ВИКОРИСТАННЯ МЕТРИК ПРОГРАМНОЇ СКЛАДНОСТІ ДЛЯ ВИЯВЛЕННЯ ШКІДЛИВОГО КОДУ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

PhD, доц. О.А. Горносталь, PhD, доц. В.В. Челак, д-р техн. наук, проф. С.Ю. Гавриленко, НТУ "ХПІ", м. Харків

Постійне зростання кількості та складності шкідливого програмного забезпечення залишається актуальною загрозою інформаційного простору. При цьому статистичні дані підтверджують, що 68% інцидентів напряму пов'язані з ризикованими діями самих користувачів через відкриття підозрілого посилання чи вкладення з невідомих джерел, які містили небезпечний скрипт [1]. У цьому контексті актуальною є розробка методів, що дозволять виявляти аномальні властивості програмного коду на основі його структурних характеристик, зокрема метрик складності.

Традиційні методи вирішення цієї проблеми передбачать порівняння сигнатур, тобто націлені на виявлення схожості певних фрагментів коду [2]. Основним недоліком такого підходу є його прив'язка до конкретних частин потенційно шкідливих програм, які можуть бути модифіковані та заплутані зловмисниками для ускладнення процесу їх виявлення. Крім того, така група методів має проблеми з виявленням нових загроз, що в сукупності з потребою аналізувати складні зв'язки призводить до потреби створення нових підходів до виявлення шкідливого програмного забезпечення.

Для вирішення цієї задачі прийнято розглядати підходи, що дозволяють математично зв'язати характеристики програмного коду для подальшої його класифікації. Одним з найперспективніших є напрямок, що передбачає поєднання метрик програмної складності, які опосередковано характеризують сам код та дії, які він виконує, а також технологію машинного навчання [3], що дозволяє виявляти складні закономірності в великій кількості статистичних даних.

Метою роботи є дослідження можливості виявлення шкідливого програмного коду на основі аналізу метрик програмної складності із застосуванням методів машинного навчання та визначення найбільш інформативних метрик для задачі класифікації.

Попередні дослідження демонструють, що шкідливий код, зазвичай, характеризується вищими показниками цикломатичної складності, більшою кількістю вкладених конструкцій порівняно з безпечними програмами, наявністю функцій із нетипово високою кількістю викликів, особливо системних, а також низькою когезією класів.

Популярні метрики складності програмного коду можна умовно поділити на декілька категорій, що представлено у табл. 1. Статистичний

аналіз розглянутих показників дозволяє формувати залежності та потенційно виявляти наявність шкідливого програмного забезпечення.

Таблиця 1. Огляд категорій метрик складності програмного коду

Категорія метрик	**Приклади метрик**	**Можливий вплив на виявлення шкідливого коду**
Структурні	Цикломатична складність	Високі значення можуть свідчити про заплутану логіку, характерну для шкідливих програм
Кількісні	Рядки коду, кількість операторів	Надмірний обсяг функції чи фрагмента може приховувати небезпечний код
Логічні	Глибина вкладеності умов/циклів	Значна вкладеність ускладнює аналіз
ООП-метрики	Зв'язність класів, когезія, вага методів	Низька когезія та висока зв'язність можуть свідчити про підозрілу структуру
Стилістичні	Щільність коментарів	Мінімальна кількість коментарів може вказувати на намір приховати логіку коду

Схематично запропонований метод ідентифікації потенційно небезпечного програмного коду представлено на рис. 1.

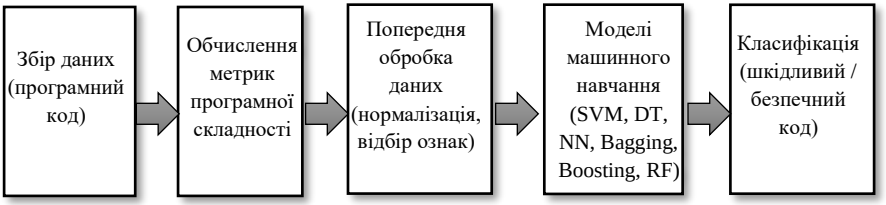

Рис. 1. Схема запропонованого методу

Таким чином, в роботі розглянуті особливості застосування метрик програмної складності у поєднанні з методами машинного навчання для підвищення ефективності систем виявлення шкідливого коду. Такий підхід дозволяє виявляти нові та модифіковані фрагменти шкідливого програмного забезпечення, яке не вдається ідентифікувати сигнатурними методами. Подальші дослідження будуть спрямовані на вивчення ефективності розглянутого методу при використанні різних метрик складності.

Список літератури: 1. Proofpoint, Inc. "2024 State of the Phish Report", 2024. **2.** Лисенко С. М. Аналіз методів виявлення шкідливого програмного забезпечення в комп'ютерних системах / С. М. Лисенко, Р. В. Щука // Вісник Хмельницького національного університету. Технічні науки. – 2020. – № 2. – С. 101-107. **3.** S. Gavrylenko, V. Chelak and O. Hornostal, "Research of Intelligent Data Analysis Methods for Identification of Computer System State," *2020 XXX International Scientific Symposium 'Metrology and Metrology Assurance (MMA)*, Sozopol, Bulgaria, 2020, pp. 1-5.