[12] Lippert, K.J.; Cloutier, R. Cyberspace: A Digital Ecosystem. // Systems 2021, 9, 48. URL: https://doi.org/10.3390/systems9030048.
[13] Mazurczyk,W.; Drobniak, S.; Moore, S. Towards a Systematic View on Cybersecurity Ecology. URL: https: // arxiv.org/ftp/arxiv/papers/1505/1505.04207.pdf.
[14] Gorman, S.P.; Kulkarni, R.G.; Schintler, L.A.; Stough, R.R. A Predator Prey Approach to the Network Structure of Cyberspace. URL: https://www.researchgate.net/publication/255679706_A_predator_prey_approach_to_the_network_structure_of_cyberspace.
[15] Crandall J R, Ladau J, Ensafi R, Shebaro B, Forrest S, The Ecology of Malware, Proceedings of the New security paradigms Workshop (NSPW '08), pp. 99-106, Lake Tahoe, CA, USA.
[16] Fink, Glenn A., Haack, Jereme N., McKinnon, Archibald D., and Fulp, Errin W. Defense on the Move: Ant-Based Cyber Defense. United States, 2014. Web. doi:10.1109/MSP.2014.21.
[17] Lifeng Wu and Yinao Wang. Estimation the parameters of Lotka-Volterra model based on grey direct modelling method and its application. Expert Syst. Appl. 38, 6 (2011), pp. 6412-6416. URL: http://dx.doi.org/10.1016/j.eswa.2010.09.013.
[18] Diz-Pita, É.; Otero-Espinar, M.V. Predator–Prey Models: A Review of Some Recent Advances. Mathematics 2021, 1783 p. URL: https://doi.org/ 10. 3390/math9151783.
[19] S. Pohasii and other. Development of a method for assessing the security of cyber-physical systems based on the Lotka–Volterra model. Eastern-European Journal of Enterprise Technologies. 2021. 5/9 (113). pp. 30-47.
[20] Serhii Yevseiev and other. Development of a method for assessing forecast of social impact in regional communities. Eastern-European Journal of Enterprise Technologies. 2021. 6/2 (114). pp. 30-47.
[21] O. Shmatko and other. Development of methodological foundations for designing a classifier of threats to cyberphysical systems. Eastern-European Journal of Enterprise Technologies ISSN 1729-3774 3/9 (105) 2020. pp. 6-19.

SOCIOCYBERPHYSICAL SYSTEMS' SECURITY MODELS

The object of the study is the process of building multi-contour systems for the protection infrastructure elements of socio-cyber-physical systems based on a modification of the Lotka-Volterra model. The article presents the formation of security models for socio-cyber-physical systems based on the Lotka-Volterra model, which allows determining preventive measures of the security system against targeted (mixed) attacks with integration with social engineering methods and the possibility of hybridity and synergism signs. This approach allows, based on the initial data on the socio-political (economic) component, to determine the possibility of influencing the general opinion of both a separate society and certain age groups. In addition, the identification of signs of hybridity and synergism of cyber threats in the main components of socio-cyber-physical systems: social networks, the cloud and the physical component allows to determine the basic principles of building multi-contour security systems, considering the external and internal security contour systems on each platform. For the formation of multi-contour information protection systems of socio-cyber-physical systems, possible scenarios of the implementation of targeted attacks and their directionality are considered. And also, the possibility of influencing the socio-psychological state through social networks of formal and informal "leaders" of society.

Keywords: socio-cyberphysical systems, Lotka-Volterra model, hybridity, synergy, targeted attacks.

Мілевський Станіслав Валерійович, кандидат економічних наук, доцент кафедри кібербезпеки Національного технічного університету "Харківський політехнічний інститут", Україна.
Stanislav Milevsky, Ph.D., Associate Professor, Department of Cybersecurity, National Technical University "Kharkiv Polytechnic Institute," Ukraine.
E-mail: milevskiysv@gmail.com.
Orcid ID: 0000-0001-5087-7036.

DOI: 10.18372/2410-7840.25.18225
УДК 004.056.5

ВИКЛИКИ ТА СТРАТЕГІЇ ЗБЕРІГАННЯ ВЕЛИКИХ ОБСЯГІВ ДАНИХ У СУЧАСНОМУ СВІТІ

Олег Дейнека, Олег Гарасимчук

У сучасному світі, зберігання великих обсягів даних стає надзвичайно актуальною проблемою. Споживачі та організації постійно генерують великі обсяги інформації, і ця тенденція зростає. Щоб забезпечити ефективне

та безпечне зберігання цих даних, важливо розглянути виклики та стратегії, які використовуються в цій галузі. Світовою тенденцією в зберіганні даних є розширення можливостей доступу до інформації. Організації активно впроваджують різні типи доступу, такі як хмарні рішення, віртуалізацію та розподілені системи зберігання. Це дозволяє забезпечити більшу доступність та швидкість отримання даних, що є важливим у сучасному виробничому середовищі. Організації повинні дотримуватися вимог існуючих стандартів безпеки та нормативних актів для забезпечення конфіденційності, цілісності та доступності інформації. Це означає, що вони повинні встановити належні процедури, контролі та моніторинг, щоб захистити дані від несанкціонованого доступу та інших загроз. Обсяги інформації продовжать зростати, і разом з цим зростатиме і важливість забезпечення безпеки та прозорості її зберігання. Розробка нових стратегій та технологій для забезпечення цього стане докладним завданням для галузі зберігання даних у майбутньому.

***Ключові слова**: кібербезпека, великі дані, класифікація та збереження даних, стандарти SOC2 та ISO.*

ВСТУП

В сучасному світі інформація стає найцінніш им глобальним ресурсом, що потребує надійного та безпечного зберігання, а також можливості оперативного доступу до збережених даних. Чим більше в організації вагомої та критичної інформації, тим більш важливо грамотно нею керувати та зберігати. Обсяги даних зростають експоненційно, тому безпосередня проблема довгострокового зберігання великих обсягів даних є надзвичайно актуальною на сьогоднішній день, оскільки сучасне суспільство постійно зустрічається з викликами пов'язаними з вибором надійних методів та способів такого зберігання, а також захисту цих сховищ від несанкціонованого доступу до них. Тому методи та способи зберігання великих обсягів даних стають визначальними компонентами сучасної інформаційної інфраструктури. Швидкий розвиток інформаційних технологій та засобів обчислювальної техніки постійно розширює можливості зберігання даних. На сьогоднішній день більшість організацій, включаючи державні та корпоративні структури, активно впроваджують діджиталізацію як один з основних напрямків для свого розвитку. Однією з центральних складових цих процесів є безпечне та контрольоване зберігання великих обсягів різноманітних даних та їх захист від неавторизованого доступу. Також таке зберігання повинно супроводжуватися економією ресурсів на його реалізацію та надавати можливість ефективного керування цими даними.

Під великими даними розуміють набори даних, які швидко генеруються та поступають із різноманітних джерел. Такі дані можуть накопичуватися практично в кожній організації: дані про клієнтів, товари та послуги, відгуки на сайті, різноманітні опитування та їх результати. Як наслідок відбувається накопичення великого масиву даних, які як правило мають надзвичайну цінність для організації. Підвищенню ефективності, надійності та безпеки систем збереження великих обсягів даних, а також загальному розвитку даної інфраструктури сприяє стандартизація в даному напрямку. Стандарти безпеки вимагають наявності засобів контролю доступу, а також системи класифікації та збереження даних. Основним завданням цих стандартів є розуміння, що саме володіє інституція на рівні даних та як вона забезпечує контроль над їх збереженням та доступом під час проведення бізнес-процесів.

Існує багато компаній, які пропонують готові рішення для вирішення цих проблем. Проте ці рішення, з одного боку, вартують дорого, з іншого боку, потребують складності в управлінні та підтримці.

Постановка завдання полягає в аналізі сучасних методів та способів зберігання великих обсягів даних. Основною метою є визначення потреб у зберіганні та класифікації великих обсягів даних інформації, визначення проблем та викликів, які при цьому виникають та розгляд стандартів, які регламентують діяльність у даній області. Отримані результати можуть бути використані фахівцями, сфери діяльності яких потребують надійного та безпечного зберігання великих обсягів даних.

ОСНОВНА ЧАСТИНА

Великі обсяги даних застосовуються в багатьох галузях та сферах життя, оскільки наш світ стає все більше цифровим і зв'язаним з технологіями. Ось декілька прикладів, де великі обсяги даних важливі:

1. Наука: великі обсяги даних використовуються в астрофізиці, генетиці, кліматології та ін-

ших галузях науки для аналізу складних хімічних, фізичних, біологічних і екологічних процесів;

2. Інтернет речей (IoT): великі обсяги даних генеруються в результаті отримання інформації з великої кількості різних пристроїв;

3. Електронна комерція: великі інтернет-магазини та платформи збирають дані про покупців, їхні вподобання та звички для покращення маркетингу та обслуговування клієнтів;

4. Соціальні мережі: соціальні мережі обробляють великі обсяги даних про користувачів, їхню активність, зв'язки та контент, який вони споживають;

5. Фінансова сфера: банки та фінансові установи аналізують великі обсяги фінансових даних для виявлення зловживань, шахрайств, прогнозування ризиків та ринків, а також для того, щоб приймати більш виважені рішення та виявляти нові можливості;

6. Медицина: великі обсяги медичних даних використовуються в процесі діагностики, досліджень, створення реєстрів пацієнтів, при розробці нових ліків, прогнозуванні потреб медичних послуг у населення, аналізу стану здоров'я пацієнтів на локальному рівні та на рівні держави, тощо;

7. Телекомунікації: провайдери телекомунікаційних послуг обробляють, зберігають та передають великі обсяги даних;

8. Наукові дослідження та розвідка: великі обсяги даних використовуються для аналізу інформації отриманої з різних джерел для прийняття рішень у наукових та розвідувальних проектах;

9. Транспорт та логістика: дані від GPS, давачів та маршрутизація обробляються з метою оптимізації маршрутів, управління логістикою та підтримки надійного сполучення в транспортній інфрастуктурі;

10. Енергетика: великі обсяги даних використовуються для моніторингу та оптимізації електромереж, генерування та розподілу енергії;

11. Екологічні дослідження. дані про зміну клімату, екосистеми та біорізноманіття можуть бути збережені для екологічних досліджень та моніторингу;

12. Мережева безпека: логи та дані подій доволі часто зберігаються для моніторингу та аналізу роботи систем та мереж;

13. Міждисциплінарні дослідження. для зберігання інформації про складні системи різної природи;

14. Машинне навчання та штучний інтелект: для навчання та розробки алгоритмів машинного навчання та штучного інтелекту необхідні великі обсяги даних;

15. Збереження медіаконтенту: компанії в медіа та розважальній галузях зберігають великі обсяги медіаконтенту, такий як зображення аудіо та відеофайли;

16. Архівація даних: інколи є потреба у тривалому зберіганні даних, які рідко використовуються, але повинні бути збережені у відповідності із законодавчими вимогами;

17. Резервне копіювання та відновлення: великі обсяги даних можуть зберігатися для резервного копіювання та відновлення інформації у випадку збоїв чи втрат даних;

18. Розваги та хобі: для зберігання приватних колекцій медіафайлів, фотографій та інших даних.

Збільшення доступності до великих даних і розвиток аналітики дозволяють ефективно використовувати ці дані для реагування на певні виклики в режимі реального часу, відкриття нових можливостей в управлінні і покращення прийняття рішень в різних галузях.

Дослідженням проблем зберігання великих даних, впровадження нових методів, способів та заходів, підходів та стандартів присвячена велика кількість праць різних дослідників [1-8]. Існують різні підходи та принципи, які мають багато спільного між собою, а інколи сильно різняться. Проте не усі з них відповідають вимогам, що висуваються до надійності на безпечності зберігання даних. А ті, що враховують так вимоги не завжди придатні для практичного застосування у всіх випадках та для всіх типів даних.

Зокрема в [9] автори пропонують фреймворк, який здійснює фрагментацію конфіденційних даних, а потім відбувається шифрування конфіденційних даних відповідно до політики власника даних. Запропоновано структуру для отримання даних від клієнтів, аналізу отриманих даних, ідентифікації конфіденційних і неконфіденційних даних, застосування фрагментації, шифрування конфіденційних даних і, нарешті, зберігання даних. В

[10-12] досліджується технологія безпечного зберігання просторово-часових великих даних на основі блокчейну. В [13] запропоновано оригінальний метод надійного зберігання даних на основі коригуючих кодів системи залишкових класів. Важливою характеристикою систем зберігання даних є висока швидкодія запису та зчитування інформації з накопичувачів. Враховуючи, що система зберігання базується на коригуючих кодах, необхідно отримати підвищену швидкодію процесів кодування та декодування даних. Перевагою запропонованого рішення є менша надлишковість для відновлення даних, порівняно з технологією RAID та інших систем, що базуються на коригуючих кодах. Важливий вплив у зберігання великих обсягів даних вніс розвиток хмарних технологій.

Питанням ефективного та надійного зберігання великих даних в хмарних технологіях також присвячена значна кількість праць та досліджень [14-20]. Зокрема в [21] пропонується підхід до безпечного зберігання захист даних у хмарі шляхом поділу великого набору даних на блоки, що містять конфіденційні дані користувача, неконфіденційні дані та загальнодоступні дані. Конфіденційні дані переміщуються до приватної хмари та добре захищені за допомогою повторного шифрування проксі. Неконфіденційні дані зберігаються у відкритому хмарному доступі та деякі блоки даних зашифровані випадковим чином. Крім того, інформація про індекс зберігання неконфіденційних блоків даних у хмарі зашифрована та надана авторизованим користувачам. Пропонований підхід показує кращі результати з меншими обчисленнями та покращеною безпекою. В [22] запропонована схема під назвою «Модель ефективного розподіленого зберігання з урахуванням безпеки» (SA-EDS), яка в основному підтримується запропонованими авторами алгоритмами, зокрема алгоритмом альтернативного розподілу даних (AD2), алгоритмом безпечного ефективного розподілу даних (SED2) та ефективного об'єднання даних (EDCon). Експериментальні оцінки авторів оцінювали як безпеку, так і ефективність, і експериментальні результати свідчать про те, що запропонований підхід може ефективно захищати основні загрози з хмар і вимагає прийнятного часу обчислення.

Науковці, які займаються питаннями організації надійного та безпечного зберігання великих даних доволі часто пропонують використовувати різноманітні методи шифрування [23-27]. Зокрема в [28] розроблений алгоритм гібридного шифрування для забезпечення безпеки великих даних перед їх зберіганням у мультихмарі. В [29] запропонована система побудована з поєднанням генерації простих чисел, генерації ключів для ECC, генерації ключів для DSA, процесу шифрування, процесу дешифрування та процесу авторизації. Автори запропонували нову техніку пошуку альтернативного простого числа, яке корисне для генерації ключів для ECC і DSA. Також, були розроблені нові алгоритми шифрування/дешифрування на основі еліптичної кривої та поліноміальної конгруенції для виконання шифрування та дешифрування даних у процесі зберігання даних і обміну даними в хмарі. У дослідницькій роботі [30] за допомогою криптографічних алгоритмів зменшено проблеми безпеки. Запропонована система покращує безпеку в структурі хмарного сховища, використовуючи різні алгоритми шифрування, такі як алгоритм AES із S-box і алгоритм Фейстеля.

Визначення потреб у зберіганні та класифікації великих обсягів даних інформації

Аналіз потреб у зберіганні великих обсягів даних інформації є важливою частиною стратегії інформаційного управління для багатьох організацій і підприємств. Усвідомлення цього є важливим, оскільки великі дані відіграють ключову роль в аналітиці даних. Саме аналітика нам дозволяє вірно зрозуміти та інтерпретувати ці дані, щоб їх можна було використовувати для прийняття вірних та обґрунтованих рішень, прогнозування тенденцій тощо. Тут важливо зрозуміти, що сховища великих даних це не просто "велика база даних". Основна відмінність полягає у тому, що бази даних, як правило зберігають структуровані дані та мають фіксовану схему, а сховища неструктурованих даних можуть також зберігати неструктуровані дані та обробляти великі обсяги інформації.

Загалом можна виділити три способи зберігання цифрових даних:

1. Традиційний. Інформація зберігається десь на своїх власних накопичувачах, локальних сховищах. RAID-масивах тощо.

Перевагами є:

- звичка – дані завжди знаходяться поруч і це заспокоює їх власників;

- швидкий доступ. як правило до локальних носіїв легше і швидше під'єднатися;

- відносно низька вартість зберігання.

Серед недоліків варто виділити:

- мала надійність, яка спричинена тим, що через моральне та фізщичне старіння накопичувачі інформації та сервери можуть вийти з ладу, а також їх можуть викрасти зловмисники чи вони можуть зазнати пошкоджень внаслідок природних катаклізмів;

- проблеми з масштабуванням, спричинені тим, що не завжди можна спрогнозувати потреби у розмірності сховища;

- проблеми з доступом до даних на віддалі, який не завжди є зручним та безпечним;

2. В публічних хмарних середовищах. Наприклад в хмарних сховищах таких гігантів як Google, Amazon, Microsoft тощо, які надають можливість зберігати дані в хмарі за певну визначену оплату, яка залежить від обсягу даних та супутніх послуг.

В якості переваг такого способу зберігання можна виділити наступні:

- низька вартість, оскільки ціни на зберігання 1 гб інформації за місяць є доволі демократичними і спостерігається очевидна тенденція до їх зниження;

- відносна безпека. більшість постачальників хмарних сховищ забезпечують захист даних не лише користувацьким паролем, але також мають власні розроблені алгоритми шифрування. а стосовно вразливості до природніх катастроф, то існує механізм географічної реплікації даних;

- зручність, оскільки постачальники стараються по максимуму спрощувати базові сценарії роботи і дозволяють здійснювати обмін такими даними та їх зберігання на значних відстанях.

Недоліки даного способу є незначними і серед них можна відзначити:

- психологічний фактор, оскільки дані далеко від їх власників, то можуть виникати думки, що до них мають доступ інші особи;

- швидкість доступу є нижчою чим при організації локального сховища, а вартість буде більшою.

3. В приватних хмарних сховищах. Даний варіант в основному використовується для корпоративного сегменту при якому сховище є частиною інфраструктури організації та є доступним лише її співробітникам. При цьому в порівнянні з попереднім способом зберігання у власників з'являється відчуття, що вони володіють більшим контролем над процесом зберігання.

Існує три типи даних, які можуть зберігатися, і технології для їх зберігання розвиваються швидко [31, 32]:

1. Структуровані дані: це дані, організовані у таблиці з фіксованими полями, такі як інформація про клієнтів у базі даних. Такі дані зазвичай зберігаються у реляційних базах даних (наприклад, MySQL, PostgreSQL) або базах даних стовпців (наприклад, Apache Cassandra);

2. Неструктуровані дані: сюди входять тексти, зображення, відео та аудіофайли. Їх можна зберігати на файлових серверах, хмарних сховищах (наприклад, Amazon S3, Google Cloud Storage), а також в спеціальних системах управління контентом;

3. Напівструктуровані дані: до цієї категорії належать дані, які мають деяку структуру, але не суворо визначену. Наприклад, дані у форматі JSON або XML. Ці дані можна зберігати в NoSQL базах даних (наприклад, MongoDB, Couchbase).

Зберігання даних може відбуватися на різних рівнях, від локальних серверів до хмарних і гібридних рішень. Вибір конкретної технології залежить від типу даних, обсягів, потреб в доступності та безпеці, бюджету та багатьох інших факторів. Розвиток технологій для зберігання даних постійно відбувається, і нові можливості стають доступними для різних видів даних.

Структуровані, неструктуровані та напівструктуровані дані можуть бути збережені в різних типах сховищ даних залежно від їх характеристик і вимог. Розглянемо, де найчастіше зберігаються ці типи даних [31].

Структуровані дані:

1. Реляційні бази даних. Структуровані дані, як правило, зберігаються в реляційних базах даних, таких як MySQL, PostgreSQL, Microsoft SQL Server і Oracle. Ці бази даних використовують таблиці з фіксованою структурою для зберігання даних;

2. Сховища даних. Особливо великі обсяги структурованих даних можуть зберігатися в спеціалізованих сховищах даних, таких як Amazon

Redshift або Google BigQuery, які спрямовані на аналітику даних.

Неструктуровані дані:

1. Файлові сервери. Неструктуровані дані, такі як тексти, зображення, відео і аудіофайли, зазвичай зберігаються на файлових серверах. Це можуть бути локальні сервери або хмарні сховища, такі як Amazon S3 або Google Cloud Storage;

2. Об'єктні сховища. Для зберігання неструктурованих даних також можна використовувати об'єктні сховища, які дозволяють зберігати об'єкти, такі як фотографії та відео, як об'єкти з метаданими.

Напівструктуровані дані:

1. NoSQL бази даних. Напівструктуровані дані, такі як дані у форматі JSON або XML, зазвичай зберігаються в базах даних NoSQL, які дозволяють гнучку схему та обробку таких даних без строгої фіксації структури;

2. Кеш-системи. Деякі напівструктуровані дані можуть бути збережені в кеш-системах, таких як Redis, для швидкого доступу та обробки;

3. Документальні сховища: Деякі бази даних, такі як MongoDB, Couchbase і Elasticsearch, спеціалізуються на зберіганні напівструктурованих даних, особливо у форматі документів.

Важливо враховувати, що сучасні організації можуть використовувати комбінацію різних технологій для зберігання різних типів даних, щоб відповідати їхнім потребам у обробці та аналізі даних. Рішення про вибір сховища даних повинні враховувати обсяги даних, їхню структуру, вимоги до доступності та швидкодії, а також бюджетні обмеження.

Ідентифікація проблем та викликів у зберіганні великих обсягів даних

Хоча, як було наведено вище, існує велика кількість різноманітних ефективних підходів, методів та способів по організації зберігання великих даних все таки існують певні проблеми в даному напрямку. В якості суттєвого недоліку можна визначити проблему пошуку необхідної інформації в неструктурованих даних.

Також можна виділити наступні важливі виклики з якими мають справу фахівці, що займаються організацією зберігання великих обсягів даних:

- зростаючі обсяги. Постійний ріст обсягів великих даних вимагає принципово нових пристроїв, методів та алгоритмів для збереження інформації, а також збільшення потужностей для надійного зберігання;

- неоднорідність даних. Оскільки дані можуть бути різними в залежності від їх важливості, швидкості оновлення, доповнення і т.п. Все це потребує різних форматів та алгоритмів зберігання;

- швидкодія. Один з ключових параметрів, коли мова йде не лише про алгоритми та способи зберігання, а й про доступ до даних. Це в першу чергу пов’язано з тим, що дані самі по собі не будуть мати тієї цінності, якщо їх не обробляти з відповідною швидкодією. Хоча тут варто зазначити, що такий параметр як швидкодія також може бути доволі відносним: тобто те що для одних даних може вважатися достатньо швидким, для інших буде надзвичайно повільним. Тому для різних категорій інформації варто виділити різні рівні швидкодії доступу. Також варто зазначити, що запити до великих обсягів даних можуть потребувати значних обчислювальних ресурсів та оптимізації для забезпечення швидкого доступу;

- безпека. Із зростанням обсягів даних зростає і їхня цінність. Дані не можуть бути втрачені, а також необхідно виключити несанкціонований доступ до них, вплив різних видів атак, які можуть пошкодити дані чи спричинити їх витік. Тому для забезпечення безпечного зберігання необхідно використовувати шифрування, використання безпечних протоколів та реалізацію багатофакторної автентифікації та інші способи для усунення усіх небезпечних ситуацій для зберігання даних та мінімізувати наслідки негативних впливів;

- вартість. Обладнання та інфраструктура для зберігання великих обсягів як правило є досить коштовними. Вимагається велика кількість сховищ, серверів і бекапів, що призводить до високих витрат;

- простір зберігання. Великі обсяги даних потребують значного фізичного простору для зберігання обладнання та серверів. А це є доволі проблематично в обмежених приміщеннях;

- терміни зберігання. Дані терміни можуть бути встановлені залежно від типу даних, законодавчих вимог, політики організації, технічних об-

межень на зберігання, бізнес-правил, потреб у доступності, вимог власників та інших факторів;

- дублювання даних. Зростання обсягу даних може призвести до дублювання інформації, яку важливо підтримувати в актуальному стані і забезпечити її синхронізацію;

- резервне копіювання і відновлення. Зростаючий обсяг даних ускладнює процеси резервного копіювання та відновлення. Прогнозування можливих збоїв та відновлення даних постійно ускладняється;

- сумісність і формати даних. Великі обсяги даних можуть бути збережені в різних форматах та системах, і таке зберігання та обробка може ускладнюватися через несумісність;

- забезпечення відповідності. Зберігання великих обсягів даних повинно відбуватися у відповідності до законів, нормативних актів, галузевих стандартів, етичних принципів тощо. Варто пам'ятати, що деякі держави мають обмеження на зберігання та зберігання особистих даних, що може впливати на способи обробки великих обсягів даних;

- визначення категорії важливості даних. Більшість організацій відчувають труднощі при здійсненні ідентифікації критично важливих елементів даних з метою вибору підходів та способів до їх зберігання;

- охорона навколишнього середовища. Дата центри споживають приблизно 4% світової електроенергії. Тому центри зберігання та обробки даних впроваджують нові технології охолодження, оптимізують інфраструктуру та архітектуру машинних залів;

- забрудненість даними. Великі обсяги даних можуть містити надлишок застарілих або непотрібних даних, що потребують очищення та управління.

Для вирішення цих проблем важливо ретельно планувати інфраструктуру зберігання, використовувати ефективні методи компресії та архівації, впроваджувати стратегії резервного копіювання та моніторингу, і регулярно проводити оптимізацію даних.

Зі зростанням обсягів даних виникає також виклик, пов'язаний з ефективністю їхнього зберігання. Основна проблема полягає у тому, що традиційні методи та рішення для зберігання даних можуть бути неадекватними для роботи з великими обсягами інформації. Можна виділити наступні аспекти даної проблеми:

1. Обсяги даних. Великі обсяги даних означають, що організації повинні забезпечити достатньо місця для зберігання всієї інформації. Це може призвести до необхідності закупівлі дорогих серверів та масивів даних, а також збільшення витрат на зберігання;

2. Резервне копіювання. Зі зростанням обсягів даних стає складніше і витратніше забезпечувати регулярне резервне копіювання. Це може призвести до втрати важливої інформації в разі аварій або непередбачених ситуацій;

3. Швидкий доступ. Збільшення обсягів даних також може призвести до сповільнення швидкості доступу до інформації. Це може вплинути на продуктивність та реакційність системи;

4. Оптимізація ресурсів. Ефективне використання ресурсів, таких як пам'ять і обчислювальна потужність, стає важливим завданням. Неспроможність оптимізувати використання ресурсів може призвести до надмірного споживання енергії та збільшення витрат;

5. Складність управління. Збільшення обсягів даних може також призвести до зростання складності управління інформацією.

Це може вимагати більше ресурсів та фахівців для ефективного управління та обслуговування системи зберігання.

Загалом, проблема ефективності зберігання великих обсягів даних вимагає від організацій розробки та впровадження спеціалізованих рішень, які б забезпечили оптимальне використання ресурсів та забезпечили ефективний доступ до даних, зберігаючи при цьому їхню надійність та безпеку. Також зростаючі обсяги даних вимагають уважного ставлення до надійності їхнього зберігання. Ця проблема полягає у забезпеченні того, щоб дані були захищені від втрати або пошкодження, оскільки втрати даних можуть призвести до серйозних фінансових та репутаційних проблем для організації. Вирішити дану проблему можна застосовуючи:

1. Забезпечення доступності даних. Одним з основних аспектів надійності є забезпечення постійної доступності до даних. Це означає, що система зберігання повинна бути доступною для

користувачів у будь-який час без великих перерв у роботі;

2. Резервне копіювання та відновлення даних. Для захисту даних від втрати чи пошкодження, організації повинні впроваджувати системи резервного копіювання та механізми відновлення інформації. Це дозволяє відновити дані в разі аварій або несподіваних подій;

3. Захист від вторгнень. Надійність зберігання даних також передбачає захист інформації від несанкціонованого доступу. Системи зберігання повинні бути обладнані механізмами захисту, такими як шифрування та ідентифікація користувачів, щоб запобігти кібератакам та витокам даних;

4. Перевірка цілісності даних. Організації повинні періодично перевіряти цілісність даних, щоб виявити будь-які зміни чи пошкодження інформації. Це допомагає уникнути втрати даних через помилки чи дефекти в системі зберігання;

5. Моніторинг та планування надійності. Для забезпечення надійності зберігання, організації повинні проводити систематичний моніторинг стану систем та планування заходів з підвищення надійності. Це включає в себе аналіз ризиків та розробку планів аварійного відновлення.

Усі ці аспекти надійності зберігання даних великих обсягів вимагають від організацій постійного вдосконалення систем зберігання та впровадження сучасних технологій та методів для запобігання втраті чи пошкодженню даних, забезпечення безпеки і надійності інформації.

Стандартизація та вимоги в області зберігання та класифікація даних

Жодне суспільство не може існувати без законодавства та нормативних документів, які регламентують правила, процеси, методи виготовлення та контролю якості товарів, робіт і послуг, а також гарантують безпеку життя, здоров'я і майна людей та навколишнього середовища. Стандартизація якраз і є тією діяльністю, якій притаманні ці функції. Стандарти безпеки дозволяють краще розуміти, як саме інституція контролює доступ до даних і забезпечує їх безпеку та конфіденційність.

Стандарти та вимоги до зберігання даних організацій можуть варіюватися в залежності від країни, галузі діяльності організації, рівня чутливості інформації та інших факторів. Для конкретної організації можуть існувати специфічні стандарти і вимоги, які диктуються її потребами та вимогами законодавства.

Більшість організацій чи установ формує свою політику безпеки на основі міжнародних стандартів які переважно проходять за участю зовнішніх аудиторських компанії які проводять сертифікацію на відповідність до стандарту.

Можна виділити наступний перелік стандартів і типів щодо зберігання даних: SOC (Service Organization Control) [33] та ISO (International Organization for Standardization) [34]. Це два різних набори стандартів і вимог, пов'язаних із зберіганням даних та іншими аспектами інформаційної безпеки. Розглянемо деякі з основних вимог і принципів цих стандартів в контексті зберігання даних.

1. ISO 27001 (Інформаційна безпека):

- політика інформаційної безпеки. Вимагає розроблення та реалізації політики інформаційної безпеки, яка включає в себе правила та процедури для зберігання конфіденційної інформації;

- аналіз ризиків. Вимагає проведення аналізу ризиків, що допомагає визначити, які дані є найбільш чутливими та які заходи забезпечення інформаційної безпеки необхідно впровадити;

- заходи безпеки. Закликає встановлювати технічні і організаційні заходи безпеки для зберігання, доступу і обробки інформації;

2. ISO 27002 (Практичні рекомендації з інформаційної безпеки):

- фізична безпека. Вимагає встановлення заходів фізичної безпеки для захисту обладнання та носіїв інформації, таких як контроль доступу до приміщень і сейфів;

- аутентифікація та авторизація. Закликає встановлювати механізми аутентифікації та авторизації для контролю доступу до інформації;

- контроль доступу. Вимагає встановлення правил і політики контролю доступу до інформації, щоб забезпечити, що лише уповноважені особи мають доступ до даних;

3. SOC 2 (Service Organization Control 2):

- безпека і доступність. Вимагає оцінки і сертифікації забезпечення безпеки та доступності даних та послуг;

- обробка інформації. Вимагає контролю над обробкою інформації від її зберігання до передачі;

- моніторинг і відстеження. Вимагає ведення моніторингу та відстеження подій, що стосуються інформаційної безпеки.

ВИСНОВКИ

Сучасний світ надзвичайно високо цінує інформацію як глобальний ресурс. Зберігання цієї інформації вимагає надійності та безпеки, а також можливості оперативного доступу до даних. Постійне зростання обсягів даних створює виклики, пов'язані з вибором надійних методів та способів зберігання, а також захисту цих даних від несанкціонованого доступу.

Аналіз потреб у зберіганні великих обсягів даних є критично важливою частиною стратегії інформаційного управління для багатьох організацій і підприємств. Великі дані відіграють ключову роль в аналітиці даних і дозволяють приймати обґрунтовані рішення, прогнозувати тенденції та використовувати цінну інформацію.

На даний час існує розмаїття методів і рішень для зберігання великих обсягів даних, проте вони часто вимагають значних витрат і складних систем управління. Зберігання даних може відбуватися на різних рівнях, від локальних серверів до хмарних і гібридних рішень. Вибір конкретної технології зберігання повинен бути обдуманим і враховувати специфіку організації.

Також для даної області характерні певні виклики та проблеми, такі як зростаючі обсяги, неоднорідність даних, швидкодія, безпека, вартість, простір зберігання, терміни зберігання, дублювання, резервне копіювання, сумісність та інші. Для ефективного зберігання великих даних необхідно вдосконалювати методи та інфраструктуру, впроваджувати нові технології та стратегії, а також враховувати особливості кожного типу даних та розвивати шляхи для їх оптимізації. Тільки комплексний підхід, дотримання відповідних стандартів і нормативних документів та постійна оцінка проблем дозволять вирішити ці виклики та забезпечити ефективне зберігання великих обсягів інформації.

Отже, правильно обрані рішення для зберігання великих обсягів даних допомагають організаціям забезпечити доступність, надійність та безпеку своєї інформації, що є важливим компонентом успішного функціонування сучасних підприємств і організацій.

ЛІТЕРАТУРА

[1] Tretiak V.F. Оптимізація структури сховища даних у вузлах інфокомунікаційної мережі хмарного середовища / V.F. Tretiak, A.A. Pashnyeva // Системи управління, навігації та зв'язку. Збірник наукових праць. Полтава: ПНТУ, 2017. Т. 4 (44). С. 122-128. Режим доступу: http://journals.nupp. edu. ua / sunz / article / view / 390 (дата звернення: 15.10.2023).

[2] Kai, Z. (2021). Research on network data storage Technology based on Autonomous Controllable system. In 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), Shanghai, China, pp. 183-186. https://doi.org/ 10. 1109/icceai52939.2021.00035.

[3] Русин, Б. П., Погрелюк, Л. В., Висоцька, В. А., Осипов, М. М., Варецький, Я. Ю., & Капшій, О. В. (2019). Архітектура системи дедублікації та розподілу даних у хмарних сховищах під час резервного копіювання. Інформаційні технології та комп'ютерна інженерія, Т. 45(2), С. 40-63.

[4] Yatskiv, V., Kulyna, S., Yatskiv, N., & Kulyna, H. (2020). Protected Distributed Data Storage Based on Residue Number System and Cloud Services. 10th International Conference on Advanced Computer Information Technologies (ACIT), pp. 796-799. https: //doi.org/10.1109/acit49673.2020.9208849.

[5] Aujla GS, Chaudhary R, Kumar N, Das AK, Rodrigues JJ. SecSVA: secure storage, verification, and auditing of big data in the cloud environment. IEEE Commun Mag. 2018; 56(1): pp. 78-85.

[6] Vyas J, Modi P. Providing confidentiality and integrity on data stored in cloud storage by hash and meta–data approach. Int J Adv Res Eng Sci Tech. 2017; 4: pp. 38-50.

[7] Deibe David, Amor Margarita, Doallo Ramon. Big data storage technologies: a case study for web-based LiDAR visualization//IEEE International Conference on Big Data. 2018. pp. 3831-3840.

[8] Kaitai Liang, Willy Susilo, and Joseph K. Liu, "Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage," IEEE Transactions on Information Forensics and Security, vol.10, no.8, 2015.

[9] Jambi, Kamal & Eassa, Fathy & Alshomrani, Abdullah. (2017). A Framework to Secure Big Data Storage. Journal of Computational and Theoretical Nanoscience. 14. 5600-5605. 10.1166/jctn.2017.6990.

[10] Zhou, Bao & Zhao, Junsan & Chen, Guoping & Yin, Ying. (2023). Research on Secure Storage Technology of Spatiotemporal Big Data Based on Blockchain. Applied Sciences. 13. 7911. 10.3390/app13137911.

[11] Ren, Y.; Huang, D.; Wang, W.; Yu, X. BSMD: A blockchain-based secure storage mechanism for big spatio-temporal data. Future Gener. Comput. Syst. 2023, 138, pp. 328-338.

[12] Ali, S., Wang, G., White, B., & Cottrell, R. L. (2018). A Blockchain-Based Decentralized Data Storage and Access Framework for PingER. 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, pp. 1303-1308. https: // doi.org / 10.1109 / TrustCom/BigDataSE.2018.00179.

[13] Яцків, В.В., Кулина, С.В. (2019). Метод надійного зберігання даних на основі надлишкової системи залишкових класів. Вісник Хмельницького національного університету. Технічні науки, 6, С. 98-104. http: // journals.khnu.km.ua / vestnik / wp-content/uploads/2021/01/20-9.pdf.

[14] Gunasekaran Manogaran, Chandu Thota, M. Vijay Kumar, MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing, Procedia Computer Science, Volume 87, 2016, pp. 128-133.

[15] Hongbing C, Chunming R, Kai H, Weihong W, Yanyan L. Secure big data storage and sharing scheme for cloud tenants. Communications, China. 2015 Jun; 12(6):106-15.

[16] Rafique, A.; Van Landuyt, D.; Beni, E.H.; Lagaisse, B.; Joosen, W. CryptDICE: Distributed data protection system for secure cloud data storage and computation. Inf. Syst. 2021, 96, 101671.

[17] Premkamal, P.K.; Pasupuleti, S.K.; Singh, A.K.; Alphonse, P.J.A. Enhanced attribute-based access control with secure deduplication for big data storage in cloud. Peer-Peer Netw. Appl. 2021, 14, pp. 102-120.

[18] Yu, H.; Lu, X.; Pan, Z. An Authorized Public Auditing Scheme for Dynamic Big Data Storage in Cloud Computing. IEEE Access 2020, 8, pp. 151465-151473.

[19] Kannan, M.K.Jayanthi & Naik, Harish. (2021). A Survey on Protecting Confidential Data Over Distributed Storage in Cloud. SSRN Electronic Journal. 5. pp. 1-7. 10.2139/ssrn.3740465.

[20] Naik R, Mohan & Rao, Madala & Lai, Wen-Cheng & B D, Parameshachari & Babu, Justy & Hemalatha, Kivudujogappa. (2022). An Efficient and Secure Big Data Storage in Cloud Environment by Using Triple Data Encryption Standard. Big Data and Cognitive Computing. 6. 101 p.

[21] Reena, M. and Nargunam, A.S., 2019. Secured Storage of Big Data in Cloud. International Journal of Recent Technology and Engineering (IJRTE), 8(2S3), pp. 6-10.

[22] Li, Yibin & Gai, Keke & Qiu, Longfei & Qiu, Meikang & Zhao, Hui. (2016). Intelligent Cryptography Approach for Secure Distributed Big Data Storage in Cloud Computing. Information Sciences. 387 p. 10.1016/j.ins.2016.09.005.

[23] Kannan, M.K.Jayanthi & Naik, Harish. (2021). Protecting Confidential Data Over Distributed Storage in Cloud.

[24] Naik R, Mohan & Rao, Madala & Lai, Wen-Cheng & B D, Parameshachari & Babu, Justy & Hemalatha, Kivudujogappa. (2022). An Efficient and Secure Big Data Storage in Cloud Environment by Using Triple Data Encryption Standard. Big Data and Cognitive Computing. 6. 101 p. 10.3390/bdcc6040101.

[25] Kanna, G. & Vasudevan, V. (2022). An improved privacy aware secure multi-cloud model with proliferate ElGamal encryption for big data storage. International Journal of Information and Computer Security. 17. 1. 10.1504/IJICS.2022.10045413.

[26] Smriti, Manya & Venkatraman, Shruti & Raj, Aashish & Shukla, Vaishnavi & Cherukuri, Aswani Kumar. (2022). Secure File Storage in Cloud Computing Using a Modified Cryptography Algorithm. 10.4018/978-1-7998-8367-8.ch011.

[27] Atiewi, Saleh & Al-rahayfeh, Amer & Almi'ani, Muder & Yussof, Salman & Alfandi, Omar & Abugabah, Ahed & Jararweh, Yaser. (2020). Scalable and Secure Big Data IoT System Based on Multifactor Authentication and Lightweight Cryptography. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3002815.

[28] Viswanath, G. & Krishna, P. (2021). Hybrid encryption framework for securing big data storage in multi-cloud environment. Evolutionary Intelligence. 14 p.

[29] Nithisha, J. & Jayarin, P. (2022). A Secured Storage and Communication System for Cloud Using ECC, Polynomial Congruence and DSA. Wireless Personal Communications. 126 p. 10.1007/s11277-022-09778-9.

[30] Pronika and S. S. Tyagi, "Secure Data Storage in Cloud using Encryption Algorithm," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 136-141.

[31] https://medium.com/@varun.sja/structured-data-vs-unstructured-data-vs-semi-structured-data-what-is-the-difference-f0e88eaba560.

[32] "Data Science for Business" Foster Provost та Tom Fawcett 2013. 367 p.

[33] https://secureframe.com/hub/soc-2/compliance-documentation.

[34] https://www.iso.org/standard/27001.

THE CHALLENGES AND STRATEGIES OF STORING LARGE VOLUMES OF DATA IN THE MODERN WORLD

In the modern world, the storage of large volumes of data is becoming an extremely relevant issue. Consumers and organizations continually generate large amounts of information, and this trend is on the rise. To ensure efficient and secure storage of this data, it is important to consider the challenges and strategies used in this field. A global trend in data storage is the expansion of information access capabilities. Organizations actively implement various types of access, such as cloud solutions, virtualization, and distributed storage systems. This enables greater availability and speed of data retrieval, which is crucial in today's industrial environment. Organizations must adhere to existing security standards and regulations to ensure the confidentiality, integrity, and availability of information. This means that they need to establish proper procedures, controls, and monitoring to protect data from unauthorized access and other threats. The volumes of information will continue to grow, and along with that, the importance of ensuring the security and transparency of its storage will increase. Developing new strategies and technologies to achieve this will be a significant task for the data storage industry in the future.

Keywords: cyber security, big data, data classification and retention, SOC2 and ISO standards.

Дейнека Олег Романович, аспірант, спеціальності «Кібербезпека та захист інформації» Національного університету «Львівська політехніка».
Oleg Deineka, Postgraduate the Department of Information Security, National University "Lviv Polytechnic".
E-mail: oleh.r.deineka@lpnu.ua.
Orcid ID: 0009-0005-9156-3339.

Гарасимчук Олег Ігорович, к.т.н., доцент, доцент кафедри захисту інформації Національного університету «Львівська політехніка».
Oleh Harasymchuk, Ph.D., Associate Professorat the Department of Information Security, National University "Lviv Polytechnic".
E-mail: oleh.i.harasymchuk@lpnu.ua.
Orcid ID: 0000-0002-8742-8872.

DOI: 10.18372/2410-7840.25.18226
УДК 004.421.2:519.24

МЕТОДИ ОПТИМІЗАЦІЇ РОЗПОДІЛУ НАВАНТАЖЕННЯ НА ОБЧИСЛЮВАЛЬНИЙ РЕСУРС ІНФРАСТРУКТУРИ ХМАРНОГО СЕРВІСУ

Олександр Чижов, Андрій Фесенко, Микола Пустосвіт, Тетяна Німченко

У роботі досліджено методи та алгоритми оптимізації розподілу навантаження на обчислювальний ресурс інфраструктури хмарного сервісу. Зазначається, що балансування навантаження є основною проблемою серед хмарних мереж. Основною метою балансування навантаження є ефективне використання ресурсів та підвищення продуктивності. Поряд із цим воно видаляє вузли, які містять велике навантаження, а також вузли, які не працюють належним чином або виконують невелике завдання. Наголошується, що у якості базових критеріїв, пов'язаних з підвищенням ефективності балансування хмарного навантаження в реальному часі, можна виділити наступні: мінімізація витрат переміщення ресурсів і витрат виконання завдання, максимізація швидкості передачі та виконання задачі. Під якістю (ефективністю) балансування у роботі розуміється інтегральний критерій, що містить у собі істотні параметри роботи системи. Підкреслено, що математична модель динамічного розподілу віртуальних ресурсів на фізичні машини у хмарних обчисленнях, що забезпечує облік попереднього стану навантаження системи та вплив появи нового ресурсу на баланс навантаження в системі та відрізняється використанням коефіцієнта регулювання навантаження для досягнення балансування. Зазначається, що генетичний алгоритм оптимального розподілу нових віртуальних ресурсів, відрізняється реалізацією деревоподібної структури хромосом із збереженням високонавантажених вузлів, що забезпечує підвищення якості балансування навантаження та зменшення динамічного переміщення ресурсів. Наголошується, що багатокритеріальна оптимізаційна математична модель планування завдань у хмарних обчисленнях, забезпечує мінімізацію часу передачі завдань, часу виконання та витрат виконання, що відрізняється урахуванням параметрів каналу між користувачем і центром обробки даних.

***Ключові слова**: оптимізація, навантаження, хмарні обчислення, розподіл, ресурс, інформаційні технології.*