Способи витоку персональних даних, методи обробки та захист від витоків

УДК 004.056

Ірина Волобуєва[1], Лідія Тимошенко[2]

Національний університет «Одеська політехніка»,
[1]10252809@stud.op.edu.ua, [2]l.m.timoshenko@op.edu.ua

Витік персональних даних є однією з актуальних проблем сучасної інформаційної безпеки. Він передбачає несанкціоноване поширення, копіювання або викрадення інформації, за допомогою якої можна ідентифікувати особу. До таких відомостей належать імена, адреси, електронні адреси, номери телефонів, паролі, банківські реквізити та інші дані. Особливо небезпечними є витоки облікових даних, оскільки паролі залишаються одним із найпоширеніших засобів автентифікації, а їх повторне використання в різних сервісах підвищує ризик несанкціонованого доступу [1].

Актуальність дослідження зумовлена тим, що після витоку персональні дані часто потрапляють до публічних або напівпублічних масивів, зокрема у форматі email:password. Такі набори можуть використовуватися для підбору паролів, компрометації облікових записів, шахрайства та подальших кібератак. Тому важливим є не лише захист персональних даних, а й аналіз уже наявних масивів витоку для виявлення типових слабких місць у поведінці користувачів і політиках безпеки.

Метою роботи є аналіз способів витоку персональних даних, наборів даних формату email:password, виявлення слабких і повторюваних паролів шляхом розроблення програмного застосунку, визначення доменної концентрації записів та оцінювання рівня ризику витоку для подальшого формування рекомендацій з покращення інформаційної безпеки.

До основних способів витоку персональних даних належать технічні, організаційні та пов’язані з діями користувачів. Технічні способи витоку можуть бути пов’язані з вразливостями серверів, шкідливим програмним забезпеченням, фішинговими атаками, SQL-ін’єкціями та іншими засобами отримання несанкціонованого доступу. Організаційні способи витоку виникають у разі недостатнього контролю доступу, відсутності належних політик безпеки, слабкого журналювання дій користувачів або неналежного зберігання інформації. Також поширеним способом витоку залишається людський фактор: використання слабких паролів, повторення однакових комбінацій у різних сервісах, довіра до фішингових повідомлень і нехтування базовими правилами інформаційної безпеки [2].

Окрему увагу слід приділити шаблонності поведінки користувачів. Люди часто створюють паролі за передбачуваними схемами: використовують дати народження, прості цифрові послідовності, імена, короткі слова або однакові комбінації для різних акаунтів. Така поведінка спрощує підбір паролів і підвищує ймовірність компрометації облікових записів після витоку даних [3].

Наукова новизна роботи полягає в поєднанні аналізу способів витоку персональних даних із практичним оцінюванням ризику на основі характеристик набору email:password. На відміну від робіт, у яких переважно

розглядаються окремі аспекти захисту даних або поведінки користувачів, у цій роботі пропонується підхід, що враховує слабкість паролів, повторне використання парольних комбінацій, доменну концентрацію записів та актуальність витоку.

Для розв'язання поставленої задачі запропоновано створення програмного застосунку для аналізу публічних наборів даних формату email:password. Застосунок доцільно реалізувати мовою Python із використанням бібліотек pandas, regex, matplotlib та бази даних SQLite. Основні етапи його роботи передбачають імпорт текстового файлу, розбір рядків на електронну адресу, домен і пароль, перевірку коректності формату, підрахунок унікальних паролів і доменів, визначення поширених слабких паролів, обчислення повторного використання паролів, групування записів за доменами та формування підсумкових звітів.

Для наочного подання логіки роботи застосунку використано блок-схему. Загальна структура роботи програмного застосунку наведена на рис. 1.

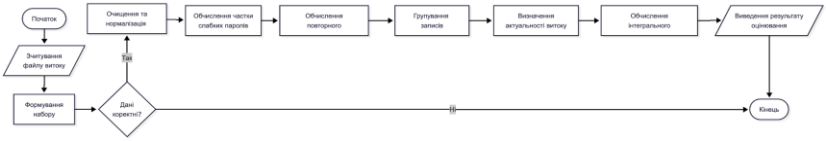

Рис.1. Узагальнена блок-схема застосунку

Вона відображає послідовність основних етапів обробки даних: зчитування файлу витоку, формування набору записів, перевірка коректності даних, очищення та нормалізація записів, обчислення показників ризику, групування за доменами та виведення результату оцінювання.

У результаті роботи обґрунтовано актуальність аналізу публічних наборів персональних даних формату email:password та запропоновано структуру програмного застосунку для їх опрацювання. Очікуваними результатами є визначення поширеності слабких паролів, рівня повторюваності парольних комбінацій, доменної концентрації записів та інтегрального ризику витоку. Отримані результати можуть бути використані для формування рекомендацій щодо посилення політик інформаційної безпеки, підвищення рівня захисту облікових даних і зменшення ризику несанкціонованого доступу.

1. Тимошенко Л., Вакуліна С., Волобуєва І. Локальний менеджер паролів із генеруванням для мобільних пристроїв, Кібербезпека в сучасному світі: актуальні виклики : матеріали VI міжнар. наук.-практ. конф., м. Одеса, 28 листопада 2025 р. Одеса, 2025. С. 40.
2. Корнєва В.Д. Способи захисту ІТ-індустрії від витоку інформації. Інформаційна безпека та комп'ютерні технології : матеріали VII міжнар. наук.-практ. конф., м. Київ: Державний торговельно-економічний університет, 2023 р. Київ, 2023. С. 31–33.
3. Заболоцький П.В. Шаблонність людського мислення та комп'ютерна безпека. Міжнародний науковий журнал «Грааль науки», №24, лютий 2023. С. 345–347.