COMPUTER SYSTEMS AND NETWORKS

Vol. 7, No. 1, 2025

DOI: https://doi.org/10.23939/csn2025.01.131

УДК 004.056

АНАЛІЗ ЕФЕКТИВНОСТІ ТА ВРАЗЛИВОСТЕЙ МЕТОДІВ ЗАБЕЗПЕЧЕННЯ ПРИВАТНОСТІ НА ПРИКЛАДІ K-АНОНІМІЗАЦІЇ, L-ДИФЕРЕНЦІАЦІЇ ТА T-ПОДІБНОСТІ

О. О. Іванюк, А. М. Вахула

Національний університет "Львівська політехніка",
кафедра комп'ютеризованих систем автоматики
E-mail: oleh.o.ivaniuk@lpnu.ua, andrii.m.vakhula@lpnu.ua

Здійснено аналіз і порівняння методів анонімізації персональних даних на прикладі k-анонімізації, ℓ-диференціації та t-близькості. Мета - оцінка ефективності цих методів щодо захисту конфіденційності та виявлення їхніх вразливостей до атак повторної ідентифікації. Дослідження виконано за допомогою інструменту ARX Anonymization Tool на тестовому наборі даних про дохід осіб.

Проаналізовано вплив зміни ключових параметрів методів на ступінь приватності й інформативність анонімізованих даних. Встановлено, що найефективнішим для захисту конфіденційності є метод t-близькості, однак його застосування призводить до суттєвого зниження деталізації інформації. Водночас метод k-анонімізації, хоча і менш стійкий до атак, забезпечує кращу практичну цінність анонімізованих даних. Метод ℓ-диференціації демонструє помірну ефективність як за показниками приватності, так і за інформативністю. Дослідження дало змогу кількісно визначити баланс між конфіденційністю та корисністю даних, що сприяє більш усвідомленому вибору оптимальних параметрів анонімізації залежно від конкретних завдань.

Ключові слова: анонімізація даних, захист персональних даних, приватність, k-anonymization, ℓ-diversity, t-closeness.

Вступ

Зі стрімким розвитком інформаційних технологій і швидким розширенням кількості цифрових сервісів, що оперують великими масивами персональних даних, особливого значення набуває проблема збереження анонімності та конфіденційності інформації користувачів. Дані, які збирають соціальні мережі, фінансові установи, медичні сервіси, маркетингові платформи та інші цифрові додатки, часто містять чутливі відомості, розголошення яких може становити загрозу приватності особи, спричинити юридичні наслідки чи завдати репутаційних збитків. У зв'язку з цим постає необхідність ефективного захисту інформації шляхом застосування методів анонімізації, що дають змогу зберегти цінність даних для аналітичних і дослідницьких цілей, водночас унеможливлюючи ідентифікацію окремих осіб.

Проте, незважаючи на значні досягнення у сфері анонімізації даних, наукова спільнота й досі стикається з труднощами щодо оцінки ефективності різних підходів. Вибір відповідного методу анонімізації залишається складним завданням через суперечливість вимог, адже необхідно знайти компроміс між достатнім рівнем приватності та максимальним збереженням інформативності набору

даних. Крім того, постійне зростання обчислювальних потужностей і доступність додаткових джерел інформації підвищують ризики повторної ідентифікації особи навіть після застосування анонімізації, що підсилює актуальність вивчення цієї проблематики.

Проаналізовано основні підходи до анонімізації, обговорено переваги та недоліки методів. Також на практиці показано метод k-анонімізації, ℓ-диференціації та t-близькості та проаналізовано їхні вразливості до атак.

Огляд літературних джерел

У першій праці [1] автори запропонували базову концепцію k-анонімності як метод узагальнення і придушення атрибутів з метою недопущення ідентифікації особистості серед набору даних. Під час дослідження було показано, що використання цього методу дає змогу суттєво зменшити ризики повторної ідентифікації. Цінність роботи полягає у тому, що вона заклала основу для подальших досліджень і стала фундаментом розвитку сучасних методів анонімізації.

У другій статті [2] розвинуто ідею k-анонімності шляхом введення поняття ℓ-диференціації. Автори здійснили аналіз вразливостей методу k-анонімності, довівши, що він не забезпечує захисту в ситуаціях, коли існує одноманітність чутливих даних. У результаті дослідження було запропоновано підхід ℓ-diversity, що передбачає забезпечення мінімального різноманіття чутливих атрибутів у кожному класі еквівалентності. Це дослідження є вагомим, бо сприяє встановленню чіткішого критерію приватності, який захищає дані від широкого спектра атак.

Третє джерело [3] було спрямоване на подолання недоліків k-анонімності та ℓ-диференціації шляхом забезпечення близькості розподілу чутливих атрибутів у групах до загального розподілу набору даних. В ході дослідження було продемонстровано, що t-близькість забезпечує більш ефективний захист даних, ніж попередні підходи, особливо в ситуаціях, коли можливі статистичні атаки. Описана робота допомогла у розширенні методологічних підходів до анонімізації шляхом впровадження статистичних критеріїв приватності.

У четвертій статті [4] автор розглянув теоретичні основи диференційної приватності як методу, що забезпечує формальні гарантії анонімності шляхом додавання випадкового шуму до запитів та даних. Дослідження показало, що диференційна приватність забезпечує найвищий рівень захисту від усіх відомих атак на ідентифікацію. Важливість цієї роботи полягає в узагальненні та систематизації результатів у галузі диференційної приватності, що дозволило широко застосовувати її у практичних задачах захисту даних.

У п’ятому дослідженні [5] наведено аналіз найсучасніших (на момент публікації) методів анонімізації даних, їх переваги та недоліки, а також запропоновано класифікацію цих методів за ефективністю. Автори визначили, що ключовим фактором у виборі методу є компроміс між приватністю і корисністю даних. Цінність статті полягає у комплексному аналізі стану проблеми та визначенні перспективних напрямів подальших досліджень у сфері анонімізації даних.

Автори шостого дослідження [6] запропонували уніфікований концептуальний фреймворк для оцінки приватності, що дає змогу оцінити ефективність різних підходів до анонімізації через критерій так званої приватності членства. Під час дослідження було проаналізовано існуючі підходи (k-анонімність, ℓ-diversity, диференційна приватність) та показано, що кожен метод може бути оцінений через єдину модель, що значно спрощує вибір оптимального рішення залежно від конкретних умов. Це дало змогу сформулювати єдиний підхід до порівняння різних методів анонімізації, що спрощує їх практичне застосування.

Постановка задачі

На прикладі тестового набору даних доходу населення проаналізувати механізм роботи алгоритмів k-анонімності, ℓ-диференціації та t-близькості, продемонструвати вплив параметрів анонімізації на практичну цінність та захищеність отриманого набору даних. Порівняти результати перевірки захищеності та практичної цінності даних із використанням різних методів.

Основні методи анонімізації

Метод **k-анонімності** є одним із перших і найбільш широко застосовуваних підходів до анонімізації даних. Вперше концепцію k-анонімності запропонували в 1998 році дослідники П'єр Самараті та Латанія Суїні як відповідь на зростаючі ризики повторної ідентифікації особистості у великих масивах даних. Основна ідея цього методу полягає у створенні набору даних так, щоб кожен запис не міг бути відрізнений від принаймні k-1 інших записів за набором певних атрибутів, які можуть сприяти ідентифікації особи. Ці атрибути називаються квазіідентифікаторами. Використовуючи цей підхід, будь-яка конкретна особа не може бути виділена з певної групи, що суттєво зменшує ймовірність її ідентифікації.

Математична модель k-анонімності формулюється так: набір даних задовольняє умові k-анонімності, якщо для кожного запису існує щонайменше k-1 інших записів, що мають ідентичні значення за квазіідентифікаторами. Формально це записується через нерівність: $|E| \geq k$, де |E| позначає кількість записів у класі еквівалентності, що визначений однаковими значеннями квазі-ідентифікаторів.

Алгоритм роботи методу k-анонімності можна описати в декілька етапів. На першому етапі відбувається ідентифікація квазіідентифікаторів – атрибутів, що можуть бути використані для ідентифікації осіб. Після цього всі записи групуються за цими атрибутами в класи еквівалентності. На наступному етапі застосовується процедура узагальнення (generalization), що полягає у заміні конкретних значень атрибутів на більш загальні, або придушення (suppression), тобто видалення окремих записів або атрибутів, які перешкоджають досягненню k-анонімності. Цей процес повторюється, поки всі класи еквівалентності не задовольнять умову k-анонімності.

Метод k-анонімності є особливо доречним у випадках, коли ключовою вимогою є простота та ефективність у досягненні певного базового рівня анонімності. Він має практичну перевагу у вигляді простої імплементації та загальної зрозумілості концепції, що дає змогу легко впроваджувати цей метод в організаціях без значних витрат на спеціалізовані ресурси. Крім того, його широка розповсюдженість дає змогу використовувати стандартні інструменти та алгоритми для ефективної обробки даних, забезпечуючи високу доступність методу.

Однак, незважаючи на свої переваги, k-анонімність має і суттєві обмеження та недоліки. Один із найважливіших недоліків – вразливість методу до атак із застосуванням фонових знань, коли додаткова інформація про особу дозволяє ідентифікувати індивіда навіть у захищеному наборі даних. Іншим недоліком є неможливість запобігти ситуації, коли всі записи в класі еквівалентності мають однакове значення чутливого атрибута, що знижує реальний рівень приватності. Додатково, процес узагальнення та придушення даних може призводити до суттєвої втрати корисності набору, оскільки що більший рівень k-анонімності, то менш детальними стають дані, які залишаються доступними для аналізу.

Метод **ℓ-диференціації** запропонував у 2006 році Ашвін Мачанаваджхала та його колеги як відповідь на відомі обмеження методу k-анонімності. Основною мотивацією для створення цього методу стало усвідомлення того, що k-анонімність сама собою не забезпечує достатнього рівня приватності в ситуаціях, коли чутливі атрибути всередині класів еквівалентності мають низьку різноманітність або одноманітність значень. Інакше кажучи, навіть коли записи неможливо відрізнити один від одного за квазіідентифікаторами, може виникати ситуація, коли всі записи класу еквівалентності мають однакове або дуже схоже значення чутливого атрибута, що дає змогу з високою ймовірністю зробити висновки про особисту інформацію індивідів.

Основна ідея ℓ-диференціації полягає у гарантуванні того, що кожен клас еквівалентності буде містити щонайменше ℓ різних значень чутливого атрибута. Це запобігає ситуаціям, коли інформація, що має бути конфіденційною, легко розкривається через одноманітність значень в межах певних груп. Математичною моделлю ℓ-диференціації є нерівність, яка визначає, що в кожному класі еквівалентності E кількість різних значень чутливого атрибута має бути не меншою за ℓ. Формально це записується як: $|\{sensitive\_value(e) \mid e \in E\}| \geq \ell$, де sensitive_value(e) – це чутливе значення запису e в класі еквівалентності E.

Алгоритм реалізації методу ℓ-диференціації є продовження методу k-анонімності. Спочатку та само потрібно визначити квазіідентифікатори та об'єднати дані у класи евівалетності. Важливою умовою є, щоб в кожному класі було не менше ℓ різних значень чутливих атрибутів. Після завершення узагальнення перевіряється умова ℓ-диференціації, і за необхідності процес повторюється до досягнення необхідного рівня різноманітності.

Метод ℓ-диференціації особливо доречний у ситуаціях, коли існує високий ризик повторної ідентифікації через одноманітність чутливих атрибутів в групах записів. Він є значно ефективнішим у таких сценаріях порівняно з k-анонімністю, оскільки забезпечує додатковий рівень захисту шляхом контролю різноманітності конфіденційних атрибутів.

Серед переваг методу ℓ-диференціації можна виділити те, що він значно знижує ризик повторної ідентифікації через чутливі атрибути, особливо порівняно з простим методом k-анонімності. Метод також досить універсальний, тому що його можна застосовувати для різних типів наборів даних, де є ризик розкриття приватної інформації. Крім того, ℓ-диференціація дає змогу більш чітко і строго регулювати рівень приватності, чим робить цей підхід дуже ефективним для забезпечення приватності. Серед недоліків підходу можна виділити те, що для забезпечення ℓ-диференціації може знадобитися значне узагальнення або придушення атрибутів, що може призвести до суттєвої втрати корисності інформації для аналітичних цілей. Також метод ℓ-диференціації не забезпечує повного захисту від атак, що базуються на знанні загальних статистичних характеристик набору даних. Наприклад, якщо в наборі даних переважає один тип чутливого значення, то навіть дотримання умови ℓ-диференціації може не гарантувати достатньо високий рівень приватності.

Метод **t-близькості** є одним із розвинених підходів до анонімізації даних, який було запропоновано у 2007 році у статті "t-Closeness: Privacy Beyond k-Anonymity and ℓ-Diversity". Створення цього методу було реакцією на певні обмеження k-анонімності та ℓ-диференціації, особливо в контексті статистичних атак, коли навіть різноманітність чутливих атрибутів не могла повністю захистити індивідуальну приватність. Основною метою введення поняття t-близькості було вирішення проблеми нерівномірності розподілу чутливих атрибутів у класах еквівалентності порівняно із загальним розподілом у всьому наборі даних.

Головною концептуальною ідеєю методу t-близькості є забезпечення того, що розподіл чутливих атрибутів у кожному класі еквівалентності не відрізнятиметься суттєво від розподілу цих самих атрибутів у всьому наборі даних. Це означає, що знання належності запису до певного класу не дає істотної переваги для виведення приватної інформації порівняно зі знанням загального розподілу.

Метод t-близькості визначається так: набір даних задовольняє умову t-близькості, якщо для будь-якого класу еквівалентності E відстань між розподілом чутливих атрибутів у класі P(E) і розподілом цих атрибутів у всьому наборі даних P(D) не перевищує порогового значення t, що записується як нерівність $D[P(E), P(D)] \leq t$, де D – це обрана функція відстані.

Метод t-близькості знаходить широке застосування у сферах, де конфіденційність особистих даних є критичною, а ризик статистичних атак – високим. Наприклад, він ефективний у медичній галузі під час публікації досліджень, де знання про розподіл захворювань може бути використане для ідентифікації особи. Метод також використовується в соціальних дослідженнях, коли важливо зберегти точність статистичних висновків, але водночас запобігти можливості витоку чутливої інформації. Іншими можливими сферами застосування є фінансова аналітика, демографічні дослідження та маркетингові аналізи, де інформація має бути достатньо детальною, але водночас гарантувати високий рівень захисту приватності.

Серед переваг методу t-близькості є те, що він забезпечує високий рівень приватності, ефективно захищаючи дані від широкого спектра статистичних атак. Цей метод дає змогу контролювати точність і приватність одночасно завдяки гнучкому налаштуванню параметра t. Однак метод t-близькості має і певні недоліки, одним з яких є його обчислювальна складність, особливо за великих наборів даних. Реалізація цього методу потребує значних ресурсів, що може обмежувати його

використання в умовах недостатніх обчислювальних потужностей. Іншим недоліком може бути складність вибору оптимального значення параметра t, оскільки це потребує додаткового аналізу специфіки набору даних і потреб приватності.

Метод **диференційної приватності** є однією з найновіших і найбільш перспективних технологій анонімізації, яка з'явилася у сфері приватності даних у 2006 році завдяки дослідниці Синтії Дворк та її колегам. Цей метод був запропонований як відповідь на обмеження класичних методів анонімізації, які не завжди здатні забезпечити достатній рівень приватності, особливо в умовах доступності додаткових фонових знань або статистичних атак. Диференційна приватність забезпечує суворі математичні гарантії приватності шляхом додавання спеціально підібраного випадкового шуму до результатів запитів або аналізу даних, що унеможливлює точну ідентифікацію будь-якого індивіда.

Головною ідеєю диференційної приватності є забезпечення того, що присутність або відсутність окремого запису в наборі даних незначно впливає на кінцевий результат аналізу. Отже, спостерігач, аналізуючи опубліковані дані, не може з високою точністю зробити висновок про участь конкретного індивіда у наборі даних. Формально метод визначається такою математичною моделлю: рандомізований механізм M задовольняє ε-диференційну приватність, якщо для будь-яких двох наборів даних D та D', що відрізняються одним записом (так звані сусідні набори даних), і для будь-якої множини можливих виходів S виконується нерівність (1).

$$Pr[M(D) \in S] \leq exp(\varepsilon) \times Pr[M(D') \in S], \quad (1)$$

де ε є параметром приватності, що визначає рівень захисту приватності; менші значення ε забезпечують вищий рівень захисту.

Алгоритм диференційної приватності розпочинається з визначення того, які саме запити чи види аналізу будуть виконуватися над набором даних. Після цього проводиться оцінка чутливості (sensitivity) запитів, яка вимірює, наскільки максимально може змінитися результат у разі додавання або вилучення одного запису з набору даних. Наступним кроком є додавання спеціально підібраного випадкового шуму, зазвичай згідно з розподілом Лапласа або Гаусса, до результатів цих запитів. Цей шум налаштовується так, щоб забезпечити виконання умови диференційної приватності. Після додавання шуму результати запитів стають публічно доступними для аналізу, водночас приватність окремих індивідів залишається захищеною.

Перевагами диференційної приватності є її сувора математична обґрунтованість, що забезпечує надійний захист від широкого спектра атак, зокрема атаки з використанням фонових знань. Цей підхід є гнучким і дає змогу налаштовувати рівень приватності шляхом зміни параметра ε, що робить його придатним для широкого спектра застосувань. Крім того, диференційна приватність забезпечує прозорість щодо рівня приватності, яку отримує кінцевий користувач, завдяки чітко визначеним параметрам та формальним гарантіям. Водночас недоліком методу є втрата точності даних через додавання випадкового шуму, що може значно ускладнювати точний аналіз, особливо під час роботи з малими наборами даних або аналізі рідкісних явищ. Ще одним недоліком є складність визначення оптимального значення параметра ε, який визначає баланс між приватністю і точністю даних, що часто потребує спеціалізованих знань та експертної оцінки. Нарешті, реалізація диференційної приватності може бути обчислювально складною, особливо для великих наборів даних, що може потребувати додаткових обчислювальних ресурсів.

Атаки на анонімізовані дані та їхні моделі

Атака типу **Re-identification (повторної ідентифікації)** являє собою процес, за якого анонімізовані набори даних зіставляються з іншими доступними джерелами інформації для встановлення справжньої особистості учасників, які були раніше деідентифіковані. Основним завданням цієї атаки є виявлення ідентифікаційних атрибутів або квазіідентифікаторів (наприклад, стать, вік, місце проживання, професія), які можуть бути використані для ідентифікації конкретних осіб

шляхом комбінації з іншими відкритими джерелами інформації або додатковими наборами даних. Також необхідно зазначити, що атаки повторної ідентифікації стають особливо ефективними у випадку використання даних з високим рівнем деталізації, де наявні численні квазіідентифікатори або унікальні характеристики учасників.

Атака повторної ідентифікації зазвичай розпочинається зі збору сторонніх, неанонімізованих наборів даних, які містять деякі характеристики, присутні також і в анонімізованих даних. Наступним етапом є порівняння та зіставлення квазіідентифікаторів з двох наборів для виявлення подібностей або збігів, що дає змогу із високою ймовірністю встановити зв'язок між анонімізованими записами та їхніми власниками. Після цього здійснюється перевірка статистичної значущості знайдених збігів, щоб підтвердити або спростувати гіпотезу про відповідність особистості та її запису. Отже, сутність атаки полягає у ретельному аналізі та комбінуванні різнорідних джерел даних з метою встановлення персональних зв'язків. Прикладом успішної атаки повторної ідентифікації є випадок із базою медичних даних штату Массачусетс, де дослідниця Суїні змогла успішно ідентифікувати губернатора штату, зіставивши анонімізовані медичні записи із загальнодоступними виборчими списками [1], використовуючи лише три квазіідентифікатори: дату народження, стать і поштовий індекс. Інший відомий випадок – це атака на анонімізований набір даних Netflix, коли аналітики змогли ідентифікувати певних користувачів, зіставляючи їх рейтинги фільмів із відкритими профілями на платформах типу IMDb.

Вразливими до таких атак є підходи до анонімізації даних, які ґрунтуються виключно на вилученні явних ідентифікаторів, таких як імена або номери паспортів. Цей підхід, відомий як "просте видалення ідентифікаторів", не гарантує захисту, оскільки квазіідентифікатори часто залишаються у базах і можуть бути легко використані для встановлення зв'язків з іншими наборами даних. Більш надійними є методи, що використовують узагальнення, шумування або k-анонімність, хоча і вони мають обмеження, якщо кількість параметрів занадто велика або деталізація даних надмірна. Отже, для надійної протидії атакам повторної ідентифікації рекомендується застосування комбінованих методів, що суттєво ускладнює процес повторної ідентифікації.

Inference-атака є специфічним видом атаки на анонімізовані набори даних, метою якої є не пряме розкриття ідентичності суб'єктів, а отримання додаткових відомостей про них через аналіз і узагальнення доступних даних. Головна особливість цього виду атак полягає в тому, що навіть під час ретельної анонімізації первинних ідентифікаторів чи квазіідентифікаторів зловмисник може отримати нові знання шляхом аналізу та зіставлення певних непрямих характеристик, які збереглися у наборі даних.

Inference-атака зазвичай розпочинається із аналізу анонімізованих наборів даних з метою визначення певних закономірностей, залежностей та кореляцій, які можуть неявно розкривати чутливі атрибути. Після цього зловмисник застосовує статистичні або машинні алгоритми, такі як регресійний аналіз, класифікація, нейромережеві методи, які дають змогу передбачати невідомі характеристики учасників за наявними у даних непрямими ознаками. Далі результати аналізу використовуються для формування висновків щодо прихованих атрибутів, таких як діагноз, рівень доходу або політичні уподобання конкретної людини. На завершальному етапі зловмисник може додатково перевірити або уточнити зроблені висновки, використовуючи допоміжні відкриті джерела, щоб переконатися в коректності отриманої інформації.

Одним із яскравих прикладів реального використання Inference-атак може бути ситуація з аналізом медичних даних, у яких чутлива інформація, така як діагноз або стан здоров'я, була формально видалена чи узагальнена. Дослідники або зловмисники можуть з високою точністю передбачити конкретні захворювання, використовуючи непрямі характеристики, наприклад, дані про вік, стать, рівень фізичної активності, місце проживання або особливості лікування. Подібний сценарій був у США, де на основі анонімних даних про закупівлю медичних препаратів у мережевих аптеках можна було легко зробити висновки про стан здоров'я окремих осіб чи навіть їх приналежність до груп ризику. Інший приклад – використання соціальних мереж, де формально захищені

відомості, наприклад інформація про расову приналежність або політичні погляди користувача, можуть бути визначені шляхом аналізу взаємодії з іншими особами, переглянутих сторінок або типу контенту, з яким взаємодіє особа.

До атак типу Inference найбільш вразливими є підходи анонімізації, засновані на простому видаленні чи замасковуванні прямих ідентифікаторів без додаткового захисту, оскільки вони зберігають численні непрямі атрибути, що містять значну інформацію для аналізу. Також недостатньо захищеними можуть бути навіть складніші підходи, такі як k-анонімність, якщо ці підходи не контролюють додаткові кореляції між атрибутами або не вводять шум, що перешкоджає точному статистичному аналізу. З цієї точки зору найбільш ефективним є використання методів диференційної приватності, де до вихідних даних додається контрольований шум, що значно ускладнює можливість отримання точних і надійних висновків про конкретних осіб.

Також варто звернути увагу на моделі атак на анонімізовані дані визначають специфіку загроз, зумовлених різними типами потенційних зловмисників, їхніми цілями, рівнем доступу до ресурсів та особливостями поведінки. Серед найпоширеніших виділяють такі три види моделей атак: prosecutor attacker model (модель прокурора), journalist attacker model (модель журналіста) і marketer attacker model (модель маркетолога) [7]. Кожна з них має свої специфічні характеристики, що визначають особливості здійснення атак, рівень складності, цілі зловмисників та потенційні наслідки.

Модель атак **прокурора** (prosecutor attacker model) передбачає ситуацію, в якій зловмисник вже знає або має підозру, що певна людина присутня в наборі анонімізованих даних, і намагається ідентифікувати конкретний запис, що належить цій особі, з метою розкриття чутливої інформації. Головною особливістю прокурорської моделі є те, що атакуючий заздалегідь має чітко визначену мету: підтвердити або спростувати присутність конкретної особи у наборі даних або визначити специфічні характеристики, пов’язані з нею. Наприклад, прокурор або слідчий може шукати підтвердження певної медичної історії підозрюваного в анонімізованій базі медичних даних або факт перебування особи у певній локації в конкретний час, що допоможе у розслідуванні кримінальної справи. Відповідно, для реалізації такої атаки прокурор має додаткові, дуже точні дані про цільову особу, які є унікальними або близькими до унікальних.

Журналістська модель (journalist attacker model) має іншу специфіку та мотивацію. Вона базується на спробах встановити особистість окремих індивідів у наборі анонімізованих даних, використовуючи зовнішні джерела інформації, доступні журналісту (наприклад, відкриті реєстри, соціальні мережі, інформацію, отриману через особисті контакти). Важливою особливістю цієї моделі є широта і різноманітність використовуваної допоміжної інформації: журналіст не має наперед конкретної людини, яку шукає, натомість він використовує загальнодоступну інформацію, щоб ідентифікувати будь-якого учасника, який є найбільш “цікавим” для публічного висвітлення. Наприклад, журналіст може зіставити загальнодоступні набори даних виборців з анонімізованими медичними записами з метою знаходження і розкриття інформації про політиків чи публічних осіб. На відміну від прокурорської моделі журналіст може мати менш точні, але більш масштабні джерела інформації.

Третя модель – модель **маркетолога** (marketer attacker model) – характеризується принципово іншою метою та підходом. Маркетолог не має завдання ідентифікувати конкретну особу як таку, натомість його цікавить встановлення точних характеристик та особливостей окремих осіб або груп людей з метою персоналізації комерційних пропозицій, реклами чи визначення таргетованих груп. У межах цієї моделі зловмисник не обов’язково має точну допоміжну інформацію про конкретних індивідів, але має доступ до узагальненої інформації, яка дає змогу визначити характеристики та звички потенційних клієнтів. Наприклад, компанія, яка отримала доступ до анонімізованих даних про покупки товарів у магазинах, може використовувати моделі машинного навчання для точного передбачення уподобань, фінансового становища, способу життя чи інших персональних рис окремих людей або груп. На відміну від прокурорської та журналістської моделей маркетингові атаки

зазвичай не призводять безпосередньо до публічного розголошення приватних даних, але можуть спричиняти маніпуляцію поведінкою людей, втручання в їх приватність шляхом таргетованої реклами та інших комерційних дій, що базуються на приватних даних.

Підсумовуючи, основна відмінність цих моделей полягає у мотивації та масштабності: прокурорська модель має вузьку, точну мету – конкретну людину; журналістська модель має більш широку ціль – ідентифікацію будь-якої особи; а маркетингова модель орієнтована на комерційне використання інформації, не обов'язково ідентифікуючи осіб, але порушуючи приватність шляхом прогнозування їхніх характеристик. Вразливість анонімізованих наборів даних до цих атак залежить від застосованого методу захисту: прості методи є найбільш вразливими, тоді як використання складніших методів суттєво знижує ризик успіху цих атак. Саме тому під час вибору методу анонімізації потрібно враховувати, яка з моделей атак є найбільш ймовірною для конкретного набору даних і обирати комплексний підхід до захисту.

Результати дослідження

Для проведення дослідження та практичного аналізу впливу вибору параметра k на анонімізовані дані було вибрано застосунок ARX Anonymization Tool [8]. Як тестовий набір даних було обрано інформацію, рівень доходу осіб, розміром 5692 записи.

ARX Anonymization Tool - це комплексне програмне забезпечення з відкритим кодом, призначене для анонімізації чутливих персональних даних. Воно підтримує широкий спектр моделей конфіденційності та ризиків, методів трансформації даних, а також інструментів для аналізу корисності отриманих даних. Застосунок використовується в різних сферах, наприклад комерційні платформи для аналізу великих даних, наукові дослідження, обмін даними клінічних випробувань та навчальні цілі. Інструмент реалізує такі ключові підходи до забезпечення конфіденційності, як k-анонімність, ℓ-різноманітність, t-близькість та диференційна конфіденційність. Завдяки цьому користувачі можуть гнучко вибирати оптимальні моделі або їхні комбінації залежно від конкретних завдань, типу даних та бажаного рівня захисту. Ще однією суттєвою особливістю ARX є можливість здійснювати детальний аналіз якості анонімізованих даних та оцінювати ризики повторної ідентифікації. Інструмент пропонує вбудовані статистичні моделі та показники оцінки втрати інформації після анонімізації, що дає змогу користувачам ефективно визначати оптимальні параметри для досягнення необхідного рівня конфіденційності та точності результатів.

У використаному тестовому наборі даних містились дані про стать, вік, расу, подружній статус, освіту, країну походження, місце роботи (приватна чи державна структура), рід занять та клас зарплати (більше або менше рівне 50 тисяч доларів на рік). Атрибути країни походження та класу зарплатні були визначені нечутливими, а рід занять – вразливим (sensitive). Усі інші атрибути були визначені квазіідентифікаторами.

Розрахунки в середовищі ARX проводились із такими налаштуваннями: межа придушення (suppression limit) – 100 %, що встановлює максимальну кількість записів, які можуть бути вилучені із початкового набору даних; вага кожного атрибуту – 0.5, позначаючи, що всі атрибути мають бути анонімізовані однаково (що більша вага, то менше втрат даних порівняно з іншими атрибутами).

У дослідженні використовуються ймовірності (у відсотках) успішної атаки, використовуючи моделі прокурора, журналіста та маркетолога. Для оцінки користі анонімізованих даних було використано значення гранулярності [9]. Гранулярність даних (data granularity) - це ступінь деталізації, на якому представлена інформація в наборі даних. Висока гранулярність означає більшу деталізацію (дрібніші одиниці інформації), а низька - меншу деталізацію (укрупнені, узагальнені дані).

Спершу було досліджено алгоритм k-анонімізації, оскільки він є базовим для наступних алгоритмів та найменш вимогливим до обчислювальних потужностей. Для аналізу анонімізованих даних обрано показники k на рівні 2, 5, 10 та 20. Результати експерименту подано на рис. 1. При значенні k=2 дані є досить корисним із показником деталізації 80.4 %, тобто дані зручно використовувати для практичних досліджень. Водночас ймовірність успішної атаки є найвищою – 8.14 % для усіх трьох моделей. Зі збільшенням показника k до значення 20 ймовірність успішності атак зменшується в рази: до 1.22 % - для моделі прокурора та 0.74 % - для моделей журналіста та

маркетолога. Хоча ці значення покращились, як стане зрозуміло згодом, вони досі залишаються доволі високими. Водночас показник гранулярності даних впав до 63.3 %, що робить їх менш деталізованими та погіршує потенційний аналіз чи використання як навчальний датасет. Підсумовуючи, збільшення параметра k забезпечує кращу захищеність даних, але ціною їх деталізації та тривалості обчислень.

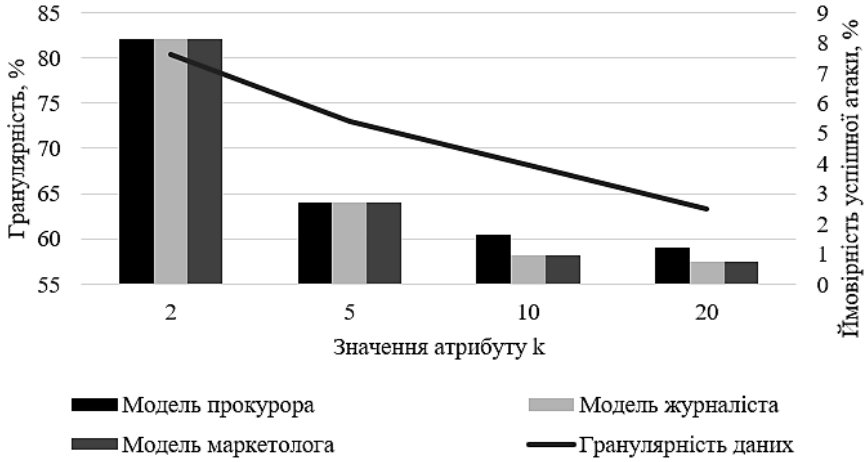

Рис. 1. Залежність успішності атаки та деталізації даних від параметра k-анонімізації

Наступним методом для дослідження став спосіб ℓ-диференціації. Початковий набір даних та інші налаштування середовища ARX залишились незмінними. Для атрибуту роду занять було обрано різні показники ℓ: 2, 4, 6, 10. Показник для основного методу k-анонімізації було встановлено на рівнях 5 і 10, оскільки в попередньому експерименті вони зберігали високу деталізацію при оптимальному захисті даних.

Із результату дослідження, яке подане на рис. 2, можна зробити висновок, що за невисоких значень параметра ℓ (2, 4) на результат досить сильно впливає параметр атрибуту k. Із його збільшенням можна помітити зниження успішності атаки із 2.8 % до 1.5 %. За подальшого підвищення параметра ℓ характеристики анонімізованих даних стають дуже схожими. Успішність ймовірних атак падає до менш як одного відсотка. Як помітно із рис. 2, цей результат викликає зменшення деталізації даних до значень 72 % та 61 %, що свідчить про низьку корисність інформації. Метод ℓ-диференціації демонструє покращений захист конфіденційності порівняно з базовою k-анонімізаіцією, однак завдяки помітній втраті деталізації.

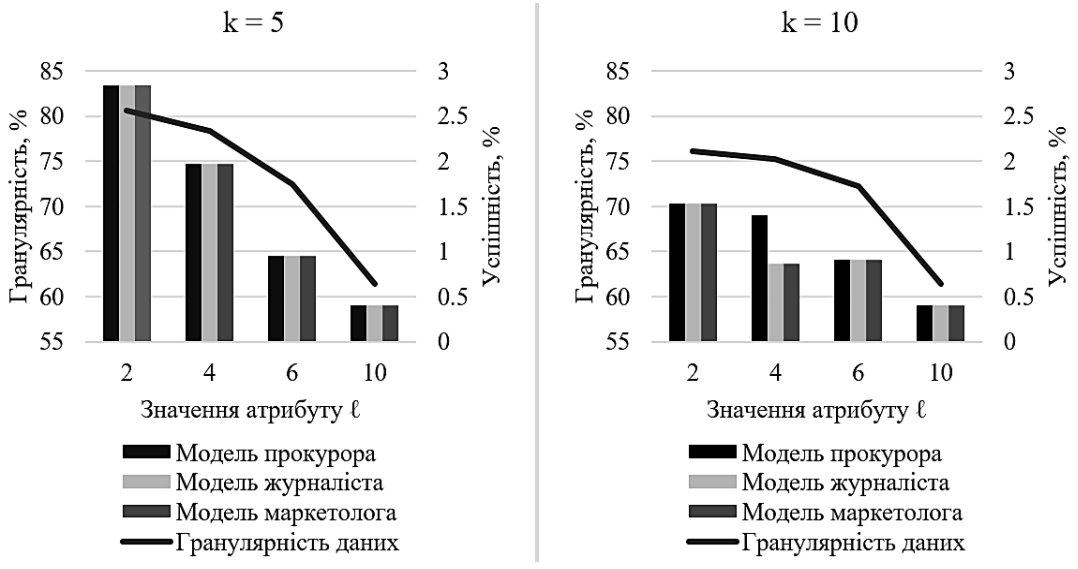

Рис. 2. Залежність успішності атаки та деталізації даних від параметрів ℓ-диференціації

Для подальшого аналізу практичної цінності та захисту даних було досліджено метод t-близькості. Його результати проілюстровані на рис. 3. Як і у попередньому експерименті, значеннями параметра k обрано 5 та 10. Далі проводилась анонімізація набору даних із показниками відстані 0.8, 0.5, 0.2 та 0.1. Можна помітити, що високі значення t (в діапазоні 1- 0.5) демонструють схожий, із незначними покращеннями, рівень захисту даних як і метод ℓ-диференціації. Подальше зниження t дозволило різко зменшити ризик вразливості даних. У разі встановленої відстані 0.2 та менше ймовірності атаки різними моделями суттєво знижуються до діапазону 0.1- 0.03 %, що є найкращим результатом серед всіх досліджених методів, що гарантує високий рівень захищеності персональних даних. Однак і деталізація записів стрімко знижується. За досягнення високого ступеня приватності значення гранулярності даних опускається нижче ніж 40 %. У разі подальшого зменшеня показника відстані значення деталізації даних продовжить падати.

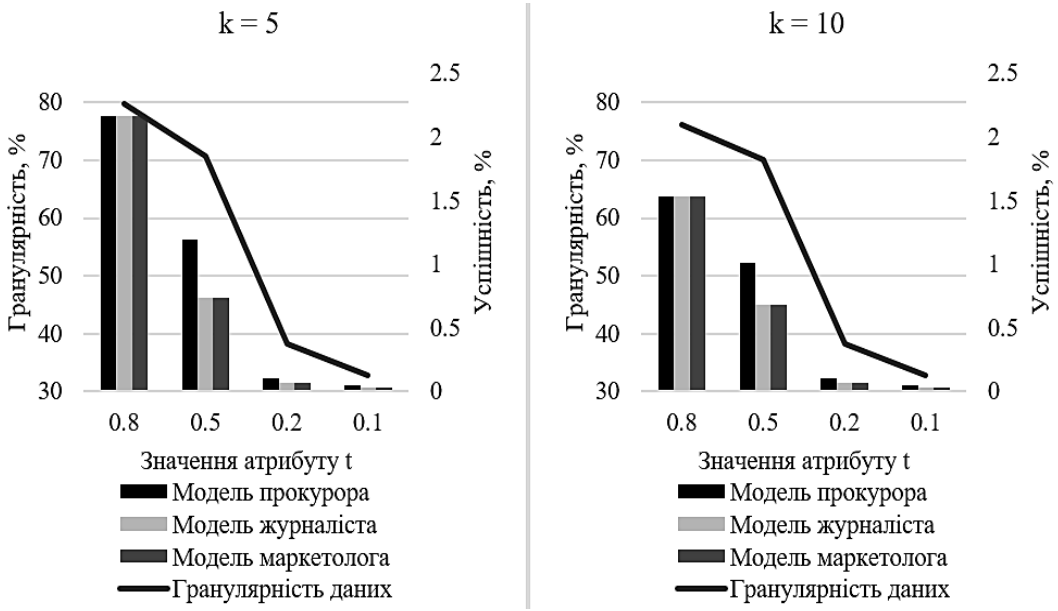

Рис. 3. Залежність успішності атаки та деталізації даних від параметрів t-близькості

Проведене дослідження показало вплив значень методів анонімізації на кінцевий набір даних. Існує тенденція до збільшення рівня захищеності даних у разі встановлення більш строгих умов до методів (збільшення показників k та ℓ, зменшення значення t), але майже завжди збільшення приватності записів приводить до втрати деталізації цих даних, їхньої корисності. Це і не дивно, бо для більшої захищеності дані мають бути більш унікальними та не належати до конкретного користувача. Варто зазначити, що в дослідженні було використано саме математичні моделі атак, що підвищило відсоток їхньої успішності. У реальному світі доволі часто людина, яка намагається деанонімізувати дані, не має точної інформації про наявність запису в наборі даних або додаткові джерела інформації можуть бути неповними або ненадійними.

Якщо враховувати лише ймовірність успішності атаки, то найкраще себе показав метод t-близькості. Найкращі показники деталізації анонімізованих даних та їхньої практичної цінності показує метод k-анонімізації. Вибір параметрів анонімізації є нетривіальною задачею, де найважливіше враховувати цілі, що стоять перед публікаторами, користувачами чи учасниками збору даних. Кожен такий вибір є балансом між захищеністю та корисністю кінцевого результату. У табл. можна побачити ймовірності успішної атаки, деталізації даних та відношення гранулярності до ймовірності компрометації.

Порівняння методів анонімізації персональних даних

Метод анонімізації	Ймовірність успішної атаки за моделлю прокурора, % ↓	Гранулярність даних, % ↑	Коефіцієнт корисності даних до їх захищеності
k(5)-анонімізація	2.71	73	26.93
k(5)-анонімізація, ℓ(4)-диференціація	1.97	78.3	39.75
k(5)-анонімізація, t(0.5)-близькість	1.19	70.6	59.32

Висновки

Здійснено аналіз ефективності популярних методів анонімізації персональних даних, а саме: k-анонімізації, ℓ-диференціації та t-близькості. Дослідження було проведено на прикладі реального набору даних про дохід населення, використовуючи інструмент ARX Anonymization Tool. Основна мета, яка полягала в оцінці ефективності згаданих методів щодо забезпечення приватності та виявлення їхніх вразливостей до атак повторної ідентифікації, була успішно досягнута, що підтверджено представленими кількісними оцінками та аналізом. У процесі дослідження було проаналізовано залежність між вибором параметрів анонімізації (k, ℓ, t) і такими ключовими показниками, як рівень деталізації інформації (гранулярність) та ймовірність успішності атак повторної ідентифікації (моделі прокурора, журналіста та маркетолога). Результати дослідження свідчать про наявність чіткого компромісу між приватністю та корисністю анонімізованих даних. Було встановлено, що метод k-анонімізації забезпечує високий рівень практичної цінності інформації завдяки меншому узагальненню, але водночас є найбільш вразливим до атак. Метод ℓ-диференціації демонструє помірний рівень захисту, знижуючи ризики повторної ідентифікації порівняно з k-анонімністю, однак завдяки помітній втраті деталізації даних. Найкращі результати щодо захисту конфіденційності були отримані під час використання методу t-близькості. Цей підхід суттєво знижує ймовірність успішності атак, роблячи дані майже непридатними для ідентифікації окремих осіб, однак застосування цього методу супроводжується значним зниженням деталізації інформації, що обмежує практичну корисність набору даних для аналітичних завдань.

Важливим практичним результатом роботи є визначення оптимальних значень параметрів анонімізації, що дає змогу знайти баланс між захистом приватності та інформативністю даних, відповідно до конкретних цілей та умов їхнього використання. Наведений у статті порівняльний аналіз може допомогти під час вибору методу анонімізації залежно від конкретних сценаріїв та вимог щодо захисту приватності.

Список літератури

1. Sweeney L. K-anonymity: a model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002. 10(5). 557–570. Doi: https://doi.org/10.1142/S0218488502001648.

2. Machanavajjhala A., Kifer D., Gehrke J., Venkitasubramaniam M. ℓ-Diversity: Privacy Beyond k-Anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD). 2007. 1(1), 3-es. Doi: https:// doi. org/10.1145/1217299.1217302.

3. Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and ℓ-Diversity. In Proceedings of the IEEE 23rd International Conference on Data Engineering (ICDE), 2007. 106–115. Doi: http://dx.doi.org/10.1109/ICDE.2007.367856.

4. Dwork C. Differential Privacy: A Survey of Results. In Theory and Applications of Models of Computation, Lecture Notes in Computer Science. 2008. Vol. 4978. 1–19. Doi: https://doi.org/10.1007/978-3-540-79228-4_1.

5. Fung B. C., Wang K., Chen R., Yu P. S. Privacy-Preserving Data Publishing: A Survey of Recent Developments. ACM Computing Surveys (CSUR)/ 2010. 42(4). 1–53. Doi: https://doi.org/10.1145/1749603.1749605.

6. Ninghui Li, Wahbeh Qardaji, Dong Su, Yi Wu, Weining Yang. Membership privacy: A unifying framework for privacy definitions. In Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, 2013. 889–900. Doi: https://doi.org/10.1145/2508859.2516686.

7. Prasser F., Kohlmayer F., Kuhn K. A. The Importance of Context: Risk-based De-identification of Biomedical Data. Schattauer GmbH. 2016. Doi: http://dx.doi.org/10.3414/ME16-01-0012.

8. Prasser F., Kohlmayer F. Putting Statistical Disclosure Control into Practice: The ARX Data Anonymization Tool. In: Medical Data Privacy Handbook. Gkoulalas-Divanis, A., Loukides, G. (eds). Springer, Cham. 2015. Doi: https://doi.org/10.1007/978-3-319-23633-9_6.

9. What Is Data Granularity? Definition, Types, and More. URL: https://www.coursera.org/articles/data-granularity (Accessed: 12.03.2025).

ANALYSIS OF EFFECTIVENESS AND VULNERABILITIES OF PRIVACY-PRESERVING METHODS USING K-ANONYMITY, L-DIVERSITY, AND T-CLOSENESS AS EXAMPLES

O. O. Ivaniuk, A. M. Vakhula

Lviv Polytechnic National University,
Department of computerized automation systems

The article analyzes and compares personal data anonymization methods using k-anonymity, ℓ-diversity, and t-closeness as examples. The aim of the research is to evaluate the effectiveness of these methods in ensuring data privacy and identifying their vulnerabilities to re-identification attacks. The study was performed using the ARX Anonymization Tool on a test dataset containing personal income information.

The authors analyzed the impact of changes in key parameters of anonymization methods on data privacy and informativeness. It was determined that the t-closeness method provides the highest effectiveness in terms of protecting confidentiality, although its application significantly reduces the granularity of information. At the same time, the k-anonymity method, despite being less resistant to attacks, provides better practical utility of anonymized data. The ℓ-diversity method demonstrates moderate effectiveness in terms of both privacy protection and informativeness. This research allowed quantitative assessment of the balance between data confidentiality and utility, facilitating a more informed choice of optimal anonymization parameters depending on specific tasks.

Keywords: data anonymization, personal data protection, privacy, k-anonymization, ℓ-diversity, t-closeness.