МАГІСТЕРСЬКА РОБОТА

МР. ІПм - 54.00.00.000 ПЗ

Група ІПм-22-4

Васильків Микола

2024

Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

Васильків Микола Романович

(прізвище, ім'я, по батькові)

УДК 004.942

(індекс)

МАГІСТЕРСЬКА РОБОТА

Криптографічні засоби реалізації концепцій безпеки даних в хмарних рішеннях

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Васильків М.Р.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Тимків Дмитро Федорович, д.т.н., професор

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

В.о. завідувача кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Рецензент

доц.

(посада) (підпис) (дата) (ініціали та прізвище)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Івано-Франківськ – 2024

Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
В.о. зав. кафедрою **ІПЗ**
доц. В.В. Бандура
“ 04 ” вересня 2023 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Васильківу Миколі Романовичу

(прізвище, ім’я, по-батькові)

1. Тема магістерської роботи “ Криптографічні засоби реалізації концепцій безпеки даних в хмарних рішеннях ”

керівник проекту (роботи) Тимків Дмитро Федорович, д.т.н., професор

затверджені наказом закладу вищої освіти від “ ____ ” листопада 2023 р. № ____

2. Строк подання студентом проекту (роботи) 15 січня 2024 р.

3. Вихідні дані до проекту (роботи) Теоретичні концепції та формальні моделі побудови та функціонування інформаційних та програмних технологій безпеки даних

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Системний аналіз предметної області виявлення атак та криптографічного захисту даних
2. Моделі та методи захисту великих даних в хмарних ресурсах
3. Реалізація інформаційної технології захисту на хмарних платформах
4. Розробка принципу розміщення великих об’ємів даних на хмарних платформах

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Класифікаційні ознаки систем виявлення і запобігання атак (рис. 1.1)
2. Загальна структура системи вивлення вторгнень (рис. 1.2)
3. Інформаційна структура великих даних (рис. 1.5)
4. Багаторівнева структура шифрування (рис. 1.7)
5. Архітектура гібридної хмари (рис. 2.1)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Нормоконтроль	доц., к.т.н. Вовк Р.Б.	
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2023 р.

Керівник _____
(підпис)

Завдання прийняв до виконання_____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	01.10.2023	виконано
2	Системний аналіз предметної області виявлення атак та криптографічного захисту даних	25.10.2023	виконано
3	Моделі та методи захисту великих даних в хмарних ресурсах	10.11.2023	виконано
4	Реалізація інформаційної технології захисту на хмарних платформах	22.11.2023	виконано
5	Розробка принципу розміщення великих об’ємів даних на хмарних платформах	01.12.2023	виконано
6	Реалізація функціональності запропонованої інформаційної технології	15.12.2023	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.01.2024	виконано

Студент – магістр _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

Магістерська робота: 84 с., 40 рис., 7 табл., 49 джерел.

Тема: Криптографічні засоби реалізації концепцій безпеки даних в хмарних рішеннях.

Об’єкт дослідження: моделі та методології безпеки даних великих об’ємів.

Мета роботи: дослідження та обґрунтування застосування методів захисту даних і методики багаторівневого розміщення даних великих обсягів на хмарних платформах різних моделей розгортання.

Предмет дослідження: методи та моделі забезпечення криптографічного захисту даних великих об’ємів з використанням хмарних сервісів.

Результати дослідження:

В роботі здійснено аналітичний огляд криптографічних методів захисту даних та запропоновано методику багаторівневого розміщення даних великих об’ємів на хмарних платформах, що функціонують на різних моделях розгортання.

Висновок

Виконано побудову архітектури та алгоритмічного забезпечення інформаційної технології, яка дозволить захистити дані великих об’ємів різних компаній використовуючи хмарні рішення різних моделей розгортання.

ВЕЛИКІ ДАНІ, ХМАРНІ СХОВИЩА, ІНФОРМАЦІЙНА БЕЗПЕКА, ШИФРУВАННЯ, БЕЗПЕКА ДАНИХ, ЗАХИСТ ДАНИХ, КРИПТОГРАФІЯ, ХМАРНІ СЕРВІСИ

ABSTRACT

Master Thesis: 84 pp., 40 fig., 7 tab., 49 sources.

Thesis Subject: Cryptographic means of implementing data security concepts in cloud solutions

Object of research: models and methodologies of security of large volumes of data.

Research goal: research and justification of the application of data protection methods and methods of multi-level placement of large volumes of data on cloud platforms of various deployment models.

Subject of research: methods and models of ensuring cryptographic protection of large volumes of data using cloud services.

The results:

In the work, an analytical review of cryptographic methods of data protection is carried out and a method of multi-level placement of large volumes of data on cloud platforms operating on different deployment models is proposed.

Conclusion

The construction of the architecture and algorithmic provision of information technology, which will allow protecting the data of large volumes of various companies using cloud solutions of various deployment models, has been completed.

BIG DATA, CLOUD STORAGE, INFORMATION SECURITY, ENCRYPTION, DATA SECURITY, DATA PROTECTION, CRYPTOGRAPHY, CLOUD SERVICES

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

IAM - Identity Access Management

ACL - Access control list

S3 - Simple Storage Service

SSE - Storage Service Encryption

RBAC - Role-based access control

NIST - National Institute of Standards and Technology

ECC - Elliptic-curve cryptography

NSA - National Security Agency

QKD - Quantum key distribution

SDK - Software development kit

GCP - Google Cloud Platform

SLA - Service Level Agreement

KMS - Key Management Service

SSO - Single Sign-on

DDR - Digital rights management

WAF - Web application firewall

HSM - Hardware Security Module

ВСТУП

Актуальність теми.

У сучасному цифровому світі, коли хмарні сервіси набули широкого поширення і стали важливою складовою багатьох організаційних інфраструктур, захист від кібератак на такі сервіси стає надзвичайно важливим завданням. Хмарні сервіси AWS та Azure є одними з найпопулярніших на ринку і використовуються мільйонами користувачів по всьому світу. Однак, зростання популярності цих сервісів також приводить до збільшення кількості кібератак, спрямованих на них.

Розвиток обчислювальних засобів та інформаційних технологій призводить до автоматизації різних процесів практично у всіх сферах життя суспільства: збільшуються обчислювальні потужності комп'ютерних засобів, удосконалюються технології мережевої взаємодії, змінюються формати і вимоги до побудови інформаційних систем. Слідом за розвитком інформаційних технологій з не меншою швидкістю з'являються нові загрози інформаційній безпеці, тому проблема захисту інформації залишається ключовим напрямком наукових досліджень.

За останні роки було створено низку шкідливих засобів, що використовують принципово нові методи і підходи, які дозволяють традиційних засобів захисту виявляти і адекватно реагувати на такі загрози. Прикладом таких засобів є поліморфні віруси, що не дозволяють виявити себе за допомогою сигнатурних антивірусів або можуть бути використані постійної великий обчислювального навантаження засобів захисту, або руткіт, що використовують апаратну віртуалізацію, повністю контролюють будь-які дії антивірусів і навіть простих антируткітів. Ці загрози стосуються як окремо взятих призначених для користувача або серверних комп'ютерів, так і мережевої безпеки. Для виявлення деяких видів сучасних мережевих атак необхідно зберігати великий обсяг сигнатур і використовувати безліч додаткових обчислень для контролю

трафіку. Останніми роками спостерігається тенденція до об'єднання обчислювальних ресурсів в розподілені обчислювальні мережі. Принципи обробки даних в розподілених обчислювальних мережах мають суттєві відмінності від роботи простих електронно-обчислювальних машин, що стосується і різних аспектів захисту інформації. Мережеві атаки є одним з основних видів порушення інформаційної безпеки в розподілених обчислювальних мережах.

В нашу епоху одними з особливостей сучасного світу є безперервне експоненціальне збільшення обсягів інформації з великою різноманітністю (Big Data), складності пов'язані з її зберіганням, забезпеченням безпеки та аналізом даних. Дослідники передбачають збільшення загальної кількості даних збережених на хмарах, які включають в себе дані з загальнодоступних хмар, хмар соціальних мереж (Apple, Google, Microsoft, Facebook, Twitter), приватних хмар, що належать середнім і великим корпораціям і постачальників хмарних послуг. Завдяки цьому експоненціальному зростанню даних можливості - для інновацій та кіберзлочинності - незліченні, оскільки дані є будівельний блок цифрової економіки. Ситуація ускладнюється тим, що не існує загальноприйнятої методології для забезпечення необхідного рівня захисту великих даних. Незважаючи на те, що дослідження в цій області не припиняються, вони є недостатньо систематизованими. Тому питання захисту великих даних на сьогоднішній день потребує подальшого дослідження й розробки в цьому напрямку.

Мета роботи - дослідження та обґрунтування застосування методів захисту даних і методики багаторівневого розміщення даних великих обсягів на хмарних платформах різних моделей розгортання.

Об’єкт дослідження – моделі та методології безпеки даних великих об’ємів.

Предмет дослідження - методи та моделі забезпечення криптографічного захисту даних великих об’ємів з використанням хмарних сервісів.

Для досягнення мети необхідно виконати наступні **завдання**:

- дослідити існуючі на сьогоднішній день методології та рекомендації захисту даних великих об’ємів;

- провести огляд криптографічних алгоритмів та сучасної криптографії;

- проаналізувати механізми забезпечення захисту даних на різних хмарних сервісах;

- провести опис критеріїв вибору хмарних платформ різних моделей розгортання з точки зору гарантування відповідності критеріям безпеки;

- розробити методику багаторівневого розміщення даних великих обсягів з різним рівнем захисту;

- розробити структуру інформаційної технології для розбивки, розподілу та реплікації даних на хмарних сервісах;

- виконати тестування інформаційної технології.

Методи дослідження. Метод системного аналізу, методи криптографічного захисту, методи порівняльного аналізу, механізми забезпечення захисту даних на хмарних платформах.

Наукова новизна отриманих результатів полягає в проведенні аналітичного огляду застосування методики багаторівневого розміщення даних великих об’ємів на хмарних платформах, що функціонують на різних моделях розгортання.

Практичне значення магістерської роботи полягає в побудові архітектури та алгоритмічного забезпечення інформаційної технології, яка дозволить захистити дані великих об’ємів різних компаній використовуючи хмарні рішення різних моделей розгортання.

Структура магістерської роботи. Робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 84 сторінки, і містить 40 рисунків, 7 таблиць, список використаних джерел із 49 позицій.

РОЗДІЛ 1. СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИЯВЛЕННЯ АТАК ТА КРИПТОГРАФІЧНОГО ЗАХИСТУ ДАНИХ

1.1. Класифікація систем виявлення атак та мережевих вторгнень

Одним з ключових засобів захисту обчислювальних систем є системи виявлення вторгнень (СОВ, Intrusion Detection System). Система виявлення вторгнень - це програма або програмно-апаратний засіб, призначений для виявлення фактів несанкціонованого доступу в обчислювальну систему або мережу. Системи виявлення вторгнень використовуються для виявлення різних видів шкідливої активності: мережевих атак проти безлічі сервісів; атак, спрямованих на підвищення призначених для користувача привілеїв, неавторизованого доступу до важливих системних і призначеним для користувача файлів, а також дій шкідливого програмного забезпечення (комп'ютерних вірусів, троянів і черв'яків).

Традиційні СОВ працюють за схожою з більшістю антивірусних асобів сигнатурної схемою і стикаються зі схожими проблемами, що й інші засоби ахисту.

З розвитком інформаційних технологій особливо актуальною стала проблема обробки великих даних. В цьому випадку недостатньо простого статистичного аналізу, що викликає перехід до більш складного інтелектуального аналізу даних (ІАД). Основне завдання методів інтелектуального аналізу даних полягає у виявленні в даних неструктурованої інформації та подання її в наочному вигляді. Безліч параметрів для виявлення мережевих атак становить значний обсяг даних, що визначає можливість їх обробки саме цими методами ІАД.

Основним засобом захисту інформаційно-телекомунікаційних систем та мереж (ІТСМ) від інформаційно-руйнівних впливів (втручань) у вигляді кібернетичних вторгнень (КВ) є системи виявлення та/або запобігання вторгненням (СВВ/СЗВ/СВА), основна задача яких зводиться до оперативної

їх ідентифікації (встановлення відповідності між об’єктом і його ідентифікатором (унікальним атрибутом) та в ідеальному випадку ініціювання ефективного захисного сценарію щодо припинення факту порушення конфіденційності, доступності та цілісності інформаційних ресурсів, сервісів.

Практика застосування СВВ сформувала два напрямки протидії КВ: виявлення зловживань (Misuse detection) та виявлення аномалій (Anomaly detection).

Перший підхід орієнтований на виявлення лише класифікованих (відомих) вторгнень на основі підходів синтаксичного порівняння відповідності структурних (сигнатур/патернів), інваріантних та кореляційних ознак виконуваного процесу (системи) з існуючою базою відомих шаблонів. Головними недоліками такого підходу є неможливість виявлення нових модифікацій КВ чи кібернетичних атак нульового дня (0-day) та неможливість автоматичного вводу нових шаблонів, що свідчить про їх достатньо малу ефективність. Другий підхід, навпаки, зводиться до задачі виявлення невідомих КВ на основі знаходження набору ознак, який не відповідає очікуваній поведінці об’єкта (користувача/системи) - шаблони характеристик, які не задовольняють визначеному поняттю нормальної поведінки фіксуються як аномалії.

Всі розробники систем виявлення атак і організації, які використовують СВА повинні розуміти й вивчати їх класифікацію, щоб вибрати кращі рішення для систем захисту інформації. При дослідженні різних аспектів таксономії і застосуванні різних варіантів ми зможемо досягти більш високого рівня безпеки інформаційних систем.

На сьогодні системи виявлення вторгнень і атак зазвичай являють собою програмні або апаратно-програмні рішення, які автоматизують процес контролю подій, що відбуваються в інформаційній системі або мережі, а також самостійно аналізують ці події в пошуках ознак проблем безпеки. Оскільки кількість різних типів і способів організації несанкціонованих

проникнень в чужі мережі за останні роки значно збільшилася, системи виявлення атак (СВА) стали необхідним компонентом інфраструктури безпеки більшості організацій.

Розглянемо сучасний погляд на таксономію систем виявлення атак з коротким поясненням та обґрунтуванням кожної ознаки в систематиці. Щоб зробити дану класифікацію всеосяжною і повною окрім звичних ознак, таких як: середовище моніторингу, метод виявлення, архітектура, характер відповіді, принцип роботи та час реакції, були включені наступні характеристики: джерело аудиту, технологія побудови, парадигма виявлення та режим збору даних (рис. 1.1).

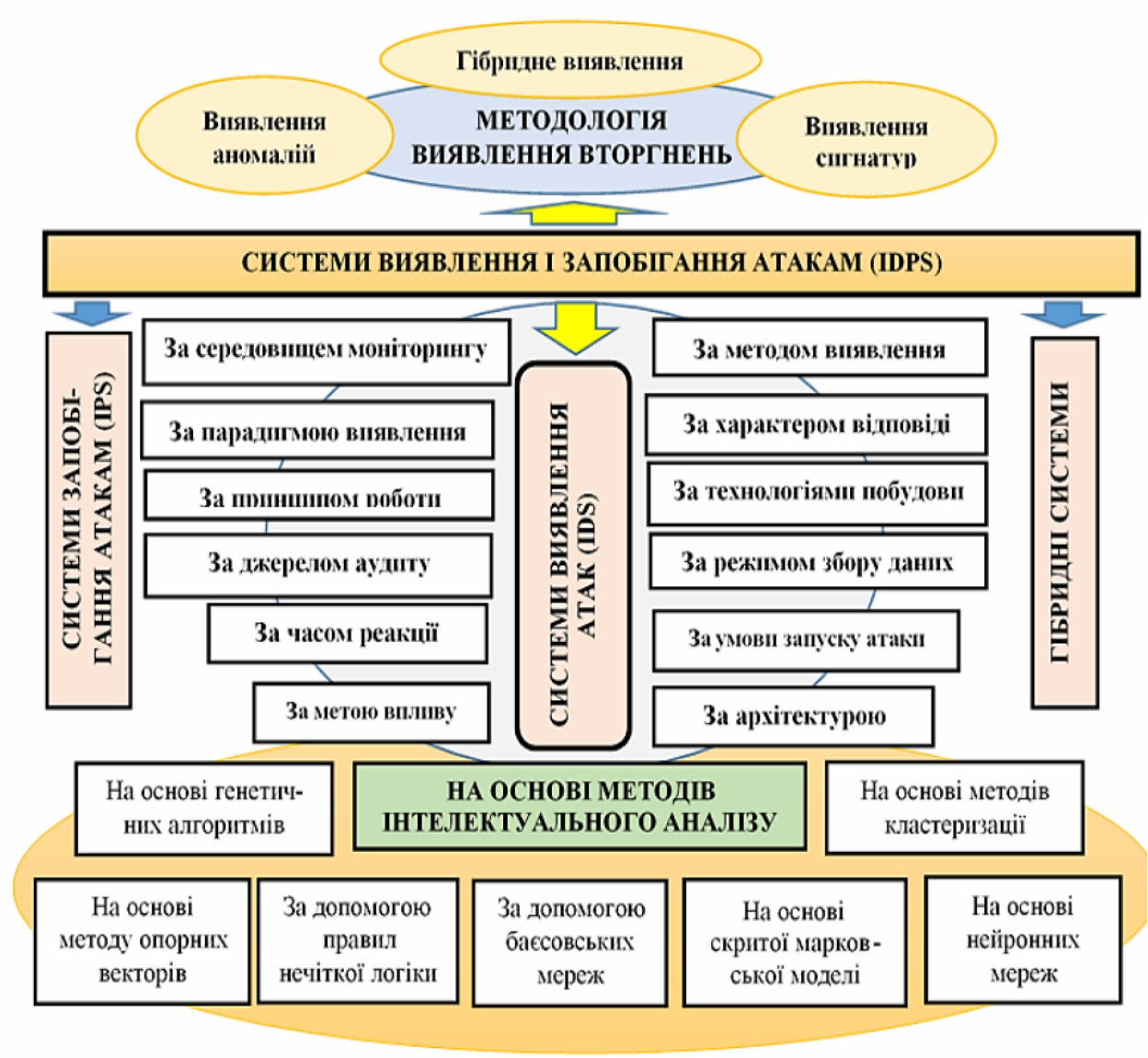


Рис. 1.1. Класифікаційні ознаки систем виявлення і запобігання атак

На сьогоднішній день можна виділити дві найбільш поширені тренувальні бази даних з відомими атаками - DARPA і KDD.

Тренувальна база даних DARPA (Defense Advanced Research Project Agency) була сформована в рамках досліджень лабораторії Лінкольна Массачусетського технологічного інституту (MIT Lincoln Laboratory) в рамках дослідження можливостей різних систем виявлення вторгнень. Під час цього дослідження використовувалися дані мережевого трафіку і відомості від файлової системи для можливості ідентифікації змодельованих вторгнень, проведених фахівцями під час запису мережевих дампов. Тренувальні дані містять як реальний потік мережевого трафіку, так і спеціально змодельований фоновий трафік. Всі атаки були спрямовані на реальні обчислювальні системи.

Після проведення дослідження роботи різних систем виявлення вторгнень, DARPA надало збережені тренувальні дані у вільному доступі. В даний час ці тренувальні бази доступні всім дослідникам, тому значна частина публікацій у науковій літературі, пов'язаних з пропозицією нових методів і підходів з виявлення мережевих атак або аномалій, спираються на ці тестові дані. Використання даної бази даних дозволяє дослідникам порівняти основні характеристики якості виявлення: ймовірності помилок пропуску (false negative) і помилкового спрацьовування (false positive).

Загальна кількість типів атак, включених в тестові дані DARPA, склало 32 атаки. З точки зору атакуючого ці атаки можна розділити на чотири категорії:

• атаки відмови в обслуговуванні (Denial of Service, DoS);

• атаки переходу від віддаленого використання до локального (Remote to Local);

• атаки отримання користувачами прав суперкористувача (User to Root);

• атаки сканування або проб (Probing/surveillance).

Інформація про атаки DARPA зберігається у вигляді текстового опису, в якому вказується час початку атаки, тривалість, адреса жертви, назва атаки, категорія атаки та інші параметри.

На відміну від тренувальних даних DARPA, база даних KDD містить не дампи мережевого трафіку, а оброблені відомості у вигляді масивів з 42 ключових значень. Дана база успішно застосовується багатьма дослідниками для аналізу застосовності різних математичних методів в завданню виявлення мережевих атак, в основному через можливість використання масивів даних з більшості програмних засобів без виконання додаткової обробки.

Найбільш ефективним способом запобігання несанкціонованому використанню інформаційних систем і мережевих ресурсів є підтримка багаторівневого захисту, коли спільно використовуються міжмережеві екрани, системи виявлення вторгнень, системи аудиту, політика безпеки і інші засоби захисту.

Найбільш загальна структура системи виявлення вторгнень, розроблена групою дослідників CIDF (Common Intrusion Detection Framework), представлена на рис. 1.2.

Рис. 1.2. Загальна структура системи вивлення вторгнень

Блок збору даних (сенсор, Event-box) - аналізує дані для обробки та прийняття рішення аналізатором. У даних можуть міститися імена контрольованих параметрів, їх особливості та значення. Сенсор може виконувати перетворення даних для перетворення в необхідний формат або для скорочення обсягу даних, що передаються.

Блок аналізу (Analyzer-box) - приймає рішення про наявність або відсутність ознак атаки або аномалії на підставі даних від сенсорів. В рамках аналізу даний блок може виконувати функції фільтрації, нормалізації, перетворення і кореляції даних. При виявленні атаки блок аналізатора може додати до вихідних даних опис виявленої атаки. Блок аналізатора може мати багаторівневу систему.

Блок бази даних (сховище даних, Database-box) - містить множини вирішальних правил і семантичний опис атак, а також накопичувальну інформацію від сенсорів. Дані можуть перебувати в текстових файлах, базі даних, і т.д.

Блок корекції (Response-box) - інформує адміністратора про зафіксовану атаку, а в випадку системи запобігання вторгнень формує активну реакцію. Системи запобігання вторгнень відстежують активність в режимі реального часу і швидко реалізовують дії щодо запобігання атак. Можливі заходи - блокування потоків трафіку в мережі, скидання з'єднань, видача сигналів оператору. Також системи запобігання вторгнень можуть виконувати дефрагментацію пакетів, упорядкування пакетів TCP для захисту від пакетів з зміненими номерами послідовності і підтвердження.

Системи виявлення мережевих атак збирають інформацію з пакетів мережевого трафіку, системних журналів і показників функціонування системи. Традиційні системи виявлення мережевих атак будуються на сигнатурному підході: за допомогою набору правил або сигнатур, що формуються експертами і розміщені в базу вирішальних правил, описуються всі можливі сценарії і особливості атак. У цього підходу існує безліч відомих недоліків. За допомогою аналізу сигнатур неможливо виявити нові види атак,

тому що база вирішальних правил не містить інформації про відповідну атаці. Процес аналізу сигнатур для розподілених атак є вкрай складним завданням. Крім того, бази вирішальних правил популярних систем виявлення вторгнень практично є загальнодоступними, тому порушник може протестувати можливості приховування атаки.

При використанні методів ІАД для виявлення мережевих атак можна виділити наступні проблеми: дані, аналізовані системами виявлення, мають високу розмірність і обсяг; вимога обробки даних в режимі реального часу; велика кількість шумів і невідповідностей в даних, що обробляються що викликають неадекватну реакцію методів інтелектуального аналізу даних.

1.2. Опис та характеристика інтелектуальних методів виявлення атак

Проаналізуємо СВА основаних на методах ІАД. Одним з таких методів є виявлення атаки за допомогою скритої марковської моделі. Скрита марковська модель представляє собою статистичну модель, де система моделюється як процес Маркова з невідомими параметрами. Задача методу полягає в оцінці скритих параметрів, що базуються на параметрах, які спостерігаються. Послідовності подій, зібрані з нормальних операційних систем, використовуються в якості навчальної вибірки для оцінки параметрів прихованої марковської моделі. Після навчання скритої марковської моделі ймовірнісні оцінки використовуються в якості порогових значень для ідентифікації мережевих аномалій в тестових даних.

Виявлення атак за допомогою **байєсовських мереж**. Байєсовська мережа являє собою модель, яка кодує імовірнісні взаємозв'язки між змінними. Основний метод застосування баєсовських мереж передбачає незалежність серед атрибутів. Кілька варіантів застосування байєсовських мереж були запропоновані для виявлення мережевих аномалій. Більшість методів направлено на формування умовних залежностей між атрибутами з

використанням складних мереж Байеса. Байєсовські методи часто використовуються в процедурі класифікації і локалізації помилкових спрацьовувань. Для виявлення вторгнень або прогнозування поведінки порушника байєсовські мережі можуть бути ефективними в деяких випадках, але в загальному випадку точність цього методу залежить від припущень, пов'язаних з поведінкою моделі цільової системи. Таким чином, будь-яке значне відхилення від припущень призведе до зменшення точності виявлення.

Виявлення атак за допомогою **методів кластеризації**. Методи кластеризації групують дані в кластери на підставі схожості об'єктів. Більшість методів кластеризації починається з вибору центральної точки для кожного кластера, а множина елементів розподіляється по кластерам. Після цього центри коригуються, а елементи перерозподіляються. Кластеризація дозволяє вивчити і виявити аномалії, не вимагаючи множини класів або типів аномалій, тобто для виявлення аномалій за допомогою методів кластеризації не виникає потреби в навчальній множині. Кластеризація досить широко застосовується для виявлення мережевих аномалій.

Виявлення невідомих мережевих атак найчастіше будується саме на методах кластеризації. Однорідні групи зі схожими характеристиками або кластери формуються шляхом розбиття набору елементів без будь-яких позначок. В системі вкрай важливо правильно визначити кластери, щоб максимально віддалити їх від викидів. Кінцева мета даних методів полягає у визначенні ступеня відхилення викидів від кластерів. За допомогою простого порівняння з пороговим значенням викиди з високим ступенем відхилення від кластерів позначаються як аномалії.

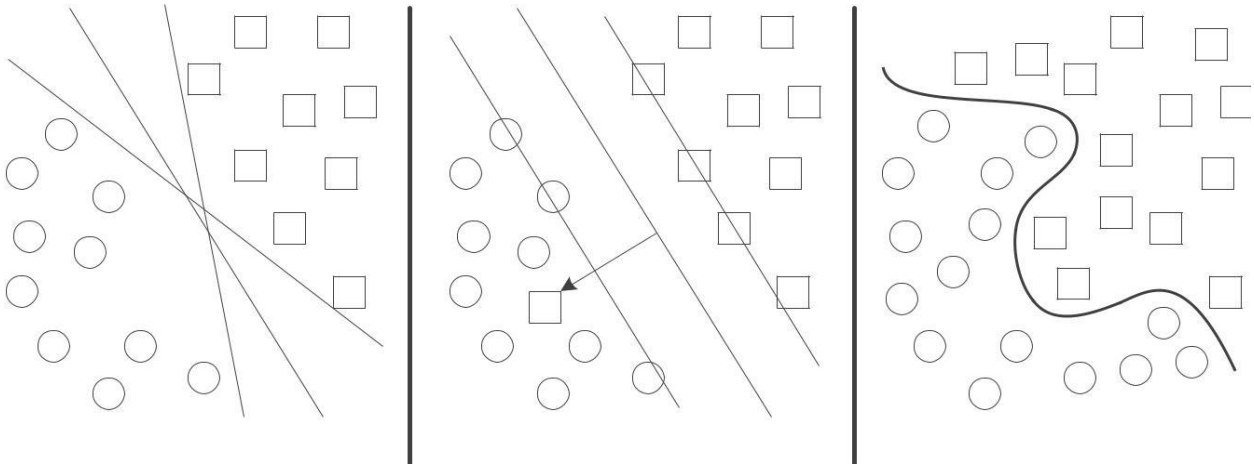
Особливу увагу заслуговує метод опорних векторів (Support Vector Machine, SVM), який представляє собою набір схожих алгоритмів категорії «навчання з учителем», застосовуваних у задачах класифікації та регресійного аналізу. Даний метод належить до сімейства лінійних класифікаторів. Характерною особливістю методу опорних векторів є

постійне скорочення емпіричної помилки класифікації і збільшення зазору між класами. Тому даний метод часто називають методом класифікатора з максимальним зазором.

Метод відшукує елементи, що знаходяться на кордонах між двома класами, які і називаються опорними векторами.

На рис. 1.3 показані різні випадки, що виникають при застосуванні SVM для двовимірних даних:

- приклади поділяють площин (а);
- розділяє площині зі штрафом (б);
- лінійна неподільність (в).



а) б) в)

Рис. 1.3. Метод опорних векторів

Метод опорних векторів здійснює пошук лінійної функції, яка дозволяє віднести елементи набору даних до одного з двох класів. Завдання бінарної класифікації може бути сформульована як пошук лінійної функції f (x), яка приймає значення менше нуля для елементів одного класу і більше нуля для елементів іншого.

Основною проблемою застосування методу опорних векторів в завданні бінарної класифікації є складність пошуку лінійної кордону між двома класами. У разі якщо таку кордон побудувати не вдається, одне з

рішень - це збільшення розмірності (перенесення даних в інший простір, більш високої розмірності), де існує можливість побудови площині, що розділяє безліч елементів на два класи.

Виявлення атак за допомогою правил нечіткої логіки. Нечіткі системи виявлення мережевих вторгнень використовують множину нечітких правил для визначення ймовірності конкретних або загальних мережевих атак. Нечітка множина може бути сформована для опису трафіку в конкретній мережі. Існує метод для побудови класифікаторів, що використовують нечіткі асоціативні правила, які застосовуються для виявлення вторгнення в мережу. Нечіткі набори правил асоціації використовуються для опису нормальних і аномальних класів. Належність запису певному класу визначається за допомогою відповідної метрики. Нечіткі асоціативні правила формуються на основі звичайних навчальних вибірок. Тестований зразок класифікується як нормальний, якщо згенерований сукупністю правил показник буде вище певного порогового значення. Зразки з більш низьким показником вважаються аномальними.

Звичайно протидіяти вторгненням і атакам основуючись тільки на одному з методів ІАД малоефективно, тому необхідно підійти до цього питання комплексно і побудувати інтелектуальну систему протидії вторгненням (рис. 1.3). При побудові такої інтелектуальної (експертної) системи пропонується вибрати нечітку модель. Це пов’язано з тим, що значна частина інформації про причини і джерела атак може бути отримана тільки експертним шляхом або у вигляді евристичних описів процесів. Для визначення джерел атак система безпеки має бути представлена моделлю тієї інформаційної мережі на яку вона орієнтується. Дана модель ділить завдання переміщення інформації між комп’ютерами через середовище мережі на кількість рівнів менш великих і легше вирішуваних підзадач. Кожна з цих підзадач вирішується за допомогою одного рівня мережі. Тому первинне завдання після фахівця безпеки може бути представлене декомпозицією завдань безпеки по окремих рівнів мережі.

Сучасний підхід до побудови систем виявлення атак на інформаційні системи сповнений недоліків і вразливостей, що дозволяють, на жаль, шкідливим впливам успішно долати системи захисту інформації. Перехід від пошуку сигнатур атак до виявлення передумов виникнення загроз інформаційної безпеки має сприяти тому, щоб докорінно змінити дану ситуацію, скоротивши дистанцію відставання в розвитку систем захисту від систем їх подолання. Крім того, такий перехід має сприяти підвищенню ефективності управління інформаційною безпекою і, нарешті, більш конкретних прикладів застосування нормативних і керівних документів, що вже стали стандартами.

Результати різних методів інтелектуального аналізу даних для виявлення вторгнень представлені в таблиці 1.1.

Метод інтелектуального аналізу даних	Показник розпізнавання (%)	Помилкове спрацьовування (%)
Метод к-найближчих сусідів	92	1
Метод опорних векторів (SVM)	95,5	1
SVM+ к-найближчих сусідів	96,3	0,84
SVM+нечітка логіка	97	0,73
SVM+ k-найближчих сусідів+ нечітка логіка	99,56	0,44

1.3. Криптографічний захист великих даних в хмарних середовищах

«Великі дані» (big data) – це серія підходів, інструментів і методів обробки, структурованих і неструктурованих даних значних обсягів для отримання результатів, що навіть не сприймаються людиною, ефективних в умовах безперервного приросту, розподілу за численними вузлами обчислювальної мережі. Сутність «великих даних» полягає в тому, що при їх обробці інформація одночасно отримується з великого обсягу, з великою швидкістю, в тому числі й з великою швидкістю приросту даних з

урахуванням одночасних – паралельних рівнів обробки, а також різноманіття даних – можливість використання різних джерел даних. Зарубіжними авторами дана концепція йменується «три V» – volume (обсяг), velocity (швидкість) variety (різноманіття).

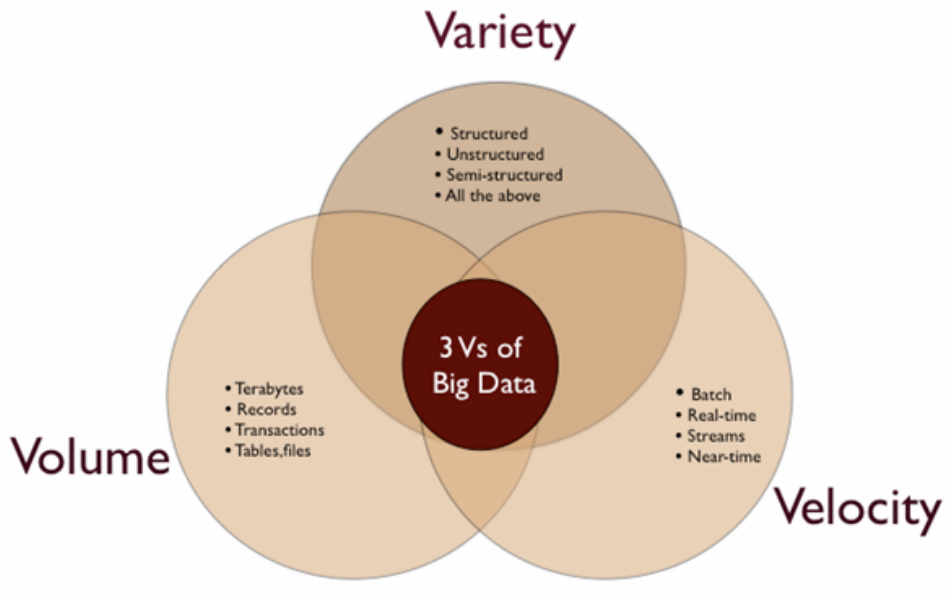


Рис. 1.4. Сутність 3V для великих даних

Під терміном Big Data прийнято розуміти будь-які об’ємні набори даних, які більшою мірою підпадають під структуровані, неструктуровані та напівструктуровані категорії даних, досить великі і складні щоб їх можна обробити традиційними засобами роботи з даними. Дані, які є найціннішим активом кожної організації та розумно використовуються для бізнесу, можуть підтримувати рішення, засновані на реальних фактах, а не на сприйнятті. В даний час значна кількість даних генерується з різних джерел, включаючи сайти соціальних мереж, різні віддалені датчики, сигнали GPS стільникового телефону, записи транзакцій та файли журналів. Завдяки Інтернету щодня виробляється терабайт структурованих, неструктурованих та напівструктурованих даних, і значна частина цієї інформації має невід’ємні бізнес-цінності. Таким чином, якщо не зафіксувати та не проаналізувати їх належним чином, значні життєво важливі дані будуть

втрачатися. Області застосування Big Data: e-commerce, телекомунікації, фінансова сфера, державне і корпоративне управління, аналітика.

Великі дані поєднують дані, представлені у різних моделях даних. Для цього повинні існувати методи їх перетворення з мінімальною втратою даних. Інформаційна структура великих даних подана на рис. 1.5.

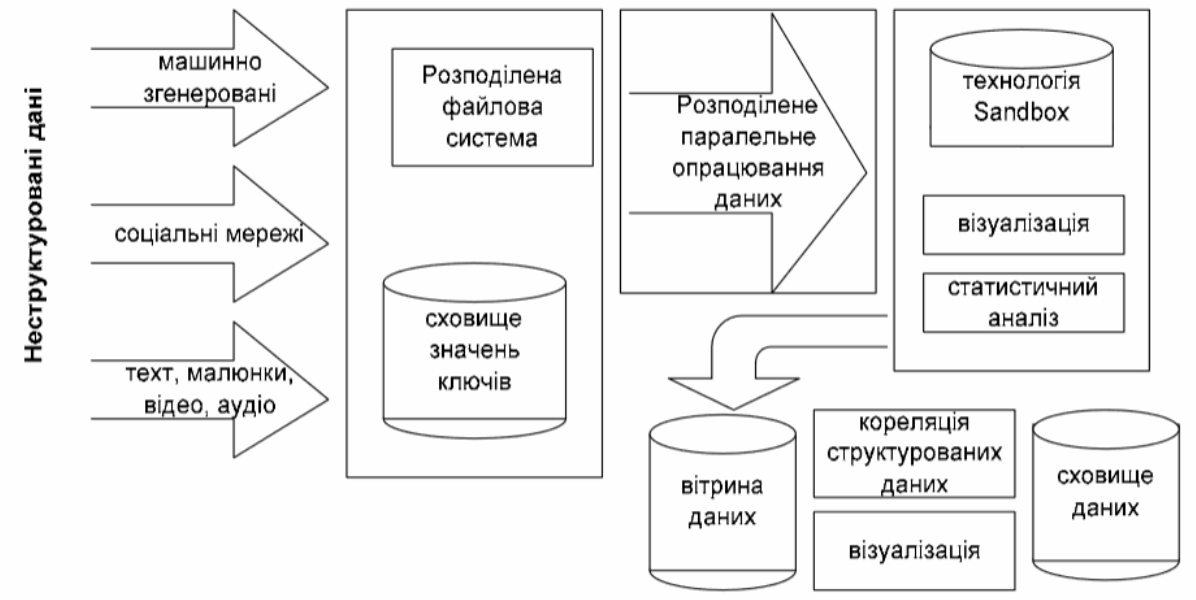


Рис. 1.5. Інформаційна структура великих даних

Окрім відомих та розповсюджених алгоритмів шифрування таких, як, наприклад, AES або RSA, на різних етапах захисту великих даних рекомендовано хешування паролів, наскрізне шифрування даних, пов'язане шифрування, шифрування на базі атрибутів (ABE), шифрування на базі ідентичності (IBE), конвергентне шифрування.

Дані повинні бути повністю зашифровані на кожному етапі обробки, зберігання або передачі. Але при цьому криптографічні процедури повинні бути швидкодіючими. Такими, що не завадять основній задачі - аналізу даних, а крім того, до них повинен бути організований ефективний та безперервний доступ. Згідно методології NIST Big Data Security and Privacy для криптографічного захисту великих даних в хмарних сховищах рекомендується використовувати одну з технологій:

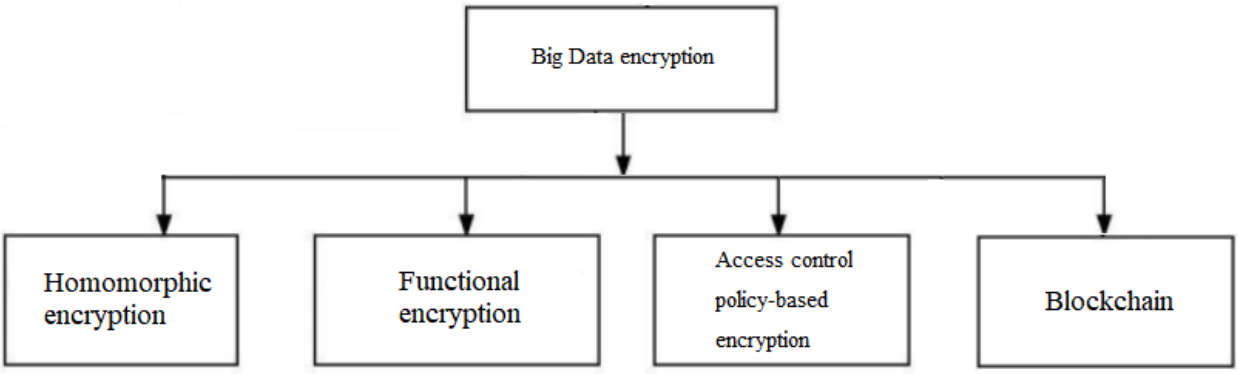


Рис. 1.6. Рекомендовані технології для шифрування великих даних за рекомендаціями NIST Big Data Security and Privacy

Також згідно методології Cloud Security Alliance Big Data Security and Privacy Handbook рекомендується використовувати fully homomorphic encryption (повне гомоморфне шифрування). Розглянемо декілька з вище наведених технологій.

1) Повне гомоморфне шифрування – модель шифрування, яка дозволяє довільне різноманіття додавання та множення, і, отже, виконує безліч видів обчислень без потреби в дешифруванні. Для поліпшення даної моделі, запропоновано наступне рішення.

Повністю гомоморфне недетерміноване шифрування - запропонована модель із використанням недетермінованої форми є формою шифрування, яка дозволяє виконувати унікальні різновиди обчислень для зашифрованих даних. Шифрування числових значень здійснюється за допомогою моделі з використанням недетермінованої схеми шифрування. Для шифрування значень потрібні два великі прості числа. Значення простих чисел має бути більшим за введені значення.

2) Функціональне шифрування – це тип шифрування з відкритим ключем, в якому власник секретного ключа може визначити функцію, за допомогою якої отримано шифр-текст. До нього можна віднести дослідження алгортиму шифрування «Калина». Шифр є прикладом блочного симетричного перетворення і підтримує розмір блоку і довжину ключа шифрування 128, 256 і 512 біт. Враховуючи різні дослідження, можна

зробити висновок що алгоритм може підходити для захисту великих даних. Були спроби створити багаторівневу структуру шифрування, яка включає в себе AES, Feistel encryption, Cross Over and Mutation та HMAC. Схема структури наведена рис. 1.7.

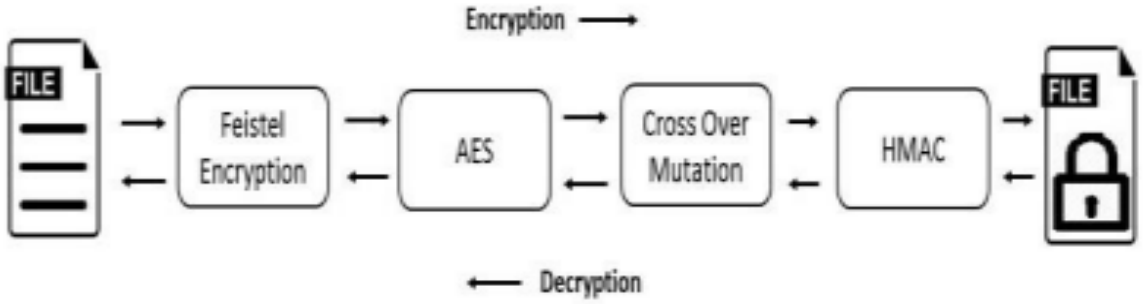


Рис. 1.7. Багаторівнева структура шифрування

Але, швидкість шифрування і дешифрування даних в такій структурі не є швидкодіюча.

3) Блокчейн (Blockchain) - це ланцюжок блоків або якщо бути точніше розподілена база даних. Вперше цей термін був застосований як назва для розподіленої бази даних криптовалюти біткоін. Blockchain – розподілена база даних, що зберігає впорядкований ланцюжок записів (так званих блоків), що постійно довшає. В даному випадку дані користувача перетворюються в блок, і вони представляються у вигляді хеш-значення, в якому кожен блок в мережі має унікальне хеш-значення, і він забезпечує захист даних, і їх не може зламати хакер, Blockchain підтримує конфіденційність даних користувача шляхом процесу автентифікації, в якому він допомагає з’ясувати, чи є користувач зловмисним чи ні. Після автентифікації дані користувача стають доступними для користувача, і це допомагає зменшити проблеми з конфіденційністю, які в основному виникають у соціальних мережах.

Також дослідження методів шифрування великих даних у хмарах може дати загальний огляд методів шифрування і експериментальні результати щодо співвідношення безпеки та ефективності.

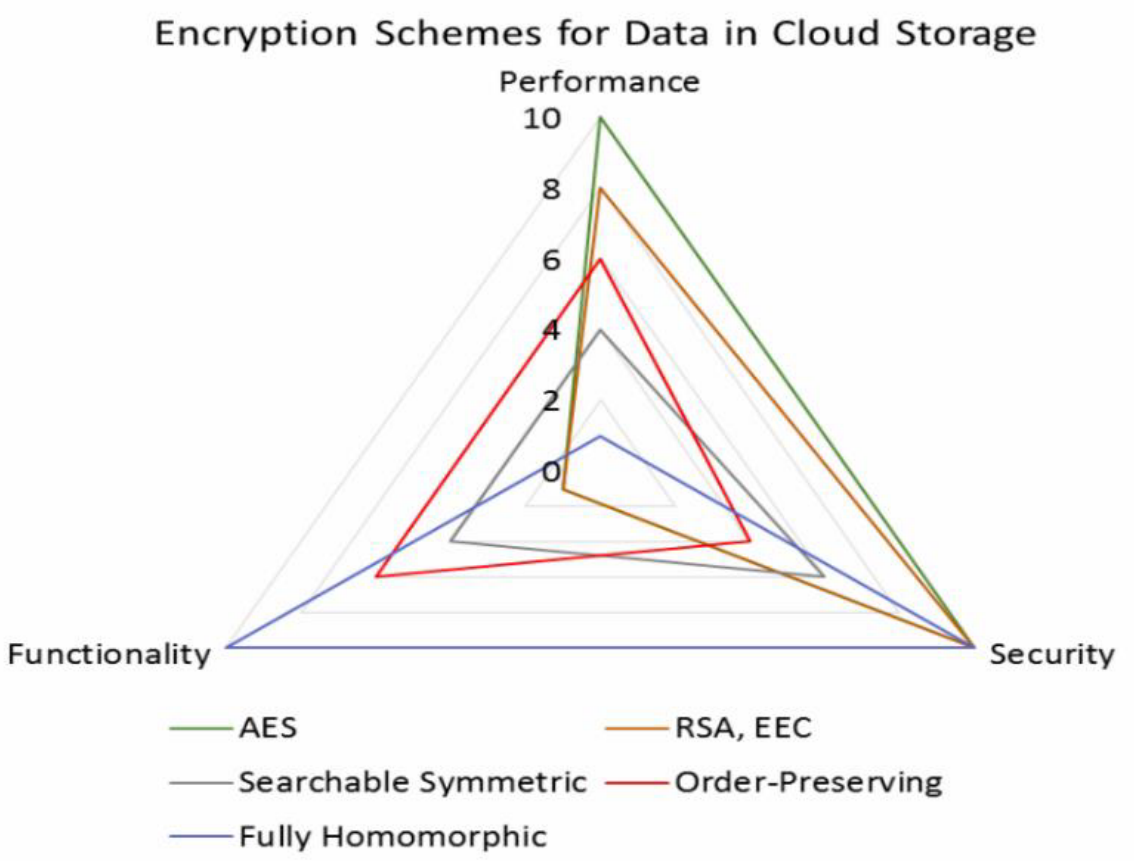


Рис. 1.8. Компроміс схем шифрування для хмарного сховища

1.4. Огляд існуючих методів шифрування

Сукупність алгоритму і множини всіх можливих вихідних текстів, шифрованих текстів і ключів називають криптосистемою. Криптосистема включає в себе криптографічні елементи, ключі, протоколи, інфраструктуру ключів.

Існує багато різних методів шифрування через різні потреби та вимоги в захисті інформації. Ось декілька причин, чому існує розмаїття методів шифрування:

– різні рівні безпеки – чим більш математично та обчислювально складніший алгоритм, тим складніше “зламати” шифрування. Різні застосування вимагають різних рівнів безпеки, і тому існують різні методи шифрування, щоб відповідати різним потребам;

– ефективність впливає на швидкість шифрування та розшифрування даних. В деяких ситуаціях важлива швидкість, тому використовуються швидкі методи шифрування, які працюють ефективно навіть на обмежених ресурсах. В інших ситуаціях пріоритет має безпека, і використовуються більш складні методи шифрування;

– ключі та управління – методи шифрування вимагають різних типів ключів та механізмів управління ними. Наприклад, симетричне шифрування використовує спільний секретний ключ, який потрібно обмінювати між комунікуючими сторонами. У асиметричному шифруванні використовуються пари ключів – публічний і приватний, що має важливе значення для безпеки та ідентифікації;

– типи даних – деякі методи шифрування оптимальні для текстових даних, тоді як інші можуть бути використані для шифрування мультимедійних файлів або мережевого трафіку;

– стандарти та сумісність – у світі криптографії існують стандарти, які встановлюються для забезпечення сумісності та взаємодії різних систем та пристроїв. Різні методи шифрування (рис. 1.9) можуть використовуватись для виконання цих стандартів і забезпечення сумісності між різними пристроями та платформами.

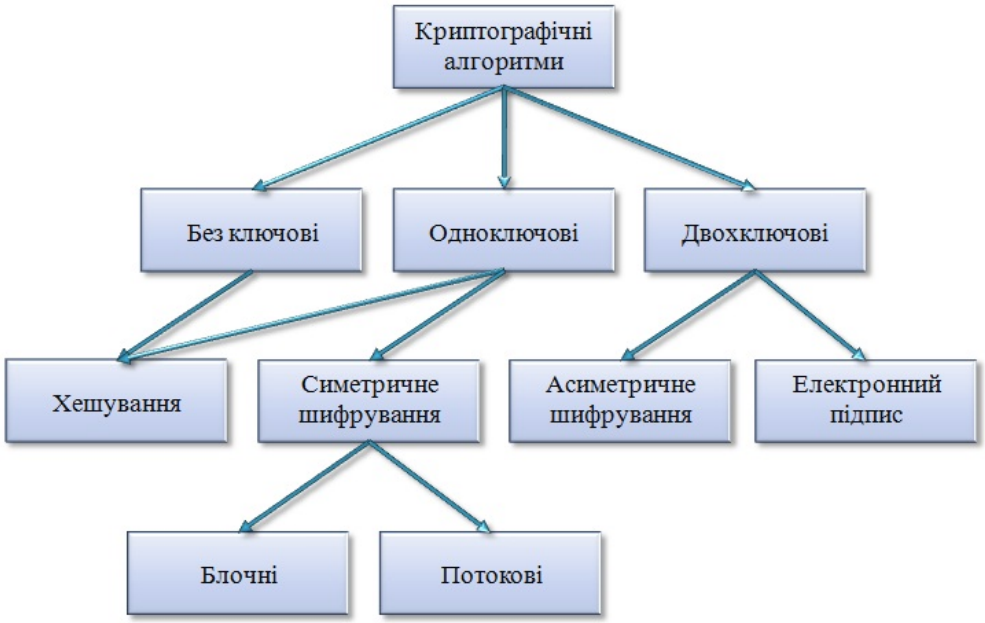


Рис. 1.9. Класифікація криптографічних алгоритмів

Отже, різноманітність методів шифрування виникає з потреб різних застосувань, вимог до безпеки та продуктивності, а також з технічних і стандартних обмежень.

Криптографічні алгоритми, засновані на ключі, поділяють на дві основні групи: симетричні (або алгоритми із секретним ключем) і асиметричні (алгоритми із відкритим ключем). Далі докладніше про особливості цих двох груп.

Симетричне шифрування – це метод шифрування, в якому для захисту інформації використовується спільний секретний ключ, як для шифрування, так і для розшифрування даних. Основний принцип симетричного шифрування полягає у використанні одного й того ж ключа для обох процесів – шифрування та розшифрування, як показано на рис. 1.10.

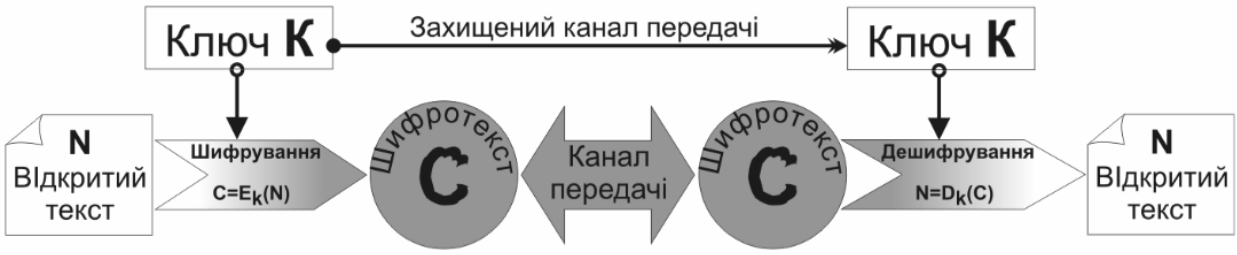


Рис. 1.10. Структура симетричної криптосистеми

У процесі симетричного шифрування, вхідні дані (повідомлення) подаються на вхід алгоритму шифрування разом з секретним ключем. Алгоритм застосовує ключ для перетворення вхідних даних в криптографічно захищену форму, яку називають шифротекстом. Потім цей шифротекст може бути переданий по незахищеному каналу передачі даних або збережений в безпечному сховищі.

При отриманні шифрованого повідомлення отримувач використовує той самий секретний ключ для виконання процесу розшифрування. Алгоритм розшифрування використовує цей ключ для відновлення вихідних даних з

шифротексту. Отримані розшифровані дані є ідентичними вхідним даним, які були використані для шифрування.

Переваги симетричного шифрування включають його швидкодію та простоту реалізації. Крім того, симетричне шифрування може бути використане для шифрування великих обсягів даних, оскільки його обчислювальні вимоги зазвичай нижчі, ніж у випадку асиметричного шифрування.

Однак, існують деякі недоліки симетричного шифрування, зокрема проблема обміну ключами між комунікуючими сторонами. Необхідність безпечного обміну секретним ключем може бути викликом, особливо якщо сторони знаходяться на віддаленій відстані. Крім того, кожна пара комунікуючих сторін повинна мати свій власний секретний ключ, що може викликати проблеми в управлінні ключами у великих системах з багатьма користувачами.

Загалом, симетричне шифрування є ефективним і широко використовуваним методом шифрування, особливо для захисту конфіденційних даних у контрольованих середовищах. Існують сотні алгоритмів симетричного типу. Найбільш поширені з них – AES, RC4, DES, 3DES, RC5, RC6 і т. д.

Асиметричне шифрування (також відоме як криптографія з відкритим ключем) – це метод шифрування, в якому використовується пара ключів: публічний ключ і приватний ключ. Цей метод шифрування відрізняється від симетричного шифрування, де використовується лише один спільний ключ.

Характеризується тим, що використовуються різні ключі для шифрування та розшифрування інформації:

– закритий ключ (en: private key) – ключ, що відомий лише своєму власнику. Власник закритого ключа єдиний, хто може розшифрувати дані, зашифровані за допомогою його публічного ключа;

– відкритий ключ (en: public key) – ключ, який можна зробити загальнодоступним з тим щоб будь-хто міг зашифровувати повідомлення для певного одержувача.

Головна властивість ключової пари: по закритому ключу легко обчислюється відкритий ключ, однак, маючи відкритий ключ практично неможливо вирахувати закритий ключ.

Класичний механізм функціонування асиметричної криптосистеми проілюстровано на рисунку 1.11.

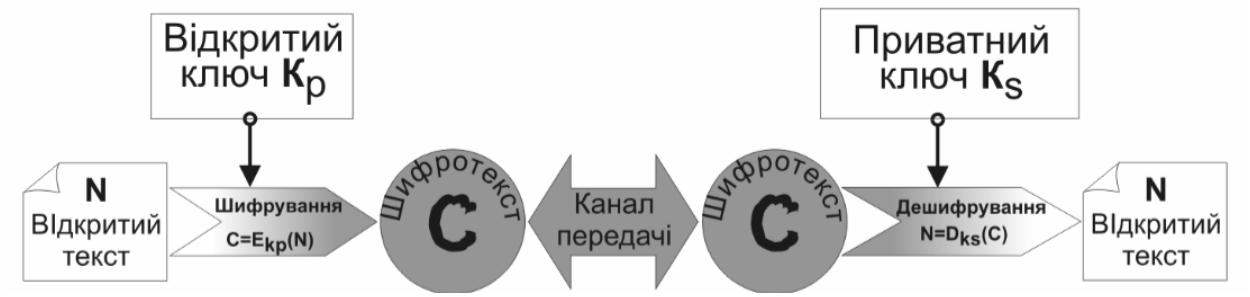


Рис. 1.11. Структура асиметричної криптосистеми

Основний принцип асиметричного шифрування полягає у наступному:

– публічний ключ використовується для шифрування повідомлень перед їх відправкою. Цей ключ може бути розповсюджений відкрито і доступний всім користувачам;

– приватний ключ зберігається в секреті і використовується для розшифрування шифрованого повідомлення. Цей ключ повинен залишатися виключно у власника ключа і не повинен розголошуватися іншим користувачам.

Коли комунікуючі сторони хочуть обмінятися зашифрованими повідомленнями, одна сторона використовує публічний ключ одержувача для шифрування повідомлення перед його відправкою. Потім відправлене зашифроване повідомлення може бути розшифроване лише за допомогою приватного ключа одержувача. Це забезпечує конфіденційність повідомлень,

оскільки тільки одержувач знає свій приватний ключ і може розшифрувати повідомлення.

Асиметричне шифрування має кілька переваг:

а) безпека – приватний ключ зберігається в секреті і не розголошується, що робить асиметричне шифрування відносно безпечним для обміну конфіденційною інформацією;

б) ідентифікація та аутентифікація – асиметричне шифрування може бути використане для підтвердження ідентичності власника приватного ключа. Приватний ключ може бути використаний для створення цифрового підпису, який може бути перевірений за допомогою відповідного публічного ключа;

в) складний обмін ключами – шифрування з відкритими ключами дозволяє уникнути проблеми обміну секретного ключа, яка виникає при симетричному шифруванні. Публічні ключі можуть бути розповсюджені відкрито, тоді як приватні ключі залишаються в таємниці.

Висновки до розділу

В даному розділі проаналізовано існуючі на сьогоднішній день провідні рішення з криптографічного захисту даних та забезпечення конфіденційності даних, отримані основні напрямки захисту даних. З’ясовано, що актуальними напрямками захисту великих даних є криптографічний захист та забезпечення конфіденційності даних.

РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДИ ЗАХИСТУ ВЕЛИКИХ ДАНИХ В ХМАРНИХ РЕСУРСАХ

2.1. Методи забезпечення конфіденційності великих даних

В рамках інформаційних технологій і сучасного інформаційного простору, моделі розгортання хмар, поділяють на три основні види: приватні; загальнодоступні (публічні); гібридні

Розглянемо проблему конфіденційності даних та зберігання даних, що надходить до гібридних хмар та необхідність забезпечення їх захисту. Гібридна хмарна архітектура (рис. 2.1) складається з двох компонентів: компонент проектування даних, який відповідає за оптимальний розподіл в гібридному хмарі, і компонент обробки запитів, який з урахуванням поділу, приймає рішення про стратегію виконання запитів.

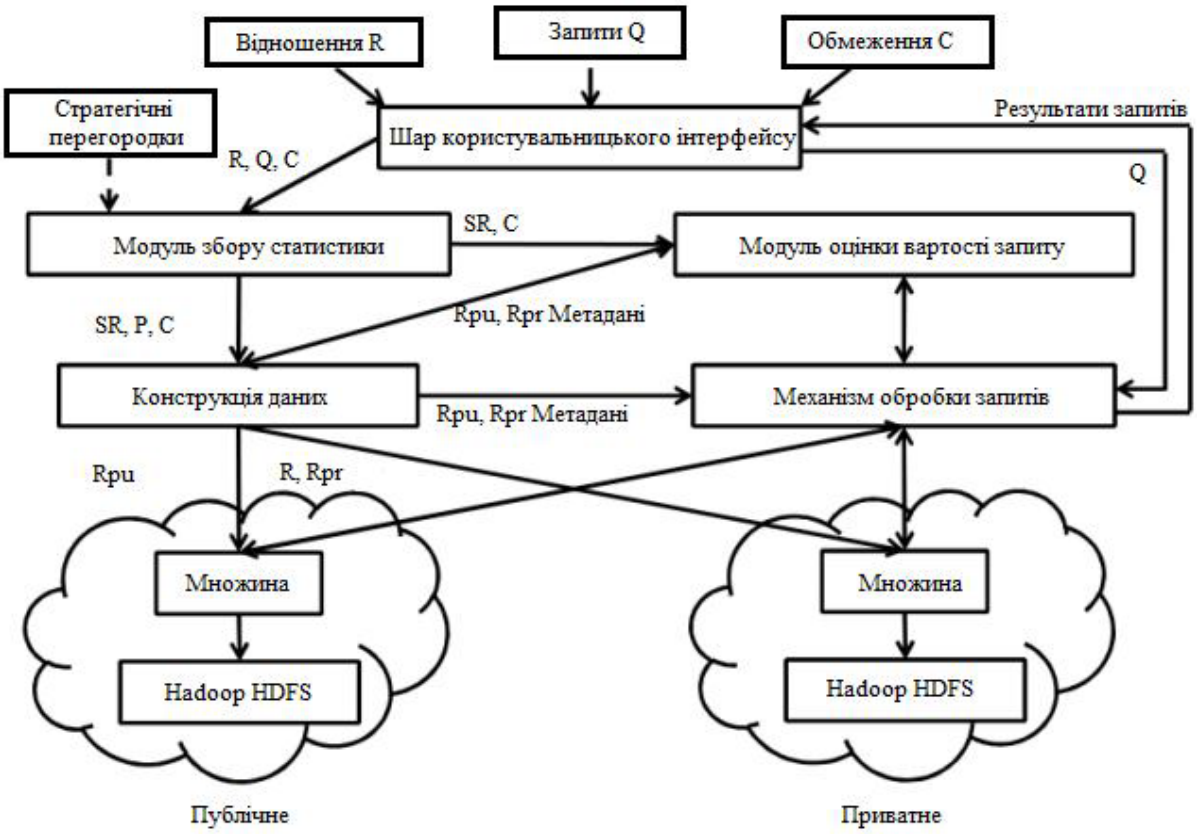


Рис. 2.1. Архітектура гібридної хмари

Оцінка витрат, необхідних для прийняття рішення про визначення оптимального розбиття залежить від стратегії обробки запитів на якій реалізований даний механізм.

Для компонента проектних даних, користувач представляє:

– сукупність відносин R;

– сукупність робочого навантаження запитуQ;

– набір розподілу ресурсів і чутливі обмеження розкриття даних, C.

Система спочатку виконує завдання збору статистики по R і Q за допомогою модуля збору статистики. Цей модуль також створює набір предикатів P, виходячи з R, Q і заданої користувачем стратегії секціонування (вертикальної або горизонтальної).

Проблема поділу даних в параметрі гібридної хмари полягає у мінімізації вартості виконання запиту навантаження і обмежена двома окремими обмеженнями, перше з яких обмежує ресурси, які можуть бути надані у публічну хмару, а друга фіксує розкриття ризику того, що користувач готовий прийняти, конфіденційні дані публічної сторони. Рішення проблеми призводить до розбиття даних між публічною та приватною сторонами. Часто пропонується виконати моделювання таких перегородок, використовуючи предикати.

У моделі поділу предиката використовуються прості предикати як фундамент, на якому можна здійснювати горизонтальний і вертикальний поділ стратегій, які лежать в основі рішень задачі розбиття даних. Використання предикатів дозволяє представити різні варіанти розбиття даних (між публічною і приватною хмарами) в рамках однієї загальної структури.

Використання даної моделі у поєднанні з обмеженнями (публічна сторона витрат і конфіденційні дані про ризики) дозволяє захопити кілька реалістичних сценаріїв у тих же рамках.

Прикладами таких сценаріїв можна назвати:

- користувачів, які не допускають зберігання конфіденційних даних у публічних хмарах, із-за законів/правил.

- користувачів, які хочуть досягти швидкості у продуктивності, і готові платити за ризик зберігання конфіденційних даних на публічній стороні. Крім того, такі загальні рамки також дозволяють вивчати різні співвідношення, які існують у предметній області на систематичній основі.

До основних видів порушень конфіденційності даних великих даних відносять:

- відстеження урядом;

- збір інформації постачальниками послуг;

- атаки повторної ідентифікації;

- порушення цілісності даних.

Існують альтернативні шляхи забезпечення конфіденційності і захисту великих даних, ці шляхи напряму пов’язані з великою кількістю хмар (multi-cloud або hybrid cloud). Наприклад застосовують асиметричну схему безпеки для великих даних в декількох хмарах для вирішення нижче наведених проблем:

- прив'язка до постачальника;

- захист даних;

- захист конфіденційності.

Для захисту файлу, власник ділить файл F на m частин, в залежності від кількості хмар. Конфіденційну інформацію цього файлу Fs власник відділяє, шифрує та відправляє на окрему хмару. Публічну інформацію файла Fp, що залишилася ділиться на рівні частини n.

Також була виведена вірогідність успішної атаки на хмар:

$$P_{MC}(Z) = \prod_{i=1}^{k} \frac{1}{n_i},$$

де $P_{MC}(Z)$ - вірогідність успішної атаки, n_i – кількість фрагментів які зберігаються в хмарному сховищі.

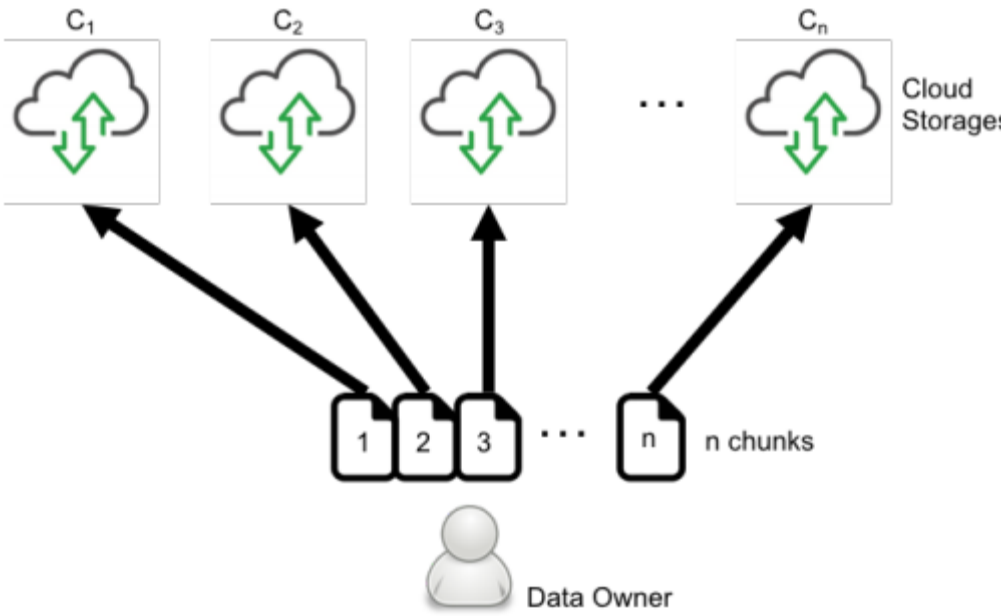


Рис. 2.2. Принцип розбиття даних та розподіл частин по хмарних ресурсах

Також запропоновано схоже рішення, в якому також приділяється увагу необхідному розподіленню публічної та конфіденційної інформації на 2 частини, які потрібно розділити на n рівних частин та розподілити між хмарними ресурсами.

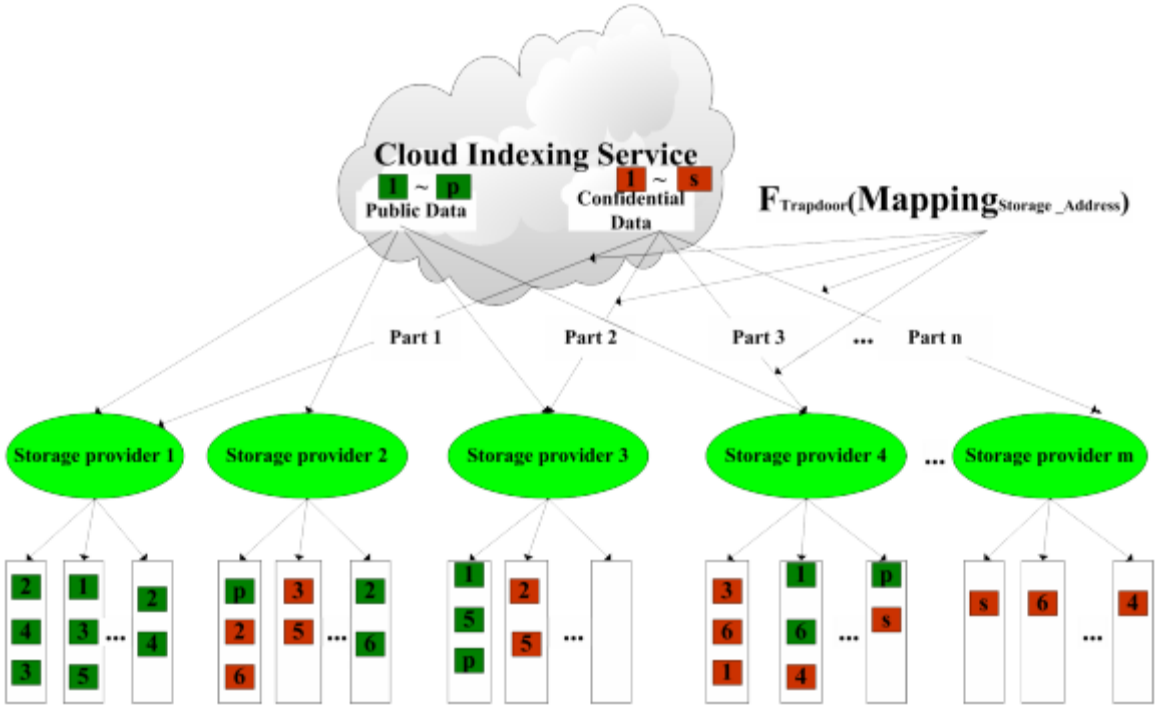


Рис. 2.3. Приклад рішення розбиття даних

Різницею від попереднього дослідження є шифрування не даних, а шляху (одностороння функція з секретом або trapdoor function) до цих даних та розподіл частин не тільки публічної, а й конфіденційної інформації по хмарам. Далі запропоновано структуру, яка включає такі модулі як: завантаження даних, нарізка, індексація, шифрування, розповсюдження, дешифрування, пошук та злиття в декількох хмарах. Якщо будь-який користувач хоче отримати доступ до даних, йому потрібно подати запит на власника даних. Власник даних надішле секретний ключ разом із іменем файлу запитуваному користувачеві через захищений канал. Запитуваний користувач передає ім'я файлу в багатохмарне середовище та отримує зашифровані частини даних. Для шифрування та дешифрування був запропонований алгоритм, який складається з двох модулів: Feistel network та AES with S-box. Виходячи з експериментів, стає зрозумілим що запропований алгоритм працює краще ніж існуючі популярні алгоритми.

2.2. Дослідження загроз хмарних обчислень та сервісів

Хмарні обчислення є більш безпечними та надійними, ніж традиційні, завдяки можливості віддаленого доступу до даних та високому рівню шифрування і протоколів безпеки, що використовуються постачальниками хмарних послуг.

Хмарна атака – це кібератака, націлена на платформи хмарних послуг, наприклад: обчислювальні служби, служби зберігання даних або програмне забезпечення.

Хмарні атаки можуть мати серйозні наслідки, такі як витік даних, втрата даних, несанкціонований доступ до конфіденційної інформації та збої в роботі служб. Оскільки все більше організацій і окремих осіб покладаються на хмарні обчислення для зберігання та обробки даних, відповідно збільшується і кількість потенційних цілей для зловмисників.

Рис. 2.4. Представлення основних видів загроз хмарних обчислень та сервісів

Найбільш значні загрози, які пов’язані з хмарними обчисленнями (рис. 2.4):

- **відмова в обслуговуванні (DDoS):** це спроба порушити нормальну роботу системи, перевантаживши її трафіком. У випадку хмарного середовища це зазвичай відбувається шляхом одночасного надсилання тисяч і тисяч з'єднань. Ці запити перевантажують сервер і заважають йому обробляти законні запити;

- **викрадення облікового запису:** це процес, під час якого хмарний обліковий запис фізичної особи або організації викрадається зловмисником. Захоплення хмарних облікових записів є поширеною тактикою в схемах крадіжки персональних даних, коли зловмисник використовує

скомпрометований обліковий запис електронної пошти або інші облікові дані, щоб видати себе за власника облікового запису;

- **внутрішні загрози:** це категорія ризику, яку становлять ті, хто має доступ до фізичних або цифрових активів організації. Такими інсайдерами можуть бути нинішні працівники, колишні працівники, підрядники, постачальники, які мають (або мали) санкціонований доступ до мережі та комп'ютерних систем організації;

- **неправильна конфігурація хмари:** неправильна конфігурація є проблемою хмарних обчислень, оскільки хмарні середовища можуть бути досить складними, а виявити та виправити помилки вручну може бути важко. Це будь-які збої, прогалини або помилки, які можуть наразити ваше середовище на ризик під час переходу на хмарні технології. Ці кіберзагрози проявляються у вигляді порушень безпеки, що можуть бути використані для несанкціонований доступу до мережі;

- **шкідливі файли cookie:** зараження файлами cookie в хмарних додатках означає несанкціоновану модифікацію або впровадження шкідливого вмісту в файл cookie, який є невеликим фрагментом даних, що зберігається на комп'ютері користувача. У SaaS та інших хмарних додатках файли cookie часто містять облікові дані, тому зловмисники можуть модифікувати ці файли, щоб отримати доступ до додатків;

- **витік даних:** це кібератака, під час якої до чутливих, конфіденційних або інших захищених даних було отримано несанкціонований доступ або вони були розголошені. Порушення даних може статися в організації будь-якого розміру, від малого бізнесу до великих корпорацій.

Небезпечні інтерфейси користувача (UI) та API: постачальники хмарних послуг відображають набір API та інтерфейсів, що дозволяє клієнтам взаємодіяти з хмарними службами та керувати ними. Доступність та безпека хмарних послуг залежить від безпеки цих API. Зламаний, погано спроектований API може призвести до неправильного використання або ще

гіршого порушення даних. Організації, які покладаються на відкриті API та слабких інтерфейси користувача, мають різноманітні проблеми безпеки, найважливішими можуть бути фінансові та регуляторні наслідки.

2.3. Дослідження криптографічних алгоритмів хмарних платформ

Проблеми безпеки в хмарної платформі можуть призвести до економічних втрат, а також до поганої репутації, якщо платформа спрямована на широку публіку і є причиною масового застосування цього нового рішення. Дані, що зберігаються у хмарі клієнтів, є життєво важливою інформацією. Ось чому порушення таких даних неавторизованою третьою стороною є неприпустимим. Таким чином, користувачі повинні бути дуже обережними при зберіганні своїх даних у хмарному сховищі. Отже, потрібно думати про способи, що перешкоджають використанню даних, навіть якщо до даних звертається третя сторона. Таким чином, усі дані мають бути зашифровані перед передачею до хмарного сховища. Безпека забезпечує конфіденційність, цілісність, справжність та доступність інформації. Розвиток технологій та їх стандартизація робить доступним набір алгоритмів та протоколів для вирішення цих проблем.

Визначено п'ять основних типів сучасної криптографії, а саме симетричну, асиметричну, гібридну, нейронну, квантову криптографію. Огляд показаний у таблиці 2.1. Безпека сучасних криптографічних систем, які у каналах зв'язку загального користування, полягає в секретності ключа (ключів), який спільно використовується чи обмінюється між набором користувачів. Для створення безпечного ключа неавторизований користувач не повинен знати або отримати доступ до нього. Це поняття створення безпечного ключа у криптографії з відкритим ключем ґрунтується на недоведених припущеннях, пов'язаних із твердістю/складністю певних математичних завдань (що означає, що невідомі алгоритми для вирішення проблем у ефективний час).

Таблиця 1.1.

Опис сучасної криптографії

Тип	Характерна риса	Опис	Переваги	Недоліки
Криптографія з публічним ключем	Ключ виходить із центру сертифікації, який видає цифровий сертифікат, що містить відкритий ключ власника сертифіката та іншу ідентифікаційну інформацію	Генеруються два набори ключів, відкритий ключ - для шифрування даних і закритий ключ - для розшифрування даних	Відкритий ключ може бути переданий для надсилання зашифрованих даних, а закритий ключ зберігається у власника та використовується для дешифрування	Нова пара ключів генерується при втраті / витоку ключа. Через підвищену складність та обчислювальну одиницю накладні витрати високі.
Криптографія із секретним ключем	Класифікуються як потокові або блокові шифри. Покладатися на математичні методи;передбачає складність факторизації великих чисел	Обидві сторони використовують один і той же ключ для шифрування та дешифрування даних	Швидкий, простий та ефективний підхід. Найменша зміна секретного ключа не призведе до розшифрування зашифрованого повідомлення	Зловмисник може використовувати витік ключа для розшифровки зв'язку між двома довіреними пристроями. Існуюча обчислювальна потужність може бути використана для злому ключа
Гібридна криптографія	Комбінація криптографії з відкритим та закритим ключем. Пара відкритого та закритого ключів використовується для шифрування та дешифрування	Має характерні особливості криптографії із закритим і відкритим ключем	Швидкість, простота обробки. Користувачі можуть генерувати власні ключі змінної довжини і можуть оновлювати ключ у будь-який проміжок часу	Працює із двійковими бітовими послідовностями. Використовує загальновідомі математичні алгоритми кодування інформації
Квантова криптографія	Заснована на принципі невизначеності Гейзенберга, поляризації та заплутаності фотонів	Використовує випадковий секретний ключ	Випадковий секретний ключ. Висока безпека: гарантує, що повідомлення не було прочитано або не було змінено.	Відсутня у квантовому каналі. Висока частота помилок у старшому біті. Працює в обмеженому просторі
Нейронна криптографія	Використовується структура, аналогічна роботи клітин мозку, і був розроблений зв'язок між виходами та входами з використанням «ваг»	Використовує випадковий секретний ключ	Важко зламати секретний ключ. Мінімізує спотворення повідомлення, що надсилається	Як для процесу шифрування, так і для процесу дешифрування, знання ключа недостатньо, якщо ключ не є одночасно вагою та мережевою архітектурою

Також для забезпечення безпеки великих даних методи шифрування поступово зростають, і врешті-решт вони підбираються до двох сильних напрямків, а саме нейронної та квантової криптографії.

AES - це алгоритм із симетричним ключем, у якому той самий ключ використовується як для шифрування, так і для дешифрування даних. AES широко використовується в практичному захисті даних у хмарі. Хоча AES є широко поширеною схемою шифрування даних у хмарному сховищі, вона обмежує багато функцій додатків, таких як пошук, логічні операції та математичні обчислення, але даний алгоритм використовують усі хмарні вендори для захисту великих даних. Масштабованість AES у хмарних середовищах також є важливою проблемою, яка потребує вирішення.

Схема Рівеста-Шаміра-Адлемана (RSA) є одним із перших успішних криптографічних алгоритмів, що відповідає вимогам до систем з відкритим ключем, є найбільш широко прийнятим та реалізованим універсальним підходом до шифрування з відкритим ключем. Схема RSA приймає відкритий і зашифрований текст з цілими числами від 0 до n - 1 для деякого n. Національний інститут стандартів і технологій США рекомендував, щоб цих систем, які зазвичай використовують 1024 біти, достатньо для використання до 2010 року. Оскільки RSA ґрунтується на арифметичному модулі великих чисел, він може бути повільним в обмежених середовищах. Для підвищення продуктивності пропонується кілька варіантів RSA, наприклад, Batch RSA, Multi-factor RSA або Rebalanced RSA. Багатофакторний RSA ґрунтується на зміні структури модуля RSA. Два багатофакторні методи RSA є багатообіцяючими в тому сенсі, що вони повністю сумісні. Алгоритм придатний як для шифрування, так і для цифрових підписів, що дозволяє використовувати його для автентифікації джерела та захисту цілісності великих даних.

Криптографія еліптичних кривих (ECC) ґрунтується на математичній теорії еліптичних кривих. Вона може забезпечити той самий рівень і тип безпеки, що і RSA, але з набагато коротшими ключами. Агентство

національної безпеки (NSA) порівнює розміри ключів для трьох різних підходів до шифрування для порівняних рівнів захисту від атак грубої сили. Порівняння розміру ключа показує, що з ECC потрібна одна шоста обчислювальних зусиль для забезпечення того ж рівня криптографічної безпеки, який виходить із 1024-бітним RSA.

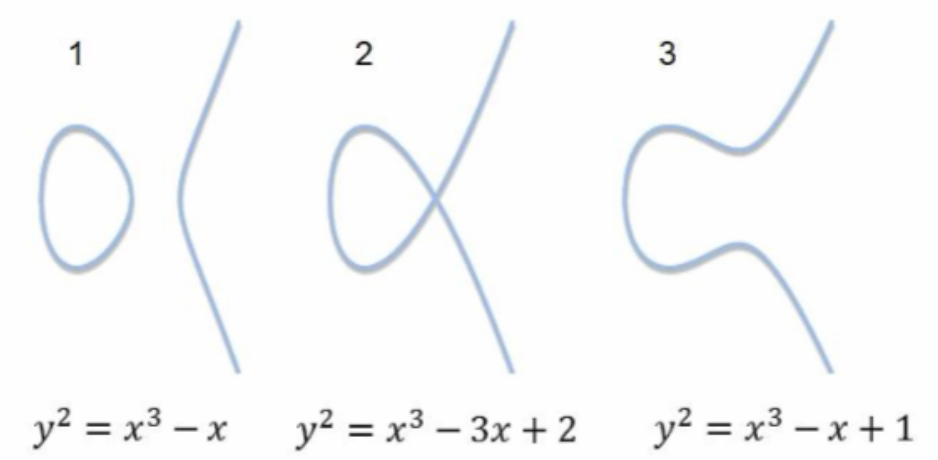


Рис. 2.5. Варіанти еліптичних кривих при D < 0, D = 0та D > 0

Через набагато менший розмір ключа алгоритми ECC можуть бути реалізовані на смарт-картах без математичних співпроцесорів. Безконтактні смарт-карти працюють тільки з ECC, тому що інші системи потребують надто багато енергії індукції. Оскільки менша довжина ключа призводить до швидшого протоколу встановлення зв'язку, ECC також стає все більш важливим для бездротового зв'язку.

Квантова криптографія - одна з найпопулярніших проблем у науці через здатність забезпечувати 100% безпеку, засновану на принципах заплутаності та нагляду. Вона була заснована на принципі невизначеності Гейзенберга, поляризації та заплутаності фотонів.

Існують різні протоколи розподілу квантових ключів, такі як BB84, BB92, COW та ін. Однак у різних протоколах розподілу ключів використовуються різні аспекти цих фізичних теорем. Наприклад, у той час як у протоколі BB84 використовуються чотири стани поляризації принципу

невизначеності Гейзенберга, у протоколі BB92 використовуються лише два стани. Більше того, вони пояснюють, чому квантова криптографія забезпечує конфіденційність та безпеку. За словами авторів, залежно від здатності QKD, який завжди створює нові та випадкові ключі, може бути забезпечена конфіденційність, проте постійна зміна ключів може викликати дешифрування обмежених фрагментів інформації. Протокол BB84, також відомий як протокол квантового розподілу ключів QKDO), виник у 1984 році на підставі невизначеності поляризації фотонів Гейзенберга. Безпека BB84 заснована на деяких принципах квантової фізики, таких як: суперпозиція, поняття міри, заплутаність та теорема про не клонування.

За допомогою різних фільтрів можна керувати напрямком неполяризованого світла. У QKD для управління напрямом електромагнітної хвилі використовуються вертикальні та горизонтальні фільтри, і ці фільтри називаються базами.

До складу квантових технологій захисту інформації (рис.2.6) входять: квантовий розподіл ключів, квантовий прямий безпечний зв'язок, квантове розділення секрету, квантовий потоковий шифр, квантовий цифровий підпис та квантова стеганографія.

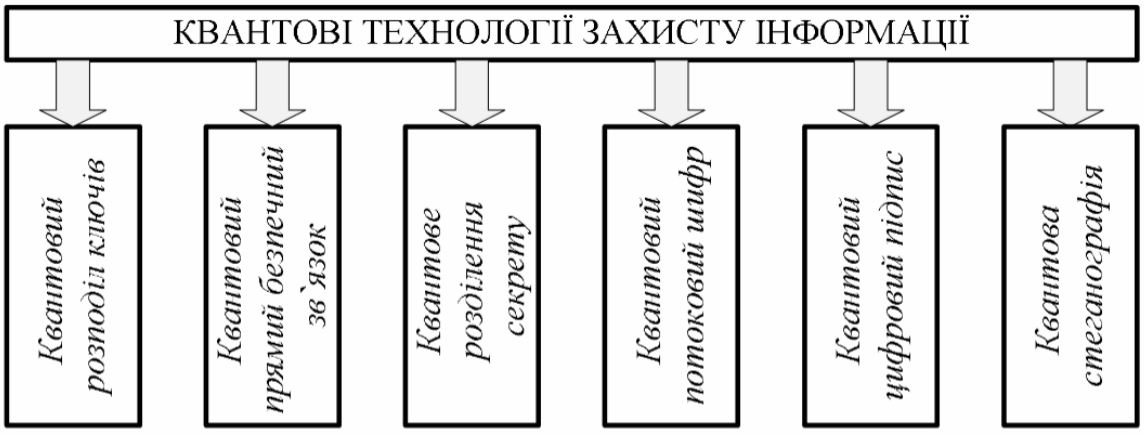


Рис. 2.6. Квантові технології захисту інформації

Квантовий розподіл ключів включає в себе наступні протоколи (рис. 2.7):

- протоколи з використанням одиничних поляризованих фотонів;
- протоколи з використанням фазового кодування;
- протоколи з використанням переплутаних станів;
- протоколи зі станами «приманки».

Рис. 2.7. Класифікація квантових протоколів розподілу ключів

В 1984 році запропоновано перший протокол квантової криптографії, що мав стати альтернативним і нетрадиційним рішенням проблеми розподілу ключів шифрування. Даний протокол отримав назву BB84, він відноситься до протоколів квантового розподілу ключів з використанням одиничних поляризованих фотонів. Основними задачами КПРК є генерація та розподіл ключів шифрування між двома абонентами, що з’єднані квантовим та класичним каналами зв’язку. У протоколах з одиничними поляризованими фотонами використовуються 4 поляризовані стани фотонів (0º, 45º, 90º, 135º), які передаються квантовим каналом зв’язку. Пошук та виправлення помилок виконується з використанням відкритого класичного каналу, який не повинен бути конфіденційним, тільки аутентифікованим. Для виявлення факту дій зловмисника використовується процедура контролю помилок, а для

забезпечення безумовної стійкості використовується класична процедура підсилення секретності (privacy amplification).

Ефективність протоколу BB84 з кубітами в ідеальних умовах дорівнює 50%. Під ефективністю розуміють відношення кількості фотонів, що використовуються для генерації ключа, до загальної кількості переданих фотонів. Крім того, запропоновано узагальнення протоколу BB84 на багаторівневі квантові системи (так званий протокол BB84 з кудитами). Цей протокол має значно більшу інформаційну місткість та стійкість до некогерентних атак, але його складніше реалізувати з технічної точки зору. Вихідними даними КПРК є ключова послідовність, яка може бути використана для подальшого шифрування даних. До вищезгаданого типу протоколів (див. рис. 2.7) крім BB84 відносяться також протокол з шістьома станами, протокол 4+2, протокол Гольденберга-Вайдмана та протокол Коаші-Імото.

Квантові протоколи розділення секрету теж мають перевагу над відповідними класичними протоколами, так як дозволяють виявити підслуховування у квантовому каналі та не потребують шифрування. Аналогічно, більш високий рівень безпеки, у порівнянні з відповідними класичними схемами, забезпечує квантовий потоковий шифр та квантовий цифровий підпис. Останній, завдяки використанню квантової односторонньої функції, має теоретико-інформаційну стійкість. Проте, практична реалізація цих квантових методів захисту інформації теж поки що зіштовхується з деякими технологічними складнощами.

Нейронна криптографія – це сфера, яка розвивається, та яка має на меті поєднати криптографію з нейронними мережами для додатків у криптоаналізі та шифруванні. На принцип роботи нейронної криптографії впливає імпульсна провідність нейронів. Подібно до робочої системи клітин мозку, штучна нейронна мережа запрограмована так, щоб розвивати математичні відносини з діапазоном входів та їх відповідних виходів і класифікувати їх.

Часто визначають нейронну криптографію як систему з кількома рівнями. На першому рівні дані передаються на другий через синапси, які можуть маніпулювати інформацією при оцінці за допомогою «вагових» параметрів, а потім другого рівня третьому.

Цей метод криптографії складається із трьох етапів: перший – це з'єднання, другий – навчання, а останній – активація. Більш того, автори заявляють, що для створення системи нейронної криптографії, яка генерується автономно, використовуються кілька складних множинних входів та безліч контурів спрямованого зворотного зв'язку.

Нейронна криптографія – це надійна мережна безпека через те, що ключі складно зламати, а шифрування виконується у три основні етапи. Які називаються з'єднанням, навчанням та активацією.

На етапі з'єднання кореляція відносини будуються між нейронами. На етапі навчання, контролюючи дані оцінки, збережені параметри переглядаються. На етапі активації вихід видається із змінених входів. Таким чином шляхом обробки цих кроків створюються інтегральні індуктивні рішення для ключового затвердження.

Для нейронної криптографії використовують різноманітні алгоритми шифрування. Вони особливо використовуються при генерації секретних ключів через те, що зламати ключ практично неможливо без його нейронно-мережевої карти. Тому даний вид криптографії може бути використано для захисту даних великих обсягів у майбутньому, в порівнянні з квантовою криптографією, вона не має обмежень на відстань.

Структура нейроподібної мережі для шифрування даних наведена на рис. 2.8, де де ПЕ - процесорний елемент, Рг - регістр, ОЗП - оперативний запам'ятовуючий пристрій, См - суматор, Вд - віднімач, У1, У2, У3 - перший, другий і третій входи управління, ВхБ - вхід даних, ВихУ - вихід результату шифрування.

Шифрування потоків даних за допомогою паралельно-потокової нейроподібної мережі вимагає попереднього обчислення вагових

коефіцієнтів Wj, формування макрочасткових добутків Рм та їх одночасно запису в усі ОЗПРм.

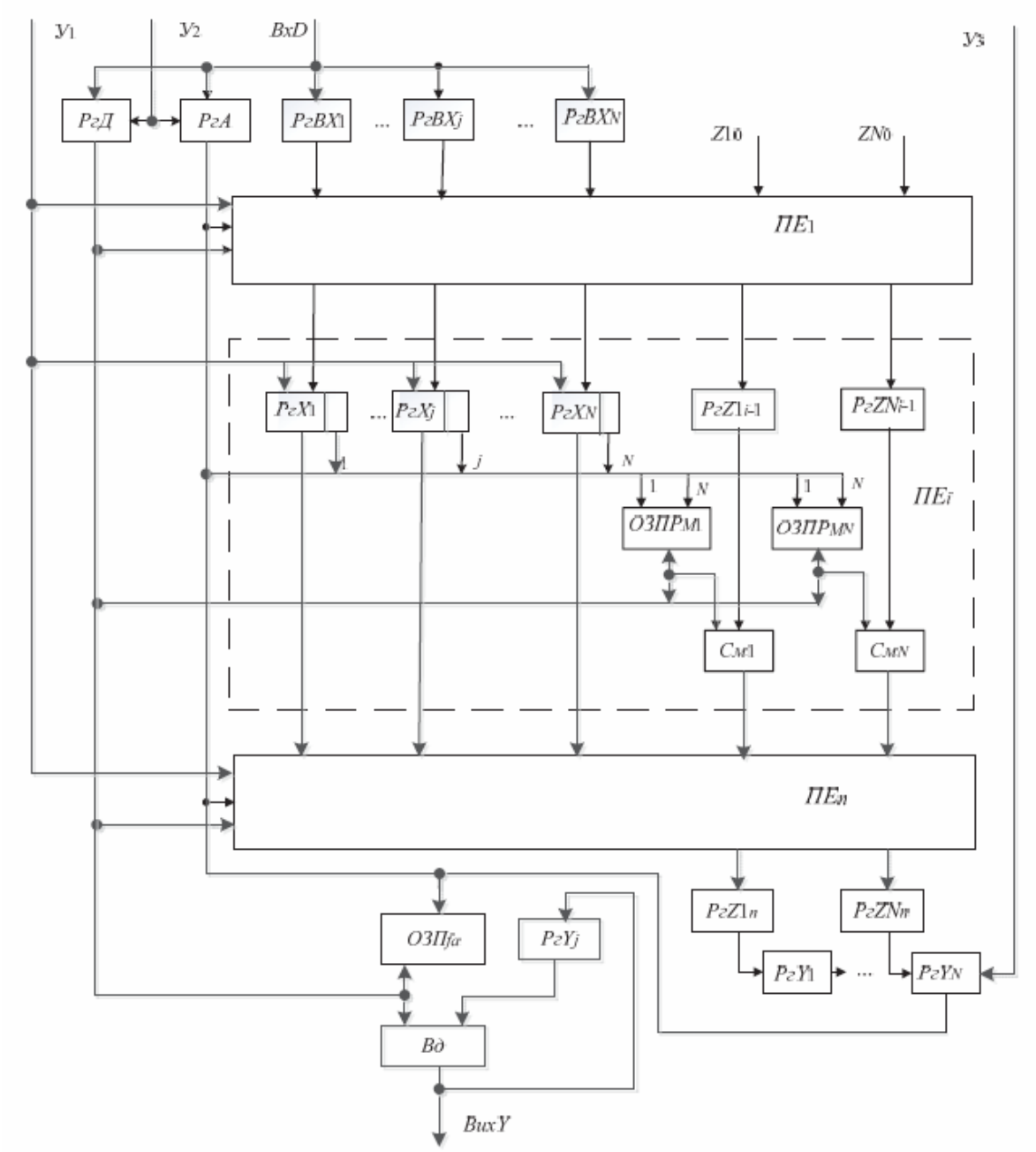


Рис. 2.8. Структура паралельно-потокової нейроподібної мережі шифрування потоків даних

Особливістю структури паралельно-потокового нейроподібної мережі шифрування є те, що дані поступають послідовно одне за одним і за допомогою вхідних РгВХі,..., РгВХм перетворюються у паралельний потік

даних, які надходять на вхід першого ПЕі. Паралельно-потокова нейроподібна мережі реалізується на базі п однотипних ПЕ, які працюють за конвеєрним принципом. Такт роботи конвеєра такої мережі рівний такту роботи нейроподібного елемента (10). У кожному такті роботи обчислені скалярні добутки записуються в регістри РгZп,..., РгZNn, а з них у регістри РгУ1,…,$РгУ_N$ за допомогою яких виконується паралельно-послідовне перетворення надходження скалярних добутків. На виході віднімана *Вд* формується потік зашифрованих даних.

2.4. Опис алгоритму шифрування великих об’ємів даних

Одним із основних алгоритмів шифрування для захисту великих даних є AES. AES – симетричний алгоритм блочного шифрування (розмір блоку 128 біт, ключ 128/192/256 біт), прийнятий в якості стандарту шифрування урядом США за результатами конкурсу AES. AES впроваджено в програмне та апаратне забезпечення по всьому світу для шифрування конфіденційних даних. Це важливо для державної комп’ютерної безпеки, кібербезпеки та захисту електронних даних.

Стандарт AES допускає тільки одне значення довжини блоку - 128 біт для 3 версій алгоритму AES. У той час як розмір ключа в різних версіях відрізняється: AES-192 використовує 192 - бітний розмір основного ключа і має 12 раундів шифрування, а AES-256 - 256 бітний розмір основного ключа і 14 раундів шифрування.

Більша кількість раундів робить шифрування складніше. Таким чином, AES-256 володіє найбільш безпечною реалізацією. Однак слід зауважити, що чим довше ключ і більше раундів, тим вище вимога до продуктивності. Враховуючи те, що в SkyProtect необхідний баланс між забезпеченням захисту даних та продуктивністю (завантаження файлів з хмар) обираємо версію алгоритму AES-128 який має 10 раундів шифрування.

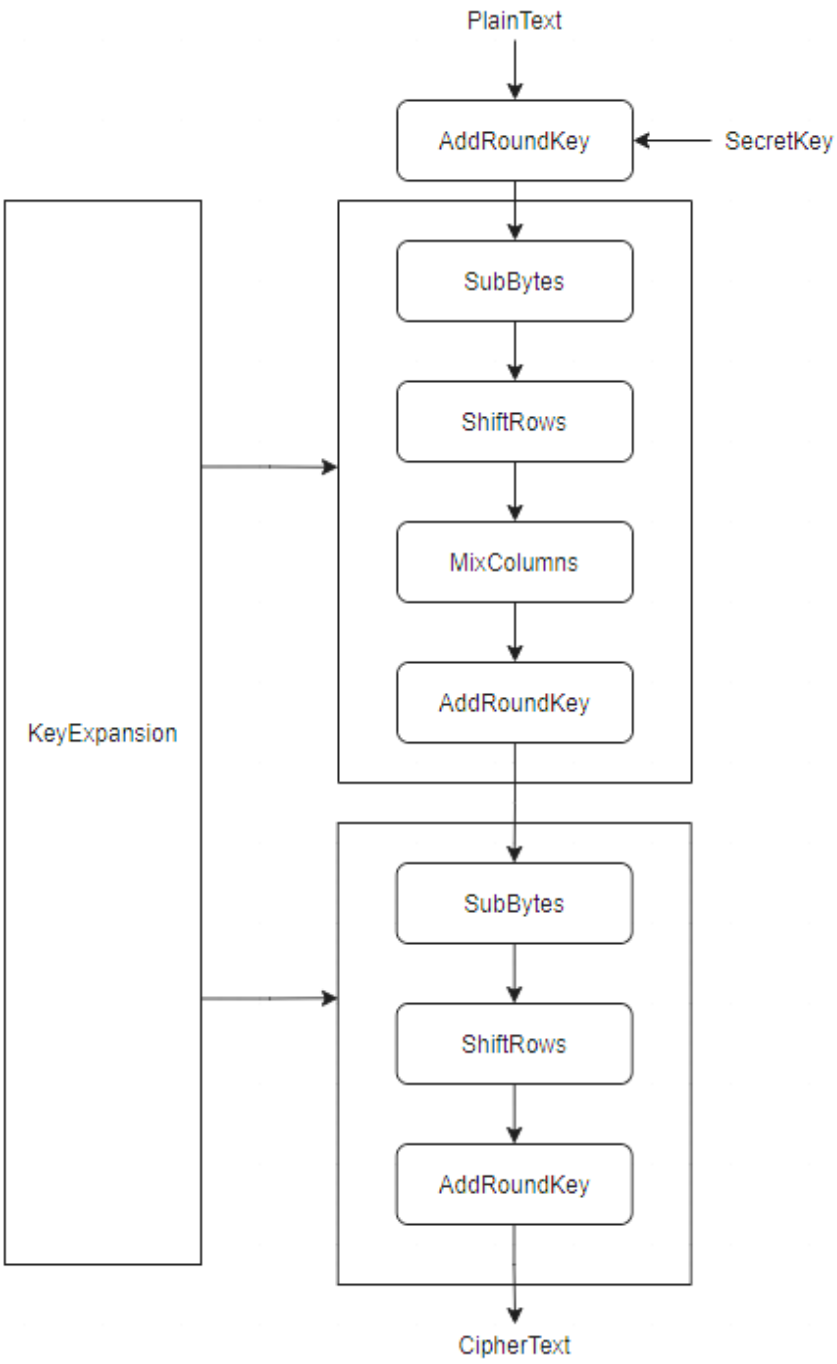


Рис. 2.9. Алгоритм шифрування AES

Для шифрування в алгоритмі AES застосовуються такі процедури перетворення даних:

1. KeyExpansion - обчислення раундових ключів для всіх раундів;

2. SubBytes - підстановка байтів за допомогою таблиці підстановок S-box;

3. ShiftRows – циклічний зсув рядків блоку. Нульовий рядок залишається на місці, перший зміщується вліво на 1 байт, другий на 2 байта і третій на 3 відповідно;

4. AddRoundKey - раундовий ключ, який поелементно додається до блоку за допомогою порозрядного XOR.

Отже, алгоритм складається з наступних кроків:

1. KeyExpansion;

2. початковий раунд – складання блоку з основним ключем;

3. раундів шифрування, кожен з яких складається з перетворень: SubBytes, ShiftRows, MixColumns, AddRoundKeys;

4. фінальний раунд, який складається з перетворень: SubBytes, ShiftRows, AddRoundKeys.

Попередньо вхідні дані розбиваються на блоки по 16 байт, якщо повний розмір не кратний 16 байтам, то дані доповнюється до розміру, кратного 16 байтам. Блоки представляються у вигляді матриці 4x4. Далі відбувається процедура розширення ключа і до кожного блоку застосовуються операції 2-4.

Так як всі перетворення шифрування виконуються однозначно, то існує зворотне перетворення, за допомогою якого CipherText перекладається у PlainText. Зворотне перетворення являє собою послідовність інвертованих операцій шифрування, які виконуються в зворотному порядку.

Алгоритм складається з наступних кроків:

1) KeyExpansion;

2) початковий раунд – складання блоку з основним ключем;

3) раундів дешифрування, кожен з яких складається з перетворень: InverseSubBytes, InverseShiftRows, InverseMixColumns, AddRoundKey;

4) фінальний раунд, який складається з перетворень: InverseSubBytes, InverseShiftRows, AddRoundKeys.

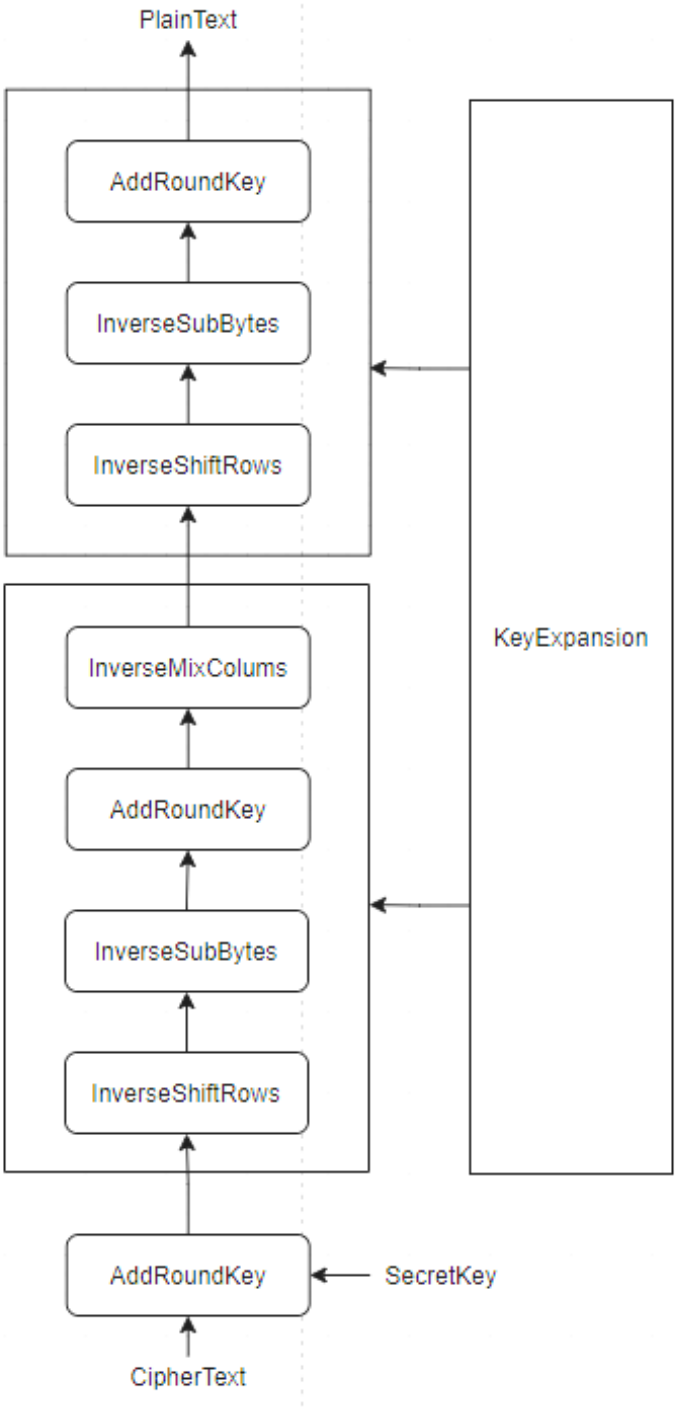


Рис. 2.10. Алгоритм дешифрування AES

Алгоритм може бути впроваджений на мові програмування. Даний клас буде мати в собі методи для генерації секретного ключа, завантаження ключа в директорію проекту або завантаження його з директорії для шифрування файлів та методи для шифрування та дешифрування. Нижче буде приведена діаграма варіантів використання.

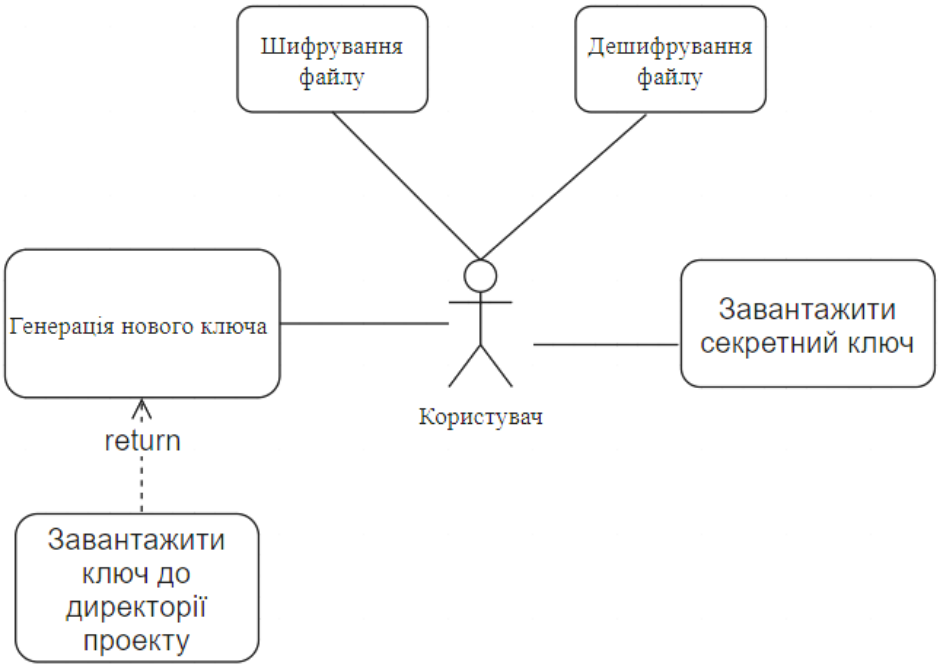


Рис. 2.11. Діаграма варіантів використання алгоритму AES

2.5. Дослідження методів захисту Amazon Web Services, Microsoft Azure і GCP

AWS є одним з провідних постачальників хмарних послуг у світі. Вони пропонують широкий спектр послуг у сфері обчислення, зберігання даних, мереж та інших інфраструктурних рішень. Безпека є невід'ємною частиною всіх цих послуг, оскільки користувачі довіряють AWS для зберігання, обробки та передачі своїх конфіденційних даних.

Щоб захистити дані в хмарному середовищі, AWS надає різноманітні інструменти та сервіси. Пропонують можливість шифрування даних на різних рівнях, від шифрування в спокої до шифрування на рівні об'єктів. AWS забезпечує управління ключами шифрування та контроль доступу до зашифрованих даних, що дозволяє користувачам зберігати дані в безпечному стані.

Веб-сервіси Amazon (AWS) надають декілька служб безпеки, щоб допомогти своїм клієнтам захистити їх хмарні дані на основі даних від втрат, корупції чи викривлення. Ці сервіси є основними складовими будь-якої

стратегії захисту даних, наприклад, керування доступом на основі ролей, автентифікація користувачів, моніторинг подій та трафіку, журнали та сповіщення тощо. Далі буде описано розширені служби безпеки AWS та те, як вони забезпечують рівень безпеки даних та додатків при стратегічному використанні із будівельними блоками безпеки AWS.

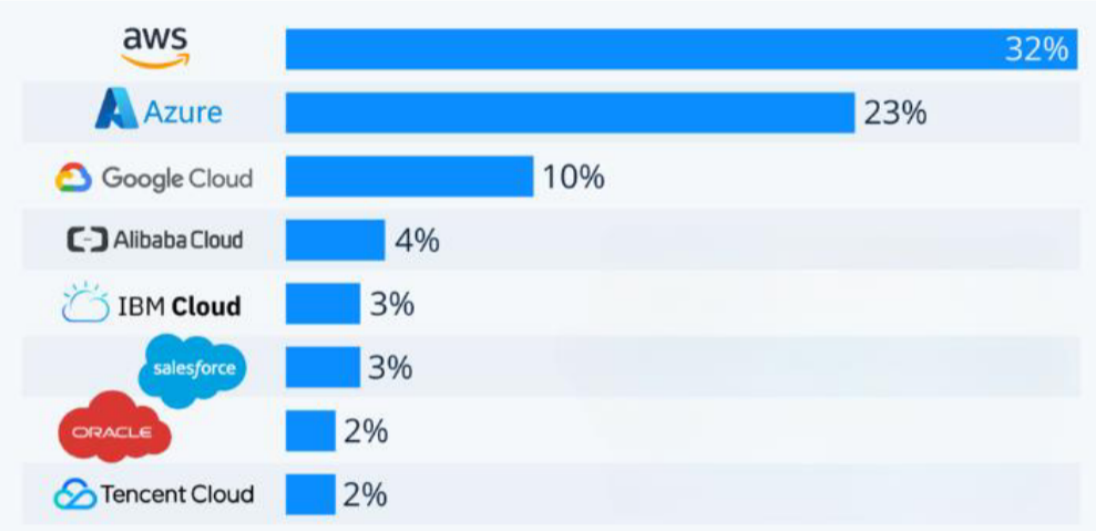


Рис. 2.12. Популярність постачальників хмарних послуг

- **Захист додатків:** AWS WAF та AWS Firewall Manager – AWS WAF з метою задоволення потреб у безпеці сьогодні широко розповсюджених додатків, брандмауер веб-додатків AWS (AWS WAF) відстежує запити HTTP / HTTPS на всіх відповідних вхідних інтерфейсах. Ці інтерфейси включають шлюз API Amazon, Amazon CloudFront (мережа доставки вмісту) та балансир завантаження програм.

- **Атака DDoS:** AWS Shield – це безкоштовна послуга захисту від розподіленої відмови у наданні (DDoS) для всіх програм, що використовують послуги AWS. Він захищає веб-сайти та програми від найчастіших DDoS-атак.

- **Інтелектуальне виявлення загроз:** Amazon GuardDuty – застосовує всі новітні технології виявлення загроз – машинне навчання, штучний інтелект, аналітику поведінки та багато іншого – оскільки він постійно

здійснює моніторинг облікових записів AWS та навантажень для зловмисної діяльності та аномальної поведінки.

- **Автоматизована безпека даних:** Amazon Macie – це повністю керована інтелектуальна служба захисту даних. Вона автоматично виявляє, класифікує та захищає конфіденційні дані, такі як особиста інформація або інтелектуальна власність.

AWS використовує модель "Shared security responsibility model", де вони відповідають за безпеку базової хмарної інфраструктури, а клієнти - за безпеку своїх робочих навантажень, розгорнутих на платформі AWS. Клієнти мають можливість жорстко обмежувати доступ до конфіденційних даних і встановлювати різні рівні контролю для інформації, яку вони хочуть оприлюднити.

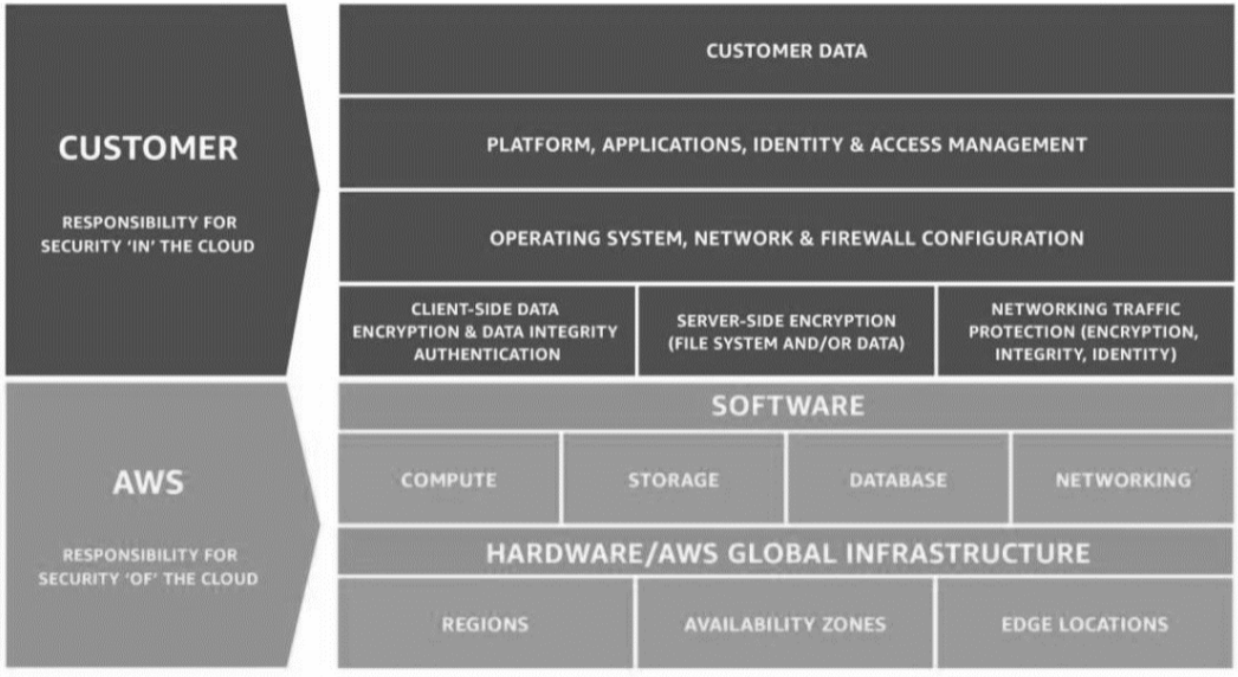


Рис. 2.13. Модель “Shared security responsibility model” від AWS

Проведемо аналіз механізмів забезпечення захисту даних для AWS, Microsoft Azure і GCP. З огляду на особливості кожного з провайдерів, можна виділити два основних напрямки:

- контроль доступу;

- обробка даних.

Amazon S3 – це сервіс зберігання об'єктів, який пропонує кращі в галузі показники продуктивності, масштабованості, доступності та безпеки даних. Amazon S3 забезпечує надійність на 99,999999999% та зберігає дані мільйонів додатків в інтересах компаній з усього світу. Нижче пояснюється, як Amazon зберігає хороші позиції з точки зору використовуваної безпеки. Amazon використовує структуру IAM для управління доступом до своїх ресурсів. IAM – це структура, яка використовується для ідентифікації, автентифікації і авторизації користувачів, процесів або груп для доступу до ресурсів AWS. Платформа підтримує централізоване подання для управління користувачами, паролями, ключами доступу і політиками. Фреймворк працює наступним чином.

Політики ідентичності (Identity Policies) визначають список дозволів, які надаються конкретній ідентичності (ролі або користувачу).

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "iam:CreateRole",
                "iam:CreateUser"
            ],
            "Resource": [
                "arn:aws:iam::123456789012:role/some-role",
                "arn:aws:iam::123456789012:user/some-user"
            ]
        },
        {
            "Action": [
                "logs:*"
            ],
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
}
```

Рис. 2.14. Приклад Identity Policy

При першій реєстрації користувача з електронною поштою і паролем, створений обліковий запис вважається кореневим обліковим записом з

повним доступом до всіх доступних ресурсів і сервісів в AWS. Вважається, що краще використовувати IAM для формування користувачів, груп і ролей.

Створений користувач IAM може бути налаштований на отримання логіну, паролю, ключа доступу і набору доступу для конкретного облікового запису. Рекомендується також використовувати ролі, для запобігання доступу користувачів до всіх ресурсів кореневого облікового запису.

Також рекомендується встановлювати мінімальні привілеї. Для отримання мінімальних привілеїв і щоб уникнути перевищення привілеїв можна використовувати роль IAM. Ролі IAM, на відміну від пароля або ключів доступу, підтримують використання тимчасових облікових даних безпеки. Політики IAM в основному налаштовані на обмеження певного джерела або часу доступу на основі інших умов.

Для забезпечення доступності даних користувачам рекомендується використовувати управління версіями і реплікацію. Управління версіями необхідно для відновлення старішої версії об'єкта. Крім того, для управління сервісної консоллю AWS і API Amazon S3 встановлюється безпечне з'єднання SSL/TLS. Amazon також використовує перевірку цілісності для автентифікації запиту і забезпечення цілісності даних, застосовуючи призначені для користувача настройки для одного з наступних методів: коди автентифікації повідомлень (SHA-1 / SHA-2) або автентифіковане шифрування (AES-GCM) або хешування.

Контроль доступу на Microsoft Azure Сховище BLOB-об'єктів Azure – це рішення Microsoft для зберігання об'єктів в хмарі. Сховище BLOB-об'єктів оптимізовано для зберігання великих обсягів неструктурованих даних. Microsoft стверджує, що дотримується найвищого кількості стандартів безпеки. Відповідно до цього Microsoft надає єдиний сервіс входу для доступу до величезної кількості сервісів і додатків. Служба є багато користувальницьким хмарним каталогом і управлінням посвідченнями, відоме як Azure AD. Azure AD дозволяє розробникам встановлювати централізований контроль доступу за допомогою налаштувань політик і

правил. Поряд з Azure AD Microsoft надає різні методи безпечного доступу до сховища Azure. Azure AD поставляється з керуванням доступом на основі ролей (RBAC) для надання доступу користувачам, групам або додатків до ресурсів Azure. Адміністратор може надати дозвіл через ролі RBAC і Azure AD для облікового запису зберігання та авторизувати операції управління ресурсами. Коли адміністратор призначає доступ, облікового запису буде надана роль для доступу. Адміністратор може контролювати доступ до операцій використовуваного облікового запису зберігання, але не для самих об'єктів даних в обліковому записі. Користувачі можуть отримати доступ до облікового запису зберігання за допомогою підписів загального доступу, які надають дозвіл на доступ до певних об'єктів даних протягом певного інтервалу часу для певного користувача. Підпис загального доступу - це рядок, що складається з токена безпеки, який прикріплений до URI, щоб дозволити і делегувати доступ до сховища великих двійкових об'єктів, і перераховує всі обмеження і дозволи, такі як діапазон дат / часу доступу.

Контроль доступу GCP Google Cloud Storage – це корпоративна загальнодоступна хмарна платформа для зберігання даних, яка може зберігати великі неструктуровані набори даних. Компанії можуть придбати сховище для первинних або рідко доступних даних. Google Cloud Storage пропонує дві системи для надання користувачам дозволів на доступ до ваших контейнерів та об'єктам: IAM і списки управління доступом (ACL). Ці системи діють паралельно - для того, щоб користувач міг отримати доступ до ресурсу хмарного сховища, тільки одна з систем повинна надати користувачеві дозвіл. У більшості випадків IAM – це рекомендований метод управління доступом до ресурсів користувача. IAM контролює надання дозволів у всьому Google Cloud Platform і дозволяє надавати дозволи на рівні контейнера і проекту. Користувач повинен використовувати IAM для будь-яких дозволів, які застосовуються до декількох об'єктів в контейнері, щоб знизити ризики ненавмисного доступу. Щоб використовувати тільки IAM, підключіть єдиний доступ на рівні контейнера, щоб заборонити списки

управління доступом для всіх ресурсів хмарного сховища. При створенні контейнера користувач повинен вирішити, чи хоче він застосовувати дозвіл з використанням уніфікованого або детального доступу:

- Уніфікований (рекомендується): уніфікований доступ на рівні контейнера дозволяє використовувати тільки управління ідентифікацією та доступом (IAM) для управління дозволами. IAM застосовує дозвіл до всіх об'єктів, що містяться всередині контейнера, або групам об'єктів із загальними префіксами імен. IAM також дозволяє використовувати функції, недоступні при роботі зі списками ACL, такі як умови IAM і журнали хмарного аудиту;

- Докладний: Докладний параметр дозволяє використовувати IAM і списки управління доступом (ACL) разом для управління дозволами. ACL – це застаріла система контролю доступу до хмарного сховища, призначена для взаємодії з Amazon S3.

Висновки до розділу

В даному розділі наведено дослідження криптографічних алгоритмів та сучасної криптографії для великих даних, зокрема квантову та нейронну криптографію які є перспективними з точки зору захисту даних великих обсягів та набирають популярності. Розглянуто та проаналізовано механізми забезпечення захисту великих даних на хмарних платформах, захист даних при завантаженні на всіх об’єктних хмарних сховищах здійснюється завдяки алгоритму шифрування AES-256.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ЗАХИСТУ НА ХМАРНИХ ПЛАТФОРМАХ

3.1. Формальна модель безпеки хмарної платформи

Першочерговою задачею при аналізі будь-якої системи є побудова моделі цієї системи з завданим рівнем деталізації. Модель хмари, запропонована NIST, включає в себе п’ять основних ролей: користувач хмари, провайдер хмарних послуг, провайдер доступу до хмарних послуг, аудитор хмари, хмарний брокер (рис. 3.1).

Рис. 3.1. Модель хмарного сервісу

Розглянемо окремо кожну з ролей та функції, що вона виконує.

1) Користувач хмари – особа або організація, яка підтримує ділові відносини з постачальниками хмарних послуг і використовує їх сервіси. Взаємодія користувача з постачальником хмарних послуг відбувається через брокера, використовуючи постачальника доступу до хмарних послуг. В залежності від потреб користувач використовує різні рівні сервісів, що пропонуються

постачальником, у зв’язку з чим має доступ до програмних додатків в хмарі (SaaS), операційної системи та розробки програмних додатків (PaaS), віртуальних комп’ютерів та компонентів мережі (IaaS).

2) Провайдер хмарних послуг – особа або організація відповідальна за створення та управляння хмарою та її службами. Провайдер також займається підтримкою інфраструктури та програмного забезпечення, яке забезпечує роботу хмари. Провайдер, що надає послуги на рівні SaaS виконує більшість обов'язків з управління та контролю додатків та інфраструктури, в той час як споживачі мають обмежений адміністративний контроль додатків. На рівні PaaS користувачу надається доступ до середовища розробки додатків для хмарних сервісів (IDE) та набору компонентів програмного забезпечення для розробки (SDK), при цьому користувач має доступ до налаштувань програмного забезпечення та деяких налаштувань хмарного середовища, можливий також доступ до деякихналаштувань нижнього рівня: операційної системи, мережі, файлового сховища.

3) Аудитор хмари – особа або організація, яка може виконувати незалежну експертизу хмарного сервісу на основі перевірки відповідності побудованої хмари стандартам, оцінці послуг, що надаються провайдером хмари з точки зору контролю безпеки, недоторканності приватного життя, продуктивності і т.д. Аудит безпеки також включає перевірку дотримання політики безпеки, регулюючих документів та відповідності до чинних законів про конфіденційність, цілісність та доступність інформації на всіх етапах розробки та експлуатації хмари.

4) Хмарний брокер – особа або організація, яка керує використанням, продуктивністю і доставкою хмарних послуг, а також веде переговори між провайдерами хмари і споживачами.

Основною задачею брокера – є полегшення взаємодії користувачів з провайдерами хмарних послуг за рахунок поліпшення управління доступом до хмарних сервісів, ідентифікацією користувачів, отримання результатів звітності, підвищення рівня безпеки.

5) Провайдер доступу до хмарних послуг – посередник, який забезпечує підключення і транспортування хмарних послуг від хмари до споживачів. Зазвичай, постачальниками доступу до хмарних послуг виступають провайдери телекомунікаційних мереж, що фізично з'єднують та забезпечують передачу інформації між провайдерами та їх користувачами.

На основі моделі хмари (рис. 3.1) NIST було запропоновано формальну модель безпеки хмари, яка визначає компоненти безпеки для кожної з ролей в хмарі. Компоненти безпеки було розмішено відповідно функції та областей діяльності ролі в хмарі. У випадку, коли ролі (або компоненти формальної моделі хмари) виконують ідентичні функції безпеки або виконують їх разом, компонент безпеки охоплює декілька ролей (компонентів формальної моделі хмари).

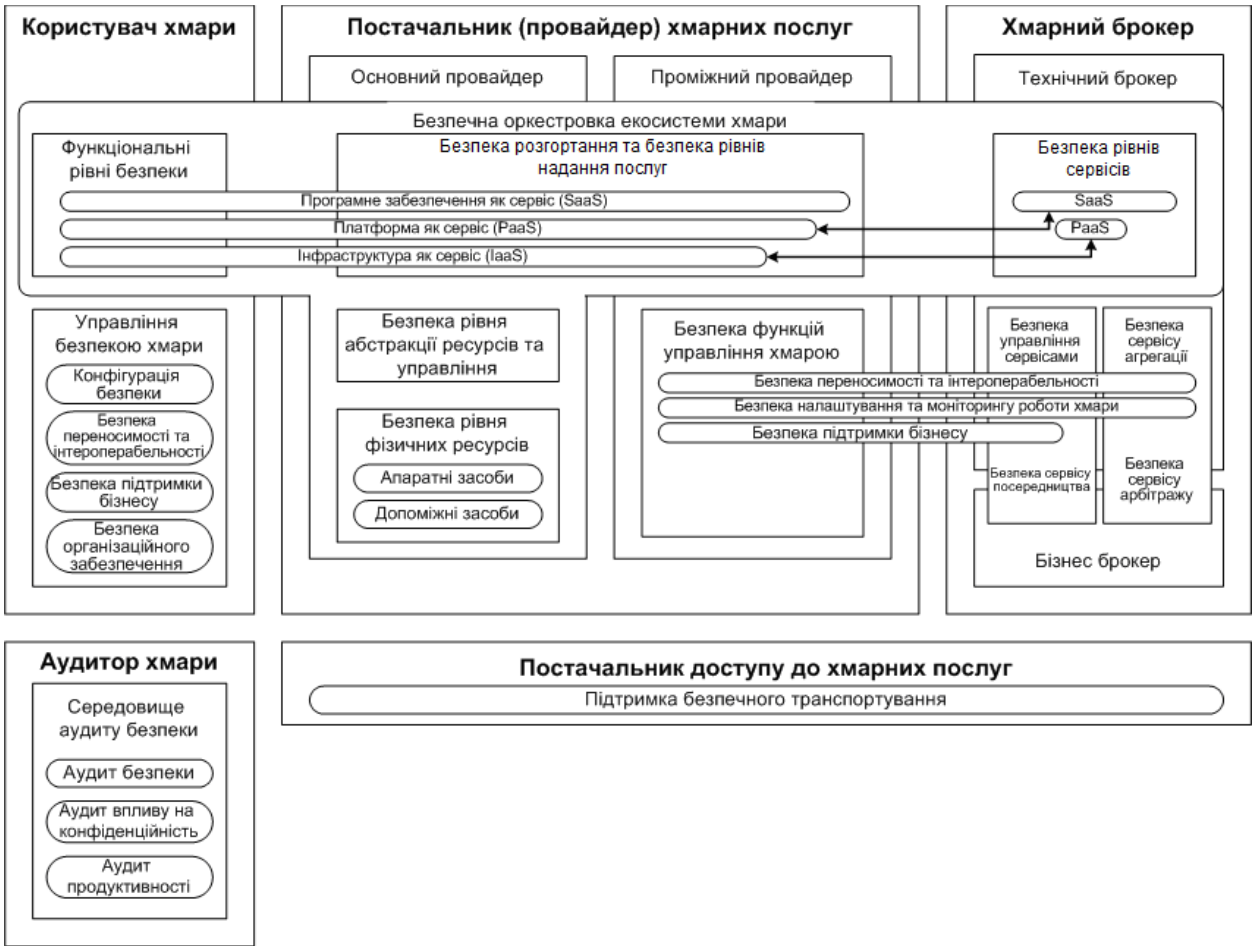


Рис. 3.2. Формальна модель безпеки в хмарі

Проблему довіри до постачальника послуг в запропонованій моделі пропонується вирішувати за рахунок реалізації механізму аудиту хмари та ролі аудитору. Це дозволяє підвищити загальний рівень довіри до постачальника послуги, але остаточно не вирішує проблему. Наприклад, дані користувача зберігаються та обробляються на стороні провайдера, що може призвести до несанкціонованого витоку або блокування.

3.2. Опис принципу багаторівневого розміщення великих об'ємів даних на хмарних платформах

В даний час хмарні обчислення, як ефективна та економічна модель обслуговування та зберігання даних, можуть надати своїм клієнтам найкраще керування даними та обслуговування. Останнім часом багато великих компаній зосереджуються на проблемах, пов'язаних з обробкою та обслуговуванням їх даних, при цьому витрачають багато часу керування даними, але наслідки іноді розчаровують. Це особливо вірно, коли ці компанії чи підприємства не мають великої ділової кореляції з управлінням даними, у такій ситуації ці компанії чи підприємства хотіли б знайти надійного та професійного постачальника, щоб мати справу зі своїми даними, при збереженні, керування та обслуговування, а не робити це самостійно. Звичайно, що передача великих даних і їх віддалене зберігання можуть спричинити багато проблем, наприклад, занадто довгий час обробки, час від часу система виходить з ладу та можуть виникати деякі невизначені системні збої. Передача великих даних безпосередньо в центр обробки даних хмарного сховища може бути вразливою до збою системи через величезні розміри даних.

Однак у запропонованій схемі дані великих обсягів будуть розділені на менші блоки даних, і ці менші блоки даних будуть зберігатися в хмарних носіях даних один за іншим. Оскільки ці блоки даних набагато менші, ніж примітивні великі дані, вони дуже ефективні для віддаленої передачі та

зберігання даних. У запропонованій схемі розділені блоки даних зберігаються на різних носіях даних, доки ми вибираємо певну кількість резервної стратегії резервного копіювання даних, навіть якщо якась служба зберігання вийшла з ладу, це не вплине на дані орендарів. Тому запропонована схема дозволяє уникнути ризику «покласти всі яйця в один кошик». При завантаженні даних великих обсягів потрібно враховувати рівень важливості вхідних даних. Поділимо їх на 3 рівня (рис. 3.3).

Рис. 3.3. Рівні захисту інформації на хмарних платформах

Перший рівень – без додаткового захисту – дані, що не містять у собі конфіденційних (або важливих) даних, захист покладаємо на хмарних провайдерів (Amazon Web Services, Microsoft Azure, Google Cloud Platform, NextCloud).

Другий рівень – з додатковим захистом (шифруванням). Це дані, що містять конфіденційну інформацію. Перед завантаженням на хмари вона шифрується та розподіляється серед хмарних провайдерів.

Третій рівень з додатковим захистом (шифруванням) та реплікацією даних. Це дані, що містять не тільки конфіденційну, але й критичну інформацію, перед завантаженням на хмари вона шифрується, розподіляється серед хмарних провайдерів та додатково реплікується на приватну хмару.

Декілька рівнів важливості даних великих обсягів додано через необхідність мінімізації витрат на завантаження не критичної та не конфіденційної інформації на об'єктні хмарні сховища та підвищення ефективності самого програмного забезпечення.

3.3. Реалізація інформаційної технології захисту даних

Для досягнення описаного принципу розміщення даних на рівнях розроблено архітектуру інформаційної технології, основна мета якої – поділ, розподіл та опціонне шифрування або реплікації даних великих обсягів серед об’єктних сховищ хмарних провайдерів (Amazon Web Services, Microsoft Azure, Google Cloud Platform) та на раніше обрану серед існуючих рішень шляхом проведення оцінки у заздалегідь підготовленому програмному модулі та самостійно розгорнуту приватну хмару NextCloud.

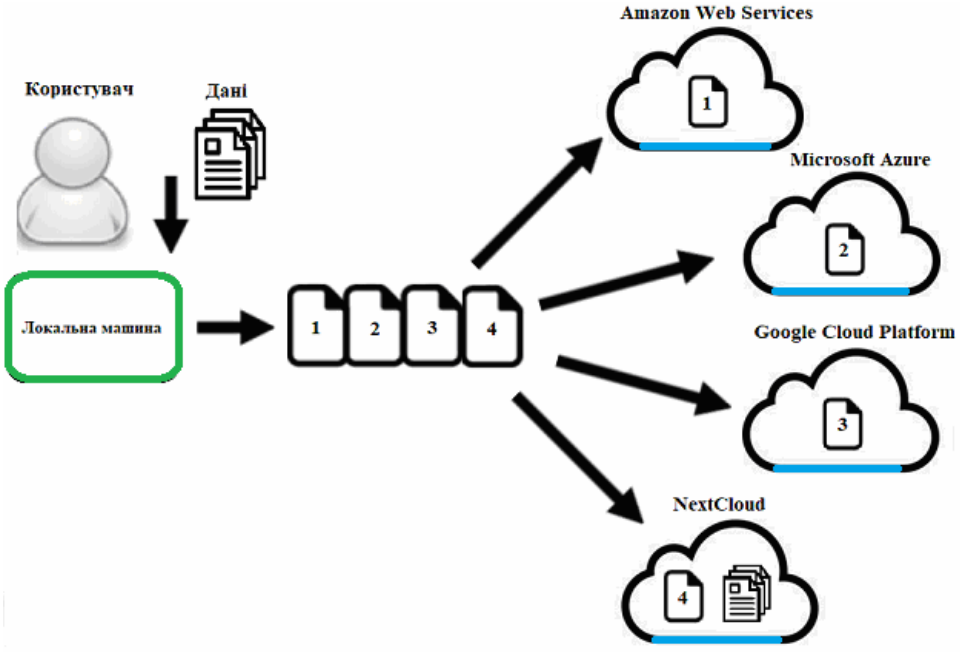


Рис. 3.4. Схема інформаційної технології захисту даних

Для початку роботи з програмним продуктом, необхідно створити акаунти на AWS, Microsoft Azure, GCP та встановити NextCloud. Після цього на кожній платформі необхідно отримати ключі доступу до API.

Для знаходження ключів Amazon S3, потрібно перейти до Консолі керування AWS, натиснути на службу S3, після цього у верхньому правому куті обрати ім’я свого облікового запису та перейти до облікових даних безпеки. Після цього згенерувати свої ключі доступу та секретні ключі. Приклад знаходження на рисунку 3.5.

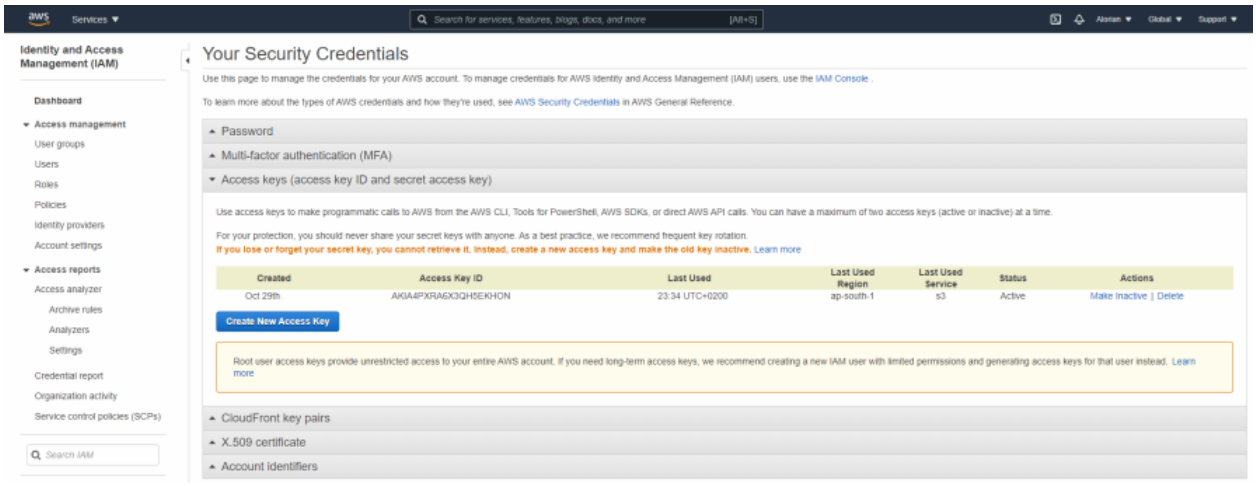


Рис. 3.5. Отримання ключів доступу на платформі AWS

Для знаходження ключів Google Storage, потрібно перейти до Google API Console, потім до роздільника Google Cloud Storage. Обираємо API & Services, у розділі знаходимо підрозділ Credentials і там знаходимо ключі доступу. Приклад на рисунку 3.6.

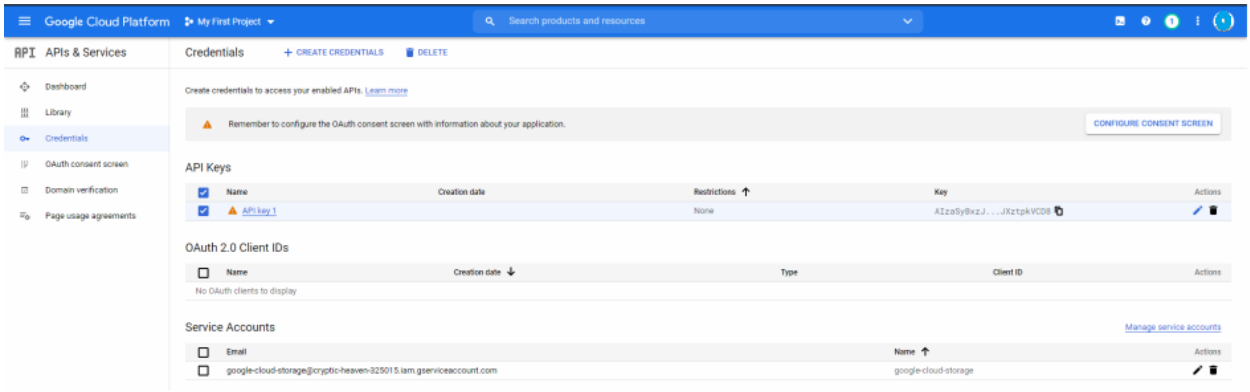


Рис. 3.6. Отримання доступу до платформи Google Cloud Storage

Для знаходження ключів Windows Azure, потрібно перейти на портал Windows Azure. Спочатку потрібно створити новий проект сховища. Після вибору цього нового проекту внизу сторінки можна знайти керування ключами. У цьому випадку ключ доступу — це назва проекту сховища, а секретний ключ — первинний ключ в управлінні ключами (рис. 3.7).

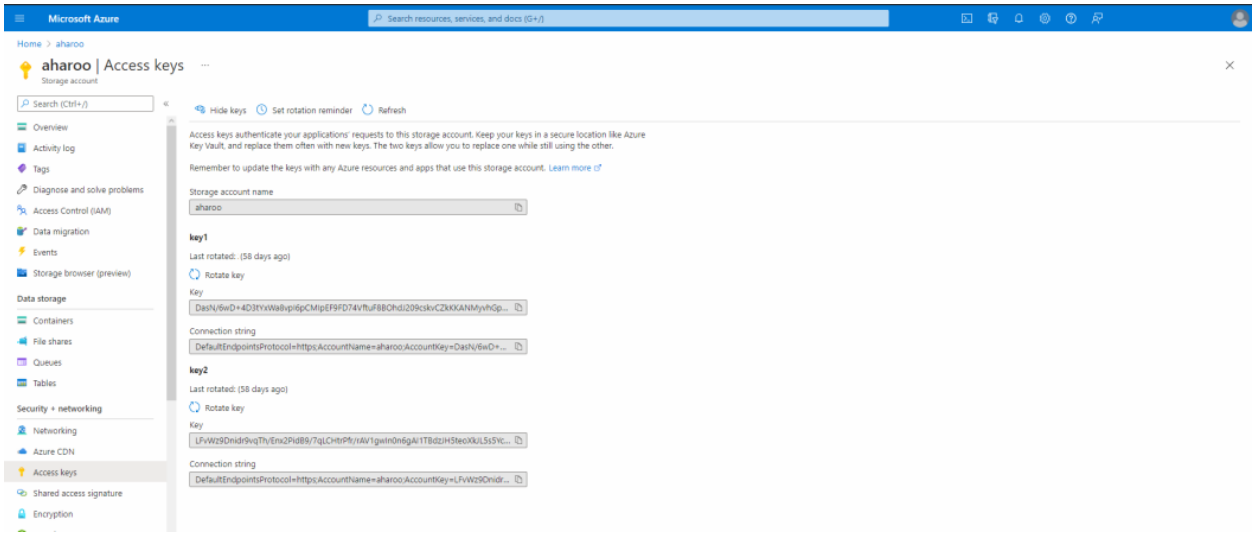


Рис. 3.7. Отримання доступу до Microsoft Azure

Для підключення до NextCloud необхідно отримати URL-адресу до WebDAV протоколу, який знаходиться в налаштуваннях у лівому нижньому кутку.

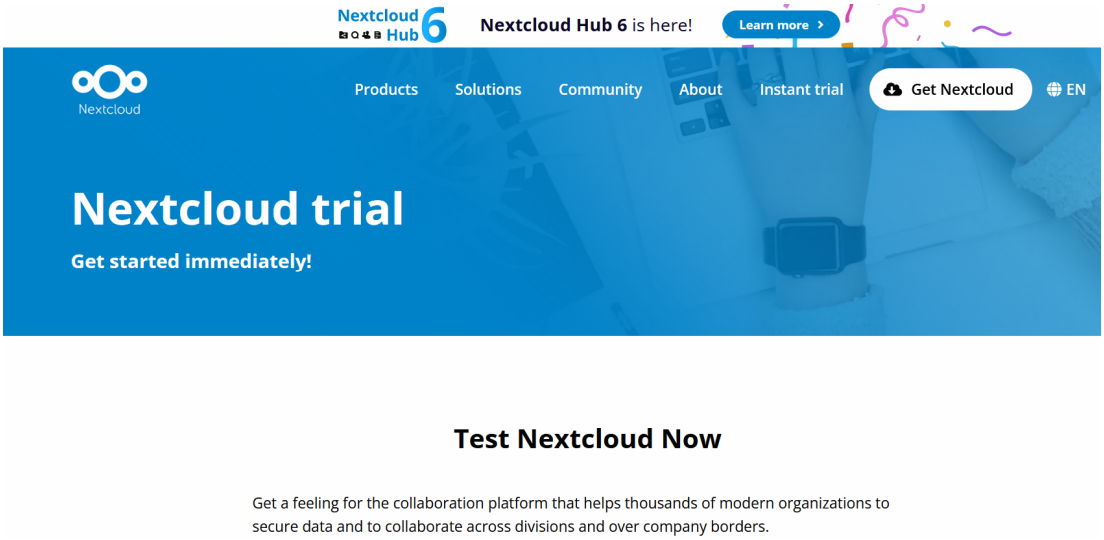


Рис. 3.8. Сервіс NextCloud

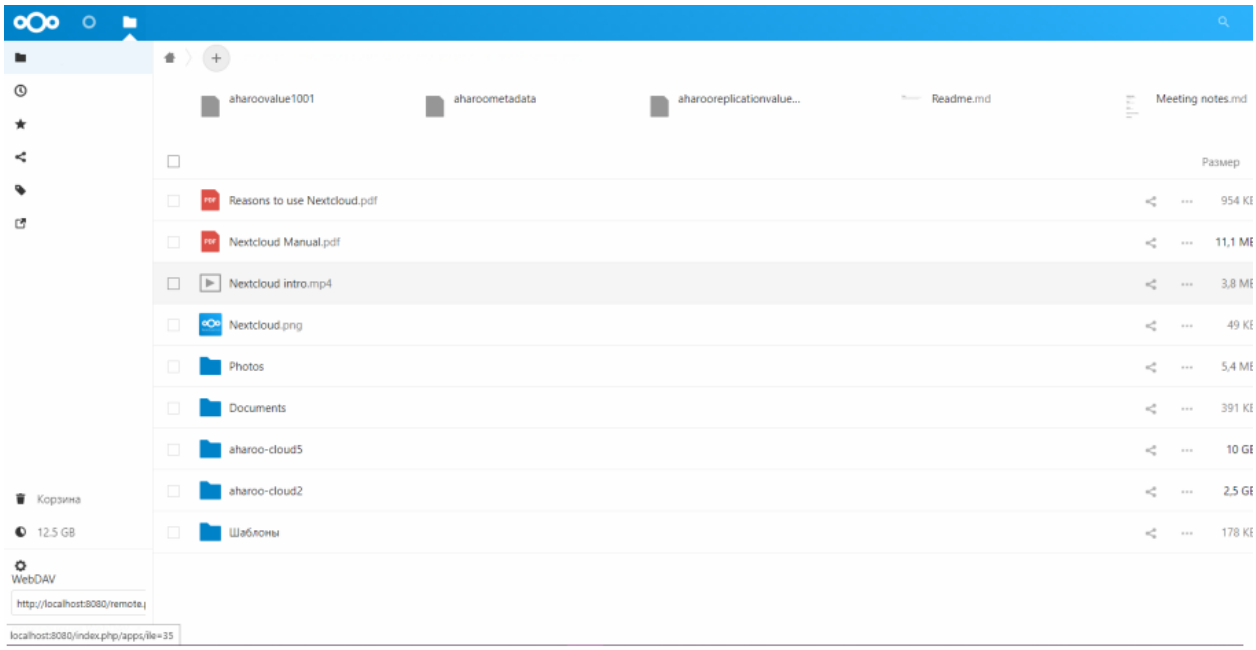


Рис. 3.9. Отримання URL-адреси засобами NextCloud

WebDAV – це протокол, який дозволяє веб-серверу діяти як файловий сервер і підтримувати спільне створення вмісту в Інтернеті. WebDAV (Web Distributed Authoring and Versioning) — це розширення протоколу HTTP, яке дозволяє клієнтам виконувати віддалені операції створення веб-вмісту, такі як завантаження та завантаження файлів, їх редагування та видалення на сервері.

Незважаючи на те, що він замінений більш сучасними механізмами, він все ще зустрічається на багатьох різних серверах, клієнтах і програмах.

В системі передбачено два режими: основний та тестовий. При виборі основного режиму використовуються чотири різні хмари. Це вже згадані Amazon S3, Azure Blob Storage, Google Storage та NextCloud. Крім того, при тестовому режимі будуть використовуватися чотири папки в приватній хмарі NextCloud.

Використання різноманітних хмар вимагає, щоб система мала справу з неоднорідністю інтерфейсів кожного хмарного сховища. Для реалізації драйверів до кожного об’єктного сховища хмарного провайдера або приватної хмари був написаний загальний інтерфейс CloudDriver.

Кожне хмарне сховище моделюється як пасивний об'єкт зберігання, який підтримує чотири операції:

- downloadData (завантажити дані з хмари в визначену директорію);
- uploadData (завантажити дані на хмару з визначеної директорії);
- deleteData (видалити окремий файл);
- deleteContainer (видалити DataUnit).

Для додавання нових файлів або завантаження існуючих необхідно обрати потрібний контейнер (DataUnit). Контейнер повинен мати заголовний блок, в який включено: унікальну назву, номер версії (для підтримки завантаження додаткових версій файлів), дані для перевірки (хеш код усієї інформації всередині файлу), цифровий підпис (для додаткової перевірки цілісності файлу). Все вищезазначене буде зберігається в файлі metadata. Сам файл розміщуються в тому ж контейнері після заголовного блоку. При завантаженні на хмари, разом з файлами завантажуються метадані на кожну хмару. В разі, якщо метадані на одній хмарі пошкодяться, їх завжди можна буде зчитати з інших хмар. Для ініціалізації та перевірки цифрового підпису буде використовуватися алгоритм RSA з хеш функцією SHA-256. Зміст метаданих та контейнера DataUnit приведено на рисунку 3.8.

Дані, що зберігаються в data unit, можуть мати довільний розмір, і цей розмір може бути різним для різних версій. Кожен data unit підтримує звичайні операції зберігання об'єктів: створення (створення контейнеру і файлу метаданих з версією 0), знищення (видалення або видалення доступу до даних з версією 0), завантаження.

Окремим питанням забезпечення захисту великих обсягів є криптографічний захист. В якості опціонного шифрування реалізовано алгоритм шифрування AES на стороні користувача як найбільш відповідний для роботи з великими даними.

На Amazon S3, Microsoft Blob Storage, Google Cloud Storage та NextCloud підключаємо шифрування на стороні сервера, в кожній з хмар автоматично підключене шифрування AES-256, додатково на Microsoft Blob

Storage можливе підключення другого рівня шифрування до даних – інфраструктурне шифрування.

Рис. 3.10. Структура контейнера DataUnit

3.4. Розробка алгоритмічного забезпечення інформаційної технології захисту даних

Діаграми поведінки поділяються на діаграму станів та діаграму діяльності. Діаграма станів зображує тільки взаємозв'язки структурного

характеру, які не залежать від часу або реакції системи на зовнішні події. Розроблена діаграма станів зображена на рисунку 3.11.

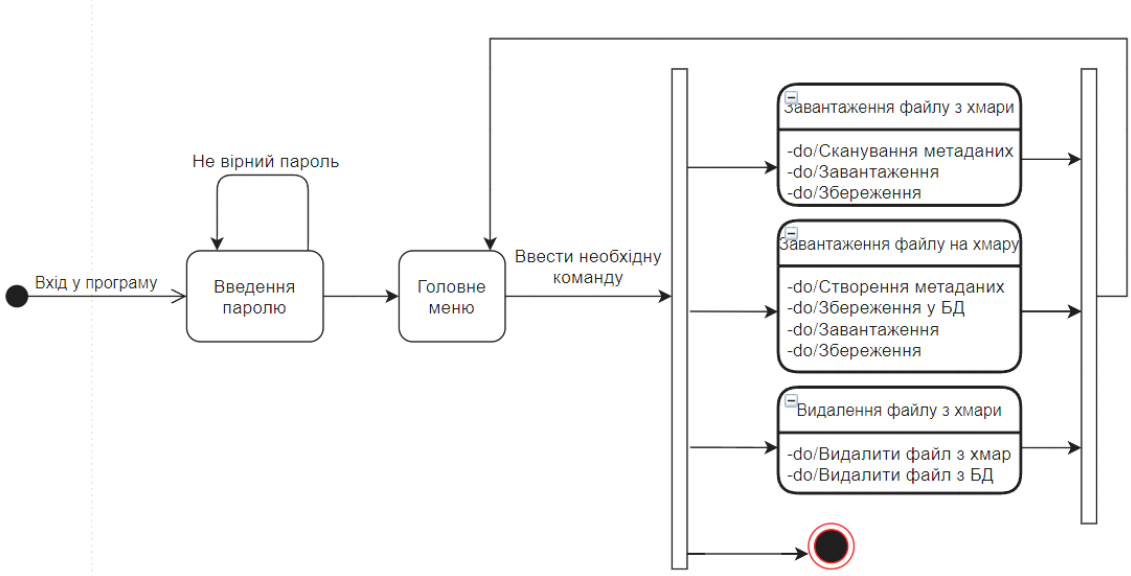


Рис. 3.11. Діаграма станів

Діаграма діяльності є графічним засобом моделювання процесів і може бути використана для аналізу процесу прийняття рішень. Діаграма візуально відображає послідовність дій, що відбуваються в процесі прийняття рішень, і може допомогти уявити потік інформації та взаємодії між різними етапами процесу. Основні елементи діаграми діяльності включають стани (елементи, які відображають конкретну діяльність або дію), роботу (стрілки, які вказують на зв'язок між станами) і рішення (елементи, які відображають вибір або альтернативу). Зазвичай використовуються прямокутники для станів, стрілки з вказівкою напрямку для роботи та ромби для рішень.

Рішення в діаграмі діяльності вказують на пункти, де потік процесу може розгалужуватись в залежності від умов або вибору. Рішення зображуються у вигляді ромбів і зазвичай мають умови або параметри, які визначають, який шлях буде обраний. Наприклад, «Чи задовольняє результат вимогам?»

Діаграма діяльності – візуальне представлення графу діяльності. Граф діяльності є різновидом графу станів скінченного автомату, вершинами якого

є певні дії, а переходи відбуваються по завершенню дій. Розроблена діаграма діяльності представлена на рисунку 3.12.

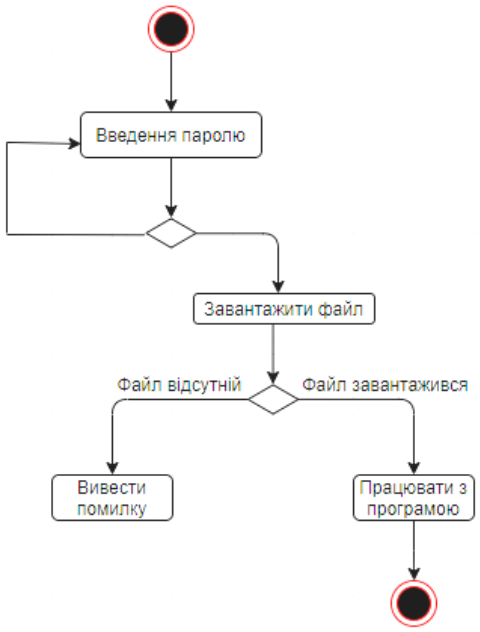


Рис. 3.12. Діаграма діяльності

Рисунки 3.11 і 3.12 показують, що робота програми починається з вводу паролю. На діаграмах видно, що програма працює з файлами та, якщо таких немає, то їх можна створити. На діаграмі станів показано, що сама робота з програмою складається з трьох основних модулів: завантаження файлу з хмари, завантаження файлу на хмару, вивести файл на екран.

3.5. Тестування інформаційної технології

В якості вхідних даних обрані текстові документи розміром від 500 МБ до 20 ГБ, враховуючи той факт, що хмарні вендори дозволяють безкоштовно користуватися об’єктними сховищами до тих пір, доки сумарний розмір файлів не перевищує 5 ГБ.

Для проведення експериментів будемо використовувати доступні команди програмного додатку: завантаження файлу на хмару, завантаження файлу на хмару з додатковою реплікацією, завантаження файлу з додатковим

шифруванням на стороні користувача та завантаження файлу з додатковим шифруванням та реплікацією.

В першому експерименті проведемо завантаження файлів на хмари, результати наведені у таблиці 3.1.

Таблиця 3.1.

Завантаження файлів на хмарної платформи

Хмарні платформи	**Розмір файлів, ГБ**			
	0,125	0,25	1,25	2,5
	Час завантаження, с			
AWS S3	18	50	328	560
Google Storage	23	45	473	867
Azure Blob Storage	60	84	538	569
NextCloud	16	54	435	622

Графічне відображення надано на рисунку 3.13.

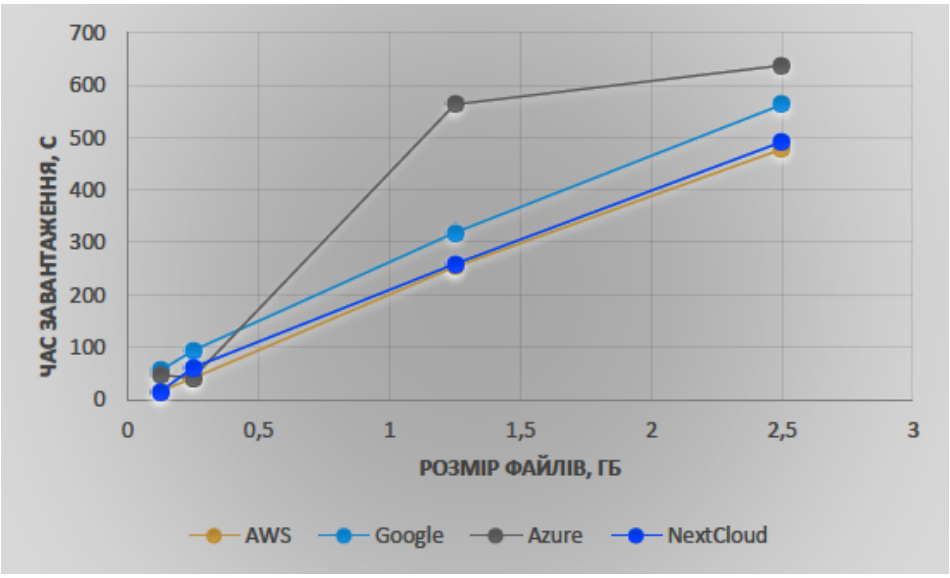


Рис. 3.13. Завантаження файлів на хмарну платформу

Як видно із результатів, Amazon S3, NextCloud та Google Storage мають достатньо схожу швидкість, й з урахуванням наявних інструментів захисту даних можуть бути застосовані для розміщення й обробки блоків даних будь я кого з обраних рівнів. У той час ресурс Azure Blob Storage слід

застосовувати для розміщення не часто змінюваних типів даних, наприклад, довідників, специфікацій або таблиць постійних значень.

В другому експерименті проведемо завантаження тих самих файлів, але вже з додатковою реплікацією, результати наведені у таблиці 3.2.

Таблиця 3.2.

Завантаження файлів на хмарної платформи з реплікацією

Хмарні платформи	**Розмір файлів, ГБ**			
	0,125	0,25	1,25	2,5
	Час завантаження, с			
AWS S3	16	42	254	478
Google Storage	55	93	319	563
Azure Blob Storage	46	42	564	638
NextCloud	15	62	259	492

Графічне відображення надано на рисунку 3.14.

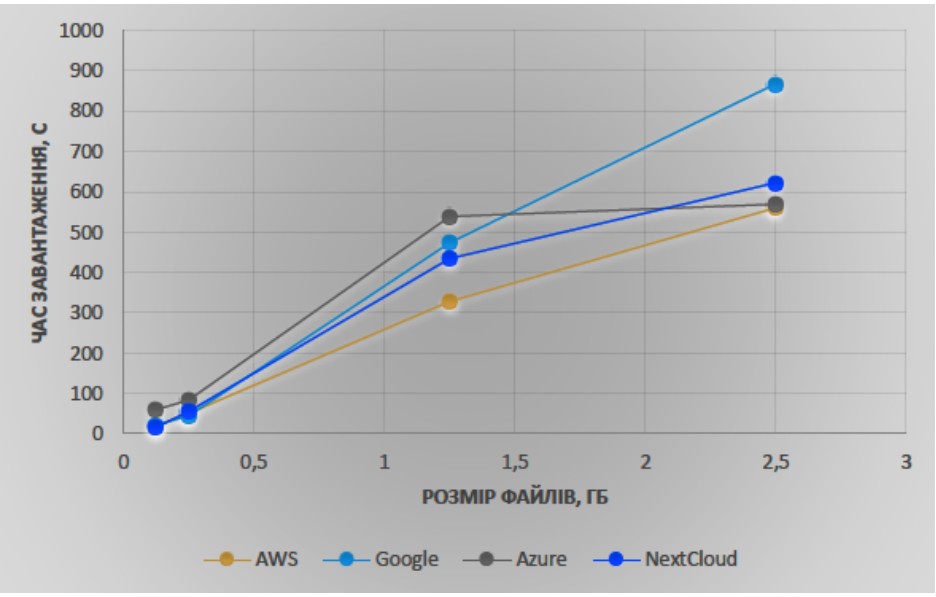


Рис. 3.14. Завантаження файлів на хмарну платформу з реплікацією

В наступному експерименті також необхідно провести тестування додаткового шифрування для файлів з боку користувача алгоритмом AES-128.

Таблиця 3.4

Швидкість шифрування

Розмір файлу, ГБ	0,5	1	5	10
Швидкість шифрування, с	1,7	2,6	43,3	152

Таблиця 3.5.

Швидкість дешифрування

Розмір файлу, ГБ	0,5	1	5	10
Швидкість дешифрування, с	1,1	2,4	106,7	237,2

Нижче на рисунку 3.15 приведено графічне відображення швидкості шифрування та дешифрування.

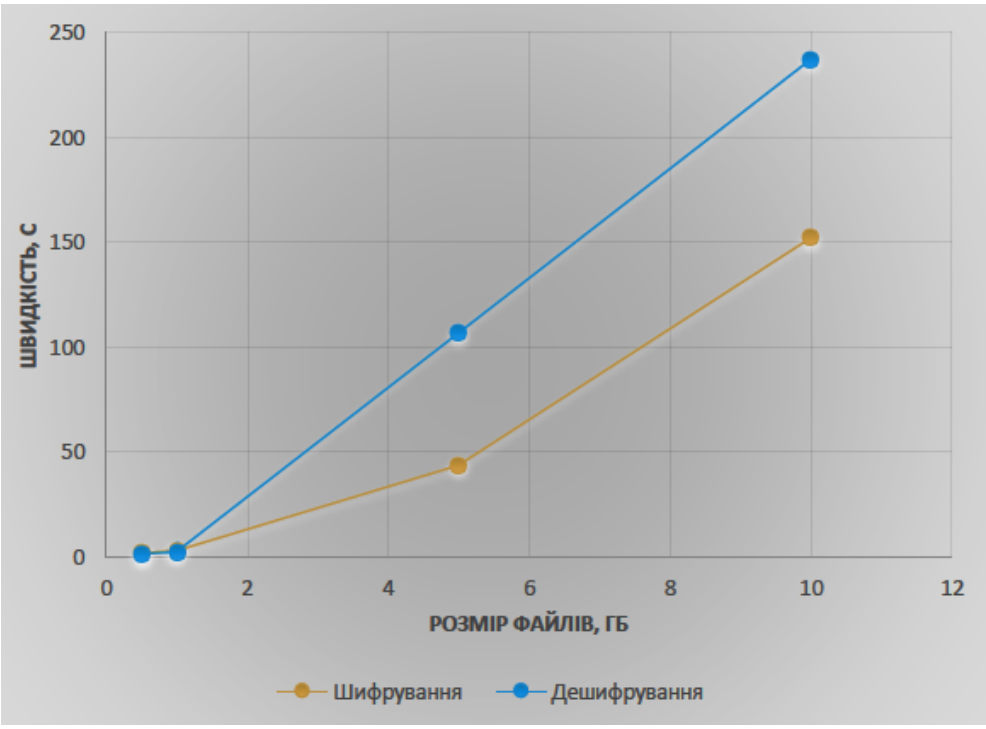


Рис. 3.15. Графічне представлення швидкості шифрування та дешифрування

Висновки до розділу

Отже, в цьому розділі представлено реалізацію іеформаційної технології захисту даних на хмарних платформах. Описано принцип

розміщення даних великих обсягів та його реалізацію методики. Побудована архітектура інформаційної технології з використанням мови візуального моделювання UML, розроблено діаграми діяльності, використання та станів. Розглянуто вхідні дані та проведено тестування інформаційної технології працездатності шляхом завантаження файлів на хмарну платформу. Проведені експерименти щодо ефективності додаткового шифрування для файлів алгоритмом AES-128.

ВИСНОВКИ

В представленій магістерській роботі досліджено криптографічні засоби реалізації концепцій безпеки даних в хмарних рішеннях.

Захист даних на хмарних ресурсах дозволяє користувачам делегувати їм частину обов'язків, що дозволяє користувачам сконцентруватися на тривалому зберіганні, не турбуючись про проблеми з конфіденційністю та цілісністю даних.

В результаті виконання магістерської роботи було зроблено наступне:

- досліджено існуючі методології, техніки та рекомендації захисту даних великих обсягів;

- проведено огляд криптографічних алгоритмів та сучасної криптографії;

- проаналізовано механізми забезпечення захисту даних засобами хмарних платформ;

- проведено опис критеріїв вибору платформ різних моделей розгортання з точки зору гарантування відповідності критеріям безпеки;

- розроблено та протестовано інформаційну технологію розміщення даних великих обсягів та їх розподіл та реплікацію на різних хмарних сервісах різних моделей розгортання.

Побудована архітектура інформаційної технології з використанням мови візуального моделювання UML, розроблено діаграми діяльності, використання та станів. Розглянуто вхідні дані та проведено тестування інформаційної технології працездатності шляхом завантаження файлів на хмарну платформу

Отже, в магістерській роботі пропонується інформаційна технологія для оцінки приватних хмар, перелік та аналіз механізмів захисту даних та критерії вибору об'єктних хмарних сховищ з точки зору гарантування відповідності критеріям безпеки.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Субач І.Ю., Фесьоха В.В. Модель виявлення аномалій в інформаційно - телекомунікаційних мережах органів військового управління на основі нечітких множин та нечіткого логічного виводу. Збірник наукових праць ВІТІ № 3 - 2017.
2. І.М. Павлов, С.В. Толюпа, В.І. Ніщенко Аналіз таксономії систем виявлення атак у контексті сучасного рівня розвитку інформаційних систем. Сучасний захист інформації №4, 2014,с. 44 -52.
3. Толюпа С.В., Штаненко С.С., Берестовенко Г. Класифікаційні ознаки систем виявлення атак та напрямки їх побудови. Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут Випуск № 3. 2018р. с. 56-66.
4. S. Report, “The 2020 Data Attack of Data by 2025 Oussama El-Hilali,” pp. 1 - 5, 2020.
5. S. Agarwal, M. Gupta, and A. Sharma, “Big Data Privacy Issues Solutions,” Proceedings of the IEEE International Conference Image Information Processing, vol. 2019-Novem, no. March, pp. 225-228, 2019, doi: 10.1109/ICIIP47207.2019.8985784.
6. G. S. Bhathal and A. Singh, “Big Data: Hadoop framework vulnerabilities, security issues and attacks,” Array, vol. 1-2, no. March, p. 100002, 2019, doi: 10.1016/j.array.2019.100002.
7. S. U. Khan and N. Ullah, “Challenges in the adoption of hybrid cloud: an exploratory study using systematic literature review,” The Journal of Engineering, vol. 2016, no. 5, pp. 107-118, 2016, doi: 10.1049/joe.2016.00.
8. Barbara, D. Detecting novel network intrusions using Bayes estimators / D. Barbara, J. Couto, S. Jajodia, N. Wu. // In: Proc. of the 1st SIAM International Conference on Data Mining. — 2001. — 17 p.
9. V. N. Inukollu, S. Arsi, and S. Rao Ravuri, “Security Issues Associated with Big Data in Cloud Computing,” International Journal of Network Security &

Its Applications, vol. 6, no. 3, pp. 45-56, 2014, doi: 10.5121/ijnsa.2014.6304.

10. Зоріна Т.І. Системи виявлення і запобігання атак в комп'ютерних мережах. Вісник східноукраїнського національного університету імені Володимира Даля № 15 (204) ч.1 2013. - с. 48-54.
11. Valdes, A. Adaptive model-based monitoring for cyber attack detection / A. Valdes, K. Skinner // In: Proc. of the Recent Advances in Intrusion Detection (Toulouse, France, 2000) — 2000. — P. 80-92.
12. Cloud Security Alliance, "Big data security and privacy handbook: 100 best practices in big data security and privacy," p. 63 pages, 2016, [Online]. Available: https://cloudsecurityalliance.org/group/big-data/
13. A. Lane, "Securing Hadoop: Security Recommendations for Hadoop Environments," Securosis, pp. 1-30, 2016, [Електронний ресурс]. – Режим доступу: https://go.thalesesecurity.com/Securosis-SecurityRecommendation-for-Hadoop-Environments.html
14. NBDPWG - Security and Privacy Subgroup, "NIST Special Publication 1500¬4: Security and Privacy," NIST Big Data Interoperability Framework, vol. 4, 2019.
15. T. B. Patil, G. K. Patnaik, and A. T. Bhole, "Big Data Privacy Using Fully Homomorphic Non-Deterministic Encryption," Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017, no. January, pp. 138-143, 2017, doi: 10.1109/IACC.2017.0041.
16. R. Oliynykov et al., "A New Encryption Standard of Ukraine : The Kalyna Block Cipher," IACR Cryptology ePrint Archive 2015, vol. 2015, pp. 1-113, 2015.
17. A. Habboush, "Multi-level encryption framework," International Journal of Advanced Computer Science and Applications, vol. 9, no. 4, pp. 130-134, 2018, doi: 10.14569/IJACSA.2018.090422.

18. Ghahramani, Z. An Introduction to hidden Markov models and Bayesian networks / Z. Ghahramani // International Journal of Pattern Recognition and Artificial Intelligence — 2001. — Vol. 15. — P. 9-42.
19. K. Rn, “Blockchain-based Secure Big Data Storage on Cloud,” International Journal of Recent Technology and Engineering, vol. 9, no. 4, pp. 37-45, 2020, doi: 10.35940/ijrte.d4744.119420.
20. D. Harinath, K. R. Babu, B. Chithra, and M. V. R. Murthy, “Encryption Techniques for Big Data in a Cloud,” pp. 413-430, 2015.
21. S. Nepal, R. Ranjan, and K. K. R. Choo, “Trustworthy processing of healthcare big data in hybrid clouds,” IEEE Cloud Computing, vol. 2, no. 2, pp. 78-84, 2015, doi: 10.1109/MCC.2015.36.
22. J. A. Shamsi and M. A. Khojaye, “Understanding privacy violations in big data systems,” IT Professional, vol. 20, no. 3, pp. 73-81, 2018, doi: 10.1109/MITP.2018.032501750.
23. H. Dev, T. Sen, M. Basak, and M. E. Ali, “An approach to protect the privacy of cloud data from data mining based attacks,” Proceedings - 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, SCC 2012, no. November, pp. 1106-1115, 2012, doi: 10.1109/SC.Companion.2012.133.
24. Kruegel, C. Bayesian event classification for intrusion detection / C. Kruegel, D. Mutz, W. Robertson, F. Valeur // In: Proc. of the 19th Annual Computer Security Applications Conference — 2003. — P. 14-23.
25. P. Suwansrikham and K. She, “Protection of big data privacy on multiple cloud providers by asymmetric security scheme,” ACM International Conference Proceeding Series, no. November, pp. 47-53, 2019, doi: 10.1145/3354153.3354156.
26. C. Rong, H. Cheng, and M. G. Jaatun, “Securing big data in the Cloud by protected mapping over multiple providers,” 2016 Digital Media Industry and Academic Forum, DMIAF 2016 - Proceedings, pp. 166-171, 2016, doi: 10.1109/DMIAF.2016.7574925.

27. G. Viswanath and P. V. Krishna, “Hybrid encryption framework for securing big data storage in multi-cloud environment,” Evolutionary Intelligence. 2020. doi: 10.1007/s12065-020-00404-w.
28. Universitas Bina Nusantara. School of Information Systems, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, ICIMTech 2020 : proceedings of2020 International Conference on Information Management and Technology (ICIMTech) : 13-14 August 2020, Indonesia.
29. A. O. Odedoyin, H. O. Odukoya, and A. O. Oluwatope, “A Quantum Cryptography Protocol for Access Control in Big Data,” International Journal on Cryptography and Information Security, vol. 8, no. 2, pp. 01-12, Jun. 2018, doi: 10.5121/ijcis.2018.8201.
30. C. Baru, Institute of Electrical and Electronics Engineers, and IEEE Computer Society, 2019 IEEE International Conference on Big Data : proceedings : Dec 9 - Dec 12, 2019, Los Angeles, CA, USA.
31. Криптографія на еліптичних кривих та її практичне застосування. – Електронний ресурс. – Режим доступу: https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/493/402
32. Collier M. Microsoft Azure Essentials Fundamentals of Azure, second edition / M. Collier, R. Shahan. - Washington: Microsoft Press, 2016. - 540 p.
33. Google Cloud Platform, “Google Cloud Security Whitepapers”. - 2018. – 3 - 75 p.
34. Nextcloud Server Administration Manual. The Nextcloud developers - 2021. - 317 p.
35. What is Java? A Beginner’s Guide to Java and its Evolution | Edureka. [Електронний ресурс]. - Режим доступу: https://www. edureka. co/blog/what-is-java/
36. Security, Identity, and Compliance on AWS. [Електронний ресурс]. - Режим доступу: https://aws.amazon.com/products/security/?nc1=h_ls

37. Docker: Empowering App Development for Developers. [Електронний ресурс]. - Режим доступу: https://www.docker.com
38. WebDAV: What it is, where it turns up, and its alternatives. [Електронний ресурс]. – Режим доступу: https://www.comparitech.com/net-admin/webdav/
39. Rath, A., Spasic, B., Boucart, N., & Thiran, P. (2019). Security pattern for cloud SaaS: From system and data security to privacy case study in AWS and azure. Computers, 8(2). https://doi.org/10.3390/computers8020034
40. AES encryption. [Електронний ресурс]. - Режим доступу: https://aesencryption.net/
41. Amazon Simple Storage Service (Amazon S3). [Електронний ресурс] - Режим доступу: https://softprom.com/ua/vendor/amazon-web-services/product/amazon-simple-storage-service-amazon-s3
42. Modesitt, Dylan, Tim Henry, Jon Coden and Rachel Lathe. “Neural Cryptography : From Symmetric Encryption to Adversarial Steganography.” - 2018. - 1 p.
43. Advanced Encryption Standard (AES). [Електронний ресурс] – Режим доступу: https://www.techtarget.com/searchsecurity/definition/Advanced-Encryption-Standard
44. What is private cloud? [Електронний ресурс]. - Режим доступу: https://www.ibm.com/cloud/learn/introduction-to-private-cloud.
45. Introduction to Azure Blob Storage. [Електронний ресурс]. - Режим доступу: https: //docs. microsoft. com/uk-ua/azure/storage/bl obs/storage - blobs-introduction
46. Google Cloud Storage. [Електронний ресурс]. - Режим доступу: https://searchstorage.techtarget.com/definition/Google-Cloud-Storage
47. Bhattacharyya, D. K. Network Anomaly Detection. A Machine Learning Perspective / D. K. Bhattacharyya, J. K. Kalita. — CRC Press, 2014. — 364 p.

48. Pedrycz W., Chen S.-M. (eds.), Information Granularity, Big Data, and Computational Intelligence, Studies in Big Data 8, DOI: 10.1007/978-3-319-08254-7, Springer International Publishing Switzerland 2015.
49. Srinivasa, S., Bhatnagar, V. (eds.): Big data analytics. In: Proceedings of the First International Conference on Big Data Analytics BDA'2012. Lecture Notes in Computer Science, vol. 7678. Springer, New Delhi, 24–26 Dec 2012.