DOI: https://doi.org/10.36910/6775-2524-0560-2021-43-29
УДК 004.9
Григоренко Володимир Андрійович, провідний науковий співробітник
https://orcid.org/0000-0003-0511-3402
Український науково-дослідний інститут спеціальної техніки та судових експертиз Служби безпеки України

ОСОБЛИВОСТІ ОРГАНІЗАЦІЇ АЛГОРИТМІВ ПОШУКУ ЗА ЗРАЗКАМИ КОДОВИХ ПОСЛІДОВНОСТЕЙ

Григоренко В. А. Особливості організації алгоритмів пошуку за зразками кодових послідовностей. Проведено аналіз сучасних підходів, що застосовуються при побудові та оптимізації алгоритмів пошуку блоків даних відповідно дозаданих зразків кодових послідовностей.Значна увага приділена вирішенню задачі захисту «чутливих» даних та, відповідно, організації високофункціональної системи забезпечення інформаційної безпеки. З цією метою запропоновано адаптацію математичної моделі атрибутивного пошуку за ключовими словами у відповідності до алгоритмів полегшеного декодування і впровадженню режиму мультиавторизації. На основі вдосконаленого математичного апарату розроблено базові підходи для розробки, оптимізації і проведення оцінки на чисельному рівні алгоритмів пошуку за кодовими послідовностями ключових слів, що надає можливість автоматичного формування методологічних рекомендацій. Система оцінки ефективності і функціональності алгоритмів пошуку базується на обчисленні цільових функцій точності виділення блоку даних у відповідності до вхідного запиту, часу виконання запиту відповідно до середнього рівня затримки, та навантаження на обчислювальний ресурс апаратно-програмного комплексу інформаційної системи центру обробки даних.
Ключові слова:центр обробки даних, системи забезпечення інформаційної безпеки, пошук за ключовими словами, атрибутивне кодування, режим мультіавторізація, мультиваріантний пошук, цільові функції.

Григоренко В. А. Особенности организации алгоритмов поиска по образцам кодовых последовательностей. Проведен анализ современных подходов, применяемых при построении и оптимизации алгоритмов поиска блоков данных в соответствии с заданными образцов кодовых последовательностей. Значительное внимание уделено решению задачи защиты «чувствительных» данных и, соответственно, организации высокофункциональной системы обеспечения информационной безопасности. С этой целью предложено адаптацию математической модели атрибутивного поиска, по ключевым словам, в соответствии с алгоритмами облегченного декодирования и внедрению режима мультиавторизации. На основе усовершенствованного математического аппарата разработаны базовые подходы для разработки, оптимизации и проведения оценки на численном уровне алгоритмов поиска покодовыми последовательностями ключевых слов, предоставляет возможность автоматического формирования методологических рекомендаций. Система оценки эффективности и функциональности алгоритмов поиска базируется на вычислении целевых функций точности выделения блока данных в соответствии с входящего запроса, времени выполнения запроса в соответствии с средним уровнем задержки, и нагрузка на вычислительный ресурс аппаратно-программного комплекса информационной системы центра обработки данных.
Ключевые слова: центр обработки данных, системы обеспечения информационной безопасности, поиск, по ключевым словам, атрибутивное кодирования, режим мультиавторизация, мультивариантный поиск, целевые функции.

Hryhorenko Volodymyr. Peculiarities of the search algorithms organization based on the samples of the code sequences. The analysis of modern approaches used in the construction and optimization of algorithms for searching data blocks in accordance with the given samples of code sequences is carried out. Considerable attention is paid to solving the problem of protecting "sensitive" data and, accordingly, organizing a highly functional information security system. For this purpose, it is proposed to adapt the mathematical model of attributive search by keywords in accordance with the algorithms for lightweight decoding and the implementation of the multi-authorization mode. On the basis of the improved mathematical apparatus, basic approaches have been developed for the development, optimization and evaluation at the numerical level of search algorithms by code sequences of keywords, it provides the possibility of automatic generation of methodological recommendations. The system for evaluating the efficiency and functionality of search algorithms is based on calculating the target functions of the accuracy of the data block allocation in accordance with the incoming request, the query execution time in accordance with the average latency level, and the load on the computing resource of the hardware and software complex of the data center information system.
Keywords: data center, information security systems, keywords search, attributive coding, multi-authorization mode, multivariate search, target functions.

Вступ. Активний розвиток у сфері інформаційних технологій (IT), що спостерігається протягом останніх двох десятиріч, зокрема, широке впровадження компактних і дешевих мобільних пристроїв реєстрації і передачі даних, експоненційне зростання параметрів перепускності інформаційних каналів, обчислювального ресурсу інформаційних систем та щільності цифрового запису, поява концепцій граничних і хмарних обчислень, поєднання інформаційних ресурсів у рамках «Інтернету речей» (Internet of Things, IoT), «Інтернету транспортних засобів» (Internet of Vehicles, IoV) і, зрештою, «Інтернету всього» (Internet of Everything, IoE), призвів до росту інформаційних потоків та необхідності впровадження для центрів обробки даних (ЦОД) принципово нових політик забезпечення інформаційної безпеки. Як зазначається у сучасних дослідженнях [1-3] відповідні політики, у першу чергу, мають базуватися на стратегіях запобігання витокам інформації (Data Leak/Loss Prevention, DLP), що пов'язано з проблемою втрати, так званих, «чутливих» даних (стратегічної інформації, яка може

бути використана зловмисниками, для порушення роботи ЦОД або здійснення противоправних дій по відношенню до користувачів відповідного сервісу), що вказує на актуальність даного дослідження.

Аналіз сучасних досліджень і публікацій у галузі побудови багатофункціональної системи забезпечення інформаційної безпеки вказав на типові недоліки схем шифрування з відкритим ключем при їх адаптації для інфраструктури сучасних ЦОД, що базуються архітектурі розподілених інформаційних систем [4-6], а також пріоритет алгоритмів атрибутивного кодування блоків даних (Attribute Based Encryption, ABE). У рамках застосування алгоритмів ABE можуть бути побудовані наступні політики захисту інформаційного середовища [3, 8-9]:політика атрибутивного кодування на основі криптографічного коду (Cipher-Text Policy, CT) та політика застосування відкритого ключа(Public Key Policy, PK).Було показано, що адаптація ABE-алгоритмів відповідно до CT-політики [1, 7] ефективно застосовуються при організації розподілених інформаційних систем ЦОД, зокрема забезпечує захист «чутливих» даних [10, 15] при роботі у режимі мультиавторизації користувачів сервісу(Multi-AuthorityService, MAS).

Крім того, для побудови математичного апарату і адекватного математичного моделювання процесу машинного пошуку за вхідним набором ключових слів (Keyword Search, KS), зокрема мультиваріантного пошуку (Multi-Keyword Search, MKS) та полегшеного декодування (Light Weight Decryption, LWD)з метою розширення функціоналу користувачів та зменшення навантаження [11, 12] на обчислювальний ресурс апаратно-програмної платформи загальної системи ЦОД [3, 13].

Проведений аналіз вказав на необхідність побудови комплексної методології, що може бути використана при розробці, оптимізації та побудові системи оцінки ABE-алгоритмів, що організовано відповідно до CT-політики, що мають бути використані при організації системи захисту інфраструктури ЦОД.

Таким чином, за **мету дослідження** була поставлена розробка методів математичного моделювання процедури атрибутивного кодування у режим імультиваріантного пошуку відповідно до кодових послідовностей ключових слів, що надається користувачем сервісу ЦОД.

1. Методика оптимізації системи захищеного мультиавторизаційного доступу до сервісу ЦОД

Модель сучасної системи захисту процедури мультиавторизаційного доступу до сервісу ЦОД як було показано вище, включає у себе застосування ABE-алгоритміввідповідно до CT-політики, а також MKS- і LWD-алгоритми, а також впровадження MAS-режиму. Узагальнена схема, що надає можливість провести математичне моделювання зазначеної системи захисту «чутливих» даних ЦОД у рамках стратегії DLPпредставлена на рис. 1.

Як показано на схемі, застосування MKS- та ABE-алгоритмів, що адаптуються до CT-політики, збільшує точність виконання запиту по вхідному набору кодових послідовностей(ключових слів) за умов забезпеченням конфіденційності. У свою чергу застосування LWD-алгоритмів призводить до зниження рівня навантаження на обчислювальні ресурси ЦОД, а мультиавторизаційний доступ надає можливість застосувати схеми балансування навантаження відповідно структури апаратно-програмної платформи ЦОД та графіку надходження запитів від користувачів сервісу.

Слід окремо зазначити, що конкретна схема системи захисту ЦОД у рамках політики DLP розробляється відповідно до поставлених задач та обмежень, таким чином, функціональні елементи, зазначені на рис. 1 можуть бути вилучені або модифіковані, що необхідно врахувати при побудові математичної моделі. Оцінка її ефективності при цьому може бути визначена на чисельному рівні відповідно до показників цільових функцій.

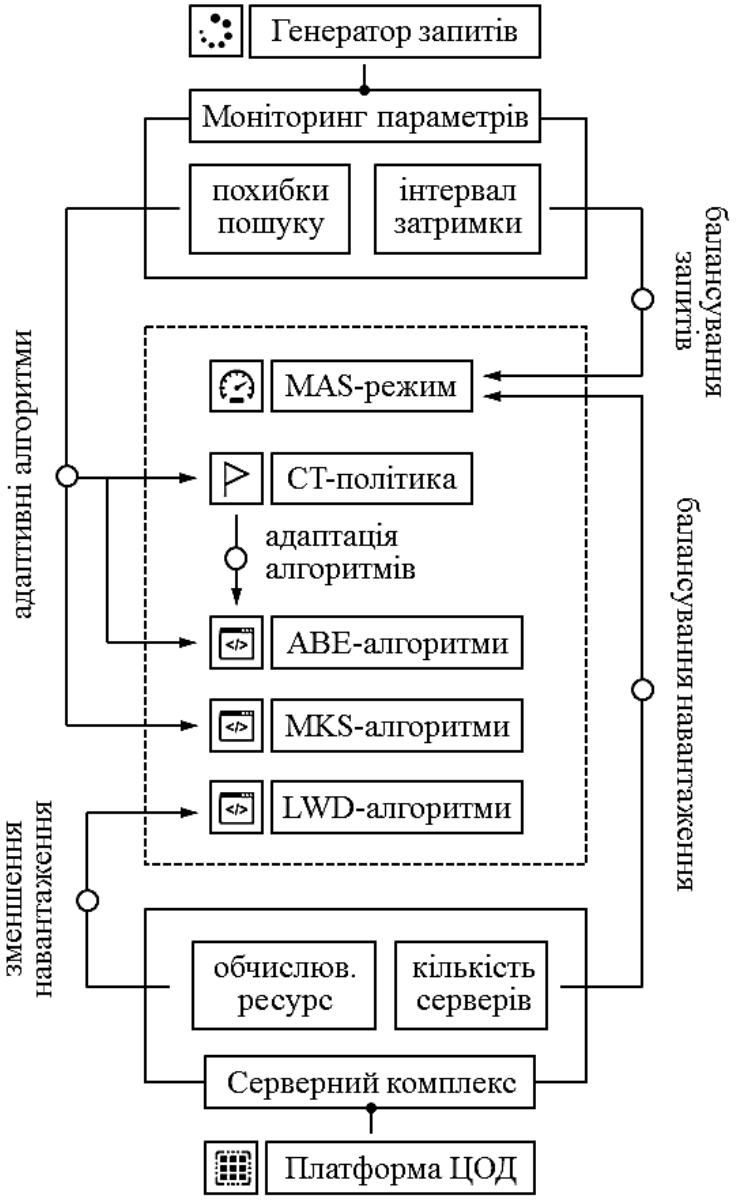

Рис. 1. Узагальнена схема побудови математичної моделі системи захисту мультиавторизаційного доступу до сервісу ЦОД

2. Побудова архітектури системи розподіленої інформаційної системи у рамках концепції IoT

У рамках даного дослідження математичну модель розподіленої інформаційної системи ЦОД пропонується побудувати відповідно загальних принципів концепції IoT. У рамках зазначеного підходу системі моніторингу відповідатиме агрегатор даних, що надходять від сенсорних вузлів інформаційної системи (Owner of Sensor Nodes Data,OSN).Аналогічно архітектура включає у себе центр ідентифікації користувачів сервісу (CertificateAuthorityCenter, CAC), що генерує адресу і глобальний ідентифікатор користувача (GlobalID, GID) та атрибутивні центри (Attribute Authority Centers, AAC), що генерують відкриті атрибутивні ключі користувача (Attribute Public Key, APK) і закриті атрибутивні ключі користувача (Attribute Secret Key,ASK). Ефективність роботи ЦОД визначається перепускністю інформаційних каналів, ємністю інформаційних сховищ та потужністю обчислювальних ресурсів, відповідно до статистичних значень (середнє, медіана, пікове, тощо) навантаження відповідно до графіку надходження запитів користувачів по пошуку за кодовими послідовностями ключових слів. Базове налаштування політики DLPзумовлює необхідність моніторингу максимального об'єму даних кожного користувача, що є доступним для аналізу відповідно початкових домовленостей у рамках угоди. Система захисту ЦОД виконує процедури декодуванню даних, що передаються разом з запитом користувача у закритому атрибутивному ключі, що у свою чергу базується на адресі користувача згідно зналаштуванням GID. На основі декодованого ASK у системі автоматично формується код доступу (рис. 2).

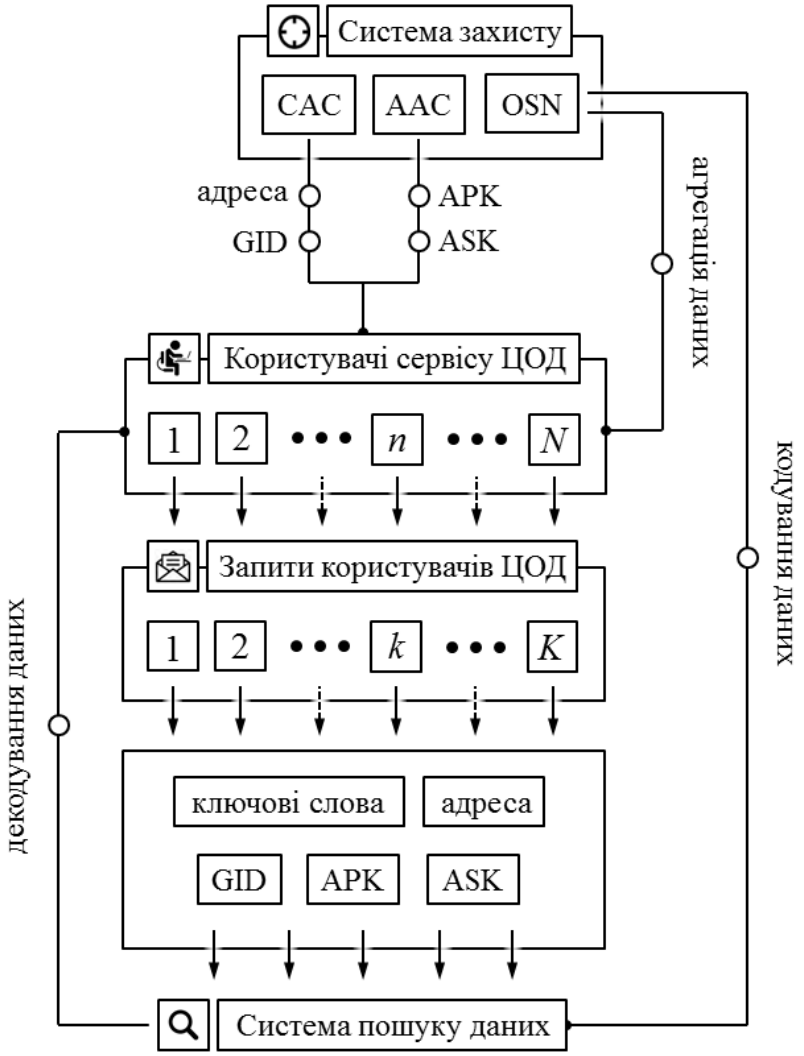

Рис. 2. Модель роботи системи захисту процесу передачі даних на базі атрибутивних центрів і центрів ідентифікації

Представлена схема вказує на необхідність формалізації процедури захисту процесу передачі даних у мультиавторизаційному середовищі ЦОД на рівні побудови відповідного математичного апарату, що базуватиметься на таких функціях і параметрах як:

F_{Σ} — функція налаштування загальної системи захисту;

S_C — параметр стійкості протоколу передачі кодування послідовностіGID від зламу сторонніми особами;

набір відкритих параметрів $\{P_m\}$, де $m \in [1; M]$;

K_S — послідовність, що відповідає закритому ключу;

P_{GID}— послідовність GID.

Таким чином основа математичного апарату складатиме визначення залежності функції налаштування загальної системи захисту, атрибутом якої є параметр стійкості протоколу передачі кодування від набору відкритих параметрів, закритого ключа та GID: $F_{\Sigma}(S_C) \sim F(\{P_m\}, K_S, P_{GID})$.

3. Математичне моделювання процедури пошуку і передачі даних у середовищі сервісу ЦОД

Як було показано у попередньому розділі, математичне моделювання процедури захисту процесу передачі даних на базі атрибутивних центрів і центрів ідентифікації базується функції налаштування загальної системи захисту, що визначається через набір відкритих параметрів, закритий ключ таGID-послідовності. Набір відкритих параметрів$\{P_m\}$, у свою чергу визначається через функцію авторизації F_A:

$$F_A(P_m) \sim F\left(K_P^{i,j}, K_S^{i,j}\right), \text{ для } \forall m \in [1; M], \text{ де } i \in [1; I], j \in [1; J]. \quad (1)$$

У рамках запропонованої схеми $K_P^{i,j}$ і $K_S^{i,j}$ — відкриті і закриті атрибутивні ключі, де через параметр J визначає загальну кількість атрибутивних центрів, а I — загальну кількість атрибутів.

Організація моделі захисту, таким чином, зумовлює необхідність побудови для кожного атрибута $i \in [1; I]$ атрибутивного ключа K_S^i, що формується на основі GID-послідовності:

$$K_S^i(P_{GID}) \sim F^i\left(K_P^{i,j}, K_S^{i,j}, \{P_m\}, P_{GID}\right) \text{ для } \forall i \in [1; I]. \quad (2)$$

Введемо функцію F_C кодування вхідного коду $\{C\}$, політику доступу P_A набір кодових послідовностей ключових слів $\{K\}$. Якщо закодований текст представити як послідовність $\{A_C\}$, що формуєтьсяна базі вхідних даних $\{A\}$та параметру захисту кодування P_C, процес кодування даних з метою захисту можна формалізувати на математичному рівні:

$$\{A_C(\{A\}, P_C)\} \sim F_C\left(\{C\}, \{K\}, \{P_m\}, K_P^{i,j}, P_A\right). \quad (3)$$

Відповідно код доступу визначається на базі параметрів закритого ключа, набору ключових слів і відкритих параметрів. Крім того модель включає у себе функцію ключ перетворення (transformationkey), що будується на основівакритого ключа, а також функції пошуку, що базується на множинах $\{A_C\}$ і $\{K\}$ та приймає набір значень $\{0; 1\}$ (де значення «1» відповідає виконанню запиту, а значення «0» блокує процедуру кодування).

Представлена математична модель може бути використана як методологічна база для побудови алгоритмів захищеного мультиваріантного пошуку наборам кодових послідовностей у середовищі ЦОД на основі визначення екстремумів цільових функцій точності виділення блоку даних у відповідності до вхідного запиту, часу виконання запиту відповідно до середнього рівня затримки, та навантаження на обчислювальний ресурс інформаційної системи ЦОД.

Висновки. В результаті проведеного дослідження було визначено та класифіковано сучасні підходи, що застосовуються при побудові та оптимізації алгоритмів пошуку блоків даних відповідно до заданих зразків кодових послідовностей ключових слів у середовищі ЦОД. Було показано, що вирішення задачі захисту «чутливих» даних користувачів ЦОДбазується на організації високофункціональної системи забезпечення інформаційної безпеки. Запропоновано методи побудови математичної моделі атрибутивного пошуку за кодовими послідовностями у відповідності до алгоритмів полегшеного декодування і впровадженню режиму мультиавторизації. Запропонований математичний апаратвключає у себе базові підходи для розробки, оптимізації і проведення оцінки алгоритмів пошуку, що надає потенційну можливість автоматичного формування методологічних рекомендацій.

References

1. Bhattacharjee, A., Borgohain, S. K., Soni, B., Verma, G., &Gao, X.-Z. (2020). Machine Learning, Image Processing, Network Security and Data Sciences Second International Conference, Mind 2020, Silchar, India, July 30 - 31, 2020, Proceedings, Part Ii. Springer Singapore.
2. Bracciali, A., Clark, J., Pintore, F., Rønne, P. B., &Sala, M. (2020). Financial Cryptography and Data Security Fc 2019 International Workshops, Voting and Wtsc, St. Kitts, St. Kitts and Nevis, February 18-22, 2019, Revised Selected Papers. Springer International Publishing.
3. Bonneau, J. (2020). Financial cryptography and data security: 24th international conference, Fc 2020, Kota Kinabalu, Malaysia, February 10-14, 2020: revised selected papers. Springer.
4. Long, J., Zhang, K., Wang, X., & Dai, H.-N. (2019). Lightweight Distributed Attribute Based Keyword Search System for Internet of Things. Security, Privacy, and Anonymity in Computation, Communication, and Storage Lecture Notes in Computer Science, 253–264. doi: 10.1007/978-3-030-24900-7_21.
5. Huang, D., Dong, Q., & Zhu, Y. (2020). Comparable Attribute-Based Encryption. Attribute-Based Encryption and Access Control, 19–42. https://doi.org/10.1201/9781351210607-2
6. Fuzzy identity and attribute based encryption for fine grained access control of encrypted data. (2018). International Journal of Modern Trends in Engineering & Research, 5 (7), 23–26. doi: 10.21884/ijmter.2018.5172.m7iz2.
7. Qiuxin, W. (2014). A generic construction of ciphertext-policy attribute-based encryption supporting attribute revocation. China Communications, 11 (13), 93–100. doi: 10.1109/cc.2014.7022531.
8. Zhenpeng, L., Xianchao, Z., &Shouhua, Z. (2014). Multi-authority Attribute Based Encryption with Attribute Revocation. 2014 IEEE 17th International Conference on Computational Science and Engineering.doi: 10.1109/cse.2014.343.
9. Li, Q., Feng, D., & Zhang, L. (2012). An attribute based encryption scheme with fine-grained attribute revocation. 2012 IEEE Global Communications Conference (GLOBECOM).doi: 10.1109/glocom.2012.6503225.

10. Yang, Y., Chen, X., Chen, H., & Du, X. (2018). Improving Privacy and Security in Decentralizing Multi-Authority Attribute-Based Encryption in Cloud Computing. IEEE Access, 6, 18009–18021. doi: 10.1109/access.2018.2820182.
11. Rane, D. D., &Ghorpade, V. (2015). Multi-user multi-keyword privacy preserving ranked based search over encrypted cloud data. 2015 International Conference on Pervasive Computing (ICPC)
12. Zhang, W., Lin, Y., Xiao, S., Wu, J., Zhou, S.: Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing. IEEE Trans. Comput. 65 (5), 1566–1577 (2016).
13. Belguith, S., Kaaniche, N., &Russello, G. (2018). Lightweight Attribute-based Encryption Supporting Access Policy Update for Cloud Assisted IoT. Proceedings of the 15th International Joint Conference on e-Business and Telecommunications.doi: 10.5220/0006854601350146
14. Wei, J., Liu, W., Hu, X.: Secure and efficient attribute-based access control for multiauthority cloud storage. IEEE Syst. J.12 (2), 1731–1742 (2018).