УДК 004.056.53

А. О. Підлісна
А. В. Грицак

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПОВЕДІНКОВОГО ВИЯВЛЕННЯ ПРИХОВАНИХ ПРОЦЕСІВ

Вінницький національний технічний університет

Анотація

Досліджено проблему виявлення прихованих процесів операційної системи, які можуть бути пов'язані з діяльністю шкідливого програмного забезпечення класу руткіт. Такі програми характеризуються здатністю маскувати власну присутність у системі, змінювати поведінку процесів та приховувати системні ресурси. Традиційні сигнатурні методи захисту є недостатньо ефективними для виявлення нових або модифікованих зразків шкідливого програмного забезпечення. Запропоновано використання поведінкового підходу до виявлення аномалій на основі алгоритму Isolation Forest. Розглянуто архітектуру системи, яка включає модуль збору телеметрії процесів операційної системи, модуль машинного навчання для оцінки аномальності поведінки процесів та аналітичний модуль для інтерпретації результатів. Показано, що використання алгоритму Isolation Forest дозволяє ефективно виявляти нетипову поведінку процесів без використання сигнатурних баз.

Ключові слова: інформаційна безпека, руткіт, машинне навчання, виявлення аномалій, Isolation Forest, поведінковий аналіз.

Abstract

This study examines the problem of detecting hidden operating system processes that may be associated with the activity of rootkit-class malware. Such programs are characterized by their ability to mask their presence in the system, alter process behavior, and conceal system resources. Traditional signature-based protection methods are insufficiently effective for detecting new or modified malware samples. The use of a behavioral approach to anomaly detection based on the Isolation Forest algorithm is proposed. The system architecture is discussed, which includes a module for collecting operating system process telemetry, a machine learning module for assessing the abnormality of process behavior, and an analytical module for interpreting the results. It is shown that the use of the Isolation Forest algorithm allows for the effective detection of atypical process behavior without the use of signature databases.

Keywords: information security, rootkit, anomaly detection, machine learning, Isolation Forest, cybersecurity.

Вступ

У сучасних інформаційних системах однією з суттєвих загроз є шкідливе програмне забезпечення, здатне приховувати свою присутність у системі. До таких типів загроз належать руткіти — програмні засоби, що забезпечують прихований доступ до операційної системи та дозволяють маскувати власну діяльність [1].

Більшість існуючих засобів захисту ґрунтується на сигнатурних методах виявлення шкідливого програмного забезпечення. Проте ефективність таких підходів значно знижується у випадку появи нових або модифікованих зразків шкідливого коду, зокрема при виникненні zero-day загроз. У зв'язку з цим актуальності набувають поведінкові методи аналізу, що базуються на застосуванні алгоритмів машинного навчання.

Результати дослідження

Метою роботи є дослідження можливості застосування алгоритму Isolation Forest для поведінкового виявлення прихованих або підозрілих процесів операційної системи на основі аналізу їхніх характеристик виконання.

Проблема виявлення шкідливого програмного забезпечення залишається однією з ключових у сфері інформаційної безпеки. Особливу складність становлять руткіти – програмні засоби, призначені для прихованого доступу до комп'ютерної системи та маскування власної активності. Руткіти можуть приховувати процеси, файли, мережеві з'єднання та інші системні ресурси, що значно ускладнює їх виявлення традиційними засобами захисту.

Більшість сучасних антивірусних систем використовує сигнатурні методи виявлення загрози, які базуються на пошуку відомих шаблонів шкідливого коду. Однак такі підходи мають обмежену ефективність у випадку нових або модифікованих атак, зокрема zero-day загроз. У зв'язку з цим актуальним напрямом досліджень є застосування методів машинного навчання для поведінкового аналізу системи [2].

Одним із ефективних підходів до виявлення аномалій є алгоритм Isolation Forest, запропонований для пошуку аномальних об'єктів у великих наборах даних. Основна ідея алгоритму полягає у тому, що аномальні спостереження ізолюються значно швидше, ніж нормальні дані. Алгоритм будує множину випадкових дерев рішень, у яких дані рекурсивно розділяються випадковим вибором ознаки та порогового значення [3].

Для кожного об'єкта визначається довжина шляху від кореня дерева до вузла, у якому цей об'єкт ізольовано. Чим коротший цей шлях, тим більш імовірно, що об'єкт є аномальним.

У межах дослідження було запропоновано систему поведінкового аналізу процесів операційної системи, яка дозволяє виявляти потенційно приховані або підозрілі процеси. Архітектура запропонованої системи включає декілька основних модулів та представлена на рисунку 1.

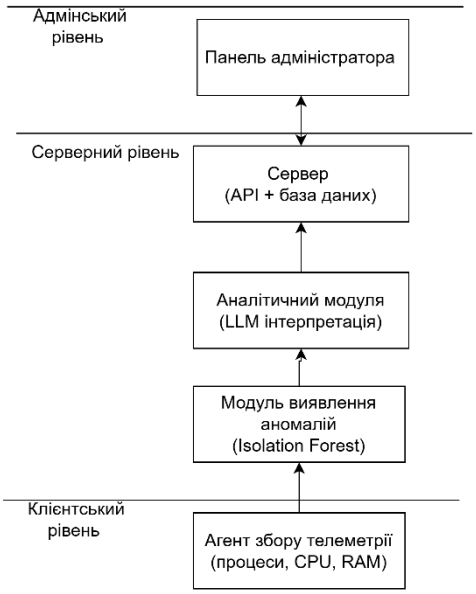

Рисунок 1 – Архітектура системи поведінкового виявлення прихованих процесів

Першим компонентом системи є агент збору телеметрії, який працює на клієнтській системі та здійснює моніторинг процесів операційної системи. Агент збирає основні характеристики виконання процесів.

На основі зібраних даних формується вектор ознак, який використовується для подальшого аналізу поведінки процесів.

До вектора ознак входять основні характеристики виконання процесу, такі як рівень використання процесорних ресурсів, обсяг використаної оперативної пам'яті, кількість потоків процесу, інформація про батьківський процес, а також шлях до виконуваного файлу. Аналіз зазначених параметрів дозволяє формувати поведінковий профіль процесу та виявляти відхилення від нормальної роботи системи.

Другим компонентом є модуль машинного навчання, який реалізує алгоритм Isolation Forest для виявлення аномалій у поведінці процесів. Алгоритм аналізує сформовані вектори ознак та визначає ступінь аномальності кожного процесу.

Модель Isolation Forest попередньо навчається на вибірці нормальної поведінки системи. Для цього протягом певного періоду роботи комп'ютера здійснюється збір телеметричних даних про процеси операційної системи за відсутності відомих шкідливих програм. На основі цих даних формується набір навчальних векторів ознак, що характеризують типову поведінку процесів.

Алгоритм Isolation Forest будує ансамбль випадкових дерев ізоляції, що дозволяє сформувати базову модель нормального функціонування системи. Після навчання модель використовується для

оцінювання нових спостережень. Якщо значення аномального скору перевищує заданий поріг, процес класифікується як потенційно підозрілий.

Отримані результати передаються до аналітичного модуля, який використовує велику мовну модель для інтерпретації результатів машинного навчання. Даний модуль формує текстові пояснення щодо виявлених аномалій, описує можливі причини підозрілої поведінки процесу та надає рекомендації щодо подальшого аналізу або реагування на інцидент інформаційної безпеки [4].

Фінальним компонентом системи є серверний модуль, який реалізує API для взаємодії між компонентами системи та забезпечує збереження результатів аналізу у базі даних. Сервер також виконує функції централізованого зберігання повідомлень про небезпеку та забезпечує доступ до інформації про виявлені інциденти [5].

Для взаємодії з системою передбачено веб-інтерфейс адміністратора, який забезпечує візуалізацію результатів аналізу. Через адміністративну панель користувач може переглядати список виявлених аномалій, детальну інформацію про підозрілі процеси, а також історію інцидентів інформаційної безпеки.

Панель взаємодіє із серверним модулем через API та отримує дані з бази даних системи. Це дозволяє здійснювати централізований моніторинг стану системи та оперативно реагувати на виявлені загрози.

Отже, використання алгоритму Isolation Forest дозволяє виявляти аномальні процеси без застосування сигнатурних методів. Перевагами запропонованого підходу є можливість роботи з нерозміченими даними, висока швидкість обробки інформації та здатність виявляти нові типи шкідливого програмного забезпечення. Подальші дослідження можуть бути спрямовані на розширення набору ознак та інтеграцію системи з інструментами автоматизованого реагування на інциденти.

Висновки

Запропоновано підхід до поведінкового виявлення прихованих процесів операційної системи на основі алгоритму Isolation Forest. Розроблена архітектура системи дозволяє поєднати автоматизоване визначення аномальної поведінки процесів із інтелектуальною інтерпретацією результатів аналізу за допомогою великих мовних моделей.

Розроблена система передбачає збір телеметрії процесів операційної системи, формування векторів ознак, застосування алгоритму машинного навчання для виявлення аномальних процесів, а також подальший аналіз результатів та їх візуалізацію через адміністративний веб-інтерфейс. Такий підхід дозволяє здійснювати централізований моніторинг стану системи та підвищує ефективність виявлення потенційно небезпечної активності.

Запропоноване рішення може використовуватися як додатковий рівень захисту інформаційних систем поряд із традиційними сигнатурними методами, забезпечуючи можливість виявлення нових або модифікованих типів шкідливого програмного забезпечення.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Руткіт | ESET Glossary [Електронний ресурс] – Режим доступу до ресурсу: https://help.eset.com/glossary/uk-UA/rootkits.html

2. Zero day: що таке вразливість нульового дня? [Електронний ресурс] – Режим доступу до ресурсу: https://itedu.center/ua/blog/articles/zero-day/

3. What is Isolation Forest [Електронний ресурс] – Режим доступу до ресурсу: https://www.geeksforgeeks.org/machine-learning/what-is-isolation-forest/

4. Що таке велика мовна модель? [Електронний ресурс] – Режим доступу до ресурсу: https://www.sap.com/ukraine/resources/what-is-large-language-model

5. Інтерфейс прикладного програмування (API) [Електронний ресурс] – Режим доступу до ресурсу: https://data.rada.gov.ua/open/main/api

Підлісна Анна Олександрівна – студентка групи 2КІТС-22б, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, Вінниця, annapidlisna370@gmail.com.

Науковий керівник: **Грицак Анатолій Васильович** – кандидат технічних наук, доцент, доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com.

Pidlisna Anna O. – student of group 2KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, annapidlisna370@gmail.com.

Supervisor: ***Hrytsak Anatoly V****. – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Management and Security of Information Systems Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com.*