УДК 004.056.5 **DOI:10.30837/rt.2024.3.218.02**

В.О. ПОДДУБНИЙ, Р.Ю. ГВОЗДЬОВ, О.В. СЄВЄРІНОВ, канд. техн. наук

ВИКОРИСТАННЯ НУЛЬОВИХ ВОДЯНИХ ЗНАКІВ ДЛЯ ПІДТВЕРДЖЕННЯ АВТОРСТВА ЗОБРАЖЕНЬ ТА БАГАТОФАКТОРНОЇ АВТЕНТИФІКАЦІЇ

Вступ

З розвитком інформаційних систем, з інтеграцією та цифровізацією держави, цифровою трансформацією суспільства, реалізацією сервісів, послуг в електронному форматі питання захисту систем, що обробляють інформацію, стає все більш актуальним. Реалізація даних послуг та процесів потребує безліч адміністраторів, операторів, обслуговуючого персоналу, які поділяються за функціями, можливостями, обов'язками тощо. Проте, такий ріст цифровізації сприяє збільшенню кількості атак на інформаційно-комунікаційні системи (далі – ІКС). Однією з найрозповсюдженіших атак є фішинг [1], який направлений на кінцевого користувача чи оператора, тобто найпростіший метод злому – це не намагатися обійти засоби захисту ІКС, знайти вразливості, заразити ІКС вірусом, а отримати облікові дані адміністратора чи користувача. Одним із методів захисту від крадіжки акаунтів та облікових засобів в ІКС є використання багатофакторної автентифікації. Нульові водяні знаки досить новий метод, що з'явився на заміну «звичайним» цифровим водяним знакам. Нульові водяні знаки першочергово повинні забезпечувати перевірку авторства, проте все частіше з'являються ідеї їх використання для автентифікації [2]. Даний метод підтвердження авторства та автентифікації досить гнучкий та універсальний, існують різні нульові водяні знаки для зображень, тексту чи програмного коду [3], засновані на різних математичних моделях.

В даній роботі висуваються наступні цілі:

– здійснення огляду нульових водяних знаків;

– модифікація існуючого алгоритму нульового водяного знаку для його тестування;

– визначення можливості використання нульового водяного знаку для підтвердження авторства зображень;

– пропозиція використання нульового водяного знаку в схемі багатофакторної автентифікації.

Водяні знаки, розповсюдження та використання для підтвердження авторства

Цифровий водяний знак – технологія, створена для захисту авторських прав мультимедійних файлів. Зазвичай цифрові водяні знаки невидимі. Однак ЦВЗ можуть бути видимими на зображенні або відео [4]. Зазвичай ця інформація являє собою текст або логотип, який ідентифікує автора.

Цифрові водяні знаки доцільно використовувати в багатьох сферах діяльності, де корисно зв'язати деяку додаткову інформацію (метадані) з контейнером даних (об'єктом, в який вбудовується цифровий водяний знак). Ці метадані можуть бути вбудованим як водяний знак. Є й інші способи асоціювати інформацію з об'єктом, наприклад, розміщення його в заголовку цифрового файлу, кодування у видимому штрих-коді на зображенні, QR-коди, цифрові підписи та печатки та ін.

В подальшому при розгляді цифрового водяного знаку буде матися на увазі цифровий водяний знак для зображення.

Цифрові водяні знаки мають дві основні характеристики для порівняння: стійкість і непомітність. Стійкість – це характеристика, яка визначає можливість вилучення водяного знака після додавання до контейнера шуму, спотворення або стиснення. Непомітність є характеристикою ефекту вбудованого знака на контейнер. Оптимальне значення невидимості – це коли водяний знак не впливає безпосередньо на контейнер. На рис. 1 показано приклад цифрового водяного знаку для перевірки авторства зображення. В такому випадку в контей-

нер (зображення) вбудовується додаткова інформація про власника, а до контролюючого органу передається кінцева інформація про зображення.

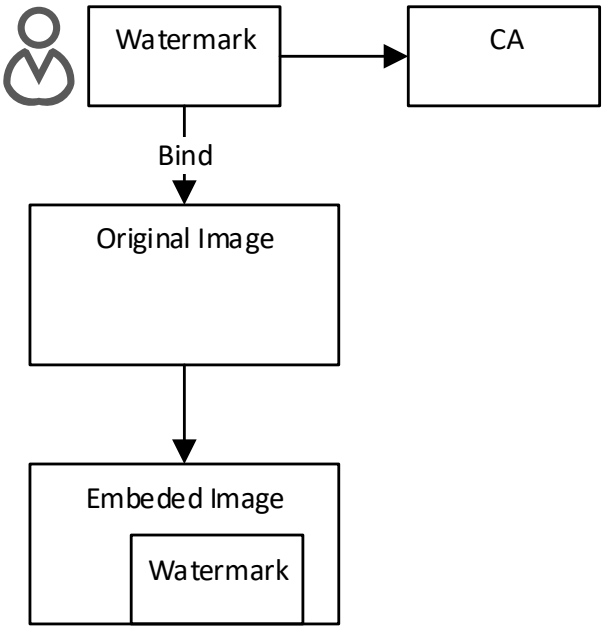

Рис. 1. Спрощена схема ідентифікації володільця за допомогою цифрового водяного знаку

Проте «класичні» водяні знаки мають такий істотний недолік, як спотворення оригінального зображення при вбудові водяного знаку. Це може бути важливо, наприклад, для медичних даних [5, 6]. Також такі водяні знаки сильно спотворюються при зміні форматів зображення, стисканні, форматуванні.

Нульові водяні знаки, їх використання для підтвердження авторства

Zero digital watermarking (нульовий водяний знак) – це метод створення водяного знаку для зображень, який має на меті максимально зменшити вплив водяного знаку на візуальну якість зображення. У цьому методі використовується підхід "zero visibility" (нульова видимість), який означає, що водяний знак невидимий для людського ока.

Цей підхід корисний в різних областях, таких як захист авторських прав на зображення, відстеження поширення зображень в Інтернеті, інформаційна безпека та інші. Zero digital watermarking дозволяє додавати ідентифікуючу інформацію до зображень, не руйнуючи їх зовнішній вигляд, і може бути корисним інструментом для вирішення проблем з підробкою та незаконним використанням зображень.

Основна перевага методу нульового водяного знаку полягає в тому, що водяний знак не вбудовано в саме зображення, тому, на відміну від традиційних методів водяних знаків, не вносить жодних змін до зображення, таким чином уникаючи будь-яких спотворень зображення. Приховані функції витягуються з головного зображення та поєднуються з водяним знаком (якась прихована інформація, наприклад логотип), після чого шифруються та створюється ключ. Потім секретний ресурс має зберігатися в довіреному органі для майбутнього вилучення водяних знаків. Таким чином, вилучення внутрішньої репрезентативної інформації про ознаки з даних зображення є основним підходом нульових водяних знаків [7].

Графічне зображення нульового цифрового знаку показано на рис. 2.

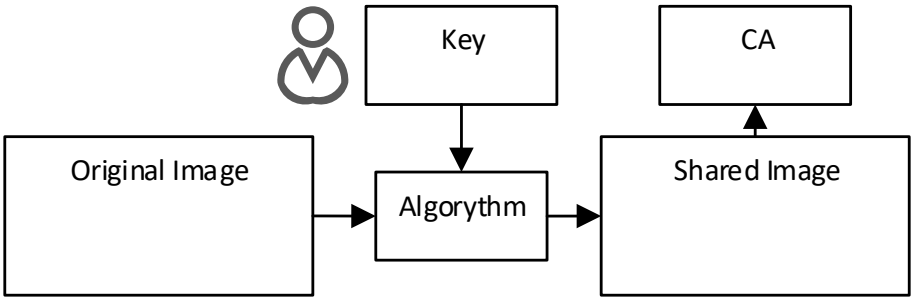

Рис. 2. Спрощена схема ідентифікації володільця за допомогою цифрового водяного знаку

Алгоритм нульового водяного знаку

В роботі [7] запропоновано дві схеми нульового водяного знаку для зображень. Для тестування та демонстрації алгоритму модифіковано першу схему. Схему даного алгоритму зображено на рис. 3.

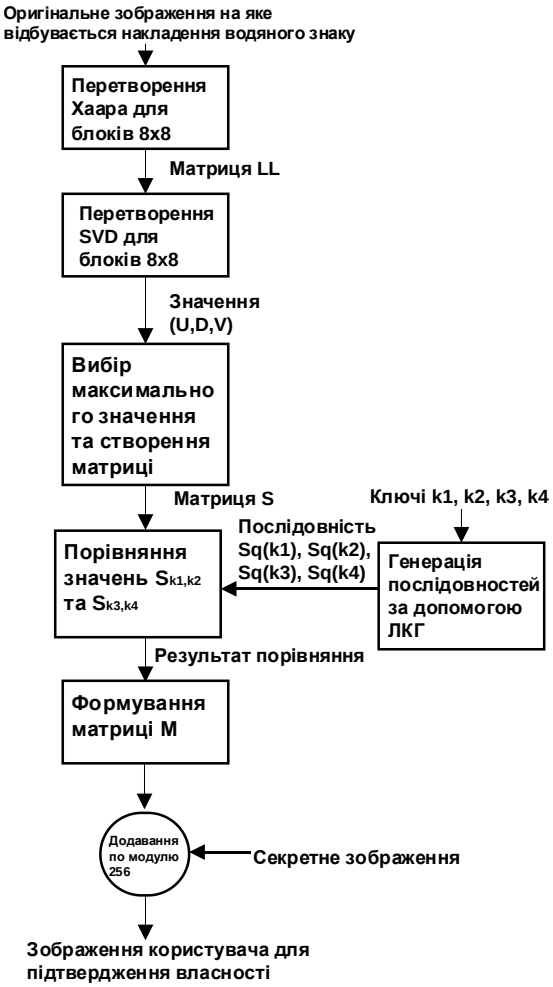

Рис. 3. Запропонована схема алгоритму

Таким чином, алгоритм складається з наступних кроків:

1) Оригінальне зображення (O) розбивається на блоки 8x8.

2) Для кожного блоку 8x8 виконується однорівневе перетворення Хаара [8], з якого в подальшому використовується високочастотний спектр LL. Спектр LL має розмір *m/2*n/2*.

3) Для кожного блоку 8x8 виконується CVD перетворення [9].

4) З отриманих коефіцієнтів перетворення CVD: U,D,V^T вибирається максимальне значення, з якого створюється матриця S.

5) Генеруються ключі, які задають стан ЛКГ key1, key2, key3, key4.

6) Відбувається генерація чотирьох послідовностей k1, k2, k3, k4. Оскільки ці коефіцієнти будуть встановлювати значення кожного пікселю, то довжина послідовностей повинна складати *m/2*n/2*. Послідовність k1, k3 складається з діапазону *[0;m/2]*. Послідовність k2, k4 складається з діапазону *[0;n/2]*, генерація відбувається за допомогою алгоритму $k_{i+1}=(a* k_i+1) \bmod C$.

7) Відбувається порівняння двох вибраних значень D_i та B_i з матриці M. Значення вибираються як $D_i = S[k1_i,k2_i]$; $B_i=S[k3_i,k4_i]$.

8) На підставі вибраних коефіцієнтів складається Master Share M. Пікселі нумеруються зліва направо, зверху вниз, кожен піксель має номер в і *[0;(m/2)*(n/2)]* Якщо значення D_i не дорівнює значенню B_i, то значення пікселя в Master Share обчислюється як

$M_i=(k1_i + k2_i) \bmod (256)$

в іншому випадку

$M_i=(k4_i + k4_i) \bmod (256)$

в кінцевому випадку отримується майстер-зображення для кожного спектру M.

9) Генерується секретне зображення (S) так, що $M+S \bmod (256) = Sig$, де *Sig* – зображення користувача, що зберігається в засвідчуваному центрі.

Використання нульових водяних знаків для підтвердження авторства

Нульові водяні знаки з'явилися як засіб підтвердження авторства, проте з часом були проведені дослідження, які показують складність використання нульових знаків для цієї мети. В роботі [10] проведено дослідження можливості використання нульових водяних знаків для захисту авторства медичних даних.

Після тестування запропонованого вище алгоритму було встановлено наступні результати при вилученні водяного знаку, що наведені в табл. 1

Таблиця 1

№	1		2		3	
	PSTR	NCC	PSTR	NCC	PSTR	NCC
1	-	1	-	1	-	1
2	23,60	0,92	20,36	0,76	24,95	0,85
3	16,05	0,90	14,14	0,66	19,05	0,82
4	14,89	0,90	13,15	0,66	18,21	0,81

Для перевірки водяного знаку використовувалося дві метрики:

– NCC;

– PSTR.

Значення NCC визначається як

$$NCC = \frac{\sum_{i=1}^{n}(A_i-\bar{A})(B_i-\bar{B})}{\sqrt{\sum_{i=1}^{n}(A_i-\bar{A})^2 \sum_{i=1}^{n}(B_i-\bar{B})^2}} \ . \quad (1)$$

Значення NCC визначає нормальну кореляцію між зображанням *A* та *B*.

Значення $\bar{A}$ та $\bar{B}$ становлять середні значення пікселів зображення *A* та *B*.

Математичне представлення *PSNR* виглядає наступним чином (1):

$$PSNR = 20log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right), \quad (2)$$

де MAX_f – максимальне значення сигналу, яке існує в вихідному зображенні; *MSE* – (середня квадратична помилка) [11].

Отримавши аналогічні результати, що і автори [10], можна зробити висновки що нульові водяні знаки мають наступні особливості:

– складність забезпечення авторства в подібних зображеннях;
– складність забезпечення авторства різних зображень одного автора.

Ці особливості, на нашу думку, не дозволяють використовувати нульові водяні знаки для підтвердження авторства.

Схема багатофакторної автентифікації на основі нульового водяного знаку

Як описано вище, нульові водяні знаки не підходять для підтвердження авторства, проте можуть бути корисними для схем авторизації.

Схема ґрунтується на двох ознаках:

1. Знання чогось: пароль.
2. Володіння чимось: зображення.

Додатковою відмінністю алгоритму є ще один фактор, що включає ознаку

1. Знання чогось: чи є фото користувача ключем.

Дана ознака включає в себе знання того, яке саме з фото є ключем. Користувач може зберігати безліч фото на своєму носії інформації, проте зловмисник не має інформації про те, чи є якесь фото на носії користувача ключем. Отримавши доступ до носія, зловмисник побачить лише фото/набір фото без додаткової інформації.

Наприклад, на телефоні користувача збережений набір фото, як на рис. 4 [12]:

Рис. 4. Приклад використання зображення в якості ключових даних

Одне з цих фото може бути ключем для водяного знаку, але й може не бути, зловмисник не знає, яке саме фото є ключем, та чи є взагалі.

Ця ознака перетворює даний алгоритм на трьохфакторний та дозволяє розмити інформацію на носії користувача.

Проте це дійсно лише для унікальних фотографій, оскільки якщо фото буде неунікальне та неконфіденційне, то ознака володіння користувачем втрачається. В такому випадку користувачу слід самому вибрати фото, що послаблює стійкість алгоритму (оскільки користувач може вибрати неунікальне фото). Теоретичним рішенням є генерація фото за допомогою штучного інтелекту, це також теоретично може включати процес відновлення фото, оскільки ШІ генеруватиме фото на основі ключових фраз та солі (наприклад ID користувача в системі або інша сіль КЗЗ), що слугуватимуть для відновлення ключа.

Запропонована схема складається з наступних компонентів:

– алгоритм розширення ключа;
– генератор геш-значень;
– алгоритм нульового цифрового знаку;
– центр сертифікації (опціонально);
– засоби зчитування зображення (опціонально).
– засіб збереження фото (опціонально).
– модуль автентифікації.

Розглянемо кожний модуль окремо.

Алгоритм розширення ключа розширює ключ на основі пароля користувача. Пароль користувача є вектором ініціації для генерації псевдовипадкової послідовності, під час тестування було використано лінійно конгруентний генератор випадкових послідовностей

проте було встановлено залежність запропонованого алгоритму від ключів згенерованих за допомогою ЛКГ. Вхідні дані: пароль користувача. Вихідні дані: розгорнутий ключ на основі паролю користувача, що відповідає вимогам алгоритму нульового водяного знаку.

Пароль користувача розподіляється на чотири частини, кожна з частин проходить через генератор геш-значень, та слугує вектором ініціації для ЛКГ, що генерує ключі key1, key2, key3, key4. Такий спосіб генерації ключових послідовностей дозволяє забезпечити стійкість від зворотної атаки на пароль. Спрощена схема розгортання ключів зображена на рис. 5.

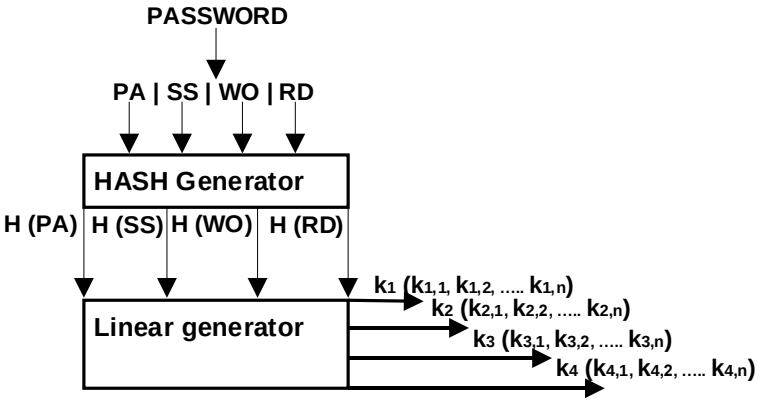

Рис. 5. Схема розгортання ключів для запропонованого алгоритму

Може бути використаний інший алогічний алгоритм. Вхідні дані для алгоритму: ключ користувача, ключове зображення користувача. Вихідні дані: зображення отримане в результаті вилучення водяного знаку.

Генератор геш-значень – використовує пароль користувача для генерації геш-послідовності, що перевіряється КЗЗ. В якості генератора можна використовувати будь-яку стійку геш-функцію, таку як SHA-2, SHA-3, Купина тощо. Вхідні дані: пароль користувача. Вихідні дані: геш-сума паролю користувача.

Центр сертифікації (опціонально). Центр сертифікації здійснює порівняння результатів вилучення водяного знаку з власними даними для кожного користувача. Центр сертифікації використовується опціонально, так як його функції може виконувати локальний модуль в ІКС, проте він може використовуватися, коли в ІКС необхідно здійснювати зовнішню ідентифікацію або коли необхідно використання довіреної третьої сторони. Центр сертифікації використовує власний ключ для генерації кінцевого водяного знаку, який порівнюється з еталонною копією водяного знаку. Після порівняння центр сертифікації повертає Модулю доступу числове значення достовірності (на скільки відсотків вірне зображення користувача). На основі цього коефіцієнту Модуль доступу приймає рішення про правильну/хибну ідентифікацію. Вхідні дані: вилучений водяний знак користувача. Вихідні дані: коефіцієнт достовірності.

Модуль доступу – є модулем, що обробляє та приймає рішення про доступ користувача до системи. Модуль доступу зберігає (або має доступ) до геш-значень паролів користувача. Модуль доступу перевіряє геш-значення паролю користувача та отриманий в результаті вилучення водяного знаку коефіцієнт схожості. Оскільки оригінальне зображення може (в залежності від типу його передачі в ІКС) містити помилки, то можливі погрішності в коефіцієнті схожості. Необхідний поріг налаштовується в самому модулю доступу, це може бути як повна схожість (значення = 1), так і якийсь коефіцієнт (наприклад, 0.95). Вхідні дані: геш-значення пароля користувача, коефіцієнт достовірності водяного знаку.

На рис. 6 зображена спрощена схема автентифікації з використанням нульового водяного знаку.

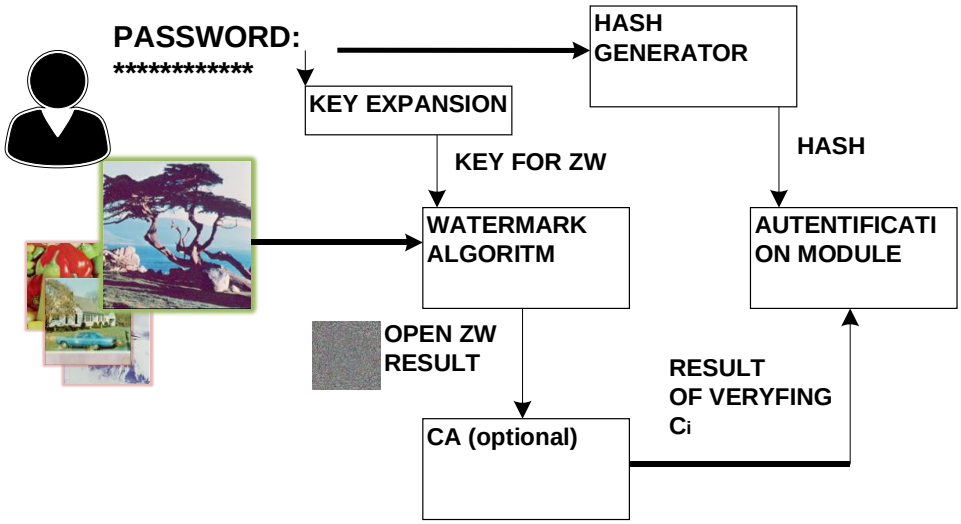

Рис. 6. Запропонована схема автентифікації (спрощена)

Схема автентифікації складається з наступних етапів:

1) Користувач має у володінні секретне зображення на будь-якому носії інформації, що має здатність відображати фото (телефон, планшет, паперова версія фото). Користувач знає, яке саме фото є ключем, пароль.

2) Користувач ініціює початок автентифікації.

3) Користувач демонструє фото засобам візуального зчитування та вводить пароль.

4) Засоби візуального зчитування цифровізують (якщо необхідно) зображення та передають до КЗЗ, в якому це зображення вноситься до вхідних даних алгоритму водяного знаку.

5) Пароль передається до КЗЗ. КЗЗ здійснює порівняння паролю з геш-значенням в БД. Пароль передається до вхідних даних алгоритму розширення ключа.

6) Алгоритм розширення ключа здійснює розгортання ключа за наявним паролем, пароль видаляється з оперативної пам'яті КЗЗ, надалі пароль не використовується, використовується лише розгорнений ключ.

7) Алгоритм розширення ключа передає розгорнутий ключ до алгоритму водяного знаку.

8) Алгоритм водяного знаку отримує розгорнутий ключ та відцифроване зображення. Алгоритм за наявними даними здійснює розгортання водяного знаку, в результаті розгортання отримується майстер зображення користувача.

9) КЗЗ передає майстер зображення до СА, СА здійснює порівняння майстер-зображення з еталонним значенням, повертаючи коефіцієнт схожості до КЗЗ. За необхідності СА може здійснювати накладання майстер-зображення з відповідним зображенням, що зберігається для даного користувача в СА.

10) КЗЗ приймає рішення на основі коефіцієнту, отриманого від СА [13].

Висновки

Дослідження та тестування алгоритму нульового водяного знаку показали, що нульові водяні знаки не підходять для підтвердження авторства, оскільки не можуть чітко «відрізнити» схоже зображення користувача, або стороннє схоже зображення при правильності введення ключових даних. Це актуально для медичних зображень [14], оскільки вони мають високу подібність. Проте нульові водяні знаки можуть бути використані в схемах автентифікації.

Запропоновано схему ідентифікації та автентифікації користувача за допомогою нульових водяних знаків для зображень. Алгоритм має як і суттєві плюси, так і декілька мінусів. До концептуальних плюсів відноситься:

– наявність мінімум двох ознак ідентифікації;

– наявність трьох ознак ідентифікації при умові унікальності зображень;

– можливість використання центрів сертифікації;

– можливість використання зовнішніх сенсорів для передачі інформації в ікс (фото камери, сканери тощо);

– можливість збереження ключа на будь-якому носії інформації, у випадку використання зовнішніх сенсорів дозволяється використання навіть паперових фото.

До концептуальних мінусів належить:

– наявність помилок першого та другого роду при використанні зовнішніх сенсорів;

– використання паролю, що не є стійким методом ідентифікації.

До проблем, що потребують вирішення, відносяться:

– розробка стійкого алгоритму водяного знаку для зображень;

– тестування запропонованого алгоритму на стійкість до атак (людина посередні, маскарад тощо)

Також можна виділити наступні вектори розвитку нульових водяних знаків:

– використання штучного інтелекту для розпізнавання зображень, їх характеристик та ключових параметрів;

– знаходження балансу між стійкістю та чутливістю алгоритму водяного знаку, алгоритм повинен бути достатньо стійким, щоб ігнорувати шуми та пошкодження зображення, та достатньо чутливим, щоб реагувати на зміну ключових параметрів;

– комбінація алгоритмів перетворення зображення для досягнення згаданих вище характеристик та швидкодії; в якості алгоритмів можуть бути використані: DWT, FWT, SVD, WFT, MWT, PCA, LDA, Fourier-Mellin, Radon-перетворення.

Підсумовуючи, можна зробити припущення, що запропонована модель є перспективною, проте потребує досліджень з реалізації, тестування й доведення ефективності. Також ускладнюючим фактором є необхідність розробки й тестування багатьох комбінацій компонентів моделі.

Список літератури:

1. Марія Огнівчук. Прогноз кіберзагроз 2024 // H-X Technologies [Електронний ресурс] https://www.h-x.technology/ua/blog-ua/cyber-threats-forecast-2024-ua
2. X. Qi and Y. Liu. Cloud Model Based Zero-Watermarking Algorithm for Authentication of Text Document // 2013 Ninth International Conference on Computational Intelligence and Security, Emeishan, China. 2013. P. 712–715. doi: 10.1109/CIS.2013.155.
3. Iwendi Celestine, Srivastava Gautam, Jo Ohyun, Javed, Abdul Rehman. KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks // 2020/04/15, IEEE Access, 10.1109/ACCESS.2020.2988160
4. Cox I., Miller M., Bloom J., Fridrich J., Kalker T. Digital Watermarking and Steganography. 2nd Edition, 2007.
5. A. Zulfiqar and M. H. Fazal-e-Amin. A Novel Fragile Zero Watermarking Algorithm for Digital Medical Images // Electronics. 2022. Vol. 11. P.710. doi: https://doi.org/10.3390/electronics1105 0710
6. Wu D., Wang M. and Zhao J. Color Zero-Watermarking Algorithm for Medical Images Basedon BEMD-Schur Decomposition and Color Visual Cryptography // Hindawi Security and Communication Networks, 2021, doi: https://doi.org/10.1155/2021/7081194
7. Asha Rani, Amandeep K. Bhullar, Deepak Dangwal, Sanjeev Kumar. A Zero-Watermarking Scheme using Discrete Wavelet Transform // Procedia Computer Science. 2015. Vol. 70. C. 603–609.
8. Nidhi Sethi, Ram KrishnaImage, Prof R.P. Arora. Image Compression Using Haar Wavelet Transform // Computer Engineering and Intelligent Systems. Vol. 2, No.3
9. Elizabeth A. Compton and Stacey L. Ernstberger Singular Value Decomposition: Applications to Image Processing // Citations Journal of Undergraduate Research. 2020. Vol. 17.

10. Roček A., Javorník M., Slavíček K. et al. Zero Watermarking: Critical Analysis of Its Role in Current Medical Imaging // Digit Imaging. 2021. Vol. 34. P. 204–211.

11. Peak Signal-to-Noise Ratio as an Image Quality Metric [Електронний ресурс]. Режим доступу: http://www.ni.com/white-paper/13306/en/

12. The USC-SIPI Image Database [Електронний ресурс]. Режим доступу: URL: https://sipi.usc.edu/database/

13. R. Gvozdov, V. Poddubnyi, O. Sieverinov, A. Buhantsov, A. Vlasov and V. Sukhoteplyi. Method of Biometric Authentication with Digital Watermarks // PIC S&T, IEEE. 2021. P.569–571. https://doi.org/10.1109/PICST54195.2021.9772134

14. Hongyan X. Digital media zero watermark copyright protection algorithm based on embedded intelligent edge computing detection // Mathematical Biosciences and Engineering. 2021. Вип. 18(5). С. 6771–6789.

Надійшла до редколегії 03.09.2024

Відомості про авторів:

Поддубний Вадим Олександрович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e-mail: vadym.poddubnyi@nure.ua; ORCID: https://orcid.org/0000-0002-4380-491X

Гвоздьов Роман Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e-mail: roman.hvozdov@nure.ua; ORCID: https://orcid.org/0000-0002-5408-943X

Сєвєрінов Олександр Васильович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри безпеки інформаційних технологій; Україна; e-mail: oleksandr.sievierinov@nure.ua; ORCID: https://orcid.org/0000-0002-6327-6405