Risks of Loss of Personal Data in the Process of Sending and Printing Documents

Oleksandr Tymchenko[1][0000-0001-6315-9375], Bohdana Havrysh[2] [0000-0003-3213-9747], Orest Khamula[3][0000-0001-7596-0813], Sergii Lysenko[4][0000-0001-7243-8747], Oleksandr O. Tymchenko[5][0000-0003-2774-2138] and Kateryna Havrysh[6][0000-0003-4155-8759]

[1] University of Warmia and Mazury Olsztyn, Poland
[2,3,5] Ukrainian Academy of Printing, Lviv, Ukraine
[4]Khmelnytskyi National University, Khmelnytskyi, Ukraine
[6] IT Step University, Lviv, Ukraine

[1]o_tymch@ukr.net, [2]dana.havrysh@gmail.com, [3]khamula@gmail.com, [4]sirogyk@ukr.net, [5]olexandr.tymch@gmail.com, [6]gavrysh.kateryna@gmail.com

Abstract. With the introduction of the General Data Protection Regulation, many companies were forced to make significant changes to their systems to meet the new requirements. General Data Protection Regulation is a regulation within the European Union legislation to protect the personal data of all people within the European Union and the European Economic Area. It also refers to the export of personal data outside the European Union and the European Economic Area. Problems of converting existing systems into conformity are analyzed, changing the use of storage systems in accordance with the standard. It has been shown that there are significant risks to the security of personal data during the printing process, including: unencrypted transmission of personal data over the network; unencrypted storage of personal data during the printing process on printers' servers or hard disks; outputting confidential documents to the wrong printers; Documents containing personal data are released to third parties from the printer. It follows that critical equipment is network printers, which transfer sensitive data to which is carried out through the corporate network. A network printer also poses a security risk. It is advisable to use a print server to centralize print processes. Not only does this simplify administration, but it also enables security technologies to be implemented. However, the applications themselves are on a dedicated server or in the cloud. The connection from a print server to a network printer can only be protected by third-party solutions. This results in a high administrative burden, as the decisions of individual printer manufacturers must be installed and managed separately on the print server. Print data that is often sent unsafe to print servers and from there to network printers must be encrypted. Certificates that require a username and password should be used for this. This paper analyzes the risks of personal data being lost in the process of sending and printing documents related to the implementation in the European Union countries of the General Data Protection Regulation.

Keywords: personal data, data storage systems, risks of personal data loss.

1 Introduction

The recently implemented General Data Protection Regulation is forcing many companies to make significant changes to their systems to meet new requirements. As Ukraine has an Association Agreement with the European Union (EU) since 2014, Ukrainian printing companies must be ready for such changes in order to be able to cooperate with European companies or provide services to foreign customers. In May 2018, a new General Regulation on Personal Data Protection (GDPR) of personal data of users, was introduced. Given this regulation, the existing data transmission processes, their processing within the information technology should be checked and all security deficiencies should be eliminated [1].

This Regulation also applies to the export of personal data outside the EU and the EEA within the framework of European Union legislation on the protection of personal data of all users within the European Union and the European Economic Area. The introduction of new functionality, which must be performed in real time (for example, synchronous recording of each user request) reduces the bandwidth of systems by 20 times. New technical challenges are emerging that need to be addressed in order to effectively achieve strict compliance with the General Data Protection Regulation. Therefore, in order to comply with the General Data Protection Regulation, it is necessary to make changes in the data processing and storage systems.

The new legislation provides new definitions on data collection and processing and raises a number of issues. What information is considered personal data? What information technology processes should affect the security of personal data? How to start restructuring information systems in accordance with the new requirements?

Therefore, existing information technology processes, in particular printing processes, should be tested for security, and any security deficiencies should be identified, corrected, and optimized. Also, the methods of using storage systems should be changed according to the GDPR standard, in particular the printing process, the transition of the document through various stages from launch to the final product.

The paper analyzes the problems of retrofitting existing systems following the law, changing the methods of using storage systems following the GDPR standard, in particular, the risks of personal data loss in the process of printing and transferring the documents for printing, the passage of the document through various stages from launch to the final product. The individual stages, existing risks, and weaknesses of the process security are illustrated [2, 4].

2 General analysis of GDPR

The General Data Protection Regulation is set out in 99 articles describing its legal requirements and 173 summaries that provide additional context and explanations to these articles. The GDPR is an expansive set of regulations covering the entire lifecycle of personal data [1, 3-5]. Thus, achieving compliance requires interoperability with infrastructure components (including computing systems, networks and storage systems) as well as operational components (processes, policies and personnel). For analysis it is necessary to use articles describing the behavior of storage systems. They fall into two broad categories: the rights of data subjects (i.e. the people whose personal data were collected) and the responsibilities of data controllers (i.e. companies that collect personal data).

Table 1 shows the impact of the articles of the General Data Protection Regulation on data storage systems, i.e. the requirements of the articles on the functions of storage systems [1].

Table 1. The main articles of the GDPR that significantly affect the design, interoperability or performance of storage systems

№	GDPR article	Key requirements	Storage functions
5.1	*Destination restrictions*	*Data must be collected and used for specific purposes*	*Metadata indexing*
5.1	*Storage restrictions*	*Only necessary data should be stored*	*Timely removal*
5.2	*Accountability*	*The controller must be able to demonstrate compliance*	*All*
13	*Terms of data collection*	*Obtain user consent for data processing*	*All*
15	*User access right*	*Provide users with timely access to all their data*	*Metadata indexing*
17	*The right to be forgotten*	*Find and delete data groups*	*Timely removal*
20	*The right to data portability*	*Transfer data to other controllers upon request*	*Metadata indexing*
21	*The right to object*	*Data should not be used for any valid reason*	*Metadata indexing*
25	*Protection by design and default*	*Protection and restriction of access to data*	*Access control, Encryption*
30	*Data processing records*	*Store audit logs of all transactions*	*Monitoring*
32	*Processing security*	*Implementation of appropriate data security measures*	*Access control, encryption*

33, 34	*Notification of personal data protection violation*	*Share information and check suspicious systems*	*Monitoring*
46	*Transmission subject to appropriate warranties*	*Control where the data is*	*Manage data location*

3 Designing systems according to requirements

Based on the analysis of GDPR articles, it is possible to identify six key features that a storage system must support in order to be compatible with the GDPR. And also, to characterize deviations of systems in support of the basic functions necessary for performance of regulations.

3.1 Storage characteristics according to the GDPR standard

Timely removal. According to the GDPR, personal data cannot be stored for an indefinite period of time. Thus, the storage system must support the mechanisms of TTL counters for personal data (the maximum period of time for which the data packet can exist until its disappearance), and then automatically remove them from all internal subsystems in a timely manner. The GDPR allows the TTL to be a static time or a policy criterion that can be objectively assessed [6].

Monitoring and logging. To demonstrate compliance, the storage system needs to check both internal and external interactions. Thus, in a strict sense, all operations, regardless of the path (e.g., read or write) or control path (say, changes to metadata or access control), must be registered.

Indexing with metadata. Storage systems must have interfaces for fast and efficient access to data groups. For example, accessing all personal data that can be processed for a specific purpose, or exporting all data that belongs to the user. In addition, you need to be able to quickly retrieve and delete large amounts of data that meet the criteria.

Access control. As the GDPR aims to restrict access to personal data only to authorized institutions, for established purposes, as well as for a predetermined period of time, the storage system must maintain fine and dynamic access control.

Encryption. The GDPR requires personal data to be encrypted in both storage and transportation. Although anonymization can help reduce the amount and size of data that needs to be encrypted, encryption is needed and is likely to degrade storage performance.

Data location management. Finally, the GDPR restricts the geographical locations where personal data may be stored. This means that storage systems must be able to find and control the physical location of data at all times [7-9, 15].

3.2 Degrees of compliance

Although the GDPR is clear in its high-level objectives, it is intentionally vague in its technical specifications. For example, the GDPR requires that personal data not be stored indefinitely and must be deleted after the expiration date. However, the regulations do not specify how soon after the data expires, they will be deleted. Seconds, hours or even days? The GDPR is silent on this, only mentioning that the data should be deleted without undue delay. What does this mean for system developers? This is that GDPR compliance should not be a fixed goal, but a spectrum. To do this, consider the variance in two dimensions: response time and capabilities [1, 8].

In real time against possible compliance. Real-time compliance is when the system completes a GDPR task (for example, deletes expired data or responds to user requests) synchronously, in real time. Otherwise, we classify tasks as those that need to be performed later. Given the harsh sanctions for violating the law (up to 4% of total revenue or € 20 million, whichever is higher), it would be advisable for companies to delete the data as soon as possible. However, the requirement to meet the requirements of the law in real time leads to significant overhead costs. This problem is exacerbated for large organizations. For example, the Google cloud platform informs its users that deleted data must be completely removed from all internal systems, but this can take up to 6 months [9-11].

Full and partial compliance. Systems that differ in response time demonstrate different levels of detail and capability. These differences are due to the fact that many of the requirements of the GDPR depend on the design principles and performance guarantees of certain systems. For example, file systems do not implement indexing to files as a basic operation, because this feature is usually supported by applications such as grep. Similarly, many relational databases only partially and indirectly support TTL, as this operation can be implemented using user-defined triggers that are inefficient. Thus, we define full compliance in order to support all GDPR functions, and partial compliance as supporting functions in combination with external infrastructures or components.

4 Characteristics of personal data protection

According to the GDPR, individuals have the right of protection of their personal data (Article 1, GDPR [10, 15]). However, the question arises as to how to recognize whether certain data or information is considered personal data. Data falls into the category of personal data when a person can be directly or indirectly identified, such as by name, telephone number, account details, postal or IP address.

Data processing is legal only if at least one of the following conditions is met (Article 6 of the GDPR - Legality of processing):

- the data subject has consented to the processing of his or her personal data for one or more special purposes;

- processing is necessary for the performance of the contract to which the data subject is a party, or for taking action at the request of the data subject prior to the conclusion of the contract;
- processing is necessary to comply with the statutory obligation that applies to the controller;
- processing is necessary to protect the vital interests of the data subject or another individual;
- elaboration is necessary for the performance of a task in the public interest or the exercise of official authority vested in the controller;
- the processing of personal data is necessary for control, except when such interests are outweighed by the interests of the fundamental rights and freedoms of the data subject, which require the protection of personal data, especially if the data subject is a child.

The right to protection of personal data is protected by Art. 5 GDPR (Principles of personal data processing) [1, 12]. By law, companies responsible for document protection are required to report data leaks in a timely manner. In addition, violations of data protection directives are subject to very high fines.

By May 25, 2018, companies had to review their IT processes, as well as document or even simplify them. Existing descriptions of IT processes, such as review processes, may need to be adjusted or even updated. One of the elements of personal data protection is the optimization of the complete printing process [13], as it often goes unnoticed that there are significant security risks in the printing process (Fig. 1). These include:

$$R_{pr} = \left\langle R_{trans}, R_{encrypt}, R_{incor}, R_{third} \right\rangle$$

where

R_{trans} - not encrypted transfer of personal data over the network;

$R_{encrypt}$ - not encrypted storage of personal data during the printing process on servers or printer hard drives;

R_{incor} - output of confidential documents to incorrect (unprotected) printers;

R_{third} - documents containing personal data are received by third parties from the printer.

The probability of losing personal data during printing will be the sum of the corresponding probabilities

$$p(R_{pr}) = p(R_{trans}) + p(R_{encrypt}) + p(R_{incor}) + p(R_{third})$$

Consider each element separately.

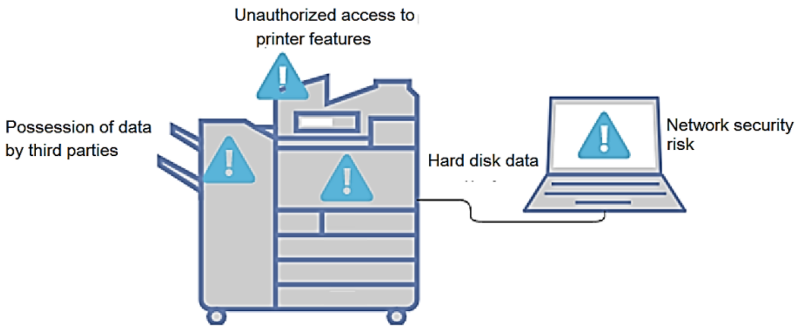
Unauthorized access to printer features
Possession of data by third parties
Hard disk data
Network security risk

Fig. 1. Security vulnerabilities of personal data during printing

4.1 Network printing

The risk of network security is quite high due to the general trend of replacing printers in the workplace with network printers. This is often used to optimize administrative and financial costs. However, this results in the transmission of sensitive data over an unsecured corporate network. If an attacker has access to network printer data, data loss is guaranteed.

The printing process begins in the application program. It runs either directly on the corresponding workstation (Fig. 2), or on the remote desktop session host, XenApp server [14-16] or virtual desktop (Fig. 3).

XenApp - software for virtualization and delivery of applications from a remote server to local devices of users through a thin client. In other words, this program allows you to run Windows programs on computers and mobile devices running other operating systems. In this case, the applications themselves are on a dedicated server or in the cloud.

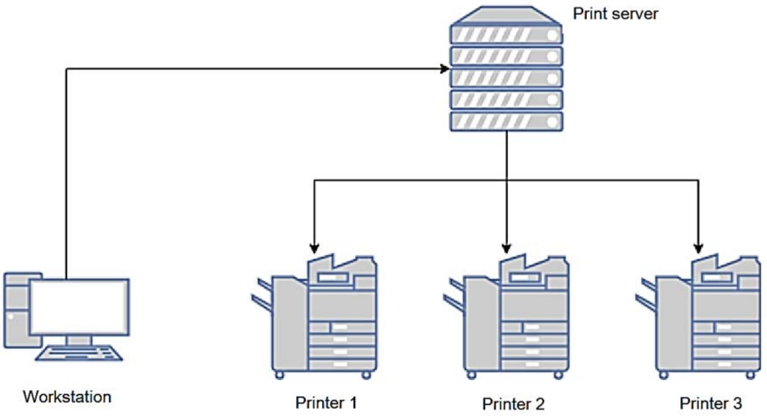

Fig. 2. Schematic of the printing process from the application, through the print server to a network printer

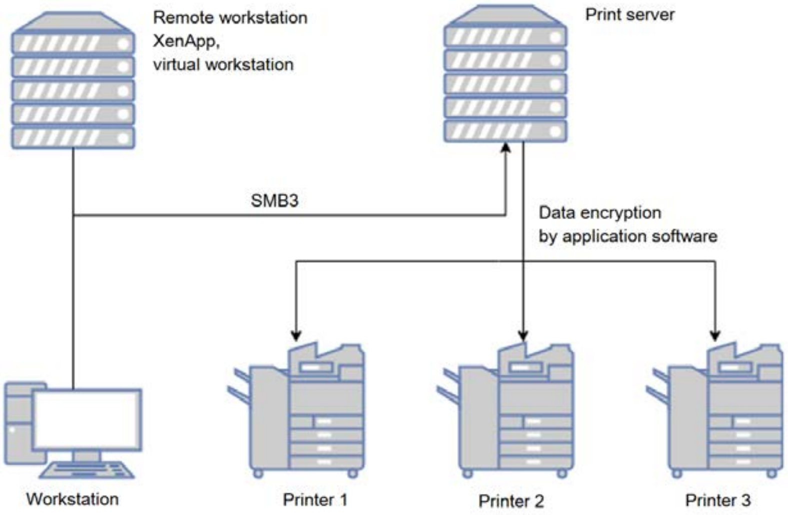

Fig. 3. Schematic of the process of encrypting print data at all stages of the program, through the print server to a network printer

You can use a print server to centralize printing processes and reduce the risk of losing personal data. This not only simplifies administration, but also makes it possible to implement security technologies, as it allows to protect data with encryption using application software [8]. In this case, the attacker needs to access a secure print server, which makes it virtually impossible to lose data, or reduces it hundreds of times [17].

4.2 Weak print points on the network

Administrators protect programs and data through access protection, and encrypt connections to servers and workstations. On the other hand, print data is often sent unsecured to print servers and from there to network printers. This leads to the following weaknesses:

- network cards of all devices covered by the print stream: workstation, desktop, network switch, router, server and network printer;
- printers share access to the print server;
- access to network printer hard drives.

5 Results of experimental studies

Encryption during printing.

There are several cases when the printing process needs to be optimized to achieve comprehensive and secure encryption of printing that meets the GDPR standard.

From the program to the print server

From the Server Message Block (SMB), the print data must be encrypted by a printer-sharing program on a print server with Windows file sharing capabilities (Fig. 2). This allows interoperability and access to shared printers on the print server only through encrypted connections, which reduces the likelihood of losing unencrypted data $p(R_{encrypt})$.

From print server to network printers

Only third-party solutions can be used to connect from the print server to network printers. The decisions of individual printer manufacturers increase the administrative burden because they must be installed and managed for each printer user on the print server and errors due to the output of data to other printers $p(R_{incor})$ That is, you need a universal, manufacturer-independent solution that is also compatible with a large number of different structures and libraries.

Network printer

On the network printer itself, the risk of data loss by third parties $p(R_{third})$ is quite high, so you need to make sure that unauthorized persons cannot enter the printer interface. To do this, use certificates that require a username and password. If you are using the printer's internal hard drive, it must be encrypted (on the hardware side). Theft of ready-made printouts from the output tray can be prevented by user authentication directly on the printer: smart card, smartphone, as well as PIN authentication.

These probabilities can be specified only if the importance of data for attackers is assessed (for example, bank details, patient medical data, ballots, etc.), but failure to comply with GDPR recommendations with the appropriate activity of attackers is likely to lead to their loss. The analysis showed that the use of encryption, even only in some stages of network printing (Fig. 3) will significantly increase the security of personal data.

The table shows the average values of information risks, which should be used to calculate the probability of data loss [9], Fig.4.

Table 2. Types of information risk

	Types of information risk	**Average rating, %**
1	*Mistakes of specialists when working with IT*	18
2	*Failures of technical means*	20
3	*Network equipment failures*	14
4	*Software failures*	15
5	*Malicious software*	12
6	*Spyware*	9
7	*Unauthorized access*	12

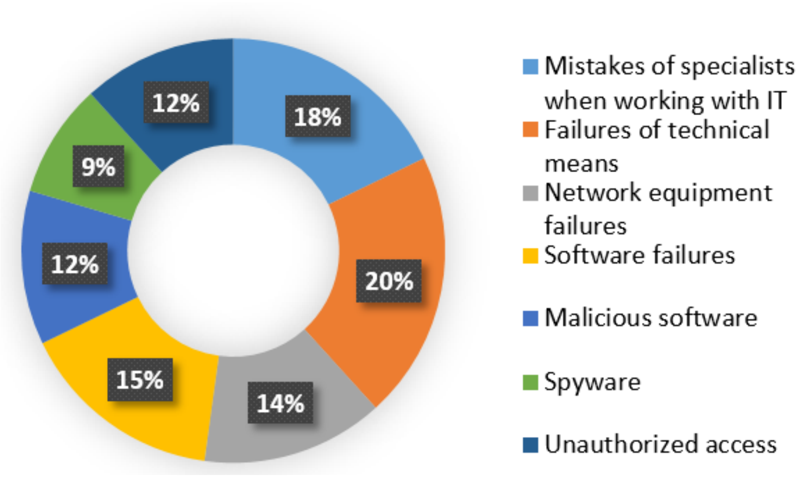

Fig.4. Types of information risk

6 Conclusion

After analyzing the impact of articles (requirements) of the GDPR on data storage, processing and transmission systems, we can conclude that achieving strict compliance with the requirements is a difficult task. Attempting to strictly comply with all regulatory requirements leads to a significant slowdown in systems. Therefore, to better study this issue, it is necessary to conduct a detailed analysis of existing data storage systems and analyze them according to the following parameters: efficient logging, deleting, indexing metadata.

As shown in this article, the printing process, which is an important element of the system of storage, processing and transmission of data, goes through various stages from launch to receipt of the final product. Therefore, the connection from the print server to the network printer can only be protected by third-party solutions. This leads to a high administrative burden, as the solutions of individual printer manufacturers must be installed and managed separately on the print server, which provides a low probability of data loss. A network printer also poses a security risk. When printing to a network printer, it is not guaranteed that users' personal data will not be disclosed to third parties. The paper proposes a reasonable scheme of the process of encrypting print data at all stages from the program, through the print server to a network printer, which minimizes losses and security of the data processing process.

Acknowledgments. The authors are appreciative to colleagues for their support and appropriate suggestions, which allowed them to improve the materials of the article.

References

1. Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)", *Op.europa.eu*, 2020. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en.
2. C. Saatci and E. Gunal, "Preserving Privacy in Personal Data Processing", *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, 2019. DOI: 10.1109/ubmyk48245.2019.8965432.
3. A. Pervushin, V. Ermachkova and A. Spivak, "Determination of loss of information during data anonymization procedure," *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT),* Baku, 2016, pp. 1-5, DOI: 10.1109/ICAICT.2016.7991650.
4. N. Elanshekhar and R. Shedge, "An effective anonymization technique of big data using suppression slicing method," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, 2017, pp. 2500-2504, DOI: 10.1109/ICECDS.2017.8389902.
5. Kyryk M., PleskankaN., TymchenkoO. Methods and models of traffic management in distributed infocommunication systems. —Lviv: Ukrainian academy of printing, 2017. — 264 p. (ISBN 978-966-322-473-2).
6. V. Gadad and C. N. Sowmyarani, "A novel utility metric to measure information loss for generalization and suppression techniques in Privacy Preserving Data publishing," *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Bengaluru, India, 2019, pp. 1-5, DOI: 10.1109/CSITSS47250.2019.9031014.
7. H. Garudadri, "Making sense of personal data in clinical settings," 2014 *48th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2014, pp. 2086-2089, DOI: 10.1109/ACSSC.2014.7094841.
8. P. C. Kaur, T. Ghorpade and V. Mane, "Analysis of data security by using anonymization techniques," *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Noida, 2016, pp. 287-293, DOI: 10.1109/CONFLUENCE.2016.7508130.
9. J. Yand, Z. Hu and J. Zhang, "Trajectory Privacy Protection Method through Active Points Hiding," *2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, Harbin, China, 2018, pp. 1101-1106, DOI: 10.1109/IMCCC.2018.00229.
10. S. Mahajan, J. Katti, A. Walunj and K. Mahalunkar, "Designing a database encryption technique for database security solution with cache," *2015 IEEE International Advance Computing Conference (IACC),*Banglore, 2015, pp. 357-360, DOI: 10.1109/IADCC.2015.7154730.
11. A. D. Yuniar and A. S. Fibrianto, "The Affect of Technical Familiarity and Consumer Protection Behavior in Using E-Commerce as Platform Online Shopping," *2019 International Seminar on Application for Technology of Information and Communication (iSemantic),* Semarang, Indonesia, 2019, pp. 300-305, DOI: 10.1109/ISEMANTIC.2019.8884265.
12. P. Prabhusundhar, V. K. N. Kumar and B. Srinivasan, "Border crossing security and privacy in biometric passport using cryptographic authentication protocol," *2013 International Conference on Computer Communication and Informatics*, Coimbatore, 2013, pp. 1-7, DOI: 10.1109/ICCCI.2013.6466144.

13. H. Zou, "Protection of Personal Information Security in the Age of Big Data," *2016 12th International Conference on Computational Intelligence and Security (CIS),* Wuxi, 2016, pp. 586-589, DOI: 10.1109/CIS.2016.0142.
14. S. Cha and K. Yeh, "A Data-Driven Security Risk Assessment Scheme for Personal Data Protection," *in IEEE Access*, vol. 6, pp. 50510-50517, 2018, DOI: 10.1109/ACCESS.2018.2868726.
15. M. Nazarkevych, I. Izonin, M. Gregus ml. and N. Lotoshynska, "An Approach towards the Protection for Printed Documents by Means of Latent Elements with Fractal Grids and Electronic Determination of Its Authenticity", *Electronics*, vol. 9, no. 4, p. 667, 2020. DOI: 10.3390/electronics9040667.
16. T. Kirkham, S. Winfield, S. Ravet and S. Kellomäki, "The Personal Data Store Approach to Personal Data Security," *in IEEE Security & Privacy*, vol. 11, no. 5, pp. 12-19, Sept.-Oct. 2013, DOI: 10.1109/MSP.2012.137.
17. L. Yuqing, "Research on Personal Information Security on Social Network in Big Data Era," *2017 International Conference on Smart Grid and Electrical Automation (ICSGEA),* Changsha, 2017, pp. 676-678, DOI: 10.1109/ICSGEA.2017.91.