

# ОСОБЛИВОСТІ ВИКОРИСТАННЯ НЕЧІТКОГО ХЕШУВАННЯ В ЗАДАЧАХ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Вінницький національний технічний університет

**Анотація.** Розглянуто функцію нечіткого хешування ssdeep в задачах інформаційної безпеки. Проведено аналіз одного з алгоритмів хешування ssdeep. Виявлено переваги та недоліки нечіткого хешування в порівнянні з традиційними методами.

**Ключові слова:** нечітке хешування, інформаційна безпека, алгоритми хешування, ssdeep.

**Abstract.** The fuzzy hashing function ssdeep in information security problems is considered. One of the ssdeep hashing algorithms is analyzed. The advantages and disadvantages of fuzzy hashing in comparison with traditional methods are identified.

**Keywords:** fuzzy hashing, information security, hashing algorithms, ssdeep.

## Вступ

У сучасному цифровому світі питання інформаційної безпеки стають все більш актуальними. Зловмисники використовують різні методи обходу традиційних засобів захисту, що вимагає впровадження новітніх технологій для аналізу даних. Останнім часом набуло перспективного розвитку нечітке хешування, що, зокрема дозволяє розв'язати проблему пошуку подібних, але не тотожних інформаційних об'єктів [1].

Метою даного дослідження є аналіз ефективності алгоритму нечіткого хешування ssdeep, для виявлення модифікованих версій шкідливого програмного забезпечення. Дослідження спрямоване на оцінку переваг та обмежень у роботі алгоритму нечіткого хешування у порівнянні з традиційними методами криптографічного хешування.

## Результати дослідження

Нечітке хешування є методом, що відрізняється від традиційного криптографічного хешування принципово іншим підходом до порівняння хеш-значень. Криптографічні функції розроблені таким чином, щоб навіть незначна зміна вхідних даних призводила до зміни вихідного значення, то нечіткі хеш-функції створюють схожі значення для подібних вхідних даних. Це дозволяє ефективно знаходити модифіковані версії файлів, що особливо важливо в контексті виявлення шкідливого програмного забезпечення.

Алгоритм ssdeep, розроблений Джессі Корнблумом [2], є одним із найбільш розповсюджених інструментів нечіткого хешування. Він базується на методі контекстно-тригерного порційного хешування, і дозволяє створювати сигнатури файлів, які можуть бути використані для виявлення подібних документів [3]. Оригінальний опис алгоритму, представлений Корнблумом включає детальний аналіз методу та його застосування для ідентифікації майже ідентичних файлів [1].

Принцип роботи ssdeep полягає в аналізі послідовності байтів вхідного файлу, використанні «rolling hash» та ідентифікації певних точок в послідовності даних та коли значення досягає заданого порогу, алгоритм генерує частину хеш-значення для поточного блоку даних. Процес повторюється для всього файлу, в результаті чого формується хеш-значення. Ssdeep використовується для виявлення модифікованих версій шкідливого програмного забезпечення, оскільки творці часто вносять незначні зміни в код для уникнення виявлення традиційними антивірусними системами. Нечітке хешування виявляє такі модифіковані версії з високою точністю, навіть якщо в оригінальний файл внесено до 30% змін [3].

У практичних умовах алгоритм нечіткого хешування ssdeep демонструє високу ефективність у боротьбі зі шкідливим програмним забезпеченням. Наприклад, він широко застосовувався для виявлення модифікованих версій вірусу WannaCry, одного з найвідоміших видів шкідливого ПЗ, який поширювався, завдяки вразливості в операційних системах. У 2018 році було ідентифіковано понад 12000 варіантів WannaCry. Саме ssdeep допоміг їх розпізнати завдяки схожості у коді. Навіть у випадках коли зміни складали до 30% від оригінального коду, алгоритм продемонстрував високий рівень точності [4]. Це підкреслює значний потенціал використання нечіткого хешування в умовах постійної адаптації зловмисників.

В умовах постійно зростаючих обсягів даних нечітке хешування дозволяє ефективно фільтрувати і виявляти потенційно небезпечні файли серед великого обсягу документів. Ssdeep демонструє високу

специфічність (низький рівень хибно-позитивних результатів), однак здатність виявляти всі подібні файли може бути недостатньою для деяких типів модифікацій. Це особливо помітно при роботі з невеликими файлами. Перспективи використання нечіткого хешування в антивірусному захисті включають інтеграцію з системами машинного навчання для підвищення точності виявлення нових загроз, а також розробку спеціалізованих алгоритмів для аналізу конкретних типів файлів або вразливостей. [3]

Ефективність ssdeep може бути значно підвищена за рахунок оптимізації параметрів тригерного порогу та розміру блоків, що аналізуються. Це дозволяє адаптувати алгоритм до конкретних типів файлів та специфіки задачі [1].

### Висновки

Нечітке хешування є потужним інструментом для розв'язування різноманітних задач інформаційної безпеки, зокрема, для виявлення модифікованих версій шкідливого програмного забезпечення, наприклад ssdeep ідентифікував понад 12000 варіантів шкідливого ПЗ WannaCry, що продемонструвало його високу ефективність і значення в сучасних умовах боротьби зі шкідливим ПЗ. Завдяки здатності виявляти структурно подібні файли, навіть при значних змінах, нечітке хешування займає важливе місце у сфері кібербезпеки. Алгоритм ssdeep незважаючи на певні обмеження, демонструє високу ефективність при виявленні структурно подібних файлів.

Перспективними напрямками подальших досліджень є розробка більш точних алгоритмів нечіткого хешування та оптимізація вже існуючих методів нечіткого хешування. Застосування нечіткого хешування в системах інформаційної безпеки дозволить підвищити загальний рівень захисту від сучасних кіберзагроз, що постійно адаптуються до традиційних засобів захисту.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Y.Zhdanova Applied and methodical aspects of using hash functions for information security // Y.Zhdanova та ін. Cybersecurity: Education, Science, Technique. – 2020. – Т. 4, № 8. – С. 85-96. URL: <https://doi.org/10.28925/2663-4023.2020.8.8596> (дата звернення: 17.03.2025).
2. J. Kornblum. Identifying almost identical files using context triggered piecewise hashing.// Digital Investigation. – 2006. – vol. 3S. –pp. 91–97. URL: <http://www.dfrws.org/2006/proceedings/12-Kornblum.pdf> (дата звернення: 18.03.2025)
3. Єремів О., Стьопчик І. Перспективи використання нечіткого хешування в антивірусному захисті. Фізико-технічний інститут «НТУУ «КПІ» – 2016. – Вип. 1(31). – С.80–84 URL: <https://ela.kpi.ua/server/api/core/bitstreams/36e76527-d105-4f88-84c5-422539227031/content> (дата звернення 18.03.2025)
4. Playbook of the Week: Uncovering Unknown Malware Using SSDeep // Palo Alto Networks Blog. – 2023. URL: <https://www.paloaltonetworks.com/blog/security-operations/playbook-of-the-week-uncovering-unknown-malware-using-ssdeep/> (дата звернення: 19.03.2025).

**Савкун Валентин Олександрович** - студент 2-го курсу групи 2БС-236, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, email: valiksavkun298@gmail.com.

**Кондратенко Наталія Романівна** – професор кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, email: kondratenko.natalia@vntu.edu.ua

**Savkun Valentyn O.** - 2nd year student, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine, email: valiksavkun298@gmail.com.

**Kondratenko Nataliia Romanivna** - Professor of the Department of Information Security, Vinnytsia National Technical University, Vinnytsia, Ukraine, email: kondratenko.natalia@vntu.edu.ua