International Journal of Information Technology and Management

ISSN online: 1741-5179 - ISSN print: 1461-4111
https://www.inderscience.com/ijitm

Multi-level secure storage method of electronic documents based on Hash function

Runhua Miao

DOI: 10.1504/IJITM.2022.10051753

Article History:
Received: 29 June 2022
Accepted: 23 September 2022
Published online: 28 January 2025

Multi-level secure storage method of electronic documents based on Hash function

Runhua Miao

Information Center of Guizhou Power Grid Co., Ltd.,
Guiyang, 550002, Guizhou, China
Email: runhmiao@mls.sinanet.com

Abstract: In order to improve the storage security and efficiency of electronic documents, a new multi-level secure storage method of electronic documents based on Hash function is proposed in this paper. Firstly, the vector space model is constructed, and the electronic document data is divided into fixed size data blocks by using the idea of linear segmentation. Then, after normalisation, Hash function is used to encrypt each data block. Finally, according to the encryption results, the compressed sensing method is used for multi-level secure storage of electronic documents. The experimental results show that, compared with the traditional storage methods, this method has strong encryption performance, and the maximum intrusion rate is no more than 0.5%, which can ensure the security of electronic documents. This method can improve the storage efficiency of electronic documents.

Keywords: Hash function; electronic document; multi-level security storage; data segmentation.

Reference to this paper should be made as follows: Miao, R. (2025) 'Multi-level secure storage method of electronic documents based on Hash function', *Int. J. Information Technology and Management*, Vol. 24, Nos. 1/2, pp.106–116.

Biographical notes: Runhua Miao received his Bachelor's in Computer Science from the Shanghai Electric Power University, in 2007 and his Master's in Electrical Engineering from the Chongqing University, in 2014. Currently, he is the Deputy Chief of Confidential Documents Department of Guizhou Power Grid Company office. His main research direction is confidential information management and document information processing, etc.

1 Introduction

For a large number of electronic document resources generated in the era of big data, there is an urgent need for an efficient way to safely store electronic document resources (Surovtseva, 2019). Traditional storage methods can no longer meet the requirements of the explosive growth of electronic documents in the internet era. It is necessary to constantly seek and explore a more secure and efficient storage method to store electronic document resources more securely in a limited space (Daudov et al., 2021; Mazin et al., 2021; Chaniago et al., 2021). With the emergence and development of big data and artificial intelligence technology, various data storage methods have developed rapidly under the background of mutual promotion. However, among many storage methods, the

security of electronic document storage is still difficult to meet the requirements. Therefore, it is urgent to study a safe and effective electronic document storage method to improve the performance and use performance of electronic document storage (Etinkaya et al., 2020).

Cheng et al. (2020) proposes a blockchain based secure storage method for electronic document data. This method first uses the signature algorithm to provide privacy protection for electronic document data, then uses the interstellar file system to store electronic document data with privacy, and finally saves the encrypted electronic document data on the alliance blockchain. However, the intrusion rate of electronic documents encrypted by this method is still high. Ren et al. (2019) proposes a secure storage method of electronic document data based on dynamic virtualisation electronic stream cipher. Firstly, the virtual server, memory and other structures are designed on the virtual machine. Secondly, the secure electronic stream cipher model is introduced to encrypt and decrypt electronic document data. Finally, the secure storage of electronic document data is realised by using Merkle Hash B+ tree. However, this method has the problem of low storage efficiency. Dong (2019) proposes a method for safe storage of electronic document data based on artificial intelligence, which is based on the basic principle of electronic document storage and defines the logical relationship between electronic document data. With the support of artificial intelligence technology, electronic document data is integrated and the location of storage points is obtained. Multiple de-noising encryption technology is adopted to realise the encrypted storage of electronic documents. However, this method requires a large number of calculations, which leads to the reduction of the efficiency of electronic document storage.

In order to solve the problems of poor encryption security and low encryption efficiency of the above traditional methods, a multi-level secure storage method of electronic documents based on hash function is proposed. The overall research technical route of this method is as follows:

1 Construct a vector space model, divide the electronic document data into fixed size data blocks by using the linear segmentation idea, and complete the hierarchical segmentation of electronic documents.

2 Based on the hierarchical segmentation results of electronic documents, hash function encrypts the segmented electronic documents. The encryption operation of electronic documents consists of encryption matrix and sequence. If the generation law of random sequence is not cracked, electronic documents are difficult to be cracked, so the security of electronic documents can be effectively improved. According to the encryption results, the compressed sensing method is used for multi-level secure storage of electronic documents.

3 Experimental verification, taking intrusion rate and storage efficiency as experimental comparison indicators, this method is compared with traditional methods.

2 Multi-level secure storage method of electronic documents based on hash function

2.1 Hierarchical segmentation of electronic documents

Considering the similarity between electronic documents, a vector space model is constructed to segment electronic documents hierarchically to improve the security of storage. In order to calculate the similarity between electronic documents, a long common substring is used to construct the similarity matrix. Define the electronic document set as $G = \{s_1, s_2, \ldots, s_m\}$, where $m = 1, 2, \ldots, i$ represents the data volume of the electronic document and s_i represents the i^{th} electronic document in the set (Tkachenko and Denisova, 2022; Azlina et al., 2019). The set of words contained in the electronic document is $s = \{w_1, w_2, \ldots, w_n\}$, and n represents the number of words in the electronic document.

The dynamic programming algorithm is used to calculate the longest common substring between two electronic documents i and j, and the calculation formula is:

$$L[i, j] = \begin{cases} 0, & i = 0 \text{ or } j = 0 \\ L[i-1, j-1]+1, & w_{1i} = w_{2j} \\ \max\{L[i-1, j], L[i, j-1]\}, & w_{1i} \neq w_{2j} \end{cases} \tag{1}$$

In the formula, $L[i, j]$ represents the longest common substring between i and j of the electronic document, and w_{1i} represents the words in the electronic document (Ragimova et al., 2020).

According to the calculation result of formula (1), a formula for calculating the similarity between electronic documents is constructed:

$$sim(s_1, s_2) = \frac{L(i, j) + (\max(i, j) - \min(i, j))}{\max(i, j)} \tag{2}$$

In the formula, $\max(i, j)$ and $\min(i, j)$ represent the maximum and minimum values of electronic documents respectively.

Before the electronic document segmentation, in order to ensure that the document sequence is not interrupted, fisher optimal segmentation method is used to segment the electronic document.

Since the division order of the electronic document needs to be ensured, the category of the divided electronic document is set to $\{s_i, s_{i+1}, s_{i+2}, \ldots, s_{i+k}\}$, and one of the categories can be represented by $\{s_i, s_{i+1}, \ldots, s_j\}$, and $1 \leq i \leq j \leq n$ exists. The expression of vector mean value of electronic documents is:

$$\overline{s}_{ij} = \frac{1}{j-i+1} \sum_{r=i}^{j} s_r \tag{3}$$

According to the calculation results of the vector mean value of the electronic document, the class diameter of the electronic document after segmentation is calculated:

$$D(i, j) = \sum_{r=i}^{j} (s_r - \overline{s}_{ij})^T (s_r - \overline{s}_{ij}) \tag{4}$$

Assuming that the electronic document needs to be divided into k segments, the split result can be expressed as $\{i_k, i_{k+1}, \ldots, n\}$, and the split loss function of the electronic document can be constructed:

$$L\left[b(n,k)\right]=\sum_{r=1}^{k} D\left(i_r, i_{r+1}-1\right) \tag{5}$$

In the formula, i_r represents the r^{th} electronic document (Kaupa and Chisa, 2020).

It can be seen from formula (5) that the smaller the value of the loss function, the more reasonable the segmentation result will be. Therefore, it is necessary to minimise the loss function shown in formula (5) to improve the segmentation accuracy. The optimal partition function in the case of minimum loss function is:

$$L\left[b(n,\ p)\right]=\min_{2\le j\le n}\left\{D(1,\ j-1)+D(j,\ n)\right\} \tag{6}$$

Through the above calculation, the optimal segmentation of electronic documents is completed, which can improve the encryption efficiency and storage efficiency of later electronic documents.

2.2 Electronic document encryption based on Hash function

Based on the above electronic document segmentation results, in order to realise the secure storage of electronic documents, hash function is used to encrypt the segmented electronic documents.

Because Hash function has two characteristics of unidirectionality and collision constraint, unidirectionality refers to the irreversibility of its operation direction. In Hash function, it refers to that the output can only be derived from the input, but the input cannot be calculated from the output; collision constraint means that one input cannot be found to make its output equal to a known output, or two different inputs cannot be found at the same time to make its output completely consistent. Therefore, using Hash function to encrypt electronic documents can effectively improve the security of electronic documents.

If the encryption key of the electronic document is defined as U and the electronic document after segmentation is S, the encryption sequence $X = [x_1, x_2, \ldots, x_n]$ of the electronic document can be obtained through operation.

Normalise the amplitude of electronic documents:

$$\left|h_B\right|_N=\frac{\left|h_B\right|}{\left|h_B\right|_{\max}} \tag{7}$$

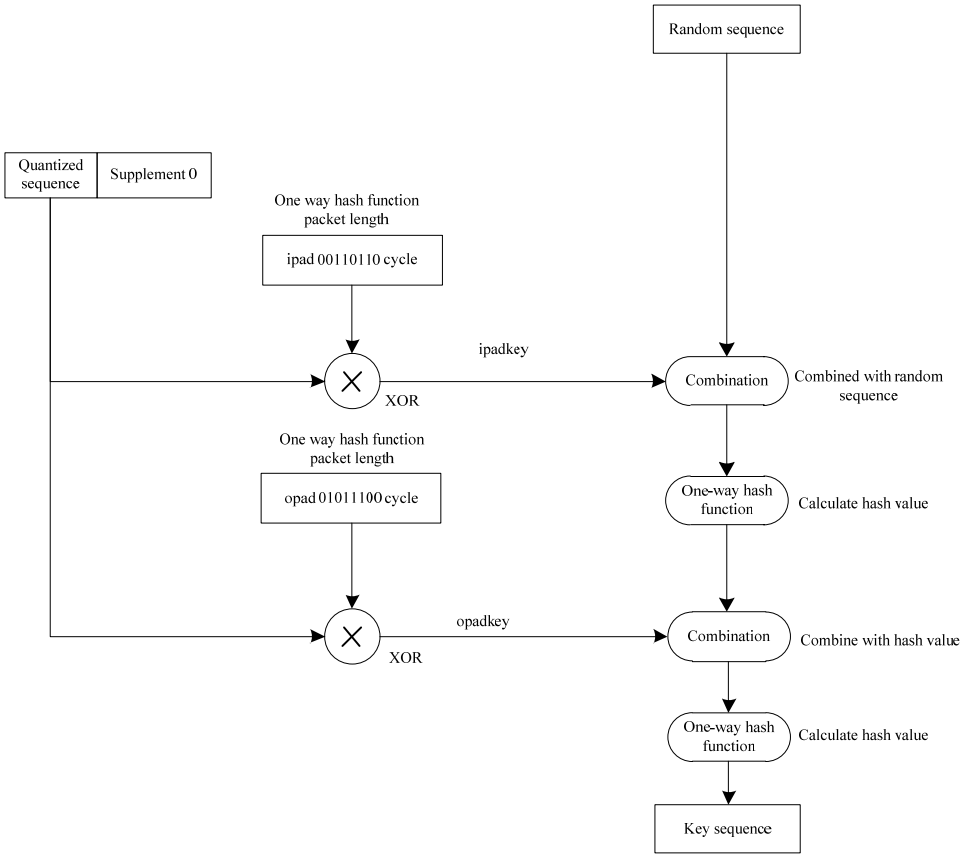
In the formula, $|h_B|$ represents the amplitude of the electronic document, and $|h_B|_{\max}$ represents the maximum value of the amplitude of the electronic document (Mikheev and Yakimov, 2019).

After the normalisation, the electronic documents are encrypted with one-way Hash. The key matrix generation process is shown in Figure 1.

According to the key matrix generation process shown in Figure 1, the hash matrix is converted:

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nm} \end{bmatrix} \tag{8}$$

Figure 1 Key matrix generation process



Perform QR decomposition on matrix H shown in formula (8):

$$H = QR \tag{9}$$

In the formula, the unitary matrix Q can represent the key matrix U:

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nm} \end{bmatrix}, n = 8 \tag{9}$$

After generating the encryption matrix of the electronic document, multiply the key matrix U and the sequence S to realise the encryption of the electronic document:

$$X = SU \tag{10}$$

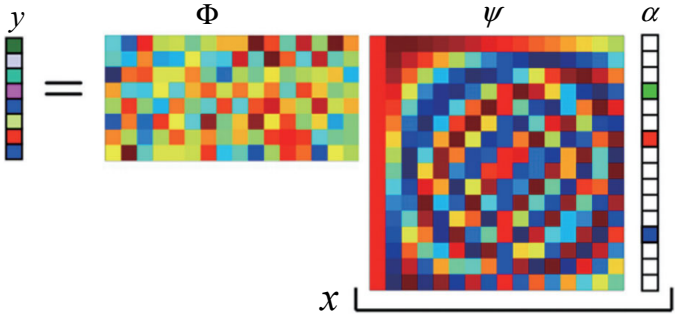
$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T = \begin{bmatrix} u_{11}s_1 + u_{21}s_2 + \ldots + u_{n1}s_n \\ u_{12}s_1 + u_{22}s_2 + \ldots + u_{n2}s_n \\ \vdots \\ u_{1n}s_1 + u_{2n}s_2 + \ldots + u_{nn}s_n \end{bmatrix}^T \tag{11}$$

It can be seen from the above formula that the encryption operation of electronic documents consists of encryption matrix and sequence (Alferov et al., 2019; Pysarenko et al., 2019). If the generation law of random sequence is not cracked, the electronic documents are difficult to be cracked, so the security of electronic documents can be effectively improved.

2.3 Multi-level secure storage of electronic documents

After completing the encryption of electronic documents, the multi-level secure storage of electronic documents is carried out by means of compression sensing. The basic principle of compressed sensing is sparse sampling, and its principle is shown in Figure 2.

Figure 2 Schematic diagram of compressed storage model (see online version for colours)



In Figure 2, ψ represents the sparse basis matrix and α represents the sparse coefficient.

The sparse sampling expression in the compressed sensing storage of electronic documents is:

$$\text{Minimise } |x|_0 \text{ subject to } y = \phi \cdot x \tag{12}$$

In the formula, x represents the original electronic document signal, y represents the electronic document information after sparse sampling recovery, and ϕ represents the observation matrix.

If the two-dimensional mathematical matrix stored in the electronic document is $\phi_{M\times N}$, the frequency domain of the original electronic document signal can be converted. If the random sampling point is q, the original signal can be reconstructed through the random sampling point is q:

$$\text{Minimise } \|x\|_1 \text{ subject to } y = \phi \cdot x \tag{13}$$

In the reconstruction process, it should be noted that the sampling of compressed sensing has high randomness, and the interference terms in different frequency domains will affect the sampling results. Therefore, it is necessary to limit the signal period:

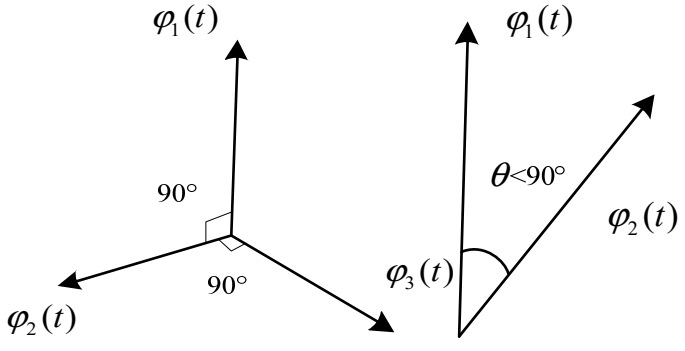
$$(1-\delta x)\|x\|_{l_2}^2 \le \|Ax\|_{l_2}^2 \le (1+\delta s)\|x\|_{l_2}^2 \tag{14}$$

The calculation formula for the coherence of the compressed sensing basis function is:

$$u(\varphi_u, \varphi_v) = \sqrt{n}\max_{1\le u,v\le n}\left|\langle \varphi_u, \varphi_v \rangle\right| \tag{15}$$

The basis function coherence diagram is shown in Figure 3.

Figure 3 Coherence of compressed sensing basis function, (a) the included angle of the basis function is 90° (b) the included angle of the basis function is less than 90°



After analysing the coherence of compressed sensing basis function, the matching pursuit algorithm is used to store electronic documents in multi-level security. Define the residual coefficient and perception matrix as r_n and Θ respectively, and calculate the absolute value of the inner product of the electronic document signal:

$$u = \left\{u_j \left| u_j = \left|\langle \Theta_j, r_n \rangle\right|, j = 1, 2, \ldots, M\right.\right\} \tag{16}$$

And there are $x = (\zeta_1, \zeta_2, \ldots, \zeta_n)^T$ and $y = (\eta_1, \eta_2, \ldots, \eta_n)^T$, so the expression for constructing inner product is:

$$\langle x, y \rangle = \zeta_1\overline{\eta}_1, \zeta_2\overline{\eta}_2, \ldots, \zeta_n\overline{\eta}_n = y^H x \tag{17}$$

According to the calculation results of inner product, the electronic documents are stored iteratively. The storage factor is defined as, the storage support set is, and the calculation formula of iterative storage is:

$$\hat{\rho} = (\tau^*, \tau)^{-1}\tau * y \tag{18}$$

Repeat iterations until the storage set reaches its maximum capacity.

Through the above research, the multi-level secure storage of electronic documents is completed. In this study, firstly, the vector space model of electronic documents is constructed, and the electronic documents are segmented. After normalisation, hash

function is used to encrypt each data block. Finally, the secure storage is completed by compression sensing method.

3 Experimental verification

In order to verify the practical application effect of the proposed multi-level secure storage method of electronic documents based on Hash function, a comparative test is carried out.

3.1 Experimental data

In order to fully verify the secure storage performance of the proposed method, it is required that the types of experimental samples and the amount of data are sufficient. Therefore, the electronic image document and electronic text document in MySQL database are selected, and the data amount of the two documents is 5 GB respectively. The experimental processing of electronic documents is carried out in MATLAB software.

3.2 Experimental scheme

In order to verify the secure storage performance of this method, the experimental scheme should be strictly set before the experiment: take the encryption security and storage efficiency of electronic documents as the experimental comparison index, and compare this method with Cheng et al. (2020) and Ren et al. (2019) methods.

- Encryption security: Encryption security refers to the intrusion rate after electronic documents are encrypted and stored. The lower the intrusion rate, the higher the encryption security of the method.
- Storage efficiency: Storage efficiency refers to the percentage of data that can be stored by different methods in the total amount of data per unit time. Therefore, the higher the storage efficiency, the stronger the storage performance of the method.

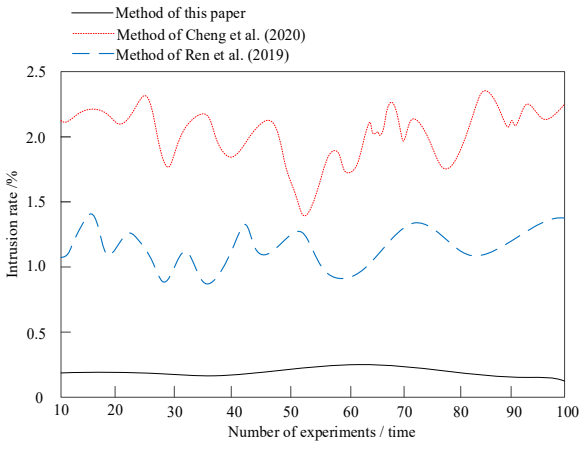
3.3 Analysis of experimental results

3.3.1 Encryption security

In the actual use of electronic documents, they are generally important documents, so the security requirements of electronic documents are higher, so the encryption security of the secure storage method of electronic documents is put forward higher requirements. Therefore, taking encryption security as the experimental comparison index, this method is compared with Cheng et al. (2020) and Ren et al. (2019) methods. The encryption security comparison results of the three methods are shown in Figure 4.

From the encryption security comparison results shown in Figure 4, it can be seen that the intrusion rate of this method is the smallest of the three methods. The maximum intrusion rate of this method does not exceed 0.5%, while the maximum intrusion rate of the two literature comparison methods reaches 2.3% and 1.4% respectively. Therefore, this method can effectively improve the security of electronic documents.

Figure 4 Encryption security comparison results (see online version for colours)



3.3.2 Storage efficiency

Due to the large amount of data in electronic documents, if the storage efficiency is too low, the use efficiency of electronic documents will be reduced. Therefore, the storage efficiency of this method needs to be verified. Set the storage efficiency as the experimental comparison index, and also compare this method with the traditional Cheng et al. (2020) and Ren et al. (2019) methods. The storage efficiency comparison results of the three methods are shown in Table 1.

Table 1 Storage efficiency comparison results

Number of experiments/time	*Storage efficiency/%*		
	Cheng et al. (2020) method	*Ren et al. (2019) method*	*Paper method*
10	79.81	72.94	95.43
20	78.75	71.73	95.35
30	79.73	72.95	96.32
40	74.77	71.69	94.75
50	80.10	70.98	94.82
60	72.31	72.73	95.75
70	79.94	73.57	94.54
80	72.85	72.29	95.69
90	73.49	74.57	94.95
100	79.43	70.98	94.73
Mean value	77.118	72.443	95.233

From the storage efficiency comparison results shown in Table 1, it can be seen that under the same experimental data and environment, the storage efficiency of this method is always higher than that of Cheng et al. (2020) and Ren et al. (2019). Under the same number of experiments, the average storage efficiency of this method is 95.233%, while the average storage efficiency of Cheng et al. (2020) and Ren et al. (2019) methods are

77.118% and 72.443% respectively. Therefore, compared with traditional methods, this method can improve the efficiency of electronic document storage.

4 Conclusions

In order to improve the security and efficiency of electronic document storage, a multi-level secure electronic document storage method based on Hash function is proposed. The performance of the method is verified from both theoretical and experimental aspects. The method has high encryption security and storage efficiency in the multi-level secure storage of electronic documents. Specifically, compared with the blockchain-based storage method, the encryption security of this method is significantly improved, and the maximum intrusion rate does not exceed 0.5%; compared with the storage method based on dynamic virtualisation electronic stream cipher, the storage efficiency of this method is significantly improved, and the average storage efficiency is 95.233%. Therefore, it fully shows that the proposed encryption method based on Hash function can better meet the requirements of multi-level secure storage of electronic documents.

References

Alferov, V.P., Drovnikova, I.G., Obukhova, L.A. et al. (2019) 'Verbal model of management of a vulnerable process of dedicating the access of users to the software of the electronic document system', *Herald of Dagestan State Technical University Technical Sciences*, Vol. 46, No. 2, pp.37–49.

Azlina, A.A. et al. (2019) 'Electronic document and records management system (EDRMS) adoption in public sector – instrument's content validation using content validation ratio (CVR)', *Journal of Physics: Conference Series*, Vol. 1196, No. 1, pp.12057–12057.

Chaniago, N., Sukarno, P. and Wardana, A.A. (2021) 'Electronic document authenticity verification of diploma and transcript using smart contract on Ethereum blockchain', *Register Jurnal Ilmiah Teknologi Sistem Informasi*, Vol. 7, No. 2, p.149.

Cheng, L., Qi, Z. and Shi, J. (2020) 'Blockchain based secure storage and sharing scheme for EHR data', *Journal of Nanjing University of Posts and Telecommunications (Natural Science)*, Vol. 40, No. 4, pp.96–102.

Daudov, K.A., Zinoviev, A.A. and Gavrilova, Z.L. (2021) 'Principles of protection, storage and movement of documents during electronic document flow within and outside the organization', *IOP Conference Series Materials Science and Engineering*, Vol. 1111, No. 1, p.12022.

Dong, Y. (2019) 'Real time storage method of electronic information resources based on artificial intelligence', *Electronic Design Engineering*, Vol. 27, No. 2, pp.144–148.

Etinkaya, Z., Erdal, E. and Ergüzen, A. (2020) 'Storage requirement estimation for electronic document management system with artificial neural networks', *Uluslararas Muhendislik Arastirma ve Gelistirme Dergisi*, Vol. 29, No. 3, pp.268–277.

Kaupa, S. and Chisa, K. (2020) 'Adoption of the electronic document records management system within the public sector in Namibia: exploring the challenges and opportunities', *International Journal of Operations Management*, Vol. 1, No. 43, pp.1231–1249.

Mazin, A., Korolev, I., Nazintsev, V. et al. (2021) 'Method for automatic control of assigning details to completed electronic document', *Issues of Radio Electronics*, Vol. 34, No. 2, pp.4–10.

Mikheev, M.A. and Yakimov, P.Y. (2019) 'Development of the documents comparison module for an electronic document management system', *Information Technology and Nanotechnology*, No. 2416, pp.527–533.

Pysarenko, V., Dorohan-Pysarenko, L. and Kantsedal, N. (2019) 'Application of new data formats for electronic document management in government bodies', *IOP Conference Series: Materials Science and Engineering*, Vol. 568, No. 1, p.12102.

Ragimova, N.A., Hajimahmud, A.V. and Soltanaga, A.V. (2020) 'Analysis of main requirements for electronic document management systems', *ScienceRise*, Vol. 1, No. 1, pp.28–31.

Ren, X., Yang, J. and Li, N. (2019) 'Secure storage based on dynamic virtualization of electronic stream cipher oriented cloud computing', *Journal of Frontiers of Computer Science & Technology*, Vol. 13, No. 8, pp.1331–1340.

Surovtseva, N.G. (2019) 'Requirements for electronic document flow systems and electronic documents storage systems: problems and solutions'.

Tkachenko, A.L. and Denisova, L.A. (2022) 'Designing an information system for the electronic document management of a university: automatic classification of documents', *Journal of Physics: Conference Series*, Vol. 2182, No. 1, p.12035.