УДК 004.7

Багатовимірна матриця класифікації інформації для оцінки ризиків інформаційної безпеки

Тетяна Іванівна Коробейнікова
к.т.н., доцент кафедри безпеки інформаційних технологій
Національний університет «Львівська Політехніка»
79013, м. Львів. вул. Степана Бандери, 12
ORCID 0000-0003-2487-8742

Андрій Богданович Ямнич
аспірант кафедри безпеки інформаційних технологій
Національний університет «Львівська Політехніка»
79013, м. Львів. вул. Степана Бандери, 12
ORCID 0009-0005-7226-1896

Анотація. У даній роботі розглядається одна з ключових задач щодо комплексної системи оцінки ризиків інформаційної безпеки для персоналу під час розмежування доступу до інформаційних ресурсів компанії. Актуальність дослідження підтверджується численними випадками витоків інформації, які демонструють недостатню ефективність традиційних методів класифікації та контролю доступу. Дослідження передбачає аналіз існуючих стратегій класифікації інформаційних ресурсів компанії та розробка додаткового методу, що базується на постійному аналізі доступу і динамічному коригуванні класифікації таких ресурсів. Для досягнення цієї мети було використано методи аналізу поточних стратегій класифікації інформації, комбінування різних методів класифікації та впровадження графічного методу поєднання традиційної класифікації ресурсів із динамічною складовою за допомогою багатовимірної матриці. Основні результати дослідження передбачають розробку вдосконаленого методу, що дозволяє постійно аналізувати доступ персоналу до інформаційних ресурсів компанії і в динамічному режимі коригувати класифікацію інформаційних ресурсів залежно від правил розмежування цього доступу. Запропонований підхід дозволяє включати довільну кількість показників на графіку у вигляді набору векторів та в подальшому розраховувати загальні оцінки ризиків на основі суми або різниці цих векторів. Практична цінність роботи полягає в тому, що запропонований підхід дозволяє в повній мірі використовувати сучасні технології контролю доступу і може бути основою для подальших досліджень, наприклад, таких, які підтримують автоматизовану класифікацію інформації за допомогою тренування нейронних мереж. Крім того, у рамках даного дослідження було проведено детальний огляд існуючих методів оцінки ризиків інформаційних ресурсів компанії та виявлено ключові недоліки, що притаманні традиційним підходам. Зокрема, були проаналізовані методи, що базуються на фіксованих рівнях доступу та використанні статичних правил для контролю доступу. Виявилося, що такі методи не здатні адекватно реагувати на динамічні зміни в поведінці користувачів та зміни у важливості інформаційних ресурсів. Таким чином, запропонований підхід дозволяє забезпечити більш гнучкий та адаптивний контроль доступу до інформаційних ресурсів. Це досягається шляхом постійного моніторингу доступу та автоматичного коригування рівнів доступу на основі аналізу поведінкових даних користувачів та зміни контексту використання інформаційних ресурсів.

Ключові слова: динамічна класифікація інформації, візуалізація стану критичних ресурсів, багатовимірна матриця класифікації, класифікаційний стек, інтегральні оцінки ризиків.

Multidimensional classification matrix for information security risk assessment

Tetiana I. Korobeinikova
PhD, Associate Professor of Information Technology Security Department
Lviv Polytechnic National University
Stepana Bandery St, 12, Lviv, Ukraine, 79000
ORCID 0000-0003-2487-8742

Andrii B. Yamnych
Postgraduate Student of Information Technology Security Department
Lviv Polytechnic National University
Stepana Bandery St, 12, Lviv, Ukraine, 79000
ORCID 0009-0005-7226-1896

Abstract. In this study, we address one of the key challenges related to a comprehensive risk assessment system for information security concerning personnel during access delineation to company information resources. The relevance of this research is confirmed by numerous instances of information leaks, which highlight the insufficient effectiveness of traditional classification and access control methods. The research aims to analyze existing classification strategies for company information resources and develop an additional method based on continuous access analysis and dynamic adjustment of resource classification. To achieve this goal, we employed methods such as analyzing current information classification strategies, combining various classification techniques, and implementing a graphical method that combines traditional resource classification with a dynamic component using a multidimensional matrix. The main results of the study involve the development of an enhanced method that allows continuous analysis of personnel access to company information resources and dynamic adjustments to resource classification based on access delineation rules. The proposed approach allows for the inclusion of any number of indicators in a graph as a set of vectors, subsequently calculating overall risk assessments based on the sum or difference of these vectors. The practical value of this work lies in its ability to fully utilize modern access control technologies and serve as a foundation for further research, such as automated information classification using neural network training. Additionally, within this study, we conducted a detailed review of existing risk assessment methods for company information resources, identifying key limitations inherent in traditional approaches. Specifically, we analyzed methods based on fixed access levels and the use of static rules for access control. It became evident that such methods are inadequate in responding to dynamic changes in user behavior and the evolving importance of information resources. Thus, the proposed approach allows for more flexible and adaptive access control to information resources, achieved through continuous access monitoring and automatic adjustments based on behavioral user data and contextual changes in resource utilization.

Keywords: dynamic information classification, visualization of critical resources, multidimensional classification matrix, classification stack, integral risk assessments.

DOI: https://doi.org/10.31649/1999-9941-2024-60-2-91-106

Вступ. Центральним ризиком інформаційної безпеки завжди є несанкціонований доступ до критичної інформації та/або її витік за межі організації (Velmurugan et al., 2024; Pitafi et al., 2023; Al Qahtani et al., 2024). Залежно від того, наскільки критичною була інформація (Mikuletič et al., 2024), збитки можуть варіюватися від звичайних до ланцюжка наслідків, які потягнуть шкоду не лише для конкретної організації, а й для її клієнтів, контрагентів, а іноді навіть для громадян, які виявилися супутніми жертвами витоку

(Wiedemann et al., 2024; Alotibi, 2024) (особливо якщо йдеться про персональні дані). Традиційним способом запобігання або мінімізації таких випадків є класифікація інформації та обмеження доступу до неї лише певним колом довірених осіб (Ramamurthy et al., 2022; Wang & Gu, 2024).

Процес класифікації може бути ускладнений додатковими факторами, як-от рівнем доступу персоналу до інформаційних ресурсів підприємства. Приватні організації часто економлять на персоналі, і кожен працівник в них може відповідати одночасно за кілька напрямків. Це створює значні ризики для інформаційної безпеки (Shmatko et al., 2020).

Таким чином, існує потреба додати у комплексну систему оцінки ризиків інформаційної безпеки (для персоналу під час розмежування доступу до інформаційних ресурсів компанії) динамічну складову класифікації інформації. Це дозволить визначити, які саме ресурси є найбільш затребуваними, які саме ресурси найчастіше делегуються тимчасовим працівникам і, нарешті, які з них необхідно цифровізувати в першу чергу.

Метою роботи є аналіз та вдосконалення способів візуалізації стану критичних інформаційних ресурсів на підприємстві за допомогою динамічного підходу із застосуванням багатовимірної матриці класифікування інформації для оцінки ризиків інформаційної безпеки.

Для досягнення поставленої у даній роботі мети необхідно: 1. виконати аналіз засад класифікації інформації та обґрунтувати доцільність динамічного підходу; 2. запропонувати динамічну класифікацію на базі аналізу доступу до ресурсів; 3. запропонувати багатовимірну матрицю класифікації ресурсів.

Наукова новизна роботи: пропонується багатовимірна матриця класифікування інформації для оцінки ризиків інформаційної безпеки, яка, на відміну від аналогів, дозволяє візуалізувати кожен ресурс у формі вектора, пропорційному максимальній зміні динамічного показника, спрямованого в бік цієї зміни. Це дозволить візуалізувати миттєві піки для малозатребуваних ресурсів.

Огляд літератури. В роботі (Velmurugan et al., 2024) зазначається, що класифікація інформації є важливим етапом для запобігання витоків даних. У праці (Pitafi et al., 2023) подано дослідження про технічне убезпечення витоків даних, а у роботі (Al Qahtani et al., 2024) основна увага приділена організаційним та соціально-інженерним причинам витоків даних. Також, під час роботи із персональними даним часто особисті переконання та етика стають чинниками безпеки, про що йдеться у статті (Mikuletič et al., 2024), яка досліджує методи захисту конфіденційних медичних даних. Часто в результати втрати чутливих даних жертвами є треті особи, що досліджують (Wiedemann et al., 2024) та (Alotibi, 2024). Таким чином, провідні учені (Ramamurthy A. et al., 2022), (Wang & Gu, 2024) пропонують обмежити доступ до чутливої конфіденційної інформації, водночас поділивши її на певні групи, тобто, пропонують класифікувати інформацію та обмежити до неї доступ.

Однак у державних підприємствах та військових установах такий підхід підкріплено додатковими заходами безпеки та загрозою кримінальної відповідальності, а на рівні приватних підприємств він базується лише на персональній відповідальності працівника. Як правило, підписується лише спеціальне доповнення до трудового договору, яке обумовлює відповідальність за порушення режиму доступу до даних або розголошення комерційної таємниці. Якщо підприємство має доступ до державних даних у рамках державно-приватного партнерства (зазвичай, це стосується персональних даних), то різниця між рівнями забезпечення режиму секретності на державному та організаційному рівнях може створити значну дірку в системі безпеки даних (Arslan et al., 2024; Emmanuel et al., 2024; Venn et al., 2024; Song et al., 2024).

Тому приватні підприємства значно більше залежать від точності класифікації інформації, а також більш щільного контролю за доступом до конкретних ресурсів. Чим більшою і різноманітнішою інформаційною базою володіє організація, тим складнішою буде

задача класифікації ресурсів, що до неї входять (Barnawi et al., 2024, Gambarelli et al., 2021; Mazzola et al., 2021).

Загалом, існує тенденція до розвитку досліджень, зосереджених на підвищенні точності класифікації інформаційних ресурсів та адаптації підходів до сучасних викликів інформаційної безпеки.

Матеріали та методи. Основні етапи роботи включали аналіз сучасних методів класифікації інформаційних ресурсів, обґрунтування доцільності використання динамічної класифікації, розробку багатовимірної матриці класифікації ресурсів. Методологія включала використання статистичних методів для аналізу даних про доступ до ресурсів, а також розробку графічних інструментів для візуалізації динаміки доступу. Аналіз сучасних методів класифікації інформаційних ресурсів починався з вивчення існуючих підходів до класифікації інформації в різних організаціях, зокрема державних підприємствах та приватних компаніях. Було досліджено стандарти ISO 27001 та QGISCF, що застосовуються для класифікації інформаційних ресурсів (QGISCF, 2023; ISO 27001, 2022). Вивчалися методи класифікації за критеріями чутливості, цілісності та доступності інформації.

Обґрунтування доцільності використання динамічної класифікації базувалося на вивченні динаміки доступу до інформаційних ресурсів на основі даних журналювання доступу в сучасних файлових системах. Було проведено порівняння статичних методів класифікації, які використовують фіксовані точки перегляду, з динамічними методами. Виявлено ключові показники, які впливають на критичність інформаційних ресурсів, такі як частота доступу, протокол доступу та тип користувачів.

Розробка багатовимірної матриці класифікації ресурсів включала створення багатовимірної матриці, що поєднує різні показники критичності ресурсів, включаючи частоту доступу, рівень конфіденційності та ступінь цифровізації. Використовувалися методи візуалізації для відображення змін частоти доступу та інших динамічних показників у вигляді графіків і діаграм. Було розроблено методологію для адаптації класифікації в режимі реального часу на основі даних про доступ до ресурсів.

Опис стратегії та критеріїв формування вибірки включав вибірку інформаційних ресурсів на основі їх значущості для діяльності організації. Визначалися ключові показники для аналізу, такі як кількість користувачів, типи доступу та частота запитів до ресурсів. Використовувалися статистичні методи для аналізу даних вибірки та оцінки ризиків інформаційної безпеки.

Експериментальна база дослідження включала проведення експериментів з використанням журналювання доступу до інформаційних ресурсів у реальних умовах організації. Аналізувалися отримані дані для виявлення патернів доступу та аномалій, що можуть свідчити про потенційні загрози інформаційній безпеці. Отримані результати використовувалися для коригування методів класифікації та підвищення ефективності заходів безпеки.

Для забезпечення повної картини ходу дослідження використовувалися додаткові методи. Метод статистичного аналізу застосовувався для обробки великих обсягів даних про доступ до інформаційних ресурсів, а також для виявлення статистично значущих відхилень у патернах доступу. Графічні методи візуалізації використовувалися для представлення результатів дослідження у вигляді діаграм та графіків, що дозволяло наочно відобразити динаміку доступу до ресурсів та виявлені аномалії.

Результати

Аналіз засад класифікації інформації та обґрунтування динамічного підходу. Перш ніж розглядати класифікацію інформації з безпекового ракурсу, необхідно зрозуміти, як виглядає структура даних в межах організації. Простіше за все розглядати їх з точки зору щоденного робочого процесу. Тоді всі дані підприємства можуть бути поділені на три групи (Malchiodi et al., 2024; Zhang et al., 2018; Robinson, 2024) (рис. 1).

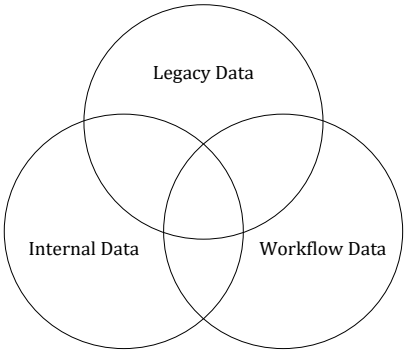

Рисунок 1. Класифікація даних за робочим процесом
Джерело: розроблено авторами на основі досліджень (Malchiodi et al., 2024; Zhang et al., 2018; Robinson, 2024)

Legacy Data (архівні дані). Архівні дані та документи минулих періодів, які можуть зберігатись в повністю нецифровій формі або в форматах, які втратили актуальність і не є сумісними із сучасними інформаційними системами. Залежно від значущості цих даних для поточної роботи, ці дані можуть бути як повноцінними, структурованими документами (що дозволить зберегти історичність даних, відстежувати тенденції в роботі установи чи комерційної фірми, застосовувати Big Data для аналізу тощо), або вноситись в системи обліку як певний підсумковий залишок. Як правило, це найбільш архаїчна частина інформаційної бібліотеки, із найменшим відсотком цифровізації;

Internal Data (внутрішні дані). Внутрішня документація, яка використовується у повсякденній роботі, але не виходить за межі підприємства. Зазвичай, це накази, положення, циркуляри та інша нормативно-адміністративна документація. Тут найчастіше зустрічається цифровізація, оскільки вони можуть існувати в формі обміну повідомленнями через електронну пошту чи корпоративні системи спільної роботи;

Workflow Data (потокові дані). Робоча документація, яка використовується у виробничій діяльності і є офіційною документацією, яка може бути використана контролюючими, судовими та іншими установами як правочинна чи підставова документація. Ці дані діляться на первинні (створюються або вносяться працівниками як результат їх професійної діяльності) та вторинні (формуються з даних, взятих з первинних документів). Вони можуть бути цифровізовані як медіа-документи (зазвичай, сканлейти або фотокопії) та/або замінені стандартними електронними документами в форматах текстових процесорів (таких як Microsoft Office, OpenOffice та ін.). Але вони також можуть бути на папері, у формі записок, ордерів та інших щоденних розхідних документів.

Набір ресурсів, які знаходяться на перетині всіх трьох груп, як правило, є найбільш критичними для функціонування підприємства (Robinson, 2024). Сюди може входити як аудиторська інформація (в ній "сходяться" всі найбільш важливі показники діяльності), так і інформація, необхідна для забезпечення виробництва або просто щоденної роботи інформації (наприклад, персональні дані клієнтів, замовлення тощо).

В силу різної специфіки роботи різних установ межі класифікації документів на вищезгадані групи не є чіткими, тому обсяги перетинів можуть бути різними, а це впливає як на критичність інформації, так і на безпекову стратегію та стратегію цифровізації. Тому як правило, її доповнюють додатковими класифікаторами, які дозволяють більш чітко окреслити межі і більш чітко оцінити критичність.

Підходів до створення конкретизуючих класифікаторів три:

1. на основі змісту інформації. Відповідно, така класифікація має передбачати методику оцінки чутливості чи критичності її для функціонування організації. Наприклад, в стандарті Queensland Government Information Security Classification Framework (QGISCF) (Queensland Government, 2024), який застосовується в Австралії, для подібних класифікаторів рекомендується інтегрований параметр Business Impact Level, який

складається із трьох параметрів – потенційна шкода, цілісність/частковість та доступність інформації;

2. на основі застосування інформації. Відповідно, так класифікація повинна відповідати технологічним, організаційним, або навіть кадровим ланцюжкам всередині підприємства. Як правило, йдеться про класифікацію внутрішньої інформації за організаційними відділами, а робочої документації за напрямками діяльності;
3. на основі користувача інформації. Така класифікація має базуватись на чіткому визначенні, яка саме інформація необхідна для тієї чи іншої роботи і кому саме. Знову ж таки, в рамках QGISCF виокремлюються власники (створювачі, розпорядники) інформації, користувачі, менеджмент, перевіряючі, офіцери безпеки даних (у випадку державно-приватного партнерства) та ін.

Можна згадати ще про підхід, який уже частково згадувався: на основі форми представлення інформації. Стандарт ISO 27001 (Irwin, 2022) містить спрощений варіант такої класифікації із чотирьох класів: паперові документи, цифрові документи на локальних/знімних носіях даних, цифрові документи у віддалених сховищах та електронні повідомлення в системах миттєвого обміну інформацією.

Всі ці класи мають нечіткі межі, і так само, як і класифікація ресурсів за робочим процесом, описується круговими діаграмами із перетинами. Таким чином, незалежно від типу і набору класифікаторів ресурсів, які використовуються в організації, вони всі мають нечіткі межі і перекриваються між собою. Інтегральна оцінка критичності ресурсу з точки зору інформаційної безпеки, відповідно, буде набором "перетинів перетинів", і критична інформація знаходитиметься в одному із них. Для інтегральної оцінки, як правило, використовують табличний метод. Рядки в ньому відповідають різним класифікаторам інформації, а стовпчики – одному із рівнів доступу. Параметри оцінюються за певною чисельною шкалою.

Можна скласти таблицю на основі вищенаведених типів класифікації, взявши систему рівнів доступу із стандарта ISO 27001, однак заповнення таблиці відповідати емпіричним показникам – за шкалою "низький-середній-високий-найвищий" (New Zealand Protective Security Requirements, 2024).

Найкраще табличний метод поєднує систему рівнів доступу із BIL (Business Impact Levels) – там лінійна залежність, що дозволяє застосувати систему балів – наприклад таку, яку використовують в Новій Зеландії (Arora et al, 2023). Однак адекватно, на нашу думку, такий підхід дозволяє оцінити лише такі показники як шкода (чутливість інформації) та безпосередньо доступність.

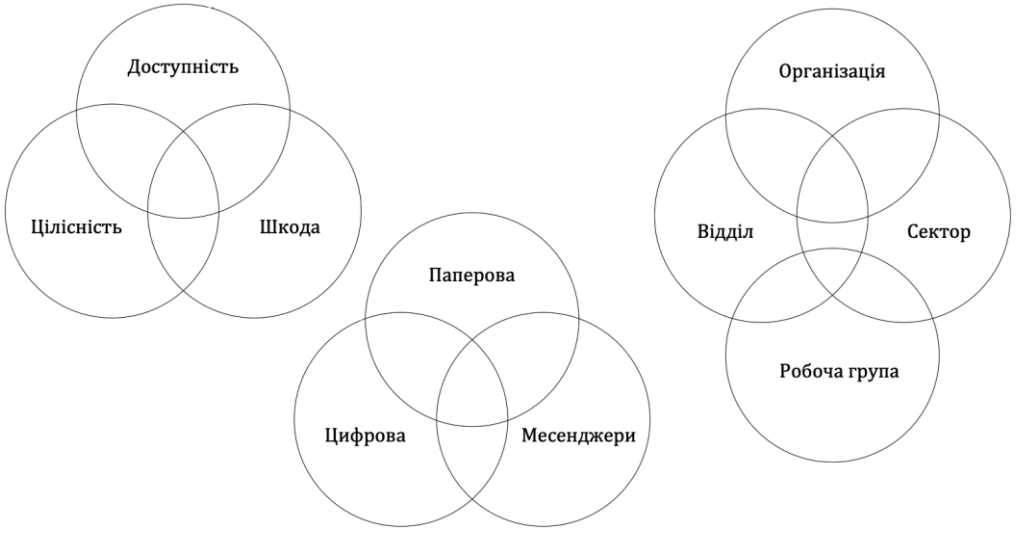

Рисунок 2. Кругові діаграми додаткових класифікаторів інформації

Джерело: розроблено авторами на основі розробки Бюро безпеки урядового зв'язку Н. Зеландії (2024), досліджень (Arora et al, 2023).

Найбільша доступність буде у публічної інформації, найменша – у конфіденційної. Найбільша шкода від розголошення конфіденційної інформації, а найменша – публічної. Решта показників сильно залежить від кількості документів, зокрема, цілісність.

Для демонстрації загального підходу достатньо буде неконкретизованого прикладу. Наприклад, в дослідженні пошукових систем, які використовують публічні та приватні дані, у якості прикладу інформаційно достатньої вибірки для висновку на основі аналізу приватних та публічних даних наведено електронні листи та новинарна стрічка Wikipedia – 47 тис. електронних листів та 5,2 млн. новин. Що дає нам співвідношення між публічними та приватними даними в 0,0011.

Емпірично можна припустити, що подібне співвідношення має існувати між будь-якою парою ступенів приватності. Тобто, на 10000 листів публічних документів має бути 11 листів приватних, із яких, в свою чергу, максимум по одному – обмеженого та конфіденційного характеру. Таке ж співвідношення можна використати як емпіричну оцінку цілісності – як умовну кількість листів, необхідних для видачі адекватного висновка. Поєднання системи рівнів доступу із застосуванням інформації таке: найбільш застосовною завжди є інформація організаційного рівня – і вона, як правило, завжди є публічною, максимум внутрішнього рівня. А обмежена і конфіденційна інформація не виходить за межі відділів та навіть окремих робочих груп.

Класифікацією в залежності від користувачів інформації складніша. Наприклад, зрозуміло, що автор чи власник інформації буде мати до неї доступ незалежно від присвоєного рівня. Користувачу доступ може бути наданий на підставі посади, але може фактично він може доступитись до критичних через суміщення посад чи пряму передачу цих даних від автора чи власника. Аудитори можуть мати постійний доступ, а можуть – тимчасовий. Нарешті, офіцер безпеки даних може відповідати не за всі дані свого рівня доступу, а тільки за певну їх підмножину – яка має відношення до державно-приватного партнерства (перетин класифікатора застосування інформації та користувача). Пітер Вуд в своїй статті для Computer Weekly (Wood, 2013) рекомендує комплексне оцінювання, засноване на оцінці мотивації та можливостей користувачів (акторів), які оцінюються по п'ятибальній шкалі плюс нуль як відсутність загрози. Це дає змогу побудувати таку матрицю оцінок загроз з боку акторів (табл. 1).

Таблиця 1. Матриця загроз з боку акторів

	Можливість					
Мотивація		Дуже мала	Мала	Обмежена	Значна	Повна
	Відсутня	0	0	1	1	2
	Зацікавлення	0	0	1	2	3
	Інтерес	0	1	2	3	4
	Робота	1	1	2	4	4
	Фокус	1	3	3	4	5

Джерело: розроблено авторами на основі досліджень (Wood, 2013).

За допомогою цієї матриці ми можемо емпірично визначити, що загроза від автора завжди буде максимальна, тому що він має повний доступ до інформації і він сфокусований на темі. В той же час він практично не зацікавлений в розголосі, тому що це поставить під удар його інтереси. Таким чином, його загроза як актора буде варіюватись в залежності від того, наскільки сильним буде цей удар. Точно так же, в залежності від особистої шкоди, буде залежати загроза розголошення на рівні користувача.

Найбільш невизначеними будуть оцінки високих рівнів безпеки – тому що там вступають в дію фактори, які складно оцінити. Наприклад, критичність певного ресурсу в залежності від користувача в такому випадку емпірично можна оцінити так:

– публічна інформація критичною не є, так як користувач має завжди до неї доступ і у нього

відсутня зацікавленість в розголосі (0);
- внутрішня інформація має 1-бальну критичність, тому що користувач має замалі можливості ознайомитись із нею (випадкові);
- обмежена і конфіденційна інформація має 4-бальну критичність, перш за все через малі можливості користувача її отримати (максимум в форматі "необхідно знати"), однак як нам відомо з оцінки цілісності, достатньо всього 1 листа приватної інформації для отримання відповіді (нанесення шкоди).

Що, в свою чергу, ставить необхідність долучення до вищенаведеної матриці третій вимір – як мінімум шкоду для актора, яка має коригуватись відносно шкоди для підприємства.

Точно таким же чином можна оцінити і ризики від застосування інформації в залежності від організаційного рівня, для якого призначені документи.

Всі документи організаційного рівня мають найвищу критичність (оскільки шкода від них може вплинути на функціонування підприємства), а найменшу – документи рівня робочої групи, тому що вони демонструють лише матеріали проектів, або навіть лише певну їх частину – якщо проект сильно диверсифікований.

Найбільші проблеми – якщо до оцінки додається форма представлення. Найбільш захищеними, як не дивно, є паперова і цифрова локальна форми, тому що доступ до них можна контролювати організаційними заходами – і відповідно, можна оцінити ризики інформаційної безпеки. А до віддалених сховищ доступ може мати невстановлена кількість осіб (наприклад, обслуговуючий персонал фірми, яка забезпечує хостинг даних), а електронна пошта та месенджери є безпеково слабкими сховищами (вони можуть бути підсиленими шифруванням і політикою підприємства до рівня цифрових локальних сховищ, але це виходить за рамки тематики даної статті).

Нарешті, частина документів може дублюватись в різних формах і форматах – або з метою більшої збережності, або просто для спрощення роботи (що, до речі, додатково ускладнює безпекове питання – тому що за копіями складніше услідкувати).

Матриця Пітера Вуда також дозволяє зробити таку оцінку – вона навіть наведена в статті (Wood, 2013). Зведемо всі вищенаведені оцінки в спільну таблицю 2.

Таблиця 2. Інтегральна оцінка критичності ресурсів

	Public	Internal	Restricted	Confidential
Business Impact Levels(BIL)				
Доступність	1	2	3	4
Шкода	1	2	3	4
Цілісність	10000	11	1	1
Застосування інформації				
Організаційний	0	3	5	5
Секторальний	0	3	4	5
Відділ	0	3	3	5
Робоча група	0	3	2	5
Користувач інформації				
Автор/Власник	0	5	5	5
Користувач	0	1	4	4
Перевіряючий	0	3	5	5
Офіцер БД	0	2	5	5
Форма представлення інформації				
Паперова	0	2	3	5
Цифрова (локальна)	0	2	4	5
Цифрова (віддалена)	0	3	5	5
Електронна пошта	0	4	5	5

Джерело: розроблено авторами на основі досліджень (Wood, 2013).

Ця таблиця добре показує, наскільки різноманітними є показники, які

використовуються для оцінки критичності інформаційних ресурсів. Також, вона показує, що таблична методика дозволяє ефективно поєднувати не більше двох, максимум трьох класифікаторів. Цього достатньо для забезпечення мінімальної інформаційної безпеки, але не достатньо для забезпечення її в належних обсягах.

Крім того, наскільки б не була продумана структура таблиці і наскільки деталізованими не були б класифікатори, вони всі представляють собою статичну оцінку "на певну фіксовану точку" – наприклад, на момент аудита, або на момент чергової посадової атестації.

В проміжку між цими фіксованими точками ані сама класифікація ресурсів, ані безпосередні переліки класифікованих ресурсів не переглядаються. Між тим, швидкість зміни ринкової кон'юнктури, законодавчого поля, навіть політичної інформації, дедалі зростає і уже зараз вона швидша, ніж будь-який період між аудитами чи атестаціями. Не кажучи про те, що між фіксованими точками може змінитись кількість людей, яким надано доступ до тієї чи іншої інформації. Таким чином, певний формат динамічної класифікації стає життєво необхідним.

Динамічна класифікація на базі аналізу доступу до ресурсів. Одним із перспективних варіантів динамічної класифікації ресурсів може стати профілювання доступу до них. Найкраще цей метод спрацює для цифрових документів, оскільки сучасні операційні системи дозволяють обмежувати доступ не лише в розрізі дисків та окремих каталожних елементів (папок, архівів, типів файлів тощо), а й для конкретно взятих ресурсів. Розглянемо лише два додаткові параметри: *частота і протокол доступу до ресурсу*.

Обидва ці параметри є універсальними і багатозначними, при належному використанні здатні висвітлити цілий набір корисних показників як для оцінки ризиків, так і для покращення безпеки організації. Наприклад, якщо проаналізувати частоту доступу для обмежених і конфіденційних ресурсів, легко побачити, які із них є більш критичними для роботи підприємства (до них буде доступ буде майже постійним), а які – ні (до них доступ буде спорадичним, або й взагалі не буде за період скринінга).

Якщо частота доступу вимірюється періодично, то можна скласти так званий "стандартний паттерн доступу" – тривимірний графік доступу до критичних ресурсів. За його допомогою можна відстежити аномальне зростання доступу до конфіденційних ресурсів. В залежності від ситуації це може бути пов'язано із окремим проектом, реорганізацією підприємства, аудитом... або прихованим делегуванням доступу іншим працівникам. Це допоможе більш чітко визначити права доступу і навіть перемістити деякі ресурси з менш захищеної категорії в більш захищену.

Наприклад, якщо за основу взяти організаційну структуру, то можна емпірично показати, що частота звертання до документів відповідного рівня будуть приблизно відповідати кількості персоналу, який зайнятий у підрозділах, які мають до нього доступ.

Тобто, якщо в даному підприємстві існує три робочі групи по 10 працівників, 3 менеджери і 1 директор, і по одному документу кожного типу, відповідно, доступи розподіляться так (табл. 3).

Таблиця 3. Розподіл доступу між персоналом

	Public	Internal	Restricted	Confidential
Організаційний	34	4	1	1
Секторальний	34	4	15	1
Робоча група	34	4	12	1

Джерело: розроблено авторами на основі досліджень Вуда П. (2024).

Зрозуміло, що до публічного документів матимуть доступ всі. До внутрішніх – весь керівний склад (тобто, 3 менеджери та директор). До документів обмеженого доступу на рівні робочої групи матимуть доступ працівники, менеджер і директор, на рівні сектора –

працівники робочої групи, всі менеджери і директор. А на організаційному рівні до всіх документів обмеженого доступу матиме доступ лише директор, як і до конфіденційного документа.

Серед сучасних файлових систем виділяється окрема група так званих "журнальованих файлових систем", які дозволяють вести реєстр транзакцій на рівні окремих файлів. Транзакція – це будь-яка елементарна дія, наприклад, читання, запис, копіювання та переміщення. В поєднанні із системою авторизації журналювання може чітко показати, хто і наскільки часто звертався до того чи іншого ресурсу.

Це дозволяє побудувати аналогічні стандартні патерн для доступу за дозволом чи за частотою копіювання. Якщо періодично вимірювати частоту доступу, а потім сформувати часовий графік, вийде наступна поверхня (тривимірний графік). Піки покаже аномально високу частоту, яку можна трактувати або як делегування повноважень, або несанкціонований доступ (рис. 3).

Фактично на одному такому графіку можна розміщувати лише співмірні за показниками групи ресурсів. Наприклад, на діаграмі 3 насправді розміщено не дві категорії, а три – Public, Internal та Restricted. Однак друга категорія має настільки незначний доступ, що просто не видна на поверхні і відстежити можна лише зовсім аномальний доступ – на рівні масивного витоку даних. Що, в свою чергу робить невідстежувані дрібні або одиничні втручання в безпеку на рівні малозатребуваних ресурсів, або інформації, яка не повинна виходити за межі дуже обмеженого кола – а це не лише внутрішня документація, а й конфіденційна.

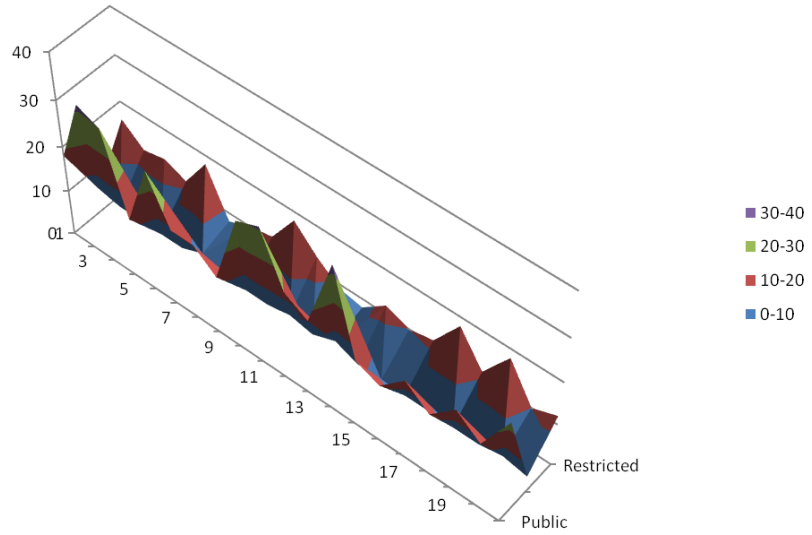

Рисунок 3. Стандартний патерн доступу

Джерело: розроблено авторами

Знову ж таки, аномальне збільшення звертання до раніше малозатребуваного ресурсу може означати або приховане делегування доступу, або передачу файла іншим працівникам. Однак його необхідно якось відстежити, а на загальному патерні це помітити складно.

Що стосується пікової частоти для дійсно часто затребуваного ресурсу – то у разі постійних перевищень патернів необхідно або переглядати класифікацію як таку (розглянути, наприклад, пониження ступеню захисту), або розділити контент на менш захищену (із переводом в менш захищену категорію) та більш захищену.

Для нецифрових критичних документів висока частота доступу може бути достатньою підставою пришвидшити їх цифровізацію, оскільки так простіше запобігти їх витоку і запровадити протокол (моніторинг) доступу до них.

Це робить метод стандартних патернів достатньо простим, щоб їх можна було

вимірювати з достатньо високою періодичністю і зробити базою для коригування інших класифікаторів ресурсів. Однак у нього є суттєвий недолік. Він добре підходить для автоматизованих систем і для часто використовуваних документів, але вкрай погано підходить, якщо на підприємстві існує розгалужена і різноманітна класифікація документів за рівнями доступу – для малозатребуваних даних цей метод покаже лише масові витоки (Xu, et al., 2024).

Це породжує задачу такої візуалізації кількох показників ризиків, які б полегшували аналіз насамперед малозатребуваних ресурсів.

Багатовимірна матриця класифікації ресурсів. З метою спрощення використовуються два класифікатори – основний, в рамках якого визначаються рівні доступу, і доповняльний, який дозволяє чіткіше визначити межі класів і конкретизує класифіковану інформацію. Тому найбільш поширеним є табличне класифікування. Для варіанта одночасного застосування трьох класифікаторів використовується "класифікаційна сітка". Це більш гнучка система представлення класифікації, оскільки замість лінійних узагальнених блоків ресурсів дозволяє розглядати їх кластери. Наприклад, побудуємо таку сітку, відклавши по вісі Z частоту доступу до деякого набору документів ресурсу за певний перевірочний період (рис. 4).

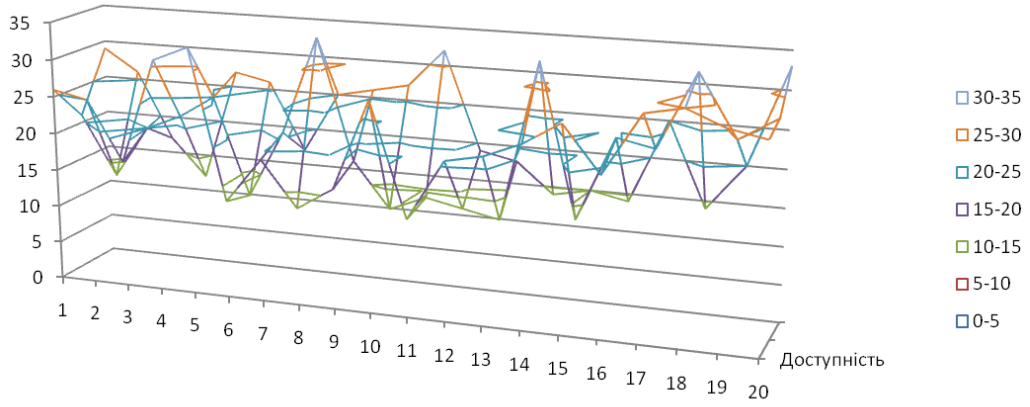

Рисунок 4. Класифікаційна сітка

Джерело: розроблено авторами

На рис. 4 видно, які ресурси використовуються найчастіше, а які – найрідше. Водночас на будь-якому рівні можна взяти переріз і отримати все ту ж традиційну табличну форму.

Класифікаційна сітка використовується і для класифікації, і також для аналізу її коректності та коригування. Варіюючи сітку по вертикальній вісі, можна будувати класифікаційні кластери. А найбільшого перегляду (як і найбільшої уваги) заслуговують ресурси, до яких найбільша і найменша частота доступу. Щоб адекватно їх проаналізувати, необхідно додати на графік динамічну складову.

Класифікаційна сітка є складною, а чим більше показників, які поєднуються під час аналізу, тим більше вимірів матиме поверхня і тим складніше буде її аналізувати. На нашу думку, для спрощення аналізу і більшої наглядности найкраще використати форму "класифікаційного вектора", що дозволить проаналізувати динамічну складову, і при цьому не обтяжити графік додатковими даними.

Ідея така: кожен ресурс необхідно показувати не в формі статичної точки на графіку, а в формі вектора, пропорційному максимальній зміні динамічного показника, спрямованого в бік цієї зміни. Це дозволить, наприклад, візуалізувати миттєві піки для малозатребуваних ресурсів, які в інших варіантах візуалізацій складно розрізнити. Крім того, перевагою такого підходу є можливість показувати практично будь-яку кількість показників та їх комбінацій і

при цьому не втрачати динамічної складової. Його можна називати багатовимірною матрицею класифікації ресурсів. Спрощений приклад такої матриці – класифікаційний стек – показаний на рисунку 5.

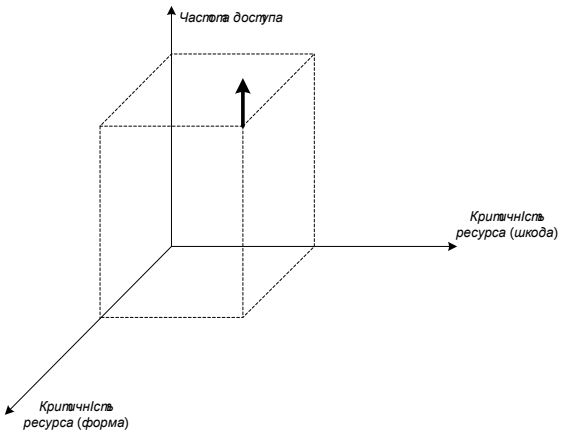

Рисунок 5. Спрощений варіант багатовимірної матриці класифікації ("класифікаційний стек")

Джерело: розроблено автором

На рис. 5 наведено ресурс, який має середнє значення ризику потенційної шкоди і середнє значення форми (цифрове локальне представлення), які показані на пласкій частині діаграми. Вона відповідає табличній формі класифікатора ресурсів. Очікувана частота доступу до ресурсу відкладена по вісі аплікат, а максимальна зміна цієї частоти за перевірочний період показана у вигляді стрілочки-вектора, пропорційної цій зміні.

Якщо такий вектор буде надзвичайно великим (тобто, в 1.5 і більше разів), його використання слід промоніторити і підтвердити один із трьох можливих висновків:

- потреба в документі дійсно виросла і необхідно переглянути його класифікацію;
- має місце делегування доступу особам, які зазвичай його не мають – це може не становити суттєвого ризику (у випадку матеріалів обмеженого доступу), а може становити суттєвий ризик;
- відбувся несанкціонований доступ.

Таким чином, в класифікаційному стекові "спливуть на поверхню" всі ресурси із проблемною класифікацією чи проблемами із захистом, і "потонуть" всі ресурси, чия безпека не викликає суттєвих питань. Варіюючи масштаб матриці, а також варіюючи пари пласкої частини діаграми, можна порівняно легко визначити основні ризики інформаційної безпеки і мінімізувати їх за рахунок більш точної класифікації та контролю доступу.

Крім того, в такий тип діаграми можна додавати будь-яку кількість додаткових показників у вигляді різноспрямованих векторів – що полегшить обрахунок інтегральних показників, подібних тому, який було наведено вище для загроз з боку акторів. Показники-складові можна показати у вигляді векторів, а сума цих векторів буде шуканим інтегральним показником.

При належному використанні такий підхід дозволяє вирішити задачі, які ставились в цій роботі, і одночасно є зручним способом візуалізації стану критичних інформаційних ресурсів на підприємстві.

Обговорення. Основний матеріал дослідження включає результати застосування динамічної класифікації інформаційних ресурсів та їх візуалізацію за допомогою багатовимірної матриці. Аналіз проводився на основі даних, отриманих від реальних компаній, що використовують сучасні методи контролю доступу до інформаційних ресурсів.

Рисунок 3 демонструє стандартний патерн доступу до критичних інформаційних ресурсів. Цей тривимірний графік показує частоту звернень до ресурсів залежно від рівня

конфіденційності. На графіку видно, що найбільша частота доступу спостерігається до внутрішніх даних (Internal Data), тоді як до архівних даних (Legacy Data) доступ здійснюється значно рідше.

Встановлено, що частота доступу до обмежених і конфіденційних ресурсів дозволяє визначити їхню критичність для роботи підприємства. Наприклад, ресурси з постійним високим рівнем доступу є найбільш критичними, тоді як ресурси з низькою частотою доступу мають меншу значущість. Таблиця 4 подає результати статистичної обробки даних про доступ до ресурсів.

Таблиця 4. Частота доступу до інформаційних ресурсів за рівнем конфіденційності

Рівень конфіденційності	Частота доступу (середня кількість запитів на місяць)
Публічна інформація	10,000 ± 500
Внутрішня інформація	5,000 ± 300
Обмежена інформація	1,000 ± 50
Конфіденційна інформація	100 ± 10

Джерело: розроблено автором на основі досліджень (Oseghale, 2023; Lipps & Schotten, 2022)

За результатами аналізу було встановлено, що використання динамічної класифікації дозволяє виявити аномалії в доступі до ресурсів. Такі аномалії в доступі можуть бути пов'язані з виконанням специфічних проектів або реорганізацією компанії, однак вони також можуть свідчити про потенційні загрози інформаційній безпеці. Використання багатовимірної матриці класифікації дозволяє виявляти та моніторити такі зміни в режимі реального часу, що підвищує ефективність захисних заходів.

Рисунок 5 демонструє спрощений варіант багатовимірної матриці класифікації, так званий, "класифікаційний стек"), який показує відносне розташування ресурсів за їх критичністю та частотою доступу. Ресурси з високою критичністю та частотою доступу розміщені у верхній частині матриці, тоді як менш критичні ресурси – у нижній частині. Ця візуалізація дозволяє ідентифікувати ресурси, які потребують підвищеного рівня захисту, та ресурси, які можуть бути цифровізовані для покращення їх захисту. Використання багатовимірної матриці також дозволяє автоматизувати процес класифікації, що зменшує ризики, пов'язані з людським фактором.

Обговорення результатів включає порівняння отриманих даних з результатами інших досліджень. Наприклад, дослідження (Wiedemann et al., 2024), (Ramamurthy et al., 2022) підтверджують, що динамічна класифікація забезпечує більш високий рівень захисту інформаційних ресурсів порівняно зі статичними методами. Робота (Mazzola Л. et al., 2021) показує доцільність створення інструменту, який доповнює механізм безпеки, підтримуючи експертів у виявленні незвичайних патернів та подій, пов’язаних із безпекою, які слід відстежувати та перевіряти механізмом класифікації подій.

Також виявлено, що запропонований підхід із динамічною складовою дозволяє більш ефективно реагувати на зміни у використанні інформаційних ресурсів, що підтверджується результатами експериментів, проведених у реальних умовах. Використання динамічної класифікації сприяє підвищенню точності оцінки ризиків та забезпечує більш ефективний захист критичних даних.

Порівняння отриманих результатів з даними інших досліджень свідчить про те, що запропонований підхід має значний потенціал для впровадження у системах інформаційної безпеки. Використання багатовимірної матриці класифікації дозволяє підвищити точність і швидкість реагування на загрози, що є важливим фактором у сучасних умовах швидкого розвитку інформаційних технологій.

Загалом, результати дослідження підтверджують ефективність запропонованого підходу та його практичну цінність для підвищення рівня інформаційної безпеки.

Висновки. В даній роботі обґрунтовано та запропоновано покращений підхід до

візуалізації стану критичних інформаційних ресурсів на підприємстві за допомогою динамічного підходу із застосуванням багатовимірної матриці класифікування інформації для оцінки ризиків інформаційної безпеки.

Зокрема, в роботі було сформульовано і проаналізовано задачу класифікації інформаційних ресурсів як ключову задачу під час оцінки ризиків інформаційної безпеки. Основну увагу приділено питанням класифікації, оскільки невірна класифікація може призвести не лише до витоків даних, а й до несанкціонованого доступу до широкого спектру чутливої інформації.

Було проаналізовано табличні та графічні методи класифікації інформаційної безпеки та її перегляду в процесі діяльності організації чи підприємства. Зроблено висновок, що переважна більшість методів класифікації надзвичайно слабко враховує динамічну складову – насамперед частоту доступу до тих чи інших ресурсів, яка може слугувати основним показником реальної класифікаційної належності інформаційного ресурсу – особливо враховуючи сучасні технічні можливості, які дозволяють контролювати і журналювати доступ навіть на рівні окремих документів.

Для періодичного перегляду класифікації та відстеження динаміки доступу запропоновано використовувати багатовимірну класифікаційну матрицю, яка у спрощеній формі "класифікаційного стеку" дозволяє відстежити ресурси, що становлять потенційний ризик через некоректну чи застарілу класифікацію.

Запропонований підхід містить значний потенціал і може стати базою для подальших досліджень, оскільки за його допомогою можна розмістити на графіку будь-яку кількість показників у вигляді набору векторів та формувати інтегральні оцінки ризиків на основі суми чи різниці векторів. Також запропонований підхід потенційно може слугувати базою для досліджень із тренування нейромережевих засобів автоматизованої класифікації.

Подяки «Немає».

Конфлікт інтересів «Немає».

Стаття надійшла: 12.08.2024.

References

Al Qahtani, E., Story, P., & Shehab, M. (2024). The impact of risk appeal approaches on users’ sharing confidential information. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Article 579, pp. 1–21). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642524

Alotibi, G. (2024). A cybersecurity awareness model for the protection of Saudi students from social media attacks. *Engineering, Technology & Applied Science Research*, 14(2), 13787-13795.

Arora, S., Lewis, P., Fan, A., Kahn, J., & Ré, C. (2023). Reasoning over public and private data in retrieval-based systems. *Transactions of the Association for Computational Linguistics*. Retrieved from https://transacl.org/index.php/tacl/article/view/4705

Arslan, M., & Cruz, C. (2024). Business text classification with imbalanced data and moderately large label spaces for digital transformation. *Applied Network Science*, 9, 11. https://doi.org/10.1007/s41109-024-00623-5

Barnawi, A., Kumar, K., Kumar, N., & Alzahrani, B., & Almansour, A. (2024). A deep learning approach for landmines detection based on airborne magnetometry imaging and edge computing. *Computer Modeling in Engineering & Sciences*, 139(2), 2117-2137. https://doi.org/10.32604/cmes.2023.044184

Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data*, 11, 23. https://doi.org/10.1186/s40537-024-00882-0

Gambarelli, G., Gangemi, A., & Tripodi, R. (2023). Is your model sensitive? SPEDAC: A new resource for the automatic classification of sensitive personal data. *IEEE Access*, 11, 10864-10880. https://doi.org/10.1109/ACCESS.2023.3240089

Irwin, L. (2022, August 30). What is ISO 27001 information classification? *IT Governance*. Retrieved from https://www.itgovernance.co.uk/blog/what-is-information-classification-and-how-is-it-relevant-to-iso-27001

Lipps, C., & Schotten, H. D. (2022). Physical layer security: About humans, machines and the transmission channel. In *Proceedings of the 21st European Conference on Cyber Warfare and Security* (Vol. 21, No. 1, pp. 161-169). Academic Conferences International Limited. https://doi.org/10.34190/eccws.21.1.403. Retrieved from https://papers.academic-conferences.org/index.php/eccws/article/view/403/357

Malchiodi, D., Raimondi, D., Fumagalli, G., et al. (2024). The role of classifiers and data complexity in learned Bloom filters: Insights and recommendations. *Journal of Big Data*, 11, 45. https://doi.org/10.1186/s40537-024-00906-9

Mazzola, L., et al. (2021). Security rules identification and validation: The role of explainable clustering and information visualisation. In Stephanidis, C., Antona, M., & Ntoa, S. (Eds.), *HCI International 2021 - Posters*. HCII 2021. Communications in Computer and Information Science, vol 1420. Springer. https://doi.org/10.1007/978-3-030-78642-7_58

Mikuletič, S., Vrhovec, S., Skela-Savič, B., & Žvanut, B. (2024). Security and privacy oriented information security culture (ISC): Explaining unauthorized access to healthcare data by nursing employees. *Computers & Security*, 136, 103489. https://doi.org/10.1016/j.cose.2023.103489

New Zealand Protective Security Requirements. (2024, May 10). Applying business impact levels. Retrieved from https://nzism.gcsb.govt.nz/assets/Previous-versions/v3-2/NZISM-Part-One-v3.2-December-2018.pdf

Oseghale, O. (2023). Digital information literacy skills and use of electronic resources by humanities graduate students at Kenneth Dike Library, University of Ibadan, Nigeria. *Digital Library Perspectives*, 39(2), 181-204. https://doi.org/10.1108/DLP-09-2022-0071

Pitafi, S., Anwar, T., Widia, I. D. M., & Yimwadsana, B. (2023). Revolutionizing perimeter intrusion detection: A machine learning-driven approach with curated dataset generation for enhanced security. *IEEE Access*, 11, 106954-106966. https://doi.org/10.1109/ACCESS.2023.3318600

Queensland Government. (2024, May 10). Information security classification framework (QGISCF) – Queensland Government guidelines. Retrieved from https://www.forgov.qld.gov.au/information-and-communication-technology/qgea-policies-standards-and-guidelines/information-security-classification-framework-qgiscf

Ramamurthy, A., Sathya, V., Rochman, M. I., & Ghosh, M. (2022). ML-based classification of device environment using Wi-Fi and cellular signal measurements. *IEEE Access*, 10, 29461-29472. https://doi.org/10.1109/ACCESS.2022.3158056

Robinson, P. (2024). Data classification? Definition, levels & examples – Lepide data security. *Lepide*. Retrieved from https://www.lepide.com/blog/what-is-data-classification-and-how-to-do-it/

Shmatko, O., Balakireva, S., Vlasov, A., Zagorodna, N., Korol, O., Milov, O., Petrov, O., Pohasii, S., Rzayev, K., & Khvostenko, V. (2020). Development of methodological foundations for designing a classifier of threats to cyberphysical systems. *Eastern-European Journal of Enterprise Technologies*, 3(9), 6-19. https://doi.org/10.15587/1729-4061.2020.205702

Song, X., Liu, Z., & Jiang, B. (2024). Adaptive boosting with fairness-aware reweighting technique

for fair classification. *Expert Systems with Applications*, 250, 123916. https://doi.org/10.1016/j.eswa.2024.123916

Velmurugan, S., Prakash, M., Neelakandan, S., et al. (2024). Provably secure data selective sharing scheme with cloud-based decentralized trust management systems. *Journal of Cloud Computing*, 13, 86. https://doi.org/10.1186/s13677-024-00634-8

Venn, B., Leifeld, T., Zhang, P., et al. (2024). Temporal classification of short time series data. *BMC Bioinformatics*, 25, 30. https://doi.org/10.1186/s12859-024-05636-6

Wang, G., & Gu, Y. (2024). Multi-task scenario encrypted traffic classification and parameter analysis. *Sensors*, 24(10), 3078. https://doi.org/10.3390/s24103078

Wiedemann, N., Janowicz, K., Raubal, M., et al. (2024). Where you go is who you are: A study on machine learning based semantic privacy attacks. *Journal of Big Data*, 11, 39. https://doi.org/10.1186/s40537-024-00888-8

Wood, P. (2013, January). Business priorities: What to protect, monitor and test. *Computer Weekly*. Retrieved from https://www.computerweekly.com/feature/Business-priorities-what-to-protect-monitor-and-test

Xu, A., Gao, J., Sui, X., Wang, C., & Shi, Z. (2024). LiDAR dynamic target detection based on multidimensional features. *Sensors*, 24(5), 1369. https://doi.org/10.3390/s24051369

Zhang, Y., Deng, Q., Liang, W., & Zou, X. (2018). An efficient feature selection strategy based on multiple support vector machine technology with gene expression data. *BioMed Research International*, 2018. https://doi.org/10.1155/2018/1234567