

ВИКОРИСТАННЯ ЕНТРОПІЇ ДЛЯ ЗНАХОДЖЕННЯ ВНЕСЕННЯ ЗМІН КОДУ ВИКОНУВАНИХ ФАЙЛІВ

Макарова Л.М., к.т.н, доцент¹; Камінський С.С.²; Бризгалов М.В.³

^{1,2,3} Національний університет кораблебудування імені адмірала Макарова

^{1,2,3} Україна, Миколаїв

¹lidiia.makarova@nuos.edu.ua; ²191181@nuos.edu.ua; ³201166@nuos.edu.ua

Анотація. Розроблено програму для обчислення ентропії дизасембльованих виконуваних файлів і порівняння з минулими значеннями з метою виявлення авторства зміни коду виконуваних файлів.

Ключові слова: ентропія, дизасемблювання, виконувані файли, зміна коду, програмний додаток.

Вступна частина. Розвиток сучасних комп'ютерних технологій породжує безліч нових можливостей, оптимізації звичайних процесів та появ нових перспектив. В основі кожного процесу лежить передача даних і виконання коду на комп'ютері, які ми використовуємо щодня і не замислюємося про всю комплексність процесів, що проходять. Чим більшими і складнішими стають ці системи, тим вище стають вимоги до безпеки на кожному етапі і тим більше потенційних слабкостей у них виявляється. Однією з них є ін'єкція коду у виконуваний файл. Саме такі тривіальні речі можуть призвести до величезних втрат, потенційної небезпеки втрати важливих даних та їхнього потрапляння до третіх осіб [1].

Одним із способів вирішення цієї проблеми є обчислення метрик вихідного або виконаного програмного коду. Було розглянуто різні метрики програмного забезпечення з точки зору їх застосування для визначення авторства програмного коду, отже, до виявлення впровадження чужорідного коду.

Так, кількісна метрика – ентропія – добре підходить для визначення авторства програмного коду, тобто, використовуючи дану метрику, можна з великою ймовірністю стверджувати, написана нова ділянка коду автором програми чи ні [2].

Метою роботи є розробка програми для розрахунку ентропії дизасембльованих виконуваних файлів та порівняння з минулими значеннями з ціллю виявлення авторства зміни коду виконуваних файлів.

Для того, щоб досягти поставленої мети потрібно вирішити наступні завдання:

- проаналізувати існуючі методи дизасемблювання та відповідні програми;
- знайти дизасемблер для виконуваних файлів, який задовольняє заданим вимогам;
- розробити програму для аналізу виконуваних файлів та їх дизасембльованого коду;
- розрахувати метрику ентропії для дизасембльованого коду;
- провести аналіз отриманих даних і зробити висновки, базуючись на ентропії, на предмет внесення авторських чи чужорідних змін.

Основна частина. Основна складність полягає у принципі аналізу виконуваних файлів щодо знаходження чужорідного коду. Для цього будемо використовувати значення метрики ентропії.

Для аналізу коду використовується розрахунок ентропії, що дозволяє проаналізувати файли та відрізняти їх по результатам обчислення.

Інформаційна ентропія – міра невизначеності або непередбачуваності інформації, невизначеність появи будь-якого символу первинного алфавіту [3]. За відсутності інформаційних втрат ентропія чисельна дорівнює кількості інформації на символ переданого повідомлення. Наприклад, у послідовності букв, що складають будь-яке речення українською мовою, різні букви з'являються з різною частотою, тому невизначеність появи для деяких букв менша, ніж для інших. Якщо врахувати, що деякі поєднання букв зустрічаються дуже рідко, то невизначеність зменшується ще сильніше.

Ентропія також може бути застосована до програмного забезпечення, як міра взаємодії. Так, розглядаючи ентропію вихідного коду, який може бути розбитий на більш дрібні сегменти (рядки, функції, змінні, команди і т.п.), можна визначати авторство тієї чи іншої частини коду. Таким чином, можна обчислити внесок різних авторів у розроблену програму. Для розрахунку ентропії в теорії інформації використовують таку формулу [4]:

$$H(i) = -p_i * \log_2(p_i), \quad (1)$$

де i - можливі стани;

p_i - імовірність появи i -го стану.

У випадку з програмним кодом у ролі станів виступають команди, і, відповідно, підраховується ймовірність їх появи в коді. Деякі метрики програмного забезпечення, наприклад, метрики Холстеда чи Берлінгера, беруть за основу ентропію. Також ентропійний метод використовується для вирішення таких завдань, як пошук зашифрованого або запакованого шкідливого програмного забезпечення.

Алгоритм роботи програми:

1) Дизасемблювання виконуваного файлу. Дизасемблювання здійснювалося за допомогою утиліти dumpbin.exe, яка входить до складу Microsoft Visual Studio [5]. В результаті отримуємо txt-файл, в якому знаходиться код програми.

2) Розрахунок значень ентропії кожної команди для отриманого дизасембльованого файлу. Обчислюємо за формулою (1). Усі отримані значення записуються у файл.

3) Розрахунок зміни ентропії. Зчитуємо значення ентропії для кожної команди оригінального файлу та відповідного йому зміненого. Підраховуємо різницю та записуємо результат у файл.

4) Визначення порогового значення ентропії кожної команди.

5) Порівняння різниці ентропії відповідних команд із пороговим значенням для цих команд.

6) Розрахунок критерію розбіжності.

Програма дозволяє дизасемблювати виконувані файли, тим самим отримати асемблерний код, розрахувати значення ентропії для кожної команди, проаналізувати отримані дані з точки зору різниці ентропії.

Програма зберігає зліпки файлів за найменуванням. Якщо якийсь файл був дизасембльований і буде завантажений новий з такою самою назвою, програма проведе порівняння між ними та відобразить результат порівняння.

Приклад роботи розробленої програми, а саме розрахунок ентропії для поточної та попередньої версії файлу, наведено на рисунку 1.



Рисунок 1 – Приклад розрахунку ентропії для поточної та попередньої версії файлу

Висновки. Була розроблена програма для аналізу дизасембльованого коду виконуваних файлів, обчислення ентропії та порівняння з минулими значеннями для знаходження змін коду виконуваних файлів. Розроблена програма може бути використана зокрема для валідації файлів існуючих програм. В подальшому передбачається удосконалення розробленої програми за рахунок додавання математичної моделі для прогнозування авторства внесених змін.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Що таке ін'єкція коду в Windows? (2018, Серпень 27). <https://www.thefastcode.com/uk-ua/article/what-is-code-injection-on-windows>
- [2] Энтропия. Как хаос помогает искать вирусы. (2021, Січень 29). <https://haker.ru/2021/01/29/viruses-entropy/>
- [3] Формула Шеннона. (n.d.). <http://um.co.ua/8/8-16/8-168268.html>
- [4] Шарапов, О.Д., Дербенцев В.Д., & Семьонов, Д.С. (2004). Економічна кібернетика. Навч. посібник. К.: КНЕУ.
- [5] Microsoft Visual Studio disassembler “dumpbin”. (2022, Вересень 27). <https://learn.microsoft.com/ru-ru/cpp/build/reference/dumpbin-reference?view=msvc-170>

Makarova L.M., Kaminskyi S.S., Bryzghalov M.V.

The entropy metric for changes search in the code of executable files

Abstract. A program for calculating the entropy of disassembled executable files and comparing them with past values in order to identify the authorship of changes in the code was developed.

Keywords: entropy, disassembly, executable files, code change, software application.