УДК 004.67:004.62
DOI https://doi.org/10.32782/2663-5941/2024.6.2/29

Рудницький П.Є.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

МЕТОД СТВОРЕННЯ СИСТЕМИ ЗАХИЩЕНОГО ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ НА БАЗІ ТЕХНОЛОГІЇ БЛОКЧЕЙН

У статті представлено метод розробки систем захищеного електронного документообігу, розроблених на основі технології блокчейн, з акцентом на забезпеченні конфіденційності, цілісності та доступності цифрових документів. Проведено детальний аналіз існуючих рішень у цій галузі, виявлено їхні основні недоліки, такі як вразливість сторонніх сховищ, обмежена масштабованість та складність управління доступом. Особливу увагу приділено порівнянню підходів до зберігання даних з використанням популярних підходів таки як: on-chain і off-chain, а також їхніх переваг і обмежень в контексті несанкціонованого доступу, швидкодії, вартості та рівню масштабування.

Запропонована модифікація включає гібридну архітектуру, що використовує зберігання конфіденційної інформації на блокчейні та захищене off-chain зберігання самих документів із додатковим шифруванням. Впроваджено динамічне генерування хеш-ключів на основі алгоритму Proof of Work, що дозволяє значно підвищити стійкість до атак: навіть мінімальна фальсифікація даних призводить до зміни ключа, що теоретично унеможливлює його підробку. Крім того, система забезпечує перевірку автентичності документів через Open Time Stamp, що підвищує рівень захисту та цілісності файлів у режимі реального часу.

Розглянуті підходи до захисту цифрових документів демонструють перспективність застосування блокчейн технологій у різних галузях, де необхідний високий рівень безпеки та надійності даних, таких як фінанси, медицина та державний сектор. Дослідження пропонує метод модифікації Off-chain підходу проєктування подібних систем, задля усунення ключових недоліків, що в свою чергу підвищує рівень масштабованості, надійності та захищеності в умовах сучасних кіберзагроз.

Ключові слова: *захищений електронний документообіг, динамічне генерування хеш-ключів, Proof of Work, Blockchain, Off-chain, захист конфіденційної інформації.*

Постановка проблеми. Сучасні системи електронного документообігу є одним з ключових елементів цифрової трансформації, яка активно впроваджується в різних галузях, зокрема в бізнесі, державному управлінні та інших сферах. Однак, попри численні переваги таких систем, вони супроводжуються низкою викликів, серед яких загрози безпеки даних, питання автентифікації користувачів, перевірки достовірності документів та забезпечення їх незмінності. Ці проблеми часто призводять до витоку конфіденційної інформації, несанкціонованого доступу та підробок документів.

Одним з перспективних підходів до вирішення вищезазначених загроз є використання блокчейн-технології. Вона забезпечує можливість створення децентралізованих систем, де дані зберігаються у вигляді незмінних блоків, об'єднаних у послідовний ланцюг. Кожен блок містить криптографічний хеш попереднього, що гарантує цілісність та незмінність інформації: будь-яка зміна даних в одному блоці порушує всю структуру, роблячи підробку неможливою. Блокчейн також дає змогу прозоро відстежувати всі операції з документами, що впроваджує високий рівень безпеки та довіри серед користувачів.

Проте, незважаючи на очевидні переваги, поточні підходи до використання даної технології в контексті захисту цифрових документів мають низку обмежень. Перш за все, зростання кількості документів призводить до значного збільшення розміру блокчейну, що знижує швидкодію, підвищує вимоги до обчислювальних ресурсів та значно збільшує собівартість розробки та підтримки системи в цілому. Крім того, обмежена конфіденційність даних у відкритих блокчейн-системах викликає труднощі з забезпеченням захисту особистих та комерційних документів, оскільки всі учасники мають доступ до інформації . Існує також проблема складності управління криптографічними ключами, що вимагає додаткових заходів безпеки та надійного контролю доступу.

Аналіз останніх досліджень і публікацій. Сучасні методи проектування систем дозволяють

зберігати дані як всередині блокчейну (on-chain), так і поза ним (off-chain), кожен підхід має свої переваги та обмеження. Вибір між on-chain і off-chain зберіганням залежить від вимог до безпеки, конфіденційності, продуктивності та обсягу даних. Розглянемо детальніше кожен із цих підходів, їхні переваги та недоліки.

On-chain зберігання передбачає, що всі дані записуються безпосередньо в блокчейні, що забезпечує їхню повну децентралізацію та незмінність. Це означає, що інформація доступна всім учасникам мережі та підлягає перевірці.

Переваги on-chain зберігання:

– *Незмінність*: дані, записані в блокчейні, не можуть бути змінені або видалені без зміни всієї структури блокчейну, що забезпечує їхню цілісність і автентичність [1].

– *Прозорість*: усі учасники мають доступ до даних, що підвищує рівень довіри та забезпечує відкритий доступ до інформації [2].

– *Безпека*: блокчейн-система забезпечує високий рівень захисту від зовнішніх втручань завдяки децентралізації та криптографічним методам [2].

Недоліки on-chain зберігання:

– *Масштабованість*: зростання обсягу даних на блокчейні значно уповільнює його роботу та збільшує обчислювальні витрати, що обмежує застосування on-chain зберігання для великих обсягів даних [3].

– *Конфіденційність*: у відкритих блокчейнах усі дані є доступними для всіх учасників, що ускладнює захист конфіденційної інформації, навіть при використанні шифрування [4].

– *Витрати*: зберігання великих обсягів даних на блокчейні може бути вкрай дорогим, оскільки кожен вузол зберігає повну копію блокчейну [5].

Приклади використання on-chain зберігання

Деякі проєкти використовують on-chain зберігання для невеликих обсягів критично важливих даних, таких як фінансові транзакції або дані для відстеження ланцюгів поставок, де важлива прозорість і незмінність записів [5].

Off-chain зберігання передбачає, що самі документи або великі обсяги даних зберігаються поза блокчейном, тоді як у блокчейні зберігаються лише контрольні значення або посилання на ці дані, що дозволяє зберігати великий обсяг інформації без суттєвого впливу на продуктивність блокчейну.

Переваги off-chain зберігання:

– *Масштабованість*: оскільки великі дані зберігаються поза блокчейном, це дозволяє значно зменшити навантаження на блокчейн і підвищити його продуктивність [2].

– *Конфіденційність*: off-chain сховища можна захистити додатковими методами шифрування, що підвищує рівень конфіденційності даних, які не призначені для широкого доступу [6].

– *Зменшення витрат*: зберігання даних поза блокчейном значно знижує витрати на обслуговування, особливо у випадку великих масивів інформації [4].

Недоліки off-chain зберігання:

1. *Залежність від зовнішніх сховищ*: якщо зовнішнє сховище стає недоступним або скомпрометованим, це може призвести до втрати доступу до даних або їх компрометації, що негативно впливає на загальну цілісність системи [5].

2. *Втрата децентралізації*: дані в off-chain сховищах не є частиною децентралізованого ланцюга блоків, що знижує рівень надійності та прозорості [2].

3. *Складність у забезпеченні цілісності*: для перевірки цілісності документів на основі контрольних хешів потрібні додаткові обчислення та механізми контролю, які не завжди ефективні у великих масштабах [7].

Приклади використання off-chain зберігання

Off-chain зберігання часто використовується для систем документообігу, де самі документи зберігаються у захищених зовнішніх базах даних, а на блокчейні записуються лише контрольні хеші документів для їх верифікації [8].

У відповідь на обмеження on-chain і off-chain підходів було розроблено *гібридні рішення*, що поєднують переваги обох методів. Наприклад, у роботі [2] пропонується модель, яка зберігає хеші або посилання на документи on-chain, тоді як самі документи зберігаються off-chain із додатковим шифруванням для забезпечення конфіденційності та доступності. Цей підхід дозволяє зберегти продуктивність і масштабованість системи, не жертвуючи прозорістю та безпекою даних.

Переваги гібридних рішень:

1. *Баланс масштабованості та безпеки:* використання off-chain для великих даних і on-chain для критично важливих метаданих дозволяє підвищити продуктивність і знизити витрати.

2. *Покращений захист даних*: шифрування off-chain даних і зберігання контрольних хешів on-chain забезпечує високий рівень конфіденційності та захисту від компрометації [7].

3. *Зручність перевірки цілісності*: контрольні хеші на блокчейні дозволяють легко перевірити автентичність і цілісність off-chain даних, що є критично важливим для систем електронного документообігу [8].

Недоліки гібридних рішень:

1. *Складність реалізації*: розробка та підтримка гібридної системи вимагає комплексних знань та обчислювальних ресурсів для забезпечення синхронізації між on-chain і off-chain даними [9].

2. *Витрати на обслуговування:* навіть за умови оптимізації, гібридні системи потребують високого рівня обслуговування для підтримки конфіденційності та доступності даних [10].

Приклади використання гібридних рішень

Гібридні рішення все частіше застосовуються у фінансових системах та великих організаціях, де потрібна прозорість транзакцій у поєднанні з конфіденційністю інформації.

Проаналізувавши вищезазначені підходи, можна побачити, що основною проблемою повністю децентралізованих систем є збільшення блокчейн частини, що в свою чергу значно підвищує собівартість розробки та підтримки усієї системи. У випадку з Off-chain підходом, збільшується ризик неавторизованого доступу до зовнішніх баз даних, і, як наслідок, витоку інформації. Якщо ж розглядати гібридні рішення, то складність реалізації та підтримки системи вцілому стає зависокою, і може бути невигідною та неефективною, для застосування в малих компаніях.

Постановка завдання. Метою даної статті є аналіз існуючих підходів до використання блокчейну в системах захисту цифрових документів, виявлення основних недоліків і розробка покращеного рішення, яке підвищує рівень захисту даних.

Запропонований підхід передбачає комбінування блокчейн-технології з off-chain зберіганням, що дозволяє зберігати великі обсяги даних поза основним ланцюгом блоків, забезпечуючи їхню конфіденційність і ефективність обробки. Також, застосування механізму динамічного генерування хеш-ключа доступу, дозволить впровадити додатковий шар захисту до системи вцілому та знизить ризик компрометації конфіденційних даних у випадку неавторизовного доступу до зовнішнього джерела зберігання даних.

Виклад основного матеріалу. Сфокусуємось на вищезгаданому підході проєктування подібних систем, а саме – Off-chain, оскільки, наразі, він є одним з найоптимальніших підходів, для вирішення поставленої задачі. Як вже зазначалось, даний метод має ряд переваг, у порівнянні з іншими, проте суттєвим недоліком є ризик компрометації конфіденційних даних, в разі неавторизованого доступу до стороннього джерела зберігання даних.

Задля мінімізації подібних ризиків, пропонується модифікація даного підходу наступним чином: впровадження механізму динамічного доступу до інформації між децентралізованою та сторонньою базами даних, за допомогою генерації хеш-ключа «на льоту».

Однією з основних переваг, такого рішення є стирання прямого статичного зв'язку між екземплярами документу та інформації про нього. Блокчейн-система тут буде використана в якості джерела істинності інформації, і слугуватиме для підтвердження та автентифікації дій таких як додавання файлу, зміни доступів, збереження інформації про власника і так далі. Як наслідок, подібна архітектура суттєво збільшує ентропію між даними.

Також, варто зазначити, що ключ доступу генерується за допомогою власного алгоритму Proof of Work та алгоритмів хешування таких як SHA-256 та SHA-3. Процес хешування є одностороннім і завжди має на виході єдиний результат, на однаковому наборі даних. Аналізуючи ці властивості, можна стверджувати, що навіть при незна-

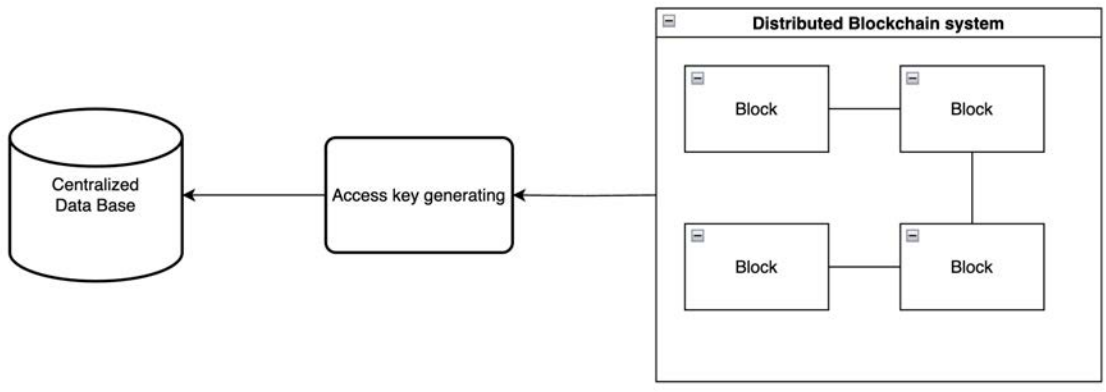

Рис. 1. Архітектура модифікованого підходу на базі Off-chain

чній фальсифікації даних в блоці, буде неможливо згенерувати валідний ключ доступу до відповідного документу.

Вищезгаданий алгоритм працює таким чином: підбираються хеш-ключі з унікальними вхідними даними до того моменту, поки не отримаємо визначену кількість нулів на початку. Для ускладнення задачі, маємо додатковий лічильник всередині циклу, який слугує для визначення чергового nonce. Тобто на кожній ітерації циклу, лічильник інкрементується та розбивається на рядок символів, з якого формується нове число, шляхом конвертування кожної цифри до її ASCII коду.

Наступним етапом є подвійне хешування за допомогою алгоритмів SHA3 та SHA256. Воно дає змогу підвищити складність генерування ключа і, як наслідок, зловмиснику доведеться обходити вже не один, а декілька алгоритмів хешування.

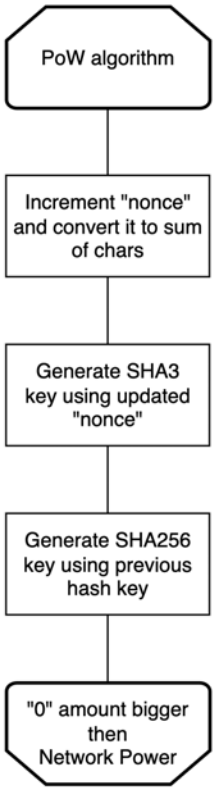

Рис. 2. Схема роботи алгоритму Proof of Work

Приймемо до уваги, що алгоритми SHA3-256 та SHA256 генерують хеш-ключ довжиною 256 бітів, а отже 64 шістнадцяткових знаки, кожен з яких представляє 4 біти і може мати значення від 0 до 15.

В найгіршому випадку, кількість ітерацій циклу може сягати $16^{64} = 2^{256}$ разів, в найкращому – 1 раз.

Введемо такі позначення:

x – кількість ітерацій циклу, де $1 < x < 2^{256}$

A – кількість нулів на початку хеш-ключа

B – кількість можливих знаків, на одному місці. Оскільки обчислення проводяться в шістнадцятковій системі, це число завжди дорівнюватиме 16

P – ймовірність випадкового підбору nonce

N – потужність мережі

Загальна формула для визначення ймовірності випадкового підбору nonce, і як наслідок, компрометації хеш-ключа матиме такий вигляд:

$$P(A) = \frac{B^{\log_B(x_{max})-A}}{X_{max}} = \frac{16^{\log_{16}(16^{64})-A}}{16^{64}} = \frac{16^{64-A}}{16^{64}} \quad (1)$$

Таблиця 1

Ймовірність компрометації хеш-ключа

N – потужність	A – кількість нулів	B – кількість знаків	P – ймовірність
10	1	16	16^{-1}
100	2	16	16^{-2}
1000	3	16	16^{-3}
10000	4	16	16^{-4}
100000	5	16	16^{-5}

Провівши аналіз роботи даного алгоритму, можна зробити висновок, що він має досить високий рівень захищеності від зовнішніх атак. На практиці, майже неможливо виконати випадковий підбір даних та згенерувати валідний ключ доступу, не знаючи принципу його роботи.

Для перевірки автентичності самого документу, пропонується використовувати Open Time Stamp – стандарт таймштампів в блокчейн екосистемі, який дозволяє генерувати унікальне значення для файлу, використовуючи його бінарне представлення, метадані та поточну дату створення.

При успішному знаходженні екземпляру документу в базі даних, проводитиметься також перевірка вищезазначеним стандартом, що в свою чергу унеможливлює фальсифікацію.

Нижче наведено блок-схему кінцевого алгоритму отримання документу з системи.

Кінцевий алгоритм працює таким чином: до системи надходить відповідний запит на отримання документу з UUID пов'язаного блоку в блокчейн-системі. Якщо такий блок існує, перевіряється право власності документу або ж право на користування ним. У разі успішної перевірки, динамічно обчислюється хеш-ключ доступу до бінарного представлення документу, з використанням алгоритму Proof of Work. Якщо документ з таким ключем існує, перевіряємо його автентичність за допомогою стандарту Open Time Stamps. У разі успішної перевірки, інформація з блоку поєднується з самим документом і стає доступною кінцевому користувачу.

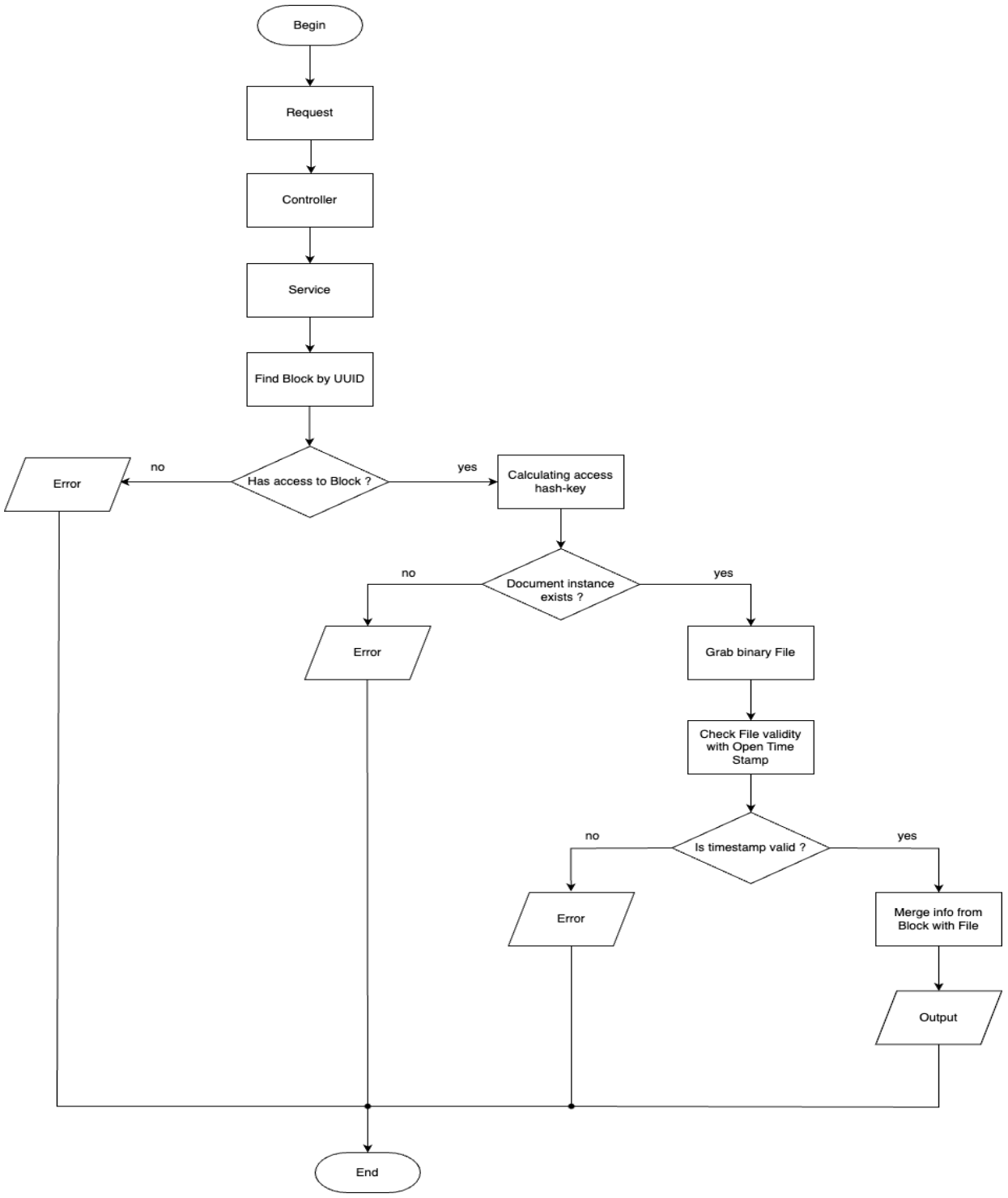

Рис. 3. Блок-схема алгоритму доступу до документу в системі

Висновки. Таким чином, запропонований підхід дозволяє ефективно запобігати фальсифікації, забезпечує надійність даних та знижує витрати на підтримку і розробку системи. Оскільки для зберігання файлів не потрібно збільшувати обсяг блоку, кожна транзакція стає менш витратною. Конфіденційна інформація про документи зберігається в блокчейні у зашифрованому вигляді, що вирішує проблеми публічності даних у блоках та захищає від несанкціонованого доступу.

Основна ідея дослідження полягала в модифікації off-chain підходу з додаванням нового рівня захисту між розподіленою та централізованою базами даних. З цією метою було впрова-

джено механізм динамічного генерування ключа доступу, що підвищує ентропію між елементами даних у системі та забезпечує високий рівень зв'язаності.

Додатково, було розроблено та застосовано власний алгоритм Proof of Work та продемонстровано теоретичні обрахунки його ефективності в контексті ймовірності компротеції хеш-ключа. Також було застосовано стандарт Open Time Stamps задля перевірки автентичності документу.

Переваги запропонованого рішення:

– Збільшена ентропія даних.

– Збережена продуктивність і ефективність off-chain підходу.

– Знижено собівартість управління системою порівняно з повністю децентралізованими рішеннями.

– У разі несанкціонованого доступу до сторонньої бази даних конфіденційна інформація про документ залишається недоступною завдяки вбудованому захисту блокчейн-модуля.

Хоча метод розроблений для систем електронного документообігу, його можна адаптувати й для інших сфер, що потребують захищеного зберігання документів, таких як: укладання різних юридичних угод, сертифікація та акредитація, перевірка ланцюгів постачання в логістиці – все це вимагає зберігання документів та потребує високого рівня захищеності з підтвердженням права власності.

Список літератури:

1. Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. An overview of blockchain technology: Architecture, consensus, and future trends // Proceedings of the IEEE. – 2017. – Vol. 105, № 9. – P. 2047–2063. URL: https://ieeexplore.ieee.org/document/8029379 (дата звернення 15.08.2024).

2. Zyskind, G., Nathan, O., & Pentland, A. Decentralizing privacy: Using blockchain to protect personal data // Proceedings of the IEEE Security and Privacy Workshops. – 2015. – P. 180–184. URL: https://doi.org/10.1109/SPW.2015.27 (дата звернення 20.08.2024).

3. Antonopoulos, A. M. Mastering Bitcoin: Unlocking Digital Cryptocurrencies. – 2nd ed. – O'Reilly Media, 2017. – 416 p.

4. Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. Bitcoin and Cryptocurrency Technologies. – Princeton University Press, 2016. – P. 2047–2063.

5. Christidis, K., & Devetsikiotis, M. Blockchain and Smart Contract for the Internet of Things // IEEE Access. – 2016. – Vol. 4. – P. 2292–2303. URL: https://doi.org/10.1109/ACCESS.2016.2566339 (дата звернення 14.10.2024).

6. Schneier, B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. – 2nd ed. – Wiley, 1996. – P. 301–315.

7. Casino, F., Dasaklis, T. K., & Patsakis, C. A systematic literature review of blockchain-based applications: Current status, classification and open issues // Telematics and Informatics. – 2019. – Vol. 36. – P. 55–81. URL: https://doi.org/10.1016/j.tele.2018.11.006 (дата звернення 29.10.2024).

8. Goyal, S. Comparison of Machine Learning Techniques for Software Quality Prediction // International Journal of Knowledge and Systems Science. – 2020. – Vol. 11, № 2. – P. 20–40.

9. Tian, F. An agri-food supply chain traceability system for China based on RFID & blockchain technology // Service Systems and Service Management (ICSSSM). – 2016. – P. 225–253. URL: https://doi.org/10.1109/ICSSSM.2016.7538424 (дата звернення 20.11.2024)

10. Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F.-Y. Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends // IEEE Transactions on Systems, Man, and Cybernetics: Systems. – 2019. – Vol. 49, № 11. – P. 62–84. URL: https://doi.org/10.1109/TSMC.2019.2899270 (дата звернення 25.11.2024).

Rudnitskiy P.Ye. METHOD FOR CREATING A SECURE ELECTRONIC DOCUMENT MANAGEMENT SYSTEM BASED ON BLOCKCHAIN TECHNOLOGY

The article presents a method for developing secure electronic document management systems based on blockchain technology, with an emphasis on ensuring the confidentiality, integrity and availability of digital documents. A detailed analysis of existing solutions in this area is carried out, identifying their main shortcomings, such as the vulnerability of third-party storage, limited scalability and complexity of access control. Particular attention is paid to comparing data storage approaches using popular on-chain and off-chain approaches, as well as their advantages and limitations in the context of unauthorized access, performance, cost and scalability.

The proposed modification includes a hybrid architecture that uses the storage of confidential information on the blockchain and secure off-chain storage of the documents themselves with additional encryption. Dynamic generation of hash keys based on the Proof of Work algorithm has been introduced, which significantly increases resistance to attacks: even minimal data falsification leads to a change in the key, which theoretically makes it impossible to counterfeit. In addition, the system provides authentication of documents through Open Time Stamp, which increases the level of protection and integrity of files in real time.

The considered approaches to digital document protection demonstrate the prospects of applying blockchain technologies in various industries that require a high level of data security and reliability, such as finance, medicine, and the public sector. The study proposes a method of modifying the off-chain approach to designing legacy systems to eliminate key shortcomings, which in turn increases the level of scalability, reliability, and security in the face of modern cyber threats.

Key words: *secured electronic document management, hash keys generation, Proof of Work, Blockchain, Off-chain, confidential information protection.*