КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.23.131143
УДК 004.855

Партика Андрій Ігорович
кандидат технічних наук, старший викладач кафедри захисту інформації
Національний університет «Львівська політехніка», Львів, Україна
ORCID 0000-0003-3037-8373
andrii.i.partyka@lpnu.ua

Михайлова Ольга Олександрівна
кандидат фізико-математичних наук, доцент кафедри захисту інформації
Національний університет «Львівська політехніка», Львів, Україна
ORCID 0000-0002-3086-3160
olha.o.mykhailova@lpnu.ua

Шпак Станіслав Володимирович
здобувач освіти спеціальності 125 Кібербезпека кафедри захисту інформації
Національний університет «Львівська політехніка», Львів, Україна
ORCID 0009-0008-0256-4850
stanislav.shpak.mkbas.2023@lpnu.ua

ВИЯВЛЕННЯ, АНАЛІЗ ТА ЗАХИСТ КОНФІДЕНЦІЙНИХ ДАНИХ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЇ МАШИННОГО НАВЧАННЯ СЕРВІСУ AMAZON MACIE

Анотація. За останні десятиліття сфера зберігання та обробки даних зазнала суттєвих змін і розширення, особливо з появою хмарних технологій та обчислень. Хмарні сервіси надають організаціям можливість зберігати великі обсяги даних та отримувати доступ до них за допомогою розподілених систем. Однак, разом з цими новими можливостями, постають і нові виклики, зокрема в області захисту конфіденційних даних. Захист конфіденційних даних є надзвичайно важливою задачею для сучасних організацій, особливо в умовах зростаючої кількості цифрових загроз та порушень безпеки. З метою забезпечення надійного захисту цінної та чутливої інформації, розробники та дослідники активно працюють над розробкою нових технологій та інструментів. Одним з потужних інструментів, який використовується для виявлення, аналізу та захисту конфіденційних даних є технологія машинного навчання сервісу Amazon Macie. Amazon Macie є сервісом хмарних обчислень AWS, який використовує штучний інтелект та алгоритми машинного навчання для автоматизованого аналізу даних та виявлення потенційних загроз безпеці даних. Основним завданням цієї роботи є проблематика виявлення, аналізу та захисту конфіденційних даних з використанням технології машинного навчання та сервісу Amazon Macie. Amazon Macie є інноваційним сервісом, розробленим компанією Amazon Web Services (AWS), який використовує передові алгоритми машинного навчання для автоматизованого виявлення та аналізу конфіденційних даних. В рамках роботи проведено аналіз основних алгоритмів машинного навчання, принципів роботи систем зберігання даних та методів захисту конфіденційної інформації. Було досліджено принципи роботи та можливості Amazon Macie, який використовує розширені алгоритми машинного навчання для автоматизованого аналізу даних та виявлення потенційних загроз для безпеки даних.

Ключові слова: машинне навчання; Amazon Macie; AWS; кібербезпека; автоматизований аналіз; конфіденційні дані.

ВСТУП

Стрімкий розвиток інтернет-сервісів також призвів до значного збільшення кількості кібератак. Кіберзагрози стають все більш складними, а автоматизація робить захист неефективним. Дослідники інформаційно-комунікаційних технологій погоджуються, що інформаційна безпека має першорядне значення. Отже, низка науковців намагалися вирішити цю проблему шляхом впровадження вдосконалених методів і технологічних артефактів; зокрема використання детекторів зловмисного програмного забезпечення, систем виявлення та запобігання вторгненням (IDPS), складних налаштувань брандмауера та алгоритмів шифрування даних [1].

Прогрес у додатках штучного інтелекту зробив можливим розробити відносно ефективні та результативні системи, які автоматично ідентифікують і запобігають зловмисним діям у кіберпросторах. Вони були прийняті для підтримки існуючих технологічних методів, оскільки вони забезпечують ефективні стандарти та механізми для кращого контролю та запобігання кібератакам. Незважаючи на всі переваги штучного інтелекту, швидка еволюція підходів ускладнює дослідникам визначення найефективнішої техніки та її впливу на безпеку кіберпростору. Загальне сприйняття серед дослідників і практиків інформаційної безпеки і кібербезпеки свідчить про те, що штучний інтелект покращив безпеку організаційної інформації [2].

Машинне навчання в кібербезпеці використовується для виявлення загроз, аналізу вразливостей та ідентифікації шкідливого програмного забезпечення. Воно базується на алгоритмах, які навчаються на основі історичних даних та створюють моделі для розпізнавання патернів та передбачення майбутніх атак [3].

Постановка проблеми. Захист конфіденційних даних є надзвичайно важливою задачею для сучасних організацій, особливо в умовах зростаючої кількості цифрових загроз та порушень безпеки. З метою забезпечення надійного захисту цінної інформації, розробники та дослідники активно працюють над розробкою нових технологій та інструментів. Одним з потужних інструментів, який використовується для виявлення, аналізу та захисту конфіденційних даних, є технологія машинного навчання сервісу Amazon Macie. Amazon Macie є сервісом хмарних обчислень AWS, який використовує штучний інтелект та алгоритми машинного навчання для автоматизованого аналізу даних та виявлення потенційних загроз безпеці даних.

Аналіз останніх досліджень і публікацій. Використання технологій машинного навчання набуває все більшої популярності, зважаючи на різноманітність сфер їх застосування. Зокрема при роботі з великим масивами даних, конфіденційними даними чи використовуючи сервісів хмарних провайдерів:

- Авторами [4] висвітлено загальні поглядами на конфіденційність даних і практики (наприклад, прийняття використання зовнішніх служб), а також теми, що висвітлюють конкретні перспективи аналізу зашифрованих даних;
- В роботі [5] розглядається модель аналізу настроїв даних великих обсягів на основі конфіденційної інформації. Зокрема отримані авторами результати показують важливість кожного слова та контекстно-залежний вектор для обчислення ваги кожного слова. Проте у разі невеликої кількості текстів цей метод не може отримати достатньо текстів в якості навчальних даних;
- Розглянуто питання безпеки даних у глибокому навчанні [6], зокрема показано потенційні загрози, а також новітні контрзаходи, засновані на різних технологіях, а також висвітлені різного роду атаки та захисту від них.

– Авторами [7] розглянуто різні алгоритми машинного навчання, які використовуються для подолання проблем безпеки в хмарі, включаючи контрольоване, неконтрольоване та навчання з підкріпленням. З подальшим порівнянням ефективності кожної техніки на основі її переваг і недоліків.

Мета статті. Основним завданням цієї роботи є проблематика виявлення, аналізу та захисту конфіденційних даних з використанням технології машинного навчання та сервісу Amazon Macie. Amazon Macie є інноваційним сервісом, розробленим компанією Amazon Web Services (AWS), який використовує передові алгоритми машинного навчання для автоматизованого виявлення та аналізу конфіденційних даних.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Метою машинного навчання є розробка машин, які автоматично навчаються приймати рішення. Навчання здійснюється за допомогою вказівок обчислювальному пристрою аналізувати деякі «існуючі» (навчальні) дані. За допомогою заданого алгоритму машинного навчання розробляється модель машинного навчання. Така модель включає в себе всі знання, отримані на етапі навчання, і реалізує функцію для прийняття рішень щодо «майбутніх» даних. Перш ніж модель машинного навчання можна буде розгорнути в робочому середовищі, необхідно оцінити її продуктивність. З цією метою деякі «перевірочні» дані обробляються моделлю машинного навчання, а її прогнози або аналізуються людьми, або порівнюються з певною відомою основною правдою. Таким чином, ми можемо визначити метод машинного навчання як «процес розробки моделі машинного навчання за допомогою алгоритмів машинного навчання на основі певних даних [8].

Моделі машинного навчання можуть виявляти аномальну або шкідливу активність в мережі або системі. Вони можуть аналізувати великі обсяги даних, щоб виявляти незвичайні патерни, використовуючи алгоритми кластеризації та класифікації. Крім того, машинне навчання використовується для виявлення вразливостей у програмному забезпеченні та мережах. Моделі можуть навчатися на основі даних про відомі вразливості, щоб автоматично виявляти нові вразливості та рекомендувати заходи для їх виправлення.

Щодо виявлення шкідливого програмного забезпечення, машинне навчання дозволяє розпізнавати характеристики шкідливих програм та виявляти нові види шкідливих програм на підставі їх сигнатур або відмінностей в поведінці. Машинне навчання також може використовуватися для прогнозування майбутніх загроз. Аналізуючи дані про попередні атаки та тренди у кібербезпеці, моделі можуть передбачати потенційні атаки та рекомендувати заходи для їх запобігання заздалегідь.

Застосування машинного навчання в кібербезпеці дозволяє автоматизувати виявлення загроз, зменшує час відновлення після атаки та покращує загальний рівень захисту системи. Проте важливо враховувати постійну необхідність навчання та підтримки моделей, оскільки кіберзлочинці постійно змінюють свої підходи та техніки атаки [9].

Amazon Macie — це послуга, надана компанією Amazon Web Services (AWS), яка використовує штучний інтелект для автоматичного виявлення, класифікації та захисту конфіденційних даних в хмарному середовищі. Вона дозволяє організаціям інтелектуалізувати та автоматизувати процеси розпізнавання та захисту конфіденційних даних, що зберігаються в Amazon S3 та інших сервісах AWS [10].

Основні можливості та функції Amazon Macie включають:

1. *Виявлення конфіденційних даних*: Amazon Macie використовує алгоритми машинного навчання для автоматичного виявлення різних типів конфіденційних даних, таких як номери кредитних карт, соціальні страхові номери, адреси електронної пошти тощо. Вона розпізнає конфіденційні дані на основі шаблонів, контексту та структури даних, що допомагає ідентифікувати потенційно чутливі інформаційні ресурси.
2. *Класифікація даних*: Amazon Macie автоматично класифікує дані, що зберігаються в Amazon S3, на основі типу інформації, яку вони містять. Вона використовує набір вбудованих класифікаторів для ідентифікації даних, таких як фінансова інформація, медичні записи, інтелектуальна власність та багато інших. Це допомагає зрозуміти, які типи даних зберігаються в системі та як з ними повинні бути пов'язані політики безпеки.
3. *Аналіз активності*: Amazon Macie відстежує та аналізує активність над конфіденційними даними, що зберігаються в Amazon S3. Вона виявляє незвичайні активності, які можуть вказувати на можливі порушення безпеки або несанкціонований доступ до даних. Наприклад, вона може спостерігати за наданням доступу до файлів, незвичайними попитами на дані, спробами несанкціонованої передачі даних тощо.
4. *Попередження та захист*: Amazon Macie надає можливість встановлювати правила попереджень та політик безпеки, що спрощує виявлення потенційних порушень безпеки та забезпечує автоматичний захист конфіденційних даних. Вона надає можливість створювати власні правила, налаштовувати оповіщення та реагувати на події, пов'язані з безпекою даних.

Amazon Macie допомагає організаціям забезпечити безпеку конфіденційних даних та виконати регуляторні вимоги щодо захисту інформації. Вона автоматизує процеси виявлення, класифікації та захисту даних, що зменшує ризики порушень безпеки та забезпечує довіру до зберігання даних в хмарному середовищі AWS. Окрім висновків, Macie надає статистику та інші дані, які дають змогу зрозуміти стан безпеки даних у сховищі Amazon S3 і те, де конфіденційні дані можуть зберігатися у файлах у сховищі. Статистика та дані надані Macie можуть допомогти прийняти рішення щодо виконання більш глибоких досліджень конкретних сегментів і об’єктів S3. Можливо переглянути й проаналізувати висновки, статистику та інші дані за допомогою консолі Amazon Macie або API Amazon Macie. [8]

МЕТОДИКА ДОСЛІДЖЕННЯ

У цій роботі ми використовували в якості сховища даних Amazon S3 [11], в якому були завантажені певні конфіденційні дані. За допомогою Macie можна автоматизувати виявлення конфіденційних даних і звітування про них двома способами: шляхом налаштування Macie для автоматичного виявлення конфіденційних даних, а також шляхом створення та запуску окремого процесу (роботи) з виявлення таких даних.

Автоматизація виявлення конфіденційних даних. Якщо Macie виявляє конфіденційні дані в об’єкті S3, він створює пошук вразливих даних. Знахідка містить детальний звіт про конфіденційні дані, які знайшов Macie. Автоматизоване виявлення конфіденційних даних забезпечує широку видимість того, де можуть зберігатися конфіденційні дані у масиві даних Amazon S3. За допомогою цієї опції Macie постійно

оцінює ваше сховище S3 інвентаризації та використовує методи вибірки для ідентифікації та вибору репрезентативних об'єктів S3 із ваших сховищ. Потім Macie отримує та аналізує вибрані об'єкти, перевіряючи їх на наявність конфіденційних даних.

Окрема робота з виявлення конфіденційних даних забезпечує більш глибокий і вузько спрямований аналіз. За допомогою цього параметра визначається ширина та глибина аналізу — сегменти S3 для аналізу, глибина вибірки та настроювання критеріїв, які випливають із властивостей об'єктів S3. Також можливо налаштувати завдання на виконання одноразового аналізу або на регулярній основі для періодичного аналізу, оцінки та моніторингу.

Виявлення різних типів конфіденційних даних. Щоб виявити конфіденційні дані Macie, використовує вбудовані критерії та методи, наприклад машинне навчання та зіставлення шаблонів для аналізу об'єктів у сховищах S3. Ці критерії та методи, які називаються ідентифікаторами керованих даних, можуть виявляти великий і постійно зростаючий список конфіденційних даних, зокрема особисту та медичну інформацію, фінансові та облікові дані. Щоб точно налаштувати аналіз, також можливо використовувати списки дозволених даних. Такі списки визначають певний текст і текстові шаблони, які Macie має ігнорувати в об'єктах S3. Зазвичай це винятки для ваших конкретних сценаріїв або середовищ, наприклад, імена публічних представників вашої організації, публічні номери телефонів або зразки тестових даних.

Оцінка та моніторинг даних для безпеки та контролю доступу. Macie автоматично генерує звіт та забезпечує переоблік ваших сховищ S3. Macie також оцінює та контролює сховище для безпеки і контролю доступу. Якщо Macie виявляє потенційну проблему з безпекою або конфіденційністю сегмента, створюється рекомендація для політик безпеки. Окрім конкретних результатів, інформаційна панель надає знімок сукупної статистики для даних Amazon S3. Це включає статистичні дані для ключових показників, наприклад кількість сегментів є загальнодоступними або доступними для інших облікових записів AWS. Кожну статистику можна деталізувати та переглянути підтверджуючі дані.

Macie також надає детальну інформацію та статистику для окремих сховищ S3 у вашому інвентарі. Дані включають розбивку налаштувань публічного доступу та шифрування сегмента, а також розмір і кількість об'єктів, які Macie може проаналізувати для виявлення конфіденційних даних у сховищі [12], [13].

Оцінювання важливості знайдених вразливостей. Коли Amazon Macie створює політику або знаходить конфіденційні дані, він автоматично призначає рівень важливості (severity) до знахідки. Важливість знахідки відображає основні характеристики знахідки та може вам допомогти оцінити і визначити пріоритетність своїх висновків. Важливість знахідки не передбачає або іншим чином не вказує на критичність або важливість ураженого ресурсу для вашої організації.

У Macie важливість знахідки представлена двома способами:

1. Рівень важливості

Це якісне уявлення про важливість. Рівні важливості коливаються від низького, для менш серйозного до високого, для найсуворішого. Рівні важливості відображаються безпосередньо на консолі Amazon Macie. Вони також доступні у форматі JSON-представлення результатів на консолі Macie. Рівні важливості також включені у знахідку події, які Macie публікує в Amazon EventBridge, і результати, які Macie публікує в AWS Security Hub.

2. Оцінка важливості

Це числове представлення важливості. Оцінки важливості варіюються від 1 до 3 і заносяться до таблиці рівнів важливості: Оцінки важливості не відображаються безпосередньо на консолі Amazon Macie. Однак вони також доступні в JSON-представлені знахідок на консолі Macie. Аналогіно як і у випадку з рівнями, оцінки важливості пошук подій також публікуються в Amazon EventBridge. Проте вони не включені до висновків, які Macie публікує в AWS Security Hub.

Налаштування ідентифікаторів даних. У нашому випадку Macie перевіряє об'єкти S3 за допомогою набору керованих ідентифікаторів даних, які ми рекомендуємо для автоматичного виявлення конфіденційних даних. Можливо налаштувати аналіз, щоб зосередитися на конкретних типах конфіденційних даних. Для цього необхідно змінити автоматизовані налаштування виявлення конфіденційних даних для вашого облікового запису будь-яким із наведених способів:

- Додати або видалити певні ідентифікатори керованих даних, наприклад в нашому випадку це номери банківських карток, медичні або паспортні чи інші особисті дані.
- Додати або видалити користувацькі ідентифікатори даних. Такого роду ідентифікатори дозволяють виявити дані, які відображають конкретні сценарії вашої організації, інтелектуальну власність або власні дані, такі як ідентифікатори працівників, електронні пошти або внутрішня класифікації даних [14].
- Додати або видалити списки дозволених. Як правило це винятки конфіденційних даних для вашого конкретного випадку чи середовища.

Macie автоматично обчислює оцінку чутливості для кожного сегмента S3, який він відстежує аналізи для вашого облікового запису. У Macie показник чутливості є кількісним показником перетину двох основних вимірів: кількість конфіденційних даних, які Macie знайшов у сховищі, і кількість даних, які Macie проаналізував у сховищі. Оцінка чутливості сегмента визначає, яку мітку чутливості призначає Macie до сховища [15]. Оцінка чутливості та мітка сегмента S3 не означає або іншим чином не вказує на критичність або значення, яке може мати сегмент або об'єкти сегмента для вашої організації. Натомість вони призначені для того, щоб надати контрольні точки, які можуть допомогти вам ідентифікувати та контролювати потенційні ризики безпеки. В процесі сканування Macie оновлює оцінку та мітку для відображення результатів аналізів. Наприклад: якщо Macie знаходить конфіденційні дані в об'єктах сегмента, Macie збільшує показник чутливості сегмента та за потреби оновлює мітку сегмента.

Ви можете налаштувати параметри оцінки чутливості для окремих сегментів S3, включивши або виключивши їх певні типи конфіденційних даних із оцінки сегмента. Ви також можете перевизначити обчислений сегмент оцінити.[16]

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для тестування ми помістили декілька різних файлів з вразливими даними у сховище S3 (рис. 1). Файли є різного типу та мають різну інформацію, зокрема імена, електроні адреси, номера карток, медичне страхування і тд. (рис. 2–3).

ISSN 2663 - 4023

id	first_nam	last_name	email	gender	password
1	Jarrid	Callar	jcallar0@g	Male	xaCA4VWdZPp
2	Geoff	Petrakov	gpetrakov	Male	JHsmRfemzOs
3	Barney	Lycett	blycett2@	Male	kKbBMptY
4	Leonhard	Jeger	ljeger3@l	Male	wkMFFGKbC
5	Lyon	Eby	leby4@to	Male	dGm1ORKRn
6	Roy	Hourican	rhourican!	Male	vJ4lBre7
7	Lora	Goodin	lgoodin6@	Female	zTtYpd0
8	Laurel	Mynott	lmynott7(	Female	RyPmbLuFg9R
9	Constanti	de Merida	cdemerid:	Female	Vcm6Wyz
10	Barth	Farloe	bfarloe9@	Agender	hR3MPOBA
11	Waylan	Jannequir	wjannequ	Male	ybVRQZxsE9e
12	Griffie	Yacobsohı	gyacobsol	Male	muxXuyRR
13	Philip	Januszew:	pjanuszev	Male	auBgu0SqVL
					w9jTR42il0
					yX9cFVCwej0S
					AnYxfL3i6n
					7ril8DQFyJ
					pwm9sCdGwnaA
					LMnOc2f1N5jS

Рис. 1. Файли з іменами і електронною поштою та файл із паролями

```
insert into cards (cards) values ('5450796528507145');
insert into cards (cards) values ('5002351932548229');
insert into cards (cards) values ('4041597751858');
insert into cards (cards) values ('4041592961615662');
insert into cards (cards) values ('4017952038424318');
insert into cards (cards) values ('5010124843190427');
insert into cards (cards) values ('5360616740043233');
insert into cards (cards) values ('5100144140585617');
insert into cards (cards) values ('4041378151070');
insert into cards (cards) values ('5320261260534102');
insert into cards (cards) values ('4041377634225');
insert into cards (cards) values ('5399906798073793');
insert into cards (cards) values ('5147830485646015');
insert into cards (cards) values ('4041592127477551');
insert into cards (cards) values ('4702754936046650');
insert into cards (cards) values ('4357613747162999');
insert into cards (cards) values ('4041376129575095');
```

Рис. 2. Файл з базою даних банківських карток

```
insert into health (names, snames, he) values ('Shelley', 'Civitillo', 'S92351K');
insert into health (names, snames, he) values ('Gretna', 'Everett', 'V559XXA');
insert into health (names, snames, he) values ('Annemarie', 'Baguley', 'S5002XS');
insert into health (names, snames, he) values ('Sylvia', 'Collicott', 'C4A11');
insert into health (names, snames, he) values ('Isa', 'Fylan', 'W12XXXA');
insert into health (names, snames, he) values ('Torrance', 'Mooring', 'S82152M');
insert into health (names, snames, he) values ('Bald', 'Bernocchi', 'S82499C');
insert into health (names, snames, he) values ('Morganne', 'Dreigher', 'Z1384');
insert into health (names, snames, he) values ('Corenda', 'Benoit', 'M8717');
insert into health (names, snames, he) values ('Dionisio', 'Mariotte', 'S72141G');
insert into health (names, snames, he) values ('Coralyn', 'Hopfer', 'M94221');
insert into health (names, snames, he) values ('Noam', 'Farries', 'S02612D');
insert into health (names, snames, he) values ('Katherina', 'Schinetti', 'S86929');
```

Рис. 3. Файл з базою даних медичного страхування

Аналіз вразливих даних з об'єктів. Для кожного завдання у зведеній таблиці відображається підсумкова інформація що включає: поточний статус роботи; чи виконується робота на плановій, періодичній основі; і чи аналізує завдання певну кількість сегментів S3, чи воно аналізує сегменти S3, які відповідають певному критерію. Якщо вибираєте завдання в таблиці, на панелі деталей відображаються параметри конфігурації та інша інформація про роботу (рис. 4).

У розділі «Безпека даних» інформаційної панелі наведено статистичні дані, які можуть допомогти вам визначити та дослідити потенційні ризики для безпеки та конфіденційності даних Amazon S3 у поточному регіоні AWS. Наприклад, ви можете використовувати ці дані для ідентифікації сегментів S3, які є загальнодоступними або доступними іншим. З результатів (рис. 5) випливає, що Macie знайшов вразливі дані та відмітив їх відповідними рівнями небезпеки. Також зі звіту видно, що у одного зі сховищ є вразливості у політиці безпеки та у шифруванні, що створює ще одне джерело вразливості.

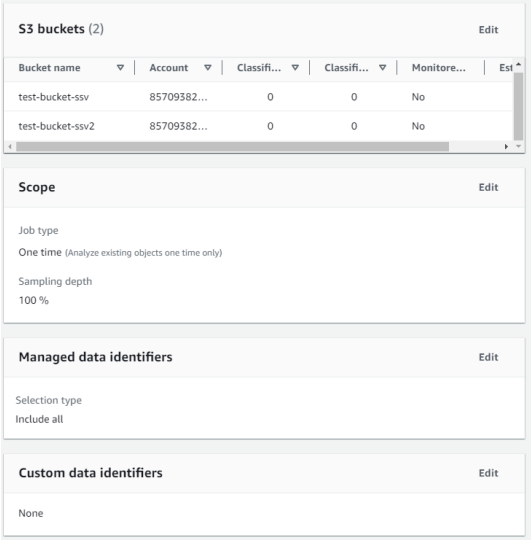

Рис. 4. Створення завдання для сканування сховища

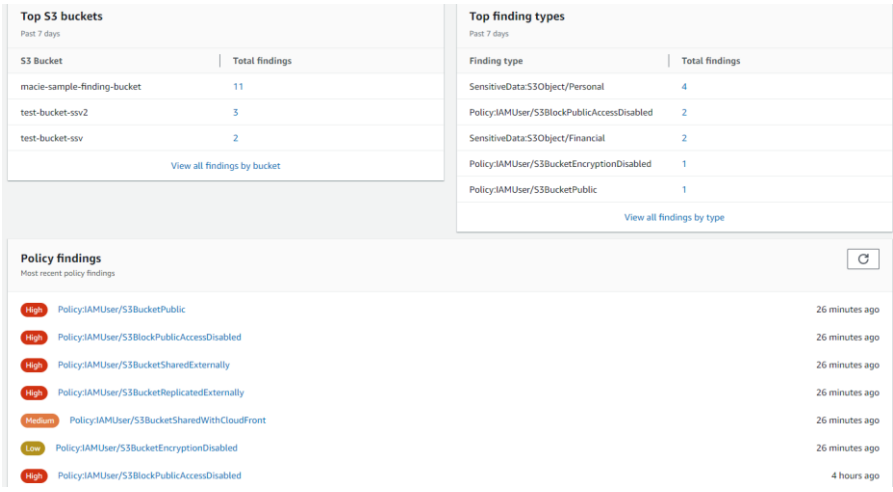

Рис. 5. Результати сканування

Також Macie генерує JSON файл, який містить усю деталізовану інформацію щодо знайдених вразливостей та наявності чутливих даних у наших файлах (рис. 6). Інформація представлена в такому вигляді є зручною для подальшого аналізу та класифікації.

ISSN 2663 - 4023

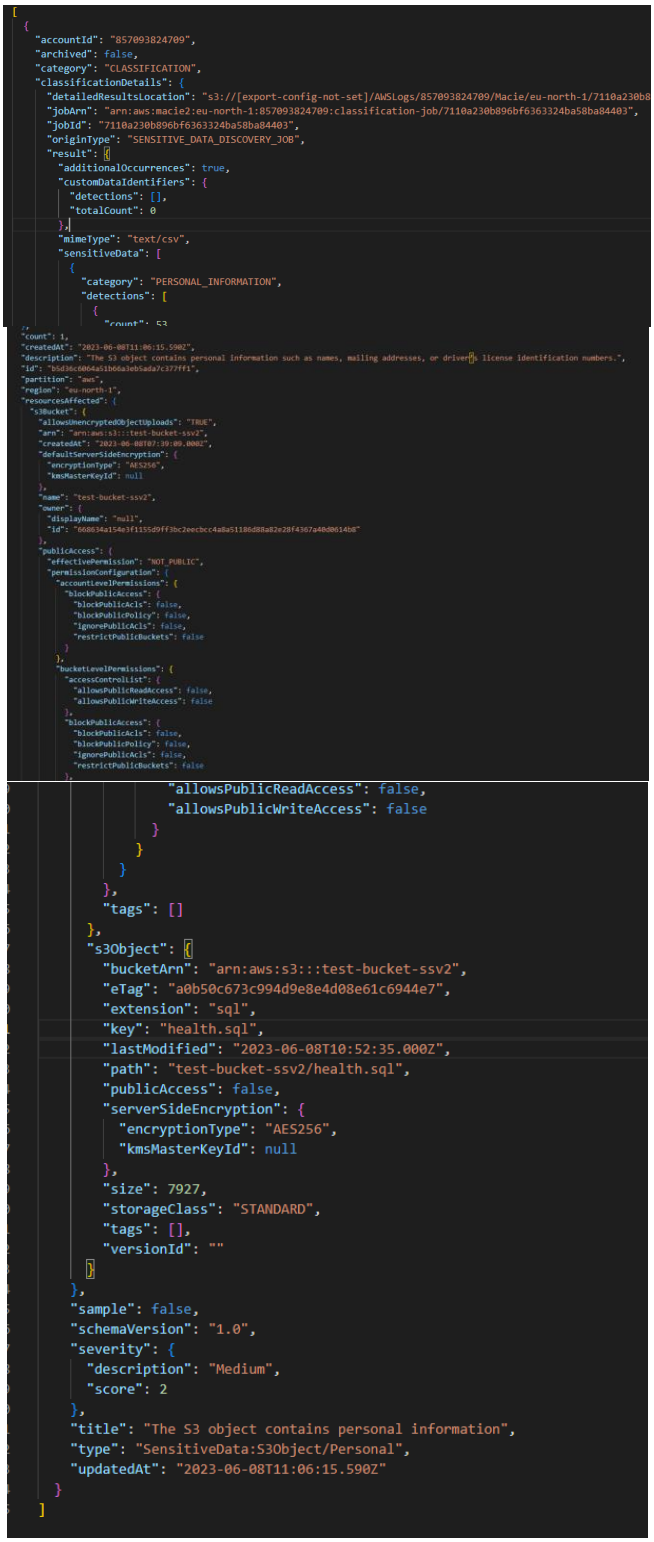

Рис. 6 Приклад згенерованого JSON файлу

Зважаючи на отримані результати варто відмітити, що рекомендується ніколи не розміщувати конфіденційну або вразливу інформацію, таку як адреси електронної пошти у теги або текстові поля довільної форми, такі як поле імені. Це потрібно враховувати також коли, ви працюєте з Macie або іншими службами AWS за допомогою консолі, API,

AWS CLI або AWS SDK. Для надання URL-адресу зовнішнього сервера, рекомендується не включати облікові дані інформацію в URL-адресі для підтвердження запиту до цього сервера.

Окрім того отримані результати можна інтегрувати AWS Security Hub, який в свою чергу надає повне уявлення про безпеку у середовищі AWS і допомагає перевірити середовище на відповідність галузевим стандартам безпеки і застосувати найкращі практики. Він робить це шляхом агрегування, організації та пріоритетності результатів із кількох служб AWS і підтримуваних рішень безпеки AWS Partner Network. Центр безпеки допомагає аналізувати тенденції безпеки та визначати проблеми безпеки з найвищим пріоритетом. З центром безпеки ви також можете об'єднати результати з кількох регіонів AWS (у випадку якщо ви агрегуєте дані з кількох сховищ S3), а потім контролювати та обробляти всі дані, які Macie публікує як результати в Security Hub.

Таким чином, ви можете використовувати Security Hub для моніторингу та обробки політик та конфіденційних даних як частину більшого зведеного набору даних про результати для вашого середовища AWS. Тоді можна проаналізувати висновки Macie, виконуючи ширший аналіз стану безпеки вашої організації, а також у разі необхідності виправити потенційні вразливості та проблеми безпеки.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У цій роботі було проведено дослідження та аналіз можливостей сервісу Amazon Macie у виявленні та захисті конфіденційних даних у різних системах зберігання даних, зокрема у хмарних сервісах AWS. В процесі дослідження було встановлено, що Amazon Macie є потужним інструментом, який використовує розширені алгоритми машинного навчання для автоматизованого аналізу даних та виявлення потенційних загроз для безпеки конфіденційних даних. Він забезпечує ефективний механізм ідентифікації чутливої інформації, такої як персональні дані, фінансові відомості чи інтелектуальна власність. Експериментальне використання Amazon Macie підтвердило його високу точність та надійність виявлення конфіденційних даних та потенційних загроз. При цьому були виявлені певні обмеження, пов'язані з типами даних та конфігурацією сервісу. На основі результатів дослідження були розроблені рекомендації щодо вдосконалення захисту конфіденційних даних у різних сервісах та системах зберігання даних. Зокрема, рекомендується регулярно оновлювати та налаштовувати правила виявлення конфіденційних даних у відповідності до специфіки організації, а також постійно оновлювати базу знань Amazon Macie з метою виявлення нових видів загроз та атак. У підсумку, використання технології машинного навчання сервісу Amazon Macie показує великий потенціал у виявленні, аналізі та захисті конфіденційних даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Galvez, R., & Gurses, S. (2018). The Odyssey: Modeling Privacy Threats in a Brave New World, *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW),* 87–94. https://doi.org/10.1109/EuroSPW.2018.00018
2. Mishra, A., et al. (2023). Artificial Intelligence based Security Solution for Data Encryption using AES Algorithm, *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS),* 1685–1690, https://doi.org/10.1109/ICSCDS56580.2023.10104702

3. Devi, S., & Bharti, T. (2021). Study of Architecture and Issues in Services of Cloud Computing. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 1578–1581. https://doi.org/10.1109/ICAC3N53548.2021.9725679
4. Alaqra, A., Kane, B., & Fischer-Hübner, S. (2021). Machine Learning–Based Analysis of Encrypted Medical Data in the Cloud: Qualitative Study of Expert Stakeholders' Perspectives. *JMIR Hum Factors, 8(3):e21810*. https://doi.org/10.2196/21810
5. Xu, G., et al. (2019). Sensitive Information Topics-Based Sentiment Analysis Method for Big Data, *IEEE Access, 7*, 96177–96190. https://doi.org/10.1109/ACCESS.2019.2927360
6. Xu, G., et al. (2019). Data Security Issues in Deep Learning: Attacks, Countermeasures, and Opportunities, *IEEE Communications Magazine, 57(11),* 116–122. https://doi.org/10.1109/MCOM.001.1900091
7. Butt, U., et al. (2020). Review of Machine Learning Algorithms for Cloud Computing Security. *Electronics, 9(9):1379*. https://doi.org/10.3390/electronics9091379
8. *Amazon Macie Documentation*. (n.d.). https://docs.aws.amazon.com/macie/
9. Kudrati, A., Peiris, C., Pillai, B. (2022). *Hunting in AWS, in Threat Hunting in the Cloud: Defending AWS, Azure and Other Cloud Platforms Against Cyberattacks*. Wiley.
10. *Monitoring data security and privacy with Amazon Macie*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/monitoring-s3.html
11. Iman, S., Sarah, B., &, Hassan, H. (2019). Security and Privacy of AWS S3 and Azure Blob Storage Services. *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 388–394. https://doi.org/10.1109/CCOMS.2019.8821735
12. Blohm, M., et al. (2019). Towards a Privacy Compliant Cloud Architecture for Natural Language Processing Platforms. *ICEIS*, 454–461. https://doi.org/10.5220/0007746204540461
13. Bermudez, I., et al. (2013). Exploring the Cloud from Passive Measurements: the Amazon AWS Case. *Proceedings - IEEE INFOCOM*, 230–234. https://doi.org/10.1109/INFCOM.2013.6566769
14. Shevchuk, D., et al. (2023). Designing Secured Services for Authentication, Authorization, and Accounting of Users. *Cybersecurity Providing in Information and Telecommunication Systems II 2023, 3550*, 217–225.
15. *Investigating sensitive data with Amazon Macie findings*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/findings-investigate-sd.html
16. *Locating sensitive data with Amazon Macie findings*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/findings-locate-sd.html

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023

Andrii Partyka
PhD, Senior lecturer department of information security
Lviv Polytechnic National University, Lviv, Ukraine
ORCID 0000-0003-3037-8373
andrii.i.partyka@lpnu.ua

Olha Mykhaylova
PhD, Associate professor department of information security
Lviv Polytechnic National University, Lviv, Ukraine
ORCID 0000-0002-3086-3160
olha.o.mykhailova@lpnu.ua

Stanislav Shpak
Student department of information security
Lviv Polytechnic National University, Lviv, Ukraine
ORCID 0009-0008-0256-4850
stanislav.shpak.mkbas.2023@lpnu.ua

DETECTION, ANALYSIS AND PROTECTION OF CONFIDENTIAL DATA USING AMAZON MACIE MACHINE LEARNING TECHNOLOGY

Abstract. Over the past decades, the field of data storage and processing has undergone significant changes and expansion, especially with the advent of cloud technologies and computing. Cloud services enable organizations to store and access large amounts of data through distributed systems. However, along with these new opportunities come new challenges, particularly in the area of protecting confidential data. Protecting sensitive data is an extremely important task for today's organizations, especially in the face of a growing number of digital threats and security breaches. In order to ensure reliable protection of valuable and sensitive information, developers and researchers are actively working on the development of new technologies and tools. One of the powerful tools used to identify, analyze and protect confidential data is the machine learning technology of the Amazon Macie service. Amazon Macie is an AWS cloud computing service that uses artificial intelligence and machine learning algorithms to automate data analysis and identify potential data security threats. The main purpose of this work is the detection, analysis and protection of confidential data using Amazon Macie machine learning technology. Amazon Macie is an innovative service developed by Amazon Web Services (AWS) that uses advanced machine learning algorithms for automated discovery and analysis of sensitive data. As part of the work, an analysis of the main machine learning algorithms, principles of data storage systems and methods of protecting confidential information was carried out. The working principles and capabilities of Amazon Macie, which uses advanced machine learning algorithms for automated data analysis and detection of potential threats to data security, were investigated.

Keywords: machine learning; Amazon Macie; cyber security; automated analysis; confidential data.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Galvez, R., & Gurses, S. (2018). The Odyssey: Modeling Privacy Threats in a Brave New World, *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 87–94. https://doi.org/10.1109/EuroSPW.2018.00018
2. Mishra, A., et al. (2023). Artificial Intelligence based Security Solution for Data Encryption using AES Algorithm, *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 1685–1690, https://doi.org/10.1109/ICSCDS56580.2023.10104702
3. Devi, S., & Bharti, T. (2021). Study of Architecture and Issues in Services of Cloud Computing. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 1578–1581. https://doi.org/10.1109/ICAC3N53548.2021.9725679

4. Alaqra, A., Kane, B., & Fischer-Hübner, S. (2021). Machine Learning–Based Analysis of Encrypted Medical Data in the Cloud: Qualitative Study of Expert Stakeholders' Perspectives. *JMIR Hum Factors, 8(3):e21810*. https://doi.org/10.2196/21810
5. Xu, G., et al. (2019). Sensitive Information Topics-Based Sentiment Analysis Method for Big Data, *IEEE Access, 7*, 96177–96190. https://doi.org/10.1109/ACCESS.2019.2927360
6. Xu, G., et al. (2019). Data Security Issues in Deep Learning: Attacks, Countermeasures, and Opportunities, *IEEE Communications Magazine, 57(11)*, 116–122. https://doi.org/10.1109/MCOM.001.1900091
7. Butt, U., et al. (2020). Review of Machine Learning Algorithms for Cloud Computing Security. *Electronics, 9(9):1379*. https://doi.org/10.3390/electronics9091379
8. *Amazon Macie Documentation*. (n.d.). https://docs.aws.amazon.com/macie/
9. Kudrati, A., Peiris, C., Pillai, B. (2022). *Hunting in AWS, in Threat Hunting in the Cloud: Defending AWS, Azure and Other Cloud Platforms Against Cyberattacks*. Wiley.
10. *Monitoring data security and privacy with Amazon Macie*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/monitoring-s3.html
11. Iman, S., Sarah, B., &, Hassan, H. (2019). Security and Privacy of AWS S3 and Azure Blob Storage Services. *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 388–394. https://doi.org/10.1109/CCOMS.2019.8821735
12. Blohm, M., et al. (2019). Towards a Privacy Compliant Cloud Architecture for Natural Language Processing Platforms. *ICEIS*, 454–461. https://doi.org/10.5220/0007746204540461
13. Bermudez, I., et al. (2013). Exploring the Cloud from Passive Measurements: the Amazon AWS Case. *Proceedings - IEEE INFOCOM*, 230–234. https://doi.org/10.1109/INFCOM.2013.6566769
14. Shevchuk, D., et al. (2023). Designing Secured Services for Authentication, Authorization, and Accounting of Users. *Cybersecurity Providing in Information and Telecommunication Systems II 2023, 3550*, 217–225.
15. *Investigating sensitive data with Amazon Macie findings*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/findings-investigate-sd.html
16. *Locating sensitive data with Amazon Macie findings*. (n.d.). https://docs.aws.amazon.com/macie/latest/user/findings-locate-sd.html