Methods of information systems protection

UDC 004.932 doi: 10.20998/2522-9052.2021.1.16

Igor Ruban, Nataliia Bolohova, Vitalii Martovytskyi, Oleh Koptsev

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

DIGITAL IMAGE AUTHENTICATION MODEL

Abstract. The development of new technologies, the growing volume of data and the total consumption of content in the digital environment are changing the ecosystem of modern media. Data can be easily and completely duplicated. It brings great convenience to life, work, scientific research and other areas of human activity. However, information security issues have appeared that have attracted a lot of attention. The purpose of this article is to present a model for digital image authentication. This article proposes a model for reliable verification of digital image authenticity with a high degree of protection and parameters for assessing the effectiveness of such systems. Reliability is achieved because the watermark is hidden not in the whole image, but in its fragment, which is most suitable for hiding the image, as well as for using anti-noise codes as a watermark. Based on the current state of watermarking methods, it is recommended to use modern algorithms and architectures of convolutional neural networks to ensure a high degree of security.

Keywords: digital watermark; authentication; copyright; digital image authentication model; steganographic; cybersecurity.

Introduction

The development of new technologies, the growing volume of data and the total consumption of content in the digital environment are changing the ecosystem of modern media. Data can be easily and completely duplicated. It brings great convenience to life, work, scientific research and other areas of human activity. However, information security issues have appeared that have attracted a lot of attention. Copyright ownership is an important aspect of information security, and the use of digital watermarks is an effective way to protect copyrights [1]. Copyright protection is achieved by embedding author information into the digital content. In recent years, many watermarking algorithms have been proposed [1-10].

The Internet is a free zone, where almost everything is open and everything is allowed. The Internet makes it easy to copy and replicate any pictures, texts, video and audio product, without thinking about the fact that these pictures, articles, songs have authors who have certain rights to this multimedia data.

Almost every site in the Internet contains photos and illustrations. At the same time, most of the images posted on the Internet are used without a valid license – unfortunately to authors.

To raise awareness of the extent of image copyright infringement, Copytrack regularly investigates how, where and to what extent images are used illegally. Copytrack's 2019 Global Infringement Report consists of a statistical analysis of more than 12,000 Copytrack user profiles. Investigation of Illegal image using was based on all searches deemed illegal by individual account holders, and on website owner data based on information gathered by internal search robots. The percentages mentioned in this report refer to the number of potential copyright infringements handled by Copytrack between December 2017 and December 2018. Geographic locations were used for the analysis.

According to the 2019 Copytrack Global Infringement Report [11], Fig. 1 shows the percentage of copyright infringement in the use of images on the global Internet.

According to the report [11], more than 2.5 billion images are stolen every day. These license violations can lead to daily damages of up to 532.5 billion euros.

The purpose of this article is to present a model for digital image authentication.

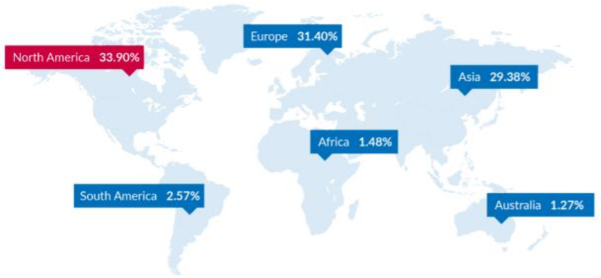

Fig. 1. Copyright Infringement Statistics by Continent [11]

Analysis of publications

A digital watermark is a special mark embedded in digital content (called a container) to protect copyrights and confirm the integrity of the document itself. A watermark can be embedded in any type of electronic document. Along with various images (photos, drawings, scanned paper documents, etc.) there are audio recordings that contain a watermark, and video (e.g., DVDs). Watermarks is actively used for placement of unique photos, videos, audio tracks in electronic form on the global Internet.

There are different methods of classification of watermarks algorithms creation. If we divide them by characteristics, there are three types of digital watermarks:

- stable (robust), such watermarks should be resistant to any influences on them [12];
- fragile, changing or collapsing with minor modification of the container [13];
- semi-fragile, resistant to some influences and unstable to others [13].

Resistant watermarks are used when it is necessary to remain the identification code, company logo and other identifiers intact with maximum distortion of the container. Fragile watermarks and electronic digital signatures are used to verify the integrity of electronic documents. Algorithms of embedding fragile watermarks are especially sensitive to any distortion and effective for integrity control and forgery protection. In the case of semi-fragile TLDs, an image, for example, can be converted to another format or compressed, but you cannot cut or paste a fragment into it; for an audio track, the sound frequencies can be changed, but the performer's voice cannot be removed.

Most of the current research focuses on a robust watermarking algorithm for copyright protection, such an algorithm usually embeds the watermark in the transform region of the image to improve reliability. Discrete Fourier transform (DFT), discrete wavelet transform (DWT) and discrete cosine transform (DCT) are the most commonly used basic transformation methods [1-3, 6].

The article [4] proposes a new method for watermarking a color image in the spatial domain for fast and efficient color image copyright protection. It discusses the direct propagation coefficient (DC) 2D-DFT, obtained in the spatial domain. Also it proves connection between the change of each pixel in the spatial domain and in the DC coefficient in the Fourier transform.

The article [5] proposes a new technique for watermarking medical images to detect damaged areas in medical images with greater accuracy by authenticating 4x4 blocks and without limiting the region of interest (ROI) size. The proposed method can label a 4 × 4 pixel block if it has at least one distorted pixel, while similar methods (which have no region of interest size limitation) label 8 × 8, 16 × 16, and 40 × 40 pixel blocks.

The article [7] proposed a reliable method for watermarking an image in a merged wavelet transform (LWT) domain. The neural network is included in the watermark extraction process to increase resistance to various attacks. The integration of the neural network with LWT makes the system resistant to various attacks, while maintaining an adequate level of stealth.

Since a reliable watermarking algorithm also has high security requirements, the article proposes a general model for image copyright protection based on LWT.

Digital Image Authentication Model

Model of digital image authentication can be considered as steganographic system that transmits an encrypted identifier, which is a digital watermark.

After analyzing of the current state of research on methods of superimposition of watermark and significant parameters of watermark, we can form the following assessment of the system of digital image authentication effectiveness:

$$EF = R \cdot \alpha_r + SR \cdot \alpha_{sr} + ER \cdot \alpha_{er} + \\ + SC \cdot \alpha_{sc} + DT \cdot \alpha_{dt}, \qquad (1)$$

where $R, R \in [0,1]$ – is an estimate of the reliability of the method of embedding a watermark;

$RS, RS \in [0,1]$ – is an estimate of the watermark invisibility in the image;

$ER, ER \in [0,1]$ – is the probability of an error of the first and second kind;

$SC, SC \in [0,1]$ – is an estimate of the fragility of the watermark;

DT – is the number of embedded watermarks;

$\alpha_r, \alpha_{sr}, \alpha_{er}, \alpha_{sc}, \alpha_{dt}$ – are significance coefficients of the corresponding parameters of the watermark method. Such coefficients are necessary because there is no universal watermark embedding method, so thanks to such coefficients you can adjust the significance of each parameter and thus influence the final effectiveness of the method for a particular watermarking task.

The process of digital image authentication with watermark is shown in Fig. 2 and consists of the following main steps:

1) defining the area for embedding;
2) generating a watermark;
3) embedding the watermark in an image fragment;
4) image preprocessing after embedding the watermarked fragment in the original image;
5) detecting the watermarked fragment;
6) extraction of the watermark from the fragment;
7) correcting errors in the watermark during extraction;
8) obtaining the label of the right holder.

Consider these steps.

Step 1 - Definition of the area for embedding.

Suppose there is an image Im [N, M] in which you want to embed a watermark $W[k,l]$.

Then, in the simplest case, the procedure for determining the optimal area for embedding a watermark can be represented using the sliding window method.

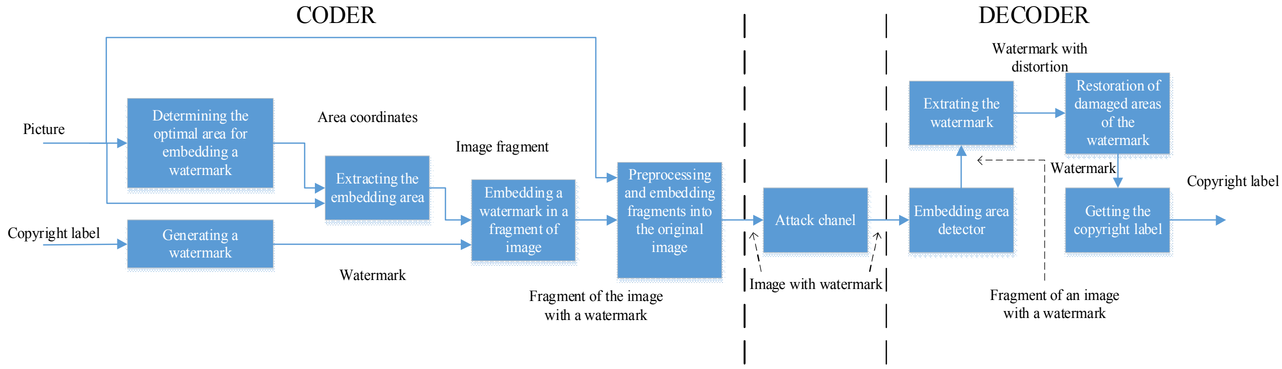

Fig. 2. Digital Image Authentication Model

The scheme of the algorithm for determining the area for embedding is shown in Fig. 3. Analytically, the sliding window processing can be represented as follows:

$$E(\mathrm{Im}_{i,j}) = Q\left|q(\mathrm{Im}_{i,j}), q(\mathrm{Im}_{i,j+1}), q(\mathrm{Im}_{i,j-1}), \ldots\right| = \\ = Q\left|q(\mathrm{Im}_{i+k}, \mathrm{Im}_{j+l})\right| \langle k,l\rangle \in S, \quad (2)$$

where $E(\mathrm{Im}_{i,j})$ – the optimality value of the embedding area; Q – function describing the rules for evaluating the pixels in the vicinity S; S – point vicinity, the set of points (pixels) surrounding the working point (usually the center pixel); k, l – the size of the sliding window, given by the S set of coordinate offsets on the abscissa axis and the ordinate axis, respectively.

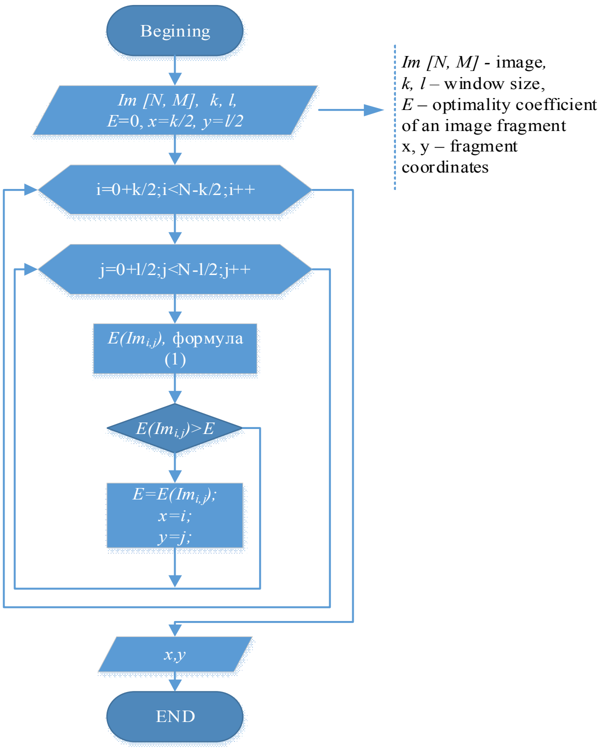

Fig. 3. Diagram of the algorithm for determining the area for embedding

Stage 2 - Generation of the watermark.

Let W', I', K', B' be the set of possible watermarks, containers (the form of watermark representation) of keys and hidden identifiers of right holders, respectively. Then the generation of the watermark can be represented as:

$$F:\ I' \times K' \times B' \to W,\ W = F(I, K, B), \quad (2)$$

where W, I, K, B – are elements of the corresponding sets. Generally speaking, the function can be arbitrary, but in practice, the robustness requirements of the watermark impose certain restrictions on it. Thus, in most cases, $F(I,K,B) \approx F(I+\varepsilon, K, B)$, i.e., a slight change in the container does not change the hidden IDs of the rights holders. The function is usually composite:

$$F = T \circ G,\ where\ G: K' \times B' \to C'\ and \\ T: C' \times I' \to W', \quad (3)$$

that is, the watermark depends on the properties of the container. The function G can be implemented using a cryptographically secure pseudorandom number generator with K as the initial value.

To improve the robustness of the watermark, authors use interference-resistant codes such as BCH codes, convolutional codes. [14, 15].

The operator T modifies the code words C', resulting in a watermark W'. It is possible not to impose irreversibility constraints on this function since the appropriate choice G already guarantees irreversibility F. The function T must be chosen such that an unfilled container I_0, a filled container I_W and a slightly modified filled container I''_W would produce the same watermark:

$$T(C, I_0) = T(C, I_W) = T(C, I''_W), \quad (4)$$

Step 3 - Generation of the watermark.

The process of embedding the watermark $W(i,j)$ in the original image $I_0(i,j)$ can generally be described as a superposition of two signals:

$$\varepsilon:\ I' \times W' \times L' \to I_W,\ I_W(i,j) \\ = I_0(i,j) \oplus L(i,j) W(i,j) p(i,j), \quad (5)$$

where $L(i,j)$ - watermark-embedding mask, which takes into account the characteristics of the human visual system, serves to reduce the visibility of the watermark; $p(i,j)$ - projecting function;

Sign $\oplus$ denotes the superposition operator, including, addition, truncation, and quantization.

The projecting function carries out the "distribution" of the watermark over the image area. Using of it can be considered as an implementation of parallel channel information spacing. In addition, this function has a certain spatial structure and correlation properties used to counteract geometric attacks.

Step 4 - Image preprocessing after embedding the watermarked fragment in the original image.

At this stage, the image fragment I_W is inserted into the original image Im. Since I_W is a modification of the original fragment I_0, which serves as a container for the watermark. When embedding it into the original image Im, the watermarked fragment will differ from the general distribution of the image brightness $I_0(i,j)+\alpha$, where α is a modification factor that occurs during the embedding process. To eliminate this drawback, after embedding a fragment with a watermark in the original image, anti-aliasing is applied along the edges of the embedded fragment to eliminate the visibility of the presence of a watermark in the image.

Step 5 - Detect the watermarked fragment.

The task of detection is to find objects (watermark) with certain properties on the image, and if the objects are detected, to determine their coordinates on the image plane. The basic principle of object detection on the image is to compare the brightness function of the image with some "reference" - a fragment of the brightness field containing the desired object. When implementing the detection procedure, the standard is sequentially moved along the image field, and at each position, its similarity to the real brightness function on the fragment is investigated. Complete coincidence of the standard and the image, as a rule, does not happen due to noise and distortion, as well as because usually there is no complete information about the shape and structure of the object, (we have to use the standard, only approximately describing the object).

Since watermarks are hidden and different in structure in the image in such a way that they are less visible, it is best to use neural networks for detection and localization.

Step 6 - Extracting the watermark from the fragment. The extraction uses the inverse of the operation in step 3 and depends on the embedding method.

Step 7 - Correcting errors in the watermark during extraction. Since it is recommended to use interference-resistant codes, such as BCH codes, convolutional codes, to improve robustness, this step is the correction of distortions in the watermark.

Step 8 - Obtaining the label of the copyright holder. This step is the reverse of step 2 and depends on the method used to generate the watermark.

There is a probability that the decoder will not detect an existing watermark and a probability of falsely finding a watermark in an empty container (false alarm probability). Decreasing one probability leads to increasing the other probability.

Reliability of the decoder is characterized by the probability of false detection. This model for verifying the authenticity of a digital image is designed to minimize the likelihood of both errors, since each of them can lead to denial of service.

Conclusion

This article proposes a model for reliable verification of digital image authenticity with a high degree of protection and parameters for assessing the effectiveness of such systems. Reliability is achieved because the watermark is hidden not in the whole image, but in its fragment, which is most suitable for hiding the image, as well as for using anti-noise codes as a watermark. Based on the current state of watermarking methods, it is recommended to use modern algorithms and architectures of convolutional neural networks to ensure a high degree of security.

REFERENCES

1. Su, Q. (2016), “Novel blind colour image watermarking technique using Hessenberg de-composition”, *IET Image Process*, pp. 817– 829, DOI: https://doi.org/10.1049/iet-ipr.2018.6040.
2. Zhang, X., Peng, F. and Long, M. (2018), “Robust Coverless Image Steganography Based on DCT and LDA Topic Classification”, *IEEE Trans. on Multimedia*, Vol. 20, No. 12, pp. 3223-3238, DOI: https://doi.org/10.1109/TMM.2018.2838334.
3. Choudhary, R. and Parmar, G. (2016), “A robust image watermarking technique using 2-level discrete wavelet transform (DWT)”, *2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, pp. 120-124, DOI: https://doi.org/10.1109/CCIntelS.2016.7878213.
4. Su, Q. (2019), “New Rapid and Robust Color Image Watermarking Technique in Spatial Domain”, *IEEE Access*, Vol. 7, pp. 30398-30409, DOI: https://doi.org/10.1109 /ACCESS.2019.2895062.
5. Ustubioglu, A. and Ulutas, G. (2017), “A New Medical Image Watermarking Technique with Finer Tamper Localization”, *J. Digit Imaging 30*, pp. 665-680, DOI: https://doi.org/10.1007/s10278-017-9960-y.
6. Hamidi, M., Haziti, M.E. and Cherifi, H. (2018), “Hybrid blind robust image watermarking technique based on DFT-DCT and Arnold transform”, *Multimed Tools Appl 77*, pp. 27181-27214, DOI: https://doi.org/10.1007/s11042-018-5913-9.
7. Islam, Mohiul, Amarjit, Roy, and Rabul Hussain, Laskar (2018), “Neural network based robust image watermarking technique in LWT domain”, *Journal of Intelligent & Fuzzy Systems*, Vol. 34.3, pp. 1691-1700, DOI: https://doi.org/10.3233/JIFS-169462.
8. Kumar, Shishir, Neha, Jain, and Steven Lawrence, Fernandes (2017), “Rough set based effective technique of image watermarking”, *Journal of Computational Science*, Vol. 19, pp. 121-137, DOI: https://doi.org/10.1016/j.jocs.2016.11.009.
9. Parah, S.A., Sheikh, J.A. and Loan, N.A. (2018), “Utilizing neighborhood coefficient correlation: a new image watermarking technique robust to singular and hybrid attacks”, *Multidim Syst Sign Process*, Vol. 29, pp. 1095-1117, DOI: https://doi.org/10.1007/s11045-017-0490-z.
10. Munib, Summuyya, and Asifullah, Khan (2017), “Robust image watermarking technique using triangular regions and Zernike moments for quantization based embedding”, *Multimedia Tools and Applications*, Vol. 76.6, pp. 8695-8710, DOI: https://doi.org/10.1007/s11042-016-3485-0.
11. (2019), *Copytrack Global Infringement Report 2019*, International Image Theft in Comparison, available at: https://www.copytrack.com/wp-content/uploads/2019/04/190328_Global_Infringement_Report_2019_EN_Online.pdf.
12. A. Ramsha, M. Mohsin Riaz, and A. Ghafoor (2018), “Attack resistant watermarking technique based on fast curvelet transform and Robust Principal Component Analysis”, *Multimedia Tools and Applications*, Vol. 77, No. 8, pp. 9443-9453, DOI: https://doi.org/10.1007/s11042-017-5128-5.
13. Shehab, Abdulaziz, et al. (2018), “Secure and robust fragile watermarking scheme for medical images”, *IEEE Access*, Issue 6, pp. 10269-10278, DOI: https://doi.org/10.1109/ACCESS.2018.2799240.
14. Yarmolik, V.N., Portyanko, S.S. and Yarmolik S.V. (2007), *Kriptografiya, steganografiya i ohrana avtorskogo prava*, monografiya, BGU, Minsk, 240 p.
15. Gribunin, V.G., Okov, I.N. and Turintsev I.V. (2002), *Tsifrovaya steganografiya. Aspektyi zaschityi*, Solon-Press, Moscow, 272 p.

Надійшла (received) 23.11.2020
Прийнята до друку (accepted for publication) 10.02.2021

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Рубан Ігор Вікторович – доктор технічних наук, професор, перший проректор, Харківський національний університет радіоелектроніки, Харків, Україна;
Igor Ruban – Doctor of Technical Sciences, Professor, The first vice-rector, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: ihor.ruban@nure.ua; ORCID ID: http://orcid.org/0000-0002-4738-3286.

Бологова Наталія Миколаївна – аспірантка, Харківський національний університет радіоелектроніки, Харків, Україна;
Nataliia Bolohova – Postgraduate student, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: nataliia.bolohova@nure.ua; ORCID ID: http://orcid.org/0000-0001-8927-0055.

Мартовицький Віталій Олександрович – кандидат технічних наук, доцент, доцент кафедри обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Vitalii Martovytskyi – Candidate of Technical Sciences, Associate Professor, Associate Professor of Electronic Computers Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: vitalii.martovytskyi@nure.ua; ORCID ID: http://orcid.org/0000-0003-2349-0578.

Копцев Олег Олегович – студент, Харківський національний університет радіоелектроніки, Харків, Україна;
Oleh Koptsev – student, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: oleh.koptsev@nure.ua; ORCID ID: http://orcid.org/0000-0002-4426-057893.

Модель аутентифікації цифрових зображень

I. В. Рубан, Н. М. Бологова, В. О. Мартовицький, О. О. Копцев

Анотація. Розвиток нових технологій, зростання обсягів великих даних і тотальне споживання контенту в цифровому середовищі змінюють екосистему сучасних медіа. Дані можна легко і повністю дублювати, що приносить велику зручність в життя, роботі, наукових дослідженнях та інших сферах діяльності людини. Однак виникли питання інформаційної безпеки, які привернули велику увагу. Право власності на авторські права - важливий аспект інформаційної безпеки, а використання цифрових водяних знаків - ефективний спосіб захисту авторських прав. **Метою даної статті** є представлення моделі перевірки справжності право власності на авторські права зображення. Цифрові водяні знаки (ЦВЗ) активно використовуються при розміщенні унікальних фотографій, відео, аудіотреків в електронному вигляді в глобальній мережі Інтернет. Оскільки надійний алгоритм водяних знаків також має високі вимоги до безпеки, в роботі запропонована загальна модель перевірки автентичності зображення на основі ЦВЗ. Модель перевірки справжності цифрового зображення може бути розглянута як стеганографічна система, в якій передається інтегрований зашифрований ідентифікатор в область зображення є цифровим водяним знаком. Існує ймовірність того, що декодер не виявить наявний ЦВЗ і ймовірність помилкового знаходження ЦВЗ в порожньому контейнері (ймовірність помилкової тривоги). Зниження однієї ймовірності призводить до збільшення іншої. Надійність роботи декодера характеризують ймовірністю помилкового виявлення. Дана модель перевірки справжності цифрового зображення побудована таким чином, щоб мінімізувати ймовірність виникнення обох помилок, так як кожна з них може привести до відмови від обслуговування. У цій статті пропонується модель надійної перевірки автентичності цифрового зображення з високим ступенем захисту і параметри оцінки ефективності роботи подібних систем. Надійність досягатися за рахунок того, що ЦВЗ ховається не в усьому зображенні, а в його фрагменті, який найбільш підходить для приховування ЦВЗ, а також застосування в якості ЦВЗ перешкодостійких кодів. Спираючись на поточний стан методів ЦВЗ для забезпечення високого ступеня захищеності рекомендується застосовувати сучасні алгоритми і архітектури згортальних нейронних мереж.

Ключові слова: цифровий водяний знак; автентифікація; авторське право; модель аутентифікації цифрових зображень; стеганографія; кібербезпека.

Модель проверки подлинности цифрового изображения

И. В. Рубан, Н. Н. Бологова, В. А. Мартовицкий, О. О. Копцев

Аннотация. Развитие новых технологий, растущие объемы больших данных и тотальное потребление контента в цифровой среде меняют экосистему современных медиа. Данные можно легко и полностью дублировать, что приносит большое удобство в жизнь, работу, научные исследования и другие сферы деятельности человека. Однако возникли вопросы информационной безопасности, которые привлекли большое внимание. **Целью данной статьи** является представление модели проверки подлинности цифрового изображения. Цифровые водяные знаки (ЦВЗ) активно используются при размещении уникальных фотографий, видео, аудиотреков в электронном виде в глобальной сети Интернет. Поскольку надежный алгоритм водяных знаков также предъявляет высокие требования к безопасности, в работе предложена общая модель проверки подлинности изображения на основе ЦВЗ. Модель проверки подлинности цифрового изображения может быть рассмотрена как стеганографическая система, в которой передается интегрированный зашифрованный идентификатор в область изображения цифровым водяным знаком. Снижение одной вероятности приводит к увеличению другой. Надежность работы декодера характеризуют вероятностью ложного обнаружения. Данная модель проверки подлинности цифрового изображения построена таким образом, чтобы минимизировать вероятность возникновения обоих ошибок, так как каждая из них может привести к отказу от обслуживания. В этой статье предлагается модель надежной проверки подлинности цифрового изображения с высокой степенью защиты и параметры оценки эффективности работы подобных систем. Надежность достигается за счет того, что ЦВЗ скрывается не во всем изображении, а в его фрагменте, который наиболее подходит для скрытия изображения, а также применения в качестве ЦВЗ помехоустойчивых кодов. Опираясь на текущее состояние методов ЦВЗ для обеспечения высокой степени защищенности рекомендуется применять современные алгоритмы и архитектуры сверточных нейронных сетей.

Ключевые слова: цифровой водяной знак; аутентификация; авторское право; модель аутентификации цифрового изображения; стеганография; кибербезопасность.