

ДОСЛІДЖЕННЯ КОМП'ЮТЕРНОЇ СИСТЕМИ ДЛЯ ВИЯВЛЕННЯ КОНФІДЕНЦІЙНИХ І ЗАХИЩЕНИХ ДАНИХ В ІНФОРМАЦІЙНОМУ СЕРЕДОВИЩІ

Прохоров М.О., науковий керівник Шкарупило В.В.

Загальноприйняте інформаційне комп'ютерне середовище зазвичай динамічно змінюється та містить певні конфіденційні дані (персональну інформацію, службові документи, фінансові звіти, результати досліджень тощо), які перебувають під загрозою перехоплення або шифрування з метою продажу чи оприлюднення. Ідентифікація конфіденційної інформації з метою запобігання її витоку стає надзвичайно актуальною проблемою.

Існуючі системи запобігання витоку даних (DLP) орієнтовані виключно на нешифровані документи, вимагають встановлення стороннього програмного забезпечення, яке має повний доступ до мереж із критичною інформацією, і, відповідно, передбачають повну довіру до постачальника — що не завжди виправдано. Людський фактор, помилки конфігурації або обмеження самої системи можуть призвести до того, що критично важлива інформація залишиться непоміченою або, навпаки, буде помилково заблокована.

Принциповим завданням цієї роботи стало дослідження доцільності та можливостей розробки власних інструментів і методик виявлення конфіденційних даних, які дозволяють більш ефективно контролювати процес аналізу та адаптувати його до конкретних потреб і загроз. Основна мета роботи — продемонструвати можливість створення власних засобів виявлення як відкритих, так і криптографічно захищених конфіденційних файлів. Виявлення стеганографічних контейнерів, що містять приховану конфіденційну інформацію, наразі виходить за межі цієї роботи та розглядається як перспективний напрям подальших досліджень.

З метою узагальнення експериментальної методики були розглянуті такі поширені формати файлів:

- текстові — як такі, що можуть містити відкриті конфіденційні дані певної структурованості;
- зашифровані PKZIP, DOCX, XLSX і PDF — як такі, що можуть містити криптографічно захищені конфіденційні дані.

Встановлено, що:

- відкриті файли мають характерні структурні ознаки, які можна використовувати як основу для ідентифікації, навіть без аналізу змісту документів;
- зашифровані файли мають жорстко задану структуру, що включає фіксовані заголовки, службові байти та типові послідовності елементів, притаманні конкретним засобам шифрування.

Для аналізу відкритих даних реалізовано демонстраційну версію агента, який ідентифікує конфіденційні документи за допомогою пакету регулярних виразів. Тобто, експериментально перевірено методику, що дозволяє виявляти документи з фінансовою, особистою або службовою інформацією за заданими шаблонами, з можливістю технологічного розширення цього набору без модифікації логіки програми.

Вочевидь, що для зашифрованих файлів така методика непридатна, оскільки їхній вміст недоступний для аналізу без процедури розшифрування. Проте було знайдено засіб, що дозволяє доволі точно ідентифікувати конкретні файли — результати криптографічних перетворень певними програмними засобами. З'ясувалося, що жорсткі структурні особливості шифрованих файлів можна описувати за допомогою так званих S-виразів. Цю тезу проілюстровано на прикладі популярного архіватора PKZIP, що працює в режимі шифрування. Відомо, що для нього існують спеціалізовані бібліотеки,

наприклад, zipfile у Python, які дозволяють легко визначити, чи ZIP-файл зашифрований. Проте для інших шифрувальників таких засобів не існує, і кожен окремий формат потребує розробки персонального інструментарію ідентифікації.

Це привело до усвідомлення потреби у створенні універсального підходу, що підтримує відкриту систему описів результатів роботи конкретних реалізацій криптографічного програмного забезпечення. Запропонована методика реалізована у вигляді демонстраційної програми, яка сканує вказаний каталог та аналізує структурні особливості файлів, співвідносячи їх із пакетом S-виразів — формальних описів результатів роботи конкретних шифрувальників. Наразі цей пакет містить єдиний S-вираз, що відповідає архіватору PKZIP у режимі шифрування. Водночас уже розпочато розробку модуля для інтерактивного створення деревоподібних шаблонів структур шифрованих файлів із автоматичною генерацією відповідних S-виразів.

У підсумку, в межах цієї роботи було закладено концептуальну основу для створення власних інструментів моніторингу конфіденційної інформації, незалежно від сторонніх рішень. Представлено демонстраційні реалізації агентів, здатних виявляти як відкриті, так і зашифровані файли за їхніми формальними ознаками та структурними шаблонами. Запропонована методика орієнтована на гнучкість, прозорість і повний контроль над конфіденційними даними — що набуває особливої важливості в умовах зростаючих загроз інформаційній безпеці.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дмитренко Г. І., Шабалтас С. А. Захист конфіденційної інформації в комп'ютерних системах. — К.: Вид-во Ліра-К, 2020. — 256 с.
2. Kent K., Souppaya M. Guide to Data Loss Prevention (DLP). — NIST Special Publication 800-214, 2023. — [Електронний ресурс]. — Режим доступу: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-214.pdf>
3. Friedl J. E. F. Mastering Regular Expressions. — 3rd ed. — O'Reilly Media, 2006. — 544 p.
4. McCarthy J. Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I // Communications of the ACM. — 1960. — Vol. 3, No. 4. — P. 184–195. (→ історичне джерело про S-вирази)
5. Petitcolas F. A. P., Anderson R. J., Kuhn M. G. Information hiding — A survey // Proceedings of the IEEE. — 1999. — Vol. 87, No. 7. — P. 1062–1078.

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



**ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ
АСПЕКТИ РОЗРОБКИ
КОМП'ЮТЕРНИХ СИСТЕМ**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів