGrew // 2017 IEEE Conference on Communications and Network Security (CNS). – 9.10.2017. – Pp. 1–9.

6. Bilge L. Riskteller: Predicting the risk of cyber incidents / L. Bilge, Y. Han, M. Dell'Amico // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (Dallas, Texas, USA). — October 30, 2017. – Pp. 1299-1311.

7. Daoud E. A. Computer Virus Strategies and Detection Methods / E.A. Daoud, I.H. Jebril, B. Zaqaibeh // International Journal of Open Problems in Computer Science and Mathematics. – 2008. – Vol. 1. – No. 2. – Pp. 29-36.

8. Pomorova O. A Technique for detection of bots which are using polymorphic code / O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, A. Nicheporuk // Communications in Computer and Information Science. – 2014. – Vol. 431. – Pp. 265-276.

9. Christodorescu M. Semantics-aware malware detection / M. Christodorescu, S. Jha, S. A. Seshia, D. Song, R. E. Bryant // Proceedings of the 15-th IEEE Symposium on Security and Privacy. – Oakland (USA), May 08–11, 2005. – Pp. 32-46.

10. Santos I. Opcode sequences as representation of executables for data mining-based unknown malware detection / I. Santos, F. Brezo, X. Ugarte-Pedrero, P.G. Bringas // Information Sciences. – 2013. – Vol. 231. – Pp. 64-82.

11. Bhatia J.S. Botnet Command Detection using Virtual Honeynet / J.S. Bhatia, R.K. Sehgal, S. Kumar // International Journal of Network Security & Its Applications. – 2011. – Vol. 3. – No. 5. – Pp. 177-189.

12. Cesare S. (2010). Classification of malware using structured control flow / S. Cesare, Y. Xiang // Proc. 8th Australasian Symposium on Parallel and Distributed Computing (AusPDC 2010), Brisbane, Australia. – 2010. URL: https://dl.acm.org/doi/pdf/10.5555/1862294.1862301

13. Structural Feature Engineering approach for detecting polymorphic malware / E. Masabo, K.S. Kaawaase, J. Sansa-Otim, D. Hanyurwimfura // IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing. – 2017. URL: https://www.academia.edu/76286606/Structural_Feature_Engineering_Approach_for_Detecting_Polymorphic_Malware?auto=download

14. Qu M. Probabilistic Logic Neural Networks for Reasoning / M. Qu, J. Tang // 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada. – 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/13e5ebb0fa112fe1b31a1067962d74a7-Paper.pdf

15. Savenko O. Approach for the Unknown Metamorphic Virus Detection / O. Savenko, S. Lysenko, A. Nicheporuk, B. Savenko // Proceedings of the 8-th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. – Bucharest (Romania), September 21-23, 2017. – Pp. 71-76.

16. VX Heavens [online] URL: http://vxheaven.org/