УДК 004.056.53 DOI 10.31673/2412-4338.2025.027860

Складанний Павло Миколайович
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0002-7775-6039

Машкіна Ірина Вікторівна
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0003-0667-5749

Рзаєва Світлана Леонідівна
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0002-7589-2045

Костюк Юлія Володимирівна
Київський столичний університет імені Бориса Грінченка, Київ, Україна
ORCID ID: 0000-0001-5423-0985

МЕТОДИ GDPR ДЛЯ ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ СХОВИЩ ДАНИХ ВІД ВИТОКІВ ТА ЗАГРОЗ

***Анотація**: У статті досліджено проблематику захисту сховищ даних від витоків і загроз у контексті впровадження у механізми безпеки сховища вимог Загального регламенту захисту персональних даних (GDPR), що є критично актуальним у світлі зростання обсягів оброблюваної інформації. Автори акцентують увагу на відсутності спеціалізованих досліджень, які б розглядали питання відповідності безпеки саме сховищ даних стандартам GDPR. Визначено, що майже всі цифрові платформи, організації та інститути використовують сховища даних, а отже, саме ці об'єкти є вразливими до різноманітних загроз – як зовнішніх (хакерські атаки, фішинг, шкідливе програмне забезпечення), так і внутрішніх (людський фактор, витоки через навмисні або ненавмисні дії співробітників установ / організацій).*

У межах даного дослідження систематизовано п'ять основних груп загроз для сховищ даних: несанкціонований доступ, внутрішні витоки, фішингові атаки, шкідливе програмне забезпечення та проблеми неналежного зберігання та резервного копіювання даних. Для кожної з категорій наведено реальні приклади порушень безпеки за останні п'ять років, що ілюструють вплив подібних загроз на безпеку персональних даних. Також у статті приділено увагу щодо застосування методів шифруванню даних (AES-256, TLS), контролю доступу, моніторингу, псевдонімізації, анонімізації, резервному копіюванню та автоматизації обробки, регламентованих статями GDPR.

Дослідження має не лише аналітичну, але й прикладну цінність, оскільки формує комплексне бачення того, як реалізувати технічну та організаційну безпеку сховищ даних у відповідності до статей GDPR. Описані моделі впровадження політик безпеки, механізмів журналювання та перевірки цілісності даних. Окремо підкреслюється важливість створення культури інформаційної безпеки та підвищення обізнаності співробітників.

***Ключові слова:** GDPR, безпека даних, витік інформації, сховища даних, шифрування, загрози, кібербезпека, атаки.*

Skladannyi Pavlo
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0002-7775-6039

Mashkina Iryna
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0003-0667-5749

Rzaeva Svitlana
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0002-7589-204

Yuliia Kostiuk
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID ID: 0000-0001-5423-0985

METHODS OF GDPR FOR ENSURING DATA STORAGE SECURITY AGAINST LEAKS AND THREATS

***Abstract**: This article explores the issue of protecting data storage systems from leaks and threats in the context of implementing the security requirements of the General Data Protection Regulation (GDPR). This topic is critically relevant given the increasing volumes of processed information. The authors emphasize the lack of specialized studies that examine the compliance of data storage security with GDPR standards. It has been determined that nearly all digital platforms, organizations, and institutions rely on data storage solutions, making them highly vulnerable to various threats—both external (hacker attacks, phishing, malware) and internal (human factor, intentional or unintentional leaks by employees of institutions or organizations).*

This study systematizes five main categories of threats to data storage security: unauthorized access, internal data leaks, phishing attacks, malware, and issues related to improper data storage and backup management. For each category, real-world examples of security breaches from the past five years are provided, illustrating the impact of these threats on the protection of personal data. The article also examines the application of security measures such as data encryption (AES-256, TLS), access control, monitoring, pseudonymization, anonymization, backup strategies, and the automation of data processing, in accordance with various articles of the GDPR.

This research holds not only analytical but also practical value, as it provides a comprehensive perspective on how to implement both technical and organizational security measures for data storage in compliance with GDPR requirements. The study describes models for enforcing security policies, logging mechanisms, and data integrity verification. Additionally, it underscores the importance of fostering a culture of information security and raising awareness among employees.

***Keywords**: GDPR, data security, information leakage, data storage, encryption, threats, cybersecurity, attacks.*

Вступ

Захист комерційних даних став обов'язковою процедурою в діяльності комерційних організацій, установ та державних структур. Прийнятий Європейським союзом Загальний регламент захисту даних (GDPR) визначає обов'язкові вимоги щодо оброблення, зберігання та передавання як персональних, так і комерційних даних. Майже всі структури державної та комерційної сфери застосовують сховища даних для зберігання даних, а отже дотримання положень прописаних GDPR є фундаментальним критерієм у розрізі забезпечення безпеки сховищ даних від витоків та загроз. У статті досліджуються основні методи та специфічні для сховищ даних заходи безпеки та аналізуються потенційні загрози, які можуть виникати в процесі зберігання, обробки та передачі даних та для підвищення надійності та цілісності сховищ даних (СД). При цьому дослідження щодо впровадження положень GDPR для забезпечення безпеки сховищ даних від витоків та загроз ще не проводились.

Аналіз літературних даних і постановка проблеми

Аналіз наукових статей та праць показує, що більшість досліджень з кібербезпеки застосовують методи захисту даних у відповідності вимог GDPR. Але немає жодної статті, яка б дослідила питання безпеки саме сховищ даних у відповідності до стандарту GDPR. Розглянемо різни напрями досліджень пов'язані з врегулюванням обробки, зберігання та передачи даних, прописаних у Загальному регламенті захисту даних.

Жаклін Ієвінас у статті «Erasure and Anonymisation of Personal Data in Context of General Data Protection Regulation» розглядається, як GDPR регулює два терміни «видалення» та «анонімізацію», а також які вимоги пред'являються до використання будь-якого з них у життєвому циклі персональних даних [4]. Зобов'язання видаляти персональні дані завжди вимагає персональних даних. У випадку анонімних даних стирання не є обов'язковим і не може бути заявлено. Авторка розглядає ключові підходи до обробки персональних даних, включаючи алгоритми деперсоналізації, та їхню відповідність статті 17 GDPR (право на видалення даних). Основна увага приділяється практичним аспектам застосування методів шифрування та маскування для мінімізації ризику повторної ідентифікації користувачів.

Стаття А. Павлова «Анонімізація даних за допомогою блокчейн-технологій» описує модель анонімізації даних на основі технологій блокчейн, яка поєднує смарт-контракти для автоматизації операцій з даними та використання криптографічних методів для створення системи, стійкої до деанонімізації, зокрема в контексті забезпечення прозорості і відповідності вимогам GDPR [5]. Модель забезпечує контроль і відповідність регуляторним вимогам, зберігаючи прозорість і безпеку всіх транзакцій. У статті також описані криптографічні методи, що можуть бути використані для псевдонімізації інформації без втрати її корисності. Робота підкреслює переваги Використання блокчейн-рішень для анонімізації даних у відповідності вимог GDPR, особливо щодо захисту даних і забезпечення прав осіб.

O. Prillard, C. Boletsis, S. Tokas досліджують та аналізуються питання конфіденційності та захисту персональних даних у віртуальному середовищі метавсесвіту. У статті «Ethical Design for Data Privacy and User Privacy Awareness in the Metaverse» робиться акцент на використанні методів деідентифікації та децентралізованих систем управління особистими даними [8].

Дотичним до нашої тематики є статті авторів О. Борисенко, А. Тимошенко та Б. Фасій, в яких розглядаються питання захисту персональних даних у хмарному середовищі в епоху цифрових технологій. У статті «Огляд методів захисту персональних даних у хмарному середовищі» авторів О. Борисенко та А. Тимошенко досліджуються методи захисту персональних даних у хмарних середовища, особливу увагу приділяючи питанням все більш зростаючим ризикам втрати персональних даних [11]. Особливий акцент автори роблять на те, що треба приділяти увагу не лише технічному захисту, але й методам управління до даних у хмарному середовищі на користувацькому рівні, а саме застосування багатофакторної автентифікації у поєднанні з сертифікацією безпеки хмарних сервісів.

Стаття «Національна безпека та захист персональних даних в епоху цифрових технологій» автора Б. Фасія вивчається зв'язок між захистом персональних даних та національною безпекою, в умовах цифровізації всіх сфер суспільного та національного устрою [10]. Автор акцентує увагу на те, що захист персональних даних є не лише питанням приватності особи, але є ще питанням національної безпеки. Тому він наголошує на необхідності впровадження у національних стандарти безпеки положень GDPR, посилення співпраці між державними та урядовими установами і комерційними / приватними закладами з метою забезпечення не лише високо рівня захисту даних, але й обміном інформації, її захисту та контролю.

Отже статей, які б досліджували питання забезпечення безпеки сховищ даних від витоків та загроз у розрізі стандартів Загального регламенту захисту даних (GDPR), прийнятий Європейським союзом, наразі немає. У даній статті зроблено спробу розв'язання даної проблеми.

Мета і задачі дослідження

Метою дослідження є аналіз методів безпеки сховищ даних у відповідності вимог GDPR та їх ефективності у розрізі запобігання витоку інформації та захисту від кібератак.

Для досягнення поставленої мети вирішено такі завдання;

- проаналізувати нормативно-правові вимоги GDPR, які стосуються захисту персональних даних у сховищах;
- ідентифікувати та класифікувати основні загрози характерні саме для сховищ даних, включаючи зовнішні та внутрішні ризики;
- вивчити та систематизувати технічні та організаційні заходи, рекомендовані GDPR для забезпечення безпеки даних;
- провести аналіз інцидентів пов’язані з витоком даних за останні п’ять років, з метою аналізу виявлених помилок у захисті сховищ та підтвердити практичну значущість застосування положень GDPR;
- сформувати рекомендації щодо впровадження політик інформаційної безпеки, які відповідають принципам GDPR та запобігають витокам інформації зі сховищ даних.

Результати дослідження

Наразі всі державні, комерційні та некомерційні структури використовують сховища даних для зберігання великих обсягів критично-важливої інформації з метою ефективного управління. Також всі сучасні інформаційні системи використовують сховища даних для зберігання структурованих даних, з подальшою обробкою для прийняття ефективних управлінських рішень, бізнес аналітики тощо. У зв’язку з цим питання безпеки сховищ даних відіграють критичного значення, адже будь-яке порушення цілісності даних, конфіденційності або доступності інформації може привести до значних втрат та численних небезпечних наслідків. Особливу роль відіграють питання дотримання стандартів GDPR у розрізі забезпечення конфіденційності, цілісності, доступності персональних даних та захисти їх від виток та загроз.

Види загроз, яким піддаються сховища даних:

- хакерські атаки та несанкціонований доступ до даних;
- витік даних через випадкові або навмисні дії внутрішніх працівників;
- фішингові атаки з метою отримання доступу до конфіденційної інформації;
- шкідливе або зловмисне програмне забезпечення;
- неналежне зберігання даних та/або відсутність резервного копіювання.

Більш детально опишимо кожну загрозу.

Хакерські атаки та несанкціонований доступ до даних

Однією з найбільш критичних загроз залишаються хакерські атаки, спрямовані на отримання несанкціонованого доступу до інформаційних ресурсів сховища даних. Метою зловмисників у даній атаці є отримання несанкціонованого доступу до персональних, службових або конфіденційних даних, яка зберігаються в сховище. Це порушує принципи конфіденційності, цілісності та доступності даних, закріплені статті 679 GDPR. Витік вказаних даних може спричинити значну шкоду для суб’єктів державного, комерційного або некомерційного суб’єктів господарювання і призвести до накладення штрафів на організацію / установу відповідно до статті 83 GDPR.

Доступ до захищеного сховища даних ґрунтується на принципі авторизованого доступу до об'єкта даних:

$$\forall u \in U, \forall d \in D : A(u, d) = 1 \Rightarrow u \in Uauth(d), \quad (1)$$

де U – множина користувачів, D – множина об'єктів даних, A(u,d) – функція, що описує дозвіл користувача u до об'єкта d, $U_{auth}(d)$ – підмножина авторизованих користувачів.

Хакерські атаки, що порушують цю відповідність, створюють умови, коли A(u, d) = 1, але $u \notin U_{auth}(d)$, тобто доступ отримує неавторизований суб'єкт.

Зловмисники застосовують низку технічних методів для обходу механізмів аутентифікації й авторизації. До таких методів належать:

Атаки типу «brute-force», коли підбираються паролі методом повного перебору. У середньому, для перебору пароля довжини n, складеного з k можливих символів, хакеру необхідно перебрати k^n комбінацій. Наприклад, для пароля, який складається з 8 латинських букв і цифр (k=62), кількість комбінацій буде $62^8 \approx 2.18 \times 10^{14}$.

Атаки за допомогою SQL-ін'єкції, де зловмисник за допомогою шкідливого SQL-коду виконує запит до сховища даних через уразливі форми введення (веб-форми або URL-запити) з метою обходу механізмів аутентифікації й авторизації та отримання несанкціонованого доступу до інформаційних ресурсів сховища. Типовий приклад SQL-ін'єкції:

*SELECT * FROM users WHERE username = ' ' OR 1 = 1; --*

В SQL-запиті, який був сформований SQL-ін'єкцією, вираз *username = ' '* містить порожній рядок. Якщо система очікує, що користувач має ввести власне ім'я у текстовому полі, то даний вираз передає ім'я користувача як порожній рядок (тобто ім'я не було введене). Логічний вираз *OR 1 = 1* завжди повертає TRUE, незалежно від значення поля *username*. У даному випадку умова *WHERE* приймає значення TRUE для всіх рядків у таблиці. Символи -- в кінці SQL-запиту означають початок коментаря і все, що йде після них – ігнорується. А це означає, що, навіть при наявності у вихідному коді перевірки паролю після *username*, дана перевірка буде ігноруватись.

Отже, використовуючи дану вразливість зловмисник може:

- обійти автентифікацію, тобто якщо вищенаведений SQL-запит використовується для перевірки логіна, то він поверне всі записи користувачів і дозволить увійти зловмиснику без пароля;
- отримати всі рядки з таблиці *users*, що може призвести до витоку конфіденційної інформації;
- виконати модифікацію або видалення даних. SQL-ін'єкції дозволяють виконувати не лише запити *SELECT*, а й *UPDATE, DELETE, INSERT*. Наприклад, наступний SQL-вираз *DROP TABLE users; --* приведе до видалення таблиці з інформацією про користувачів у сховищі даних.

Наступний метод обходу механізмів автентифікації й авторизації – це злам через відкриті порти або некоректно налаштовані інтерфейси програмного забезпечення, зокрема інтерфейси REST API, які не захищені відповідними токенами автентифікації. Архітектура REST API використовує протоколи автентифікації, такі як веб-протокол JSON (JWT), OAuth 2.0 для захисту користувача. Неправильна конфігурація, відсутність перевірки підпису токена або некоректна обробка часу його дії, створює умови для обходу механізмів автентифікації та авторизації. У цьому випадку зловмисник без підтвердження прав доступу може надсилати запити до відкритих ендпоінтів API і отримувати конфіденційну інформацію.

Прикладом хакерської атаки направленої на обхід механізмів автентифікації й авторизації є атака на компанію з охорони здоров'я Medibank, яка відбулась у жовтні 2022 року та призвела до витоку конфіденційних даних близько 9,7 мільйонів пацієнтів, а саме: повні імена та дати народження, номери паспортів, медичні записи про діагнози. Зловмисники отримали доступ до сховищ даних через вразливість у системі автентифікації адміністративного облікового запису. Розслідування показало, що облікові дані були викрадені через скомпрометований обліковий запис із доступом до Microsoft 365.

Даний інцидент був розцінений як порушення безпеки персональних даних відповідно до статті 4(12) GDPR із витоком даних у Medibank, що призвів до несанкціонованого доступу до великого обсягу чутливої медичної інформації. Оскільки обробці підлягали медичні дані, які підпадають під спеціальні категорії персональних даних згідно зі статтею 9, їхній витік вимагає особливої уваги та додаткових заходів захисту, а згідно статті 33 компанія протягом 72 годин має повідомити про порушення безпеки.

Витік даних через випадкові або навмисні дії внутрішніх працівників

Безпосередньо витік даних через випадкові або навмисні дії співробітників організації / установи відноситься до так званих внутрішніх загроз і являє собою найнебезпечнішу загрозу захисту інформації в сховищі даних, бо такі загрози дуже складно виявити. Працівники, на відміну від зловмисників, вже мають певний рівень доступу до інформації, яка зберігається у СД, можуть ненавмисно та/ або навмисно бути причасниками до витоку інформації, що є порушенням вимог GDPR, особливо у частині захисту персональних даних. Часто через відсутність контролю над ролями та обліковими записами у сховищі даних працівники можуть піднімати власні привілеї до тієї чи іншої конфіденційної інформації або надавати привілеї доступу до неї іншим співробітникам, що в свою чергу, створює додаткові ризики витоку конфіденційної інформації.

Випадкові або ненавмисні витоки виникають в наслідок відсутності чіткої інструкції обробки інформаційних даних, необережних дій співробітників, їхньої недостатньої обізнаності або через брак компетентності, нехтуванням правил інформаційної гігієни та безпеки. Прикладами таких дій можуть бути:

- помилкове надсилання електронного листа не зареєстрованому адресату з інформацією про персональні дані клієнтів, працівників або конфіденційними даними про банківські рахунки, укладені договори тощо (згідно з GDPR дана ситуація кваліфікується як data breach, навіть якщо співробітник не мав злого умислу);
- використання незахищених особистих пристроїв (таких як ноутбук чи смартфон) для роботі з даними сховища;
- розміщення файлів з чутливою інформацією у відкритих хмарних сховищах (наприклад, Google Drive або Dropbox) без належних налаштувань доступу до даного середовища.

Додаткову загрозу становить зовнішній вплив на внутрішніх працівників, зокрема через маніпуляції з боку зловмисників, які спонукають співробітників ненавмисно передавати конфіденційну інформацію або розкривати деталі внутрішньої архітектури сховища даних. Особливо вразливими є нові співробітники або ті, що працюють дистанційно й не мають належної взаємодії з командою безпеки. Такий вплив часто здійснюється за допомогою так званої «соціальної інженерії», що являє собою методику психологічного впливу, який зловмисники використовують для маніпулювання людьми з метою отримання доступу до конфіденційної інформації або здійснення певних дій, що можуть зашкодити безпеці даних. До такого методу відносяться, зокрема, фішингові електронні листи, підроблені повідомлення або дзвінки, під час яких шахраї прикидаються працівниками служби безпеки або керівниками й випитують паролі, коди доступу тощо. Крім того, ризики значно зростають у разі відсутності сформованої культури

інформаційної безпеки в організації, недостатньої кількості навчань із кібергігієни та неналагодженої процедури перевірки запитів і звернень. У сукупності ці чинники створюють сприятливі умови для витоку або компрометації важливих даних, навіть без навмисних дій з боку працівників.

Навмисні витоки трапляються у випадках, коли співробітник свідомо копіює, передає або викрадає дані з особистих мотивів чи в інтересах третіх осіб. Такі інциденти становлять серйозну загрозу для безпеки сховищ даних, оскільки зазвичай здійснюються особами, які мають або мали офіційний доступ до конфіденційної інформації. Звільнений або демотивований працівник, у якого зберігся доступ до СД та привілеї доступу до конфіденційної інформації, може навмисно пошкодити або скопіювати чутливі дані перед або після звільнення, з метою передання чи продажу зацікавленим особам. Наприклад, у 2022 році у Великобританії було виявлено випадок, коли колишній співробітник страхової компанії викрав базу контактів клієнтів і передав її конкуренту, завдавши компанії фінансових та репутаційних збитків.

У деяких випадках працівник за грошову винагороду або інші вигоди погоджується передати зловмисникам конфіденційну інформацію, паролі доступу до сховища даних або ключі шифрування. Особисті конфлікти або бажання помститися роботодавцю також можуть стати рушієм для свідомого викрадення інформації або пошкодження даних. Іноді співробітники використовують корпоративні дані у власних фріланс-проєктах або передають їх стороннім компаніям (наприклад, для маркетингових чи аналітичних цілей) не усвідомлюючи порушення політик конфіденційності та вимог GDPR або свідомо їх ігноруючи.

Несанкціоноване копіювання даних зі СД на власні пристрої, флешки або незахищені хмарні сервіси створює ризики витоку конфіденційної інформації, компрометації персональних даних та порушення вимог GDPR і може залишитися непоміченим, особливо якщо в установі / організації не впроваджено систему журналювання подій у реальному часі, а саме аудиту та / або моніторингу дій користувачів. Крім того, відсутність механізмів обмеження доступу до чутливої інформації та шифрування даних може призвести до викрадення або втрати даних у разі фізичної крадіжки даних через зовнішні пристрої або їх зараження шкідливим програмним забезпеченням. Якщо такі дії виявляться, установа / організація може зіткнутися з репутаційними втратами, фінансовими штрафами та юридичною відповідальністю.

Без ефективних механізмів контролю користувацької активності такі дії важко виявити, тому необхідно проводити аудит безпеки та управління доступом до об’єктів СД. Вказані дії дозволяють зафіксувати підозрілі дії, виявити порушення політики інформаційної безпеки та вчасно попередити виток даних. Усе вищенаведене створює необхідність побудови в установі / організації системи внутрішнього контролю користувацької активності, розробити механізм обмеження доступу за принципом «найменших необхідних прав» доступу до інформації, проведення моніторингу активності користувачів і формування етичної корпоративної культури.

Фішингові атаки

Фішингові атаки спрямовані на обмане викрадення облікових даних користувачів, їхніх паролів, ключів для шифрування / дешифрування даних або іншої конфіденційної інформації. Даний вид атак є одним із найпоширеніших та найнебезпечніших засобів соціальної інженерії, через його простоту реалізації та високу ефективність. У класичному варіанті фішингова атака реалізується шляхом надсилання електронних листів, які імітують повідомлення від керівництва, колег або представника IT-відділу, в яких зловмисники просять перейти за покликанням або оновити облікові дані. Покликання веде на підроблений вебсайт, який візуально не відрізняється від оригінального, де користувач самостійно вводить свої логін і пароль, не підозрюючи, що передає їх зловмиснику. Аби така атака була успішною зловмисник попередньо вивчає профілі співробітників у соціальних мережах, аналізує стиль спілкування компанії, її структуру,

внутрішні процеси тощо. Далі проводиться цілеспрямована атака на конкретну особу, яку зловмисник обрав для цього. У результаті атака виглядає надзвичайно достовірною, а ризик помилки користувача – дуже високим.

Особливу загрозу фішинг становить для сховищ даних, які містять у соді мільйоні записів критично важливої інформації. Отримавши доступ до облікового запису, зловмисник може безперешкодно завантажити, змінити або видалити інформацію, завдавши репутаційної шкоди установі / організації. Перш за все фішингові атаки спрямовані на системних адміністраторів або менеджерів вищої ланки управління, які мають підвищені привілеї доступу до чутливої інформації та конфіденційних даних. У сфері захисту сховищ даних фішинг виступає не просто соціотехнічним методом атаки, а початковим етапом ланцюга зловмисних дій спрямованих на викрадення, пошкодження або видалення даних сховища і дозволяє зловмиснику, на цьому етапі, пройти автентифікацію в СД, під виглядом зареєстрованого користувача. У середовищах, де реалізовано модель доступу на основі ролей (RBAC – Role-Based Access Control), компрометація облікового запису користувача із правами доступу до схем, таблиць або контрольних точок ETL-процесів може призвести до витоку або модифікації критичних даних. На другому етапі зловмисник, здобуває повний доступ до схем, таблиць, представлень, збережених процедур сховища даних, отримує можливість виконувати SELECT-запити, створювати об'єкти, експортувати результати. А у випадку слабкої політики доступу – можливість переглядати дані з усіх бізнес-підрозділів (HR, фінанси, логістика тощо).

Успішна фішингова атака може не лише призвести до витоку конфіденційної інформації, а й забезпечити несанкціонований доступ до сховища даних, що створює передумови для подальших атак, таких як встановлення бекдорів, викрадення ключів шифрування або зараження інформаційної системи шкідливим програмним забезпеченням.

Уявімо ситуацію, коли адміністратору або керівникові вищої ланки установи / організації (який має високий рівень доступу до даних сховища) надсилається спеар-фішинг (spear-phishing) лист із повідомленням про оновленням політики безпеки. Відкривши вкладення, користувач запускає скрипт на PowerShell, який автоматично зчитує токени автентифікації з кешу браузера, передає дані токени через відкритий HTTP-запит на керований зловмисником сервер (C2 – Command and Control), а разі успішної компрометації облікового запису – ініціює несанкціонований доступ до сховища даних через API або Web UI.

У результаті даної атаки зловмисник отримує дозвіл на здійснення таких дій:

1. Виконувати SQL-запити типу *SELECT * FROM sensitive_data.customers* або *EXFIL INTO EXTERNAL STORAGE.*
2. Змінювати політики безпеки для будь-яких об'єктів сховища (наприклад, *GRANT ALL ON warehouse TO attacker_user*).
3. Завантажувати конфіденційні дані до зовнішніх S3- або Azure Blob-контейнерів.

Виконання запиту *SELECT * FROM sensitive_data.customers* дозволить зловмиснику переглядати дані таблиць, які містять конфіденційну інформацію, тобто даний запит витягується вміст усіх полів із таблиці *customers*, що розміщена у схемі *sensitive_data*. Внаслідок цього система обробки даних формує відповідь і, залежно від платформи, тимчасово кешує результат в оперативній пам’яті або у внутрішньому СД. У таблиці *customers* зазвичай містяться персональні дані: ПІБ, адреси електронної пошти, номери телефонів, ідентифікаційні коди, фінансові реквізити, а також можливі логіни, хешовані або відкриті паролі, API-ключі чи токени доступу. Усе це може бути отримано в одному запиті, без перешкод з боку системи, якщо обліковий запис має відповідні дозволи, а моніторинг усіх подій пов'язаний з виконанням такого виду запитів не налаштовано. Таким чином, на цьому етапі зловмисник може повністю скопіювати зміст таблиці для подальшого використання та експортування використовуючи запит *EXFIL INTO EXTERNAL*

STORAGE до сторонніх джерел, зокрема: AWS S3 (*COPY INTO @s3_stage*), Google Cloud Storage, Azure Blob Storage.

У сучасних хмарних сховищах даних, таких як Snowflake, Google BigQuery, Amazon Redshift або Azure Synapse, реалізовано механізми експорту даних до зовнішніх джерел, таких як об'єктні сховища (AWS S3, Google Cloud Storage чи Azure Blob Storage). Зловмисник використовуючи стандартні SQL-командами, може створити зовнішнє сховище *stage* або точку призначення для експорту (*destination point*) і налаштувати її таким чином, щоб дані зі сховища передавалися на його власний сервер або хмарне сховище, а це, в свою чергу, дозволяє йому непомітно вивантажувати конфіденційну інформацію. Виконуючи команду типу *COPY INTO @external_stage/customers_data FROM (SELECT * FROM sensitive_data.customers)*, зловмисник ініціює експорт великого обсягу даних у вигляді файлів (зазвичай CSV, JSON або Parquet). У процесі такого експорту сховище, як правило, не зберігає самих файлів у внутрішньому середовищі, тому без належного моніторингу доступу до зовнішніх джерел або аудиту експорту слід даної атаки може бути непоміченим. У результаті відбувається виток даних зі сховища.

Після успішного входу до сховища даних під компрометованим обліковим записом із розширеними правами доступу до об'єктів сховища, зловмисник може змінювати політики доступу до таких об'єктів. Зокрема, це стосується SQL-команд типу *GRANT*, які дозволяють надавати або відкликати доступ до баз, таблиць, схем, функцій, обчислювальних ресурсів (кластерів warehouse), зовнішніх stage-об'єктів тощо. SQL-команда *GRANT ALL ON warehouse TO attacker_user* дає хакеру повний контроль над обчислювальними потужностями СД, а отже, над усіма запитами, які проходять через даний ресурс. Також він може змінити поточні політики доступу, знявши обмеження для себе або, навпаки, заблокувавши системних адміністраторів сховища даних, що в свою чергу, ускладнить виявлення, ліквідацію та / або блокуванню даного інциденту. Даний інцидент становить особливу загрозу тим, що зміна політик безпеки може здійснюватися за допомогою стандартних SQL-команд, які, на перший погляд, не викликають підозри в системах моніторингу, особливо у разі відсутності належного аудиту дій, пов'язаних зі зміною привілеїв. А отже, дає змогу зловмиснику забезпечити собі прихований та довготривалий доступ до сховища даних без необхідності встановлення додаткового шкідливого програмного забезпечення. У результаті відбувається захоплення керуванням доступом до об'єктів, а переадміністрування ролей і привілеїв у сховищі може бути виявлено лише аудитом журналів доступу та змін у системі авторизації.

Приклад фішингової атаки на сховище даних сталася атака, яка відбулась у квітні 2023 року на Департамент соціальних служб штату Іллінойс (IDHS), унаслідок якої зловмисники отримали доступ до облікових записів співробітників і, через них, до таблиць СД із конфіденційними даними. Це призвело до масштабного порушення безпеки, що привело до витоку персональної інформації понад мільйона осіб (імена, номери рахунків, адреси, дати народження, ідентифікаційні номери та інше). Після інциденту постраждалим особам було запропоновано послуги з моніторингу кредитної історії (звична практика у США після витоків даних: для постраждалих клієнтів або співробітників замовляють сервіси захисту особистої фінансової інформації, з метою оперативного виявлення фінансових шахрайську дій з карками постраждалих осіб), а також, з метою запобігання майбутнім фішинговим атакам, проведено навчання співробітників з інформаційної безпеки та кібергігієни.

Даний випадок демонструє руйнний вплив фішингової атаки та підкреслює необхідність запровадження багаторівневих заходів безпеки, включаючи двохфакторну автентифікацію, постійного моніторингу та аудиту процесів, які відбуваються у сховищі даних, та проведення регулярних тренінгів для співробітників з кіберзахисту та кібергігієни.

Шкідливе або зловмисне програмне забезпечення

Шкідливе або зловмисне програмне забезпечення (malware), до яких, зокрема, належать віруси, трояни, програми-стікери (stealer), кейлогери, програми-шантажисти (ransomware), являє собою небезпечний інструмент для отримання несанкціонованого доступу до сховищ даних або їх пошкодження. До сховища вони потрапляють різними шляхами через: фішингові листи, заражені вкладення або посилання, зовнішні носії, заражені бібліотеки або скомпрометоване програмне забезпечення з відкритим кодом. Сучасні malware здатне маскувати свої дії під легітимну активність. Наприклад, програма може використовувати легальні SQL-запити для витоку даних або приховування своїх слідів.

Особливо небезпечним видом шкідливого програмного забезпечення є програми-здирники (ransomware), які після проникнення до сховища шифрують великі обсяги даних, у тому числі резервні копії даних або архівні СД. Одним із прикладів таких атак є інцидент, що стався у 2021 році, коли хакерське угруповання DarkSide здійснило атаку з використанням програми-здирника (ransomware) на американську компанію Colonial Pipeline. Хоча безпосереднім об'єктом атаки були ІТ-системи управління бізнес-процесами, у тому числі білінгові сервіси, що зберігалися в централізованих СД, компанія була змушена призупинити роботу, щоб запобігти подальшому поширенню зловмисного ПЗ. Сховища даних відігравали ключову роль у функціонуванні внутрішніх облікових і фінансових систем, а порушення доступу до них призвело до дезорганізації операцій і серйозних фінансових збитків.

Також існують шкідливі програми, які використовують уразливості типу *Remote Code Execution (RCE)* у СУБД PostgreSQL або MySQL, що дозволяє хакерам виконувати довільний код на сервері та приводить до повного контролю над сховищем. Відомі випадки, коли такі експлойти застосовувалися у поєднанні з ботнетами, які автоматично сканували мережу в пошуках вразливих серверів сховищ даних.

Ще один ризик для безпеки сховища даних полягає в тому, що сховища не завжди включаються до антивирусного сканування або мають обмеження в інтеграції з EDR (Endpoint Detection and Response) чи AV-системами через побоювання негативного впливу на продуктивність. Сховища послуговують для проведення аналітичних розрахунків та задач, критично важливих, у режимі реального часу. Тому антивірусну перевірку часто виключають із-за глибокого сканування задля уникнення затримок у доступі до критично важливих для бізнесу даних. У результаті створюються так звані «сліпі зони» безпеки, в яких шкідливе програмне забезпечення протягом тривалого часу може залишатися непоміченим. Даний ризик характерний для великих аналітичних платформ, що працюють на базі реляційних СУБД (PostgreSQL, MS SQL, Oracle) або сучасних NoSQL-рішень (MongoDB, Cassandra), де традиційні агенти безпеки можуть бути несумісними або технічно складними для інтеграції. Зловмисник впроваджують бекдори, виконують скрипти шифрування або зберігають шкідливі файли прямо у базах, використовуючи приховані таблиці, функції або BLOB-об'єкти (Binary Large Objects). У 2022 році було зафіксовано низку атак, у яких шкідливі SQL-функції зберігалися безпосередньо в метаданих PostgreSQL-сховища, де їх не могли виявити стандартні сканери, адже вони не працювали на рівні бази даних. Такі приховані елементи запускалися при кожному зверненні до певної таблиці, забезпечували зловмиснику персистентність (стійке закріплення у системі) і канал збирання інформації. Таким чином, нехтування системним моніторингом і антивірусним контролем сховищ даних створює умови для впровадження зловмисниками шкідливого програмного забезпечення, з метою приховання зловмисної діяльності, дозволяючи їм тривалий час маніпулювати критично важливою інформацією, без їхнього явного виявлення та знешкодження.

Неналежне зберігання даних та/або відсутність резервного копіювання

Сучасні бізнес установи / організації неналежним чином приділяють увагу правильній організації збереження даних, а саме: використовуються застарілі формати, відсутня чітка політика доступу до даних, ігноруються вимоги щодо шифрування або географічного розміщення даних. Важливо забезпечити не тільки фізичний захист СД, а й логічний — правильну організацію доступу до інформації та зберігання даних. Неналежне зберігання даних включає в себе несанкціонований доступ до даних, неправильне зберігання чутливої інформації без шифрування або її зберігання на вразливих носіях. Також важливим аспектом є відсутність належних політик щодо резервного копіювання і архівації даних. Якщо інформація не зберігається в належному вигляді, це може призвести до серйозних порушень безпеки та фінансових втрат, оскільки компанії можуть втратити доступ до критичної інформації. Прикладом неналежного зберігання є випадок з Amazon, коли у 2019 році неправильно налаштовані сховища даних на платформі Amazon S3 стали причиною великого витоку чутливої інформації з особистими даними користувачів. Відкритий доступ до сховища дозволив будь-якій особі в інтернеті отримати персональні дані користувачів даного хмарного сховища.

Резервне копіювання даних (Backup) – один з основних елементів безпеки СД, оскільки воно дозволяє відновити дані в разі їх втрати або пошкодження. Однак багато бізнес-установ / організацій не проводять регулярне резервне копіювання даних СД або виконують копіювання лише на обмежену кількість даних. Іноді, резервні копії зберігаються на тих же носіях або в тому самому фізичному місці, що й основні дані, що в свою чергу, створює ризик одночасної втрати всіх копій у разі атаки на СД, фізичного пошкодження носія, стихійного лиха або пожежі. Тому необхідно організовувати зберігання резервних копій в декількох фізичних місцях та /або на хмарних сервісах. Також, установи / організації не проводять тестування системи відновлення даних, що, в разі інцидентів, ставить під загрозу процес відновлення інформації з резервної копії. Під час тестування важливо перевірити, чи можна відновити дані вчасно і без втрат, оскільки багато компаній з неналежно налаштованими системами резервного копіювання стикаються з труднощами при відновлені інформації після серйозних інцидентів. Прикладом такого випадку є інцидент з компанією Maersk у 2017 році, коли шкідлива програма NotPetya спричинила параліч корпоративної інфраструктури. Відновлення з резервної копії не спрацювало, оскільки резервні копії знаходились у тих самих центрах обробки даних, що й основне СД, тому даний інцидент призвів до великих фінансових втрат і тривалого відновлення даних сховища.

Ще один приклад: у 2020 році відома фінтех-компанія Robinhood повідомила про серйозну подію, пов’язану із втратою частини даних користувачів, коли через помилку конфігурації бекапів та відсутність географічного розподілу фізичного зберігання резервних копій СД, не вдалося відновити транзакційні записи за певний період. Це спричинило порушення звітності, затримки в обслуговуванні та розслідування з боку регуляторів.

Відсутність належного контролю за дотриманням стандартів зберігання та обробки даних може призвести до юридичних наслідків. Всі установи / організації мають дотримуватися нормативних вимог GDPR, які регулюють обробку, зберігання та передачу персональних даних. Порушення цих вимог може спричинити значні штрафи і репутаційні втрати, а також потенційні позови від постраждалих осіб. Сховища даних, які не відповідають вимогам безпеки або не мають належних політик щодо копіювання та зберігання даних, піддаються підвищеному ризику втрати або витоку даних, компрометації чутливої інформації. Більш того, відсутність систем перевірки даних, несанкціонований доступ до персональної інформації або її неконтрольоване зберігання можуть бути підставою для серйозних штрафів, накладених на таку установу / організацію. Щоб уникнути подібних ризиків, установам /організаціям необхідно впровадити політику регулярного резервного копіювання з географічною диверсифікацією фізичного розташування резервних копій, контрольованого шифрування даних на всіх рівнях, а також автоматизованого тестування

щодо коректного відновлення даних. Важливо впровадити контроль цілісності даних та автоматичне логування усіх дій, пов'язаних з резервним копіюванням та зберіганням. Крім того, варто розглядати використання WORM-сховищ (Write Once, Read Many), які унеможливлюють перезапис або знищення резервної копії навіть за зловмисного втручання.

Організаційні стратегії відповідності GDPR (витоку інформації та загроз)

Організаційні стратегії відповідності GDPR (General Data Protection Regulation) загрозам та витокам інформації є невід'ємною частиною управління ризиками у процесі захисту даних. Стратегічний підхід до впровадження відповідності Загальному регламенту захисту даних (GDPR) дозволяє організаціям не лише мінімізувати ймовірність витоків інформації, але й забезпечити належний рівень захисту персональних даних. GDPR визначає кілька ключових принципів, які стосуються безпеки сховищ даних, а саме: принцип мінімізації даних, конфіденційність та цілісність, обмеження терміну зберігання даних, прозорість даних. Більш детально розглянемо кожний з ключових принципів.

Принцип мінімізації даних

Принцип мінімізації даних має на меті забезпечення лише такої кількості збирання, обробки та зберігання конфіденційних даних, які необхідні для вирішення бізнесових завдань і для досягнення мети, тобто бізнес-установи / організації повинні обмежити кількість даних, звести їх до мінімуму, необхідного для виконання конкретних завдань, уникати надмірного збору інформації. Згідно з принципом мінімізації даних, інформація, яка збирається, повинна бути адекватною, відповідною й обмеженою, тобто такою необхідно для досягнення конкретної мети. Наприклад, якщо організація обробляє особисті дані клієнтів для маркетингових цілей, вона не повинна збирати такі чутливі дані, як фінансова інформація. Наприклад, у 2019 році Європейська Комісія оштрафувала компанію "H&M" на €35 мільйонів через порушення принципу мінімізації даних. Торговельна компанія збирала надмірну кількість персональних даних своїх співробітників, включаючи записи про їх особисті життя, зокрема про здоров'я, яке не мало прямого відношення до їхньої роботи. Це порушення стало результатом внутрішнього моніторингу, що перевищував необхідні межі для забезпечення ефективності роботи.

Також важливим аспектом є те, що мінімізація даних включає не лише обмеження збору інформації, але й обмеження її зберігання. Дані не повинні зберігатися довше, ніж це необхідно для досягнення мети, для якої вони були зібрані. Дана вимога є особливо критичною для установ /організацій, які зберігають великі обсяги даних у своїх сховищах та повинні регулярно оцінювати, які дані можна видалити або архівувати після того, як вони вже не потрібні для виконання конкретних завдань. Вимірювання ефективності мінімізації даних можна за наступною формулою:

$$D_{min} = \frac{D_f}{D_t}, \tag{2}$$

де D_{min} – коефіцієнт мінімізації даних; D_f – кількість даних, які фактично використовуються для досягнення конкретної мети; D_t – загальна кількість зібраних даних.

Чим вищий цей коефіцієнт, тим ефективніше здійснюється мінімізація даних. Під час впровадження принципу мінімізації даних важливо застосовувати чіткі політики збереження та оброблення даних на всіх етапах їхнього життєвого циклу, що веде за собою впровадження технологій автоматичного видалення даних після закінчення їх терміну дії, або їхнього архівування, а також використання інструментів контролю та аудиту, за терміном зберігання даних та недопущенням їхнього довгого зберігання, тобто дані не зберігаються довше, ніж це необхідно для виконання їх функцій. Відповідно до цього принципу, необхідно встановлювати конкретні часові рамки для зберігання даних. Наприклад, персональні дані клієнтів або

співробітників можуть зберігатися лише в межах часу, необхідного для виконання конкретних юридичних чи бізнесових зобов'язань. Після цього дані мають бути видалені або анонімізовані.

Принцип конфіденційності та цілісності

Принцип конфіденційності та цілісності за GDPR вимагає забезпечення захисту персональних даних від несанкціонованого доступу, їх використання, розкриття, змін, пошкодження чи знищення сторонніми особами, які не мають на те повноважень. Тобто установи / організації повинні забезпечити цілісність даних і гарантувати, що вони не будуть змінені або спотворені в процесі їхнього оброблення. Наприклад, дані не повинні бути змінені зловмисниками або через помилки в процесах обробки. Для цього використовуються технології захисту інформації, такі як шифрування, цифрові підписи та інші методи, які забезпечують достовірність і цілісність даних.

Один з аспектів даного принципу – запобігання несанкціонованому доступу до персональних даних, що можна забезпечити через надійні системи аутентифікації та контролю доступу. Мультифакторна автентифікація (MFA) є одним з найефективніших способів забезпечення вказаного принципу, оскільки вона вимагає декілька рівнів перевірки та підтвердження справжності користувача і значно знижує ймовірність несанкціонованого доступу. Наприклад, для доступу до сховищ даних співробітники установи / організації повинні проходити, окрім автентифікації та авторизації, через додаткові етапи перевірки, такі як одноразовий код або біометричні дані. Принцип конфіденційності також включає в себе заходи захисту даних для безпечної передачі даних по мережі, з використання протоколів безпеки, таких як TLS (Transport Layer Security). Протокол TLS гарантує, що всі дані, передані між сервером і клієнтом, будуть зашифровані і не можуть бути перехоплені або змінені під час передачі. Враховуючи зростання з кожним роком обсягів переданих даних і складність сучасних мереж, важливо дотримуватись високого рівня шифрування даних для забезпечення конфіденційності інформації. Наприклад, у 2020 році компанія Facebook була оштрафована на $5 мільярдів за порушення принципів конфіденційності та цілісності згідно вимогам GDPR. Проблеми виникли через недостатню реалізацію політик захисту персональних даних і неправомірне використання персональних даних користувачів для рекламних цілей без їхньої згоди. Це призвело до витоку даних, що порушило принципи несанкціонованого доступу та маніпуляції з інформацією.

Принцип конфіденційності та цілісності можна описати через фактори захисту даних і ймовірність витоку, якщо P_c – ймовірність того, що дані будуть збережені конфіденційними, а P_i – ймовірність, що дані будуть точними, то загальний коефіцієнт захисту P_t можна визначити як:

$$P_t = P_c \times P_i, \quad (3)$$

Вираз 3 показує, що чим вище обидва показники, тим надійніше будуть захищені персональні дані.

Обмеження терміну зберігання даних

Принцип обмеження строку зберігання даних визначає, що персональні дані повинні зберігатися лише протягом часу, необхідного для досягнення цілей, для яких вони були зібрані. Після того, як дані більше не є необхідними для зазначених цілей, вони повинні бути видалені або анонімізовані, щоб захистити права та свободи суб'єктів даних. Даний принцип є важливим для забезпечення мінімізації обробки даних та захисту особистої інформації від надмірного використання або зловживання. Для забезпечення вказаного принципу необхідно впроваджувати політики і процедури, які чітко визначають строки зберігання різних типів даних. Вони також повинні враховувати нормативно-правові вимоги, які можуть накладати додаткові вимоги щодо зберігання даних, таких як фінансова або податкова звітність. Важливою частиною принципу обмеження строку зберігання даних є регулярне оцінювання необхідності зберігання даних, яке

дозволяє здійснювати ефективне управління даними. Використання автоматизованих інструментів для моніторингу строків зберігання даних є ефективним способом забезпечення дотримання даного принципу. Наприклад, системи автоматичного видалення або архівації даних, на основі визначених політик, дозволяють знизити ризики, пов'язані з перевищенням термінів їхнього зберігання, можуть бути налаштовані для виконання регулярних перевірок, з подальшим автоматичим видаленням або анонімізацією даних, у разі закінчення терміну зберігання.

З точки зору математичного моделювання, можна розглядати обмеження строку зберігання даних як задачу оптимізації з метою мінімізації ризику зберігання даних, після закінчення терміну зберігання. Якщо позначити T_{max} – як максимальний дозволений строк зберігання для даних, а T_a – як фактичний строк зберігання, тоді показник ефективності R_s буде обчислюватись за формулою:

$$R_s = \frac{|T_a - T_{max}|}{T_{max}}. \quad (4)$$

Легко помітити, що чим менший показник ефективності R_s, тим ефективніше реалізовано обмеження терміну зберігання, що дозволяє знизити ризики, пов'язані з надмірним зберіганням даних.

За умовами GDPR, кожен суб'єкт має право вимагати видалення своїх даних за певних обставин (наприклад, якщо дані більше не потрібні для цілей, для яких вони були зібрані), що створює додаткову відповідальність для установ /організацій, які повинні мати чітко визначену процедуру оброблення такого типу запитів. Прикладом даного принципу стало те, що у 2021 році британська компанія British Airways була оштрафована на 20 млн фунтів стерлінгів за порушення принципу обмеження строку зберігання даних після того, як з'ясувалося, що вона зберігала персональні дані клієнтів набагато довше, ніж було необхідно для виконання власних послуг. В результаті інциденту стало відомо, що багато даних зберігалися більше 2 років після завершення бізнесових взаємодій з клієнтами і порушувало вимоги GDPR. Така ситуація підкреслює важливість реалізації політик обмеження строку зберігання і регулярного перегляду даних щодо відповідності вимогам GDPR та мінімізації можливих юридичних і фінансових наслідків такого порушення.

Прозорість даних

Принцип прозорості, згідно положень GDPR, передбачає, що обробка персональних даних має бути зрозумілою, чесною та доступною для суб'єктів даних. Тобто даний принцип означає, що установа / організація зобов'язана інформувати користувачів, зрозумілою для них мовою, про те, які дані збираються, як, ким і з якою метою вони обробляються, хто має до них доступ, а також які права мають користувачі щодо власних даних. Прозорість є ключовим фактором у встановленні довіри між установою / організацією та її клієнтами. Для реалізації даного принципу установа / організація повинна мати чітко сформульовану, легко доступну та написану простою мовою політику конфіденційності. Наприклад, користувач має отримати повідомлення про обробку його даних ще до того, як вони будуть оброблятись, включаючи інформацію про строки зберігання інформації, підстави обробки та контактні дані відповідальної особи із захисту даних (DPO). З технічного боку реалізація принципу прозорості може бути підтримана за допомогою інтерфейсів керування згодами (Consent Management Platforms – CMP), які дозволяють користувачам управляти своїми налаштуваннями щодо обробки даних.

Математично ступінь прозорості можна оцінити через коефіцієнт доступності інформації T_i, де A – кількість пунктів, які користувач може самостійно переглянути/змінити, а T – загальна кількість пунктів, що стосуються обробки:

$$T_i = \frac{A}{T}. \quad (5)$$

Чим ближче T_i до 1, тим вищий рівень прозорості, а отже краще відповідає принципам GDPR. Яскравим прикладом порушення принципу прозорості стала справа проти компанії Google, за результатами якої французький наглядовий орган CNIL у 2020 році наклав штраф у розмірі 50 мільйонів євро. Регулятор встановив, що інформація про обробку персональних даних у Google була розпорошена по кількох сторінках, подана складною мовою та не дозволяла користувачам чітко зрозуміти, яким чином і з якою метою відбувається обробка їхніх даних для персоналізованої реклами. Крім того, механізм надання згоди на обробку не забезпечував належного рівня поінформованості до моменту її отримання. Даний випадок став важливим прецедентом у практиці застосування GDPR і змусив чимало цифрових компаній переосмислити свої підходи щодо інформування користувачів. Таким чином, принцип прозорості є не лише юридичною вимогою, але й практичним інструментом побудови довіри, забезпечення контролю над персональними даними та уникнення серйозних штрафів наглядовими органами, а його ігнорування призводить не лише до фінансових, але й до репутаційних втрат.

Розглянемо, які існують методи безпеки сховищ даних від витоків та загроз відповідно до GDPR.

Шифрування даних

Першим методом безпеки сховищ даних від витоків та загроз відповідно до GDPR розглянемо метод шифрування даних. Згідно 32 статті GDPR одним з найефективнішим спосіб захисту персональної інформації в межах сховищ даних є шифрування даних. Зашифровані дані для зловмисників представляють собою набір нечитабельних символів, які можна дешифрувати у читабельний формат лише відповідним ключем. Загальний регламент захисту даних визначає, що шифрування даних виступає не тільки як засіб забезпечення конфіденційності, але і як спосіб мінімізації збитків. Якщо відбувся виток зашифрованих даних, то вони не будуть вважатись загрозливим і цей факт зменшує юридичні та фінансові наслідки такого витоку для установи / організації. На практиці для шифрування даних у сховищах використовують алгоритми симетричного шифрування, зокрема AES-256 (Advanced Encryption Standard з довжиною ключа 128, 192 або 256 біт) та криптографічний протокол, який забезпечує захищену передачу даних між клієнтом і сервером, а також між вузлами в інфраструктурі TLS (Transport Layer Security), який гарантує захист від атак типу «man-in-the-middle» та перехоплення трафіку.

Алгоритм AES являє собою симетричний метод шифрування даних у сховищах та використовують для забезпечення конфіденційності даних. Алгоритм працює шляхом перетворення відкритого тексту (plain text) у зашифрований (ciphertext) за допомогою ключа фіксованої довжини (128, 192 або 256 біт). Процес шифрування можна представити за допомогою математичного виразу:

$$C = AESk(P) \quad (6)$$

де P – відкритий текст (дані користувача), k – симетричний ключ, C – зашифровані дані, які зберігається у сховищі.

У GDPR підкреслено, що установи / організації повинні використовувати технічні заходи, такі як шифрування, для захисту персональних даних, а отже застосування алгоритму симетричного шифрування AES дозволяє забезпечити відповідність даному положенню, особливо для даних, що зберігаються в дата-центрах, хмарних сховищах або резервних копіях.

З технічного погляду, шифрування слід поєднувати з управлінням ключами KMS (Key Management Systems), які дозволяють створювати, зберігати та видаляти ключі. GDPR також наголошує, що управління ключами має бути відокремлене і зберігатися окремо від основної

інфраструктури установи / організації і при зламі сховища зловмисник не зможе легко отримати ключ шифрування. Такий вид захисту ключів знижує ризик як зовнішніх, так і внутрішніх атак.

Приклад витоку зашифрованої інформацій стався у 2021 році в Ірландії. у компанії Workhuman (провайдер HR-платформи) відбувся витік чутливої і конфіденційної інформації, який зачепив дані тисячі клієнтів. Однак завдяки тому, що всі дані були зашифровані відповідно до вимог GDPR (використовувався AES-256 з апаратною підтримкою шифрування на рівні серверів), регулятор не визнав інцидент критичним і компанія уникла штрафу. Даний випадок підтверджує важливість впровадження шифрування як частини стратегії мінімізації наслідків.

Криптографічний протокол TLS потрібен для захисту даних при передачі між користувачем і сховищем або між серверами (наприклад, між сховищем даних і вебзастосунком). GDPR вимагає захисту даних не лише в сховищі, а й при їх передачі (наприклад, у розподілених обчисленнях або API-запитах). TLS працює на основі асиметричного шифрування, при якому дані шифруються відкритим ключем, а розшифровуються приватним. Ключовий етап – це обмін сесійним ключем:

$$k = C^d \ mod \ n \tag{7}$$

де C – відкритий ключ, d, n – приватний ключ приймача.

Після узгодження ключа k, дані шифруються вже симетрично згідно формули 6, а це, в свою чергу, дає змогу зашифрувати дані «на льоту» без суттєвих втрат продуктивності.

Інцидент, який довів важливість TLS стався у 2020 році: популярна платформа для відеоконференцій Zoom зазнала критики через використання нестандартного шифрування. Після тиску регуляторів ЄС (включно з вказівками щодо відповідності статтям GDPR) компанія повністю перейшла на TLS 1.3 і AES-GCM 256-біт, щоб уникнути штрафів і втрати довіри з боку користувачів платформи.

Загалом, впровадження шифрування має бути обов'язковим елементом безпеки сховищ даних, особливо якщо йдеться про обробку великих обсягів чутливої інформації. Його ефективність зростає при інтеграції з іншими методами – такими як контроль доступу, логування та псевдонімізація. Без належного шифрування навіть найкращі організаційні заходи можуть виявитися недостатніми, якщо чутливі та конфіденційні дані виявляться в руках зловмисників.

Другим методом безпеки сховищ даних від витоків та загроз відповідно до GDPR є метод псевдонімізації та анонімізації даних.

Псевдонімізація даних – це процес обробки персональних даних, у результаті якого вони втрачають можливість безпосередньої ідентифікації особи без додаткової інформації. Згідно зі статтею 4 GDPR, дана додаткова інформація має зберігатися окремо та та бути захищеною за допомогою технічних і організаційних заходів, щоб запобігти несанкціонованому доступу або повторній ідентифікації особи. Метод псевдонімізації активно застосовується у сховищах даних, де необхідно обробляти великі масиви даних для аналітики або тестування, не розкриваючи ідентичності користувача. Процес псевдонімізації можна подати у вигляді математичної функції:

$$f : D \rightarrow P(D) \tag{8}$$

де D – множина персональних даних, P(D) – множина псевдонімізованих даних, f – псевдонімізуюча функція, яка є оборотною, але вимагає ключа або таблиці відповідності k, яка окремо зберігається. Наприклад: $UserID_{реальний} = 123456 \rightarrow UserID_{псевдонім} =$"a8s73d2". Без доступу до відповідної таблиці або ключа ідентифікація $UserID_{реальний}$ неможлива, а це, у свою чергу, мінімізує ризики у витоку даної конфіденційної інформації з сховища даних.

Приклад, у 2022 році медична компанія Roche запровадила псевдонімізацію пацієнтських даних у клінічних дослідженнях, що проводилися в ЄС. У відповідності статті 4 GDPR, персональні ідентифікатори пацієнтів були замінені унікальними хешами, що дозволило використовувати дані пацієнтів для аналітики без розкриття особистості. Даний приклад підтверджує важливість впровадження псевдонімізацію даних з метою забезпечило захисту персональних даних у разі їхнього витоку з аналітичного сховища.

Анонімізація даних – процес обробки персональних даних таким чином, щоб ідентифікація фізичної особи стала неможливою за жодних умов, навіть за допомогою додаткової інформації. Важливо: після анонімізації набір даних більше не підпадає під дію GDPR, оскільки він перестає бути персональним. Статті 4 GDPR розглядає дані як анонімізовані, лише якщо вони не відносяться до ідентифікованої, або такої, що може бути ідентифікованою, особи, прямо чи опосередковано. У сховищах анонімізація використовується для формування відкритих масивів або статистичних звітів, де немає потреби зберігати ідентифікаційні атрибути.

Нехай наявна база даних $D = \{r_1, r_2, ..., r_n\}$, де кожен запис r_i — це вектор атрибутів:

$$r_i = \left(a_1^i, a_2^i, \dots, a_m^i\right). \tag{9}$$

Атрибути поділяються на:

Ідентифікатори (ID): ПІБ, номер телефону – завжди видаляються.

Квазі-ідентифікатори (QI), тобто атрибути, які в комбінації можуть ідентифікувати особу (вік, місто, стать тощо).

Чутливі атрибути (SA): діагнози, зарплата тощо – залишаються для аналізу.

Позначимо $QI = \{q_1, q_2, ..., q_k\} \subseteq A$ – підмножину атрибутів, які використовуються для забезпечення анонімності. Набір даних D вважається k-анонімним, якщо для кожного можливого поєднання квазі-ідентифікаторів існує принаймні k записів з однаковими значеннями цих атрибутів:

$$\forall \vartheta \in \pi_{QI}(D),\ \left|\sigma_{QI=\vartheta}(D)\right| \geq k, \tag{10}$$

де $\pi_{QI}(D)$ – проєкція таблиці D на атрибути QI, $\sigma_{QI=\vartheta}(D)$ – вибірка записів із значеннями QI рівними ϑ, $|\cdot|$ – кількість записів.

Це означає, що жодну людину не можна відрізнити з точністю більше, ніж 1 до k на основі квазі-ідентифікаторів. Формально анонімізацію можна подати як перетворення функцією:

$$f: D \rightarrow A(D) \tag{11}$$

де D – початкові персональні дані, A(D) – анонімізовані дані, в яких неможливо відновити DDD, f – функція, яка задовольняє умову:

$$\forall \vartheta \in \pi_{QI}(D),\ \left|\sigma_{QI=\vartheta}(D)\right| \geq k,\ r_i \in D' \rightarrow ID(r_i) = \emptyset. \tag{12}$$

А це гарантує, що: жоден запис не містить ідентифікаторів, і усі значення QI не унікальні менш ніж для k записів.

Приклад застосування анонімізації даних: у 2020–2021 рр. Євростат публікував статистику по кількості захворювань на COVID, зібрану з багатьох лікарень ЄС. Дані пройшли багатоступеневу анонімізацію, тобто імена, адреси, точні дати народження замінювались категоріями або повністю видалялися. Таким чином, навіть при повному доступі до масиву інформації щодо захворювань на COVID, неможливо було відновити персональні дані пацієнтів лікарень.

Висновки і перспективи подальших досліджень

У результаті дослідження було підтверджено, що сховища даних є об'єктом підвищеного ризику в контексті витоків і загроз. Стаття визначає п'ять критичних типів загроз для сховищ та пропонує інструменти їхнього подолання на основі вимог GDPR, зокрема, шифрування, псевдонімізація, контроль доступу та аудит розглядаються як необхідні (а не опціональні) механізми, які повинні бути інтегровані на всіх рівнях сховищ даних. Недотримання навіть одного з принципів GDPR (конфіденційність, мінімізація, обмеження строку зберігання, прозорість) може призвести до серйозних фінансових, юридичних та репутаційних наслідків для установ / організацій. Практичні приклади порушень підтверджують, що навіть технічно розвинуті компанії зазнають хакерські атаки при відсутності належного внутрішнього контролю або культури інформаційної безпеки.

Подальші дослідження доцільно зосередити на: розробці автоматизованих засобів контролю, у відповідності GDPR , із застосуванням методів машинного навчанням, з метою виявлення аномалій у поведінці користувачів; оцінці економічної ефективності впровадження політик безпеки GDPR для малих та середніх підприємств, зокрема у хмарних середовищах.

Список використаної літератури

1. Стандарт General Data Protection Regulation (GDPR) – Режим доступу: https://gdpr-info.eu.

2. Стандарт ISO/IEC 27001:2022 – Режим доступу: https://www.iso.org/standard/27001.

3. Стандарт Payment Card Industry Data Security Standard (PCI DSS) – Режим доступу: https://www.pcisecuritystandards.org.

4. Žaklīna Ieviņa (2022). Erasure and Anonymisation of Personal Data in Context of General Data Protection Regulation. Socrates. Rīga Stradiņš University Faculty of Law Electronic Scientific Journal of Law.Volume 2022 (2022): Issue 1-22 (June 2022). 114–126. https://doi.org/10.25143/socr.22.2022.1.114-126.

5. Andriy Pavliv (2024). Anonymization of Data Using Blockchain Technology: a Model for Data Lifecycle Management to Ensure Transparency and Compliance With GDPR. CSN. Volume 6, Number 2 : pp. 179-188. https://doi.org/10.23939/csn2024.02.179.

6. Samson Yoseph Esayas (2022). The GDPR and unstructured data: is anonymization possible? International Data Privacy Law, Volume 12, Issue 3, August 2022, Pages 184–206, https://doi.org/10.1093/idpl/ipac008.

7. Starchoň, P., & Pikulík, J. (2019). GDPR principles in Data protection encourage pseudonymization through most popular and full-personalized devices - mobile phones. Procedia Computer Science. Volume 151, 2019, Pages 303-312. https://doi.org/10.1016/j.procs.2019.04.043.

8. Ophelia Prillard, Costas Boletsis, Shukun Tokas (2024). Ethical Design for Data Privacy and User Privacy Awareness in the Metaverse. ICISSP 2024 - 10th International Conference on Information Systems Security and Privacy. 333-341. DOI:10.5220/0012296500003648.

9. Василенко В.Ю., Буряк А.М. (2024). Безпека даних в епоху цифрової трансформації: проблеми та виклики. Інформація та соціум, 2024, с. 103-106.

10. Фасій Б. (2024). Національна безпека та захист персональних даних в епоху цифрових технологій. Society and Security, 2024, с. 76-82. DOI: https://doi.org/10.26642/sas-2024-6(6)-76-82.

11. Борисенко О.С., Тимошенко А.Г. (2024). Огляд методів захисту персональних даних у хмарному середовищі. Інфокомунікаційні та комп'ютерні технології, 1(07), 31-34. DOI: https://doi.org/10.36994/2788-5518-2024-01-07-04.