Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RESEARCH AND DEVELOPMENT OF PERSONAL DATA PROTECTION SYSTEMS IN ENTERPRISES

[1]ZHANAT ABDUGULOVA, [2]ZHULDYZ TASHENOVA, [3]ELMIRA NURLYBAEVA, [4]AMANDOS TULEGULOV, [5]DASTAN YERGALIYEV

[1,2]PhD, L. N. Gumilyov Eurasian National University, Department of Information technology, Nur-Sultan, Kazakhstan

[3] PhD, The Kazakh National Academy of Arts named after T. Zhurgenova, Almaty, Kazakhstan

[4]assoc.professor, Civil Aviation Academy, Almaty, Kazakhstan
[5]professor, Civil Aviation Academy, Almaty, Kazakhstan

E-mail: [1]zhuldyz_tm@mail.ru, [2]janat_6767@mail.ru, [3]nuremuk@mail.ru, [4]tad62@yandex.kz, [5]DES-67@yandex.kz

ABSTRACT

The article proposes to develop software to protect against the spread of personal information through information channels. The research paper considers the classification and channels of information dissemination, the concept of functioning of DLP systems, the analysis of the concept of DLP systems and their types, the development of a DLP system and its comparison with similar types. At the same time, the task of creating information security is envisaged, to solve which the information system of the enterprise is characterized by threats to information security and requirements for the information security protection system. General problems of the functioning of DLP systems, their role and features of use are considered, and a comparative overview of existing software products is provided. This article is devoted to the development of an additional subsystem that prevents the leakage of confidential information, in particular, as part of the analysis of network packets at border communication nodes.

Keywords: Data Leak Prevention system, printer (LPT), modem (COM) ports, DataScanner, SMS services

1. INTRODUCTION

Relevance of the research topic. Today, information security is one of the most popular concepts. Modern life is closely linked to information technology in the modern sense, and each of us has to protect our personal data. One of the main directions in the implementation of information security is security when using remote services, which is very important when performing financial transactions.

Security in information technology is understood as a set of measures and is perceived as a single system. There can be different aspects of computer security, among them more or less important, everything is important here. It is not possible to remove some of the measures, otherwise the system will not work.

Computer security is not much different from security in the original sense. Both the files themselves and the entire network must be secure. Access to any data becomes a link in a chain of mechanisms that is securely organized and all employees who have access to information are responsible for the operation of the integrated security system.[1]

The project in question describes the information assets of the enterprise in the event of various threats and their vulnerabilities. At such moments, it became clear that the company will suffer not only financially, but also in terms of the reputation of the enterprise, which casts doubt on all future activities of the enterprise.

As the main solution for maintaining the security of the information system against information leaks, it was decided to deploy the Data Leak Prevention system (DLP).

The system for secure storage and transfer of confidential data includes the following functions:

- Detection-search for confidential and personal data from anywhere, keep records of confidential information, and automatically manage data transfer or destruction;

- Control-study of the nature of the use of confidential data, regardless of the employee's presence in the corporate network;

- Security-automatic implementation of security rules to permanently protect sensitive data and prevent them from disappearing from the organization;

- Coordination-implementation of a single policy throughout the company, reporting events and minimizing consequences, detailed content analysis, and all this is carried out on the basis of a single platform.

Such a system supports the solution of the following tasks:

- Monitoring and monitoring of the transfer of confidential and personal data through network communication channels (ftp, Web, chat, mail);

- Recording user actions on each workstation (used only in operations related to the transfer of confidential content to portable disks, printers;

- Scan the corporate network of the enterprise (including file servers, portals, distance learning systems and workstations) to detect confidential data and chaotic storage of personal data.[2]

When writing a work, an object is considered that uses data related to personal secrecy in the process of document management. Based on all this, it is necessary to check this object for compliance with the specially set requirements, and if a nonconformity is detected, we can conclude that modernization will be carried out.

The subject of research of the work is protection against leakage of confidential information from the enterprise.

The relevance of the chosen topic is due to the increasing leakage of personal data at enterprises and the importance of protection against such leaks. Similar tasks can be performed using DLP systems.[3]

When creating such systems, you can solve the following problems. Reducing confidential data leaks through key data channels:

current web traffic (FTP, P2P, HTTP, etc.);

email and internal email used;

SMS services, chats, network and local printing;

control access to the input devices and output ports of drives, USB devices, portable hard drives, infrared, printer (LPT) and modem (COM) ports.

Today, there are many products that can effectively help you control and reduce the loss of confidential data through certain channels. But there are not many complicated solutions that cover all existing channels. In such cases, it is important to choose the right technology that provides maximum efficiency and a small percentage of counterfeit positives to prevent confidential data leaks.[4]

The topic is well studied in the work of information security specialists, there are many software packages to prevent the leakage of information from enterprises.

The purpose of the work is to develop software to protect against the spread of personal information through information channels.

Objectives of scientific work:

1) classification and consideration of information distribution channels;

2) consider the concept of operation of DLP systems;

3) consider the concept of DLP systems and their types;

4) Development of a DLP system and its comparison with similar types.

Based on the novelty of the work, after the introduction of software to protect against information leaks, it can be concluded that in the current situation, the number of such leaks will be sharply reduced or stopped altogether.

Methods of scientific work, such as research, synthesis and analysis, were used in the preparation of the work.[5]

2. SAMPLES AND ANALYTICAL METHODS

The purpose of the research was to develop software to protect against the spread of personal information through information channels. The scientific work consists of three chapters.

In the first chapter, the task of creating information security was considered, for the solution of which a description of the enterprise's Information System, information security threats were given, and requirements were set for the information security protection system.[6]

In the second chapter, general problems of the functioning of DLP systems, their role and features of use were considered, and a comparative overview of existing software products was made.

The third chapter is devoted to the development of an additional subsystem that prevents the leakage of confidential information,

Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in particular, as part of the analysis of network packets at border communication nodes.

By analyzing the requirements for the DLP system, you can formulate the criteria by which the comparison and selection of the DLP system will be carried out. These goals can be considered additional requirements that the DLP system must meet.[7]

Such requirements are divided into the following three groups:

1. increase the application level;
2. increase the chances of evaluating and correlating events;
3. increasing technical capabilities.

An increase in the level of use of the DLP system indicates that this area includes the service level (servicelayer). The main problem here is multi-domain management of events and data (crosslayer). Standard requirements in such a group:

- Increase the use of the DLP system, which supports the implementation of a multi-domain approach to high-level services and processes;
- Use mechanisms for accessible and effective inter-level correlation of events.[8]

Before starting the process of launching packages, you need to select the device on which the capture will take place. The device is characterized by the" LivePacketDevice " class. After that, a communicator is created for the selected device (an object of the"PacketCommunicator" class) – this is the object that performs packet capture. There are a number of methods that implement capture on objects of this class: "ReceivePackets" – performs capture of packets, "ReceiveStatistics" – performs capture of statistics of received packets. These methods are blockers (they block the stream to be run and accept packets before the communicator's "Break" method is called), so the packet capture method is run in the "DoWork" method of the "BackgroundWorker" (background stream) component. The "ReceivePackets" methods accept the callback procedure as a parameter, including the analysis of the package and its storage in the database. Packet analysis is performed by the" PacketManager " class.[9] The main method of the class "AddPacket" is to store the package in the database and run the procedure for searching for data in the specified directory files. The search for binary data of protected files is performed by an object of the DataScanner class, which accepts the contents of the package and checks the access of the content to the signatures of protected files.

To update the statistics and list of captured packages, a timer is responsible, which is located in the main form of the application (class "MainForm").

All data from captured packages is stored in the SQLite database.[10]

Figure 1 shows the database diagram, and Figure 2 shows the class diagram.

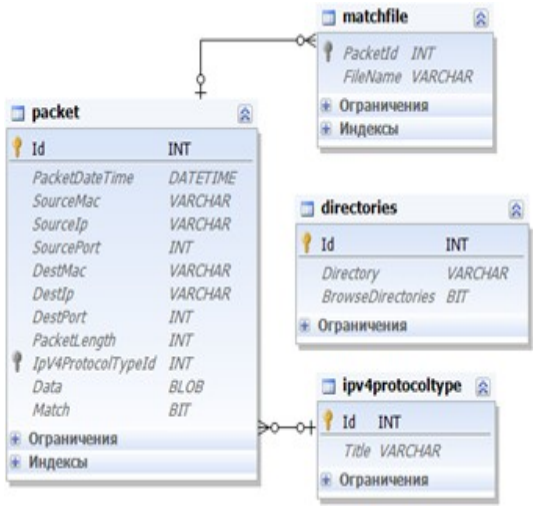

Figure 1: Ddatabase diagram

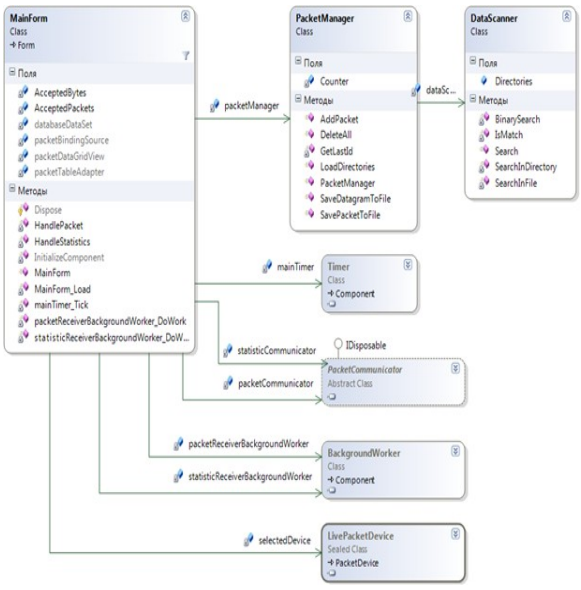

Figure 2: App class diagram

Class "MainForm".

Class of the main application form. It is the basis of the entire application and serves as the main point of the application's user interface.[11]

Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The "PacketManager" class.

A class that performs analysis and storage of captured packets. Actively interacts with the database. Basic class method – «AddPacket».

Class "DataScanner".

A class that performs a search of the contents of the package by binaries stored in the user-specified directories.

To start the program "WinPcap_4.1.3.exe" installation is required.[12]

For compilation in Visual Studio, click "sqlite-netFx40-setup-bundle - x86-2010-1.0.101.0.you need to install" exe " (you don't need to install it to run the program). Then we launch the "WindowsFormsApplication exe" application. The launch window is shown in Figure 3.

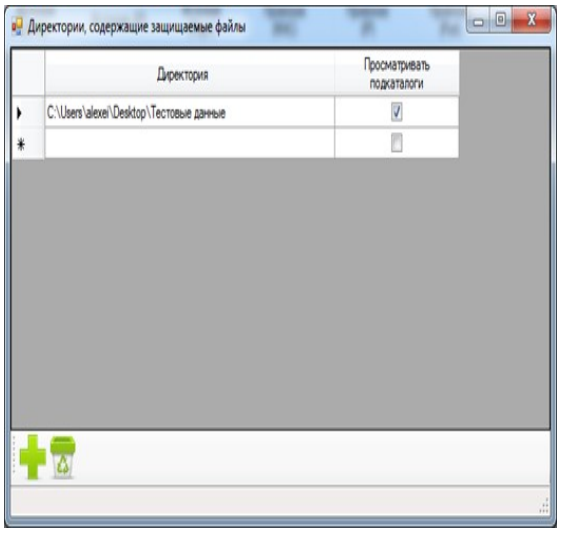

Figure 3: Launch of the project

We configure folders where protected files are stored as shown in Figure 4 (we do not select folders with multiple files or large files, as this may cause the program to freeze).

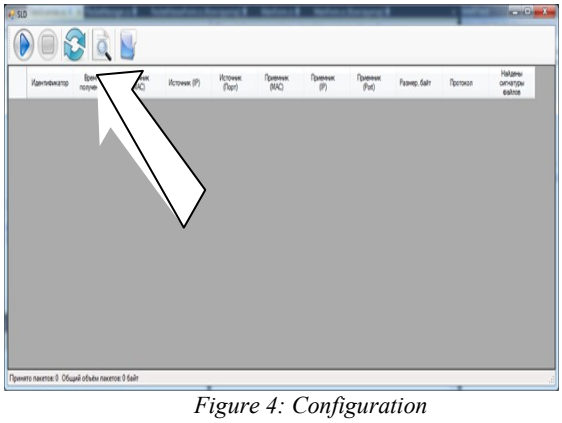

Figure 4: Configuration

The data contained in the test folder is shown in Figure 5:

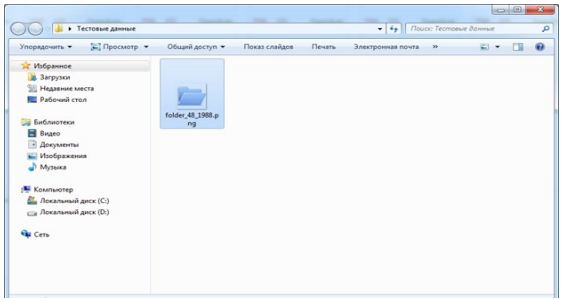

Figure 5: File to be tested

Enabling the scanning process is shown in Figure 6:

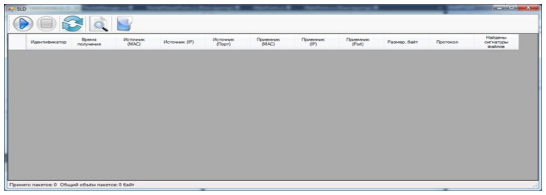

Figure 6: Starting the scanning process

Select the network interface as shown in Figure 7:

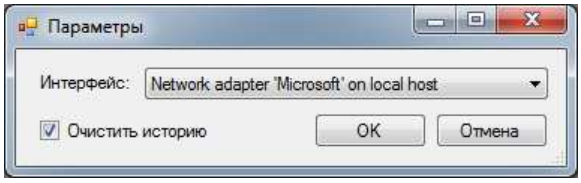

Figure 7: Network interface selection

After activation, we will try to download the same file from the browser as in Figure 8.

Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

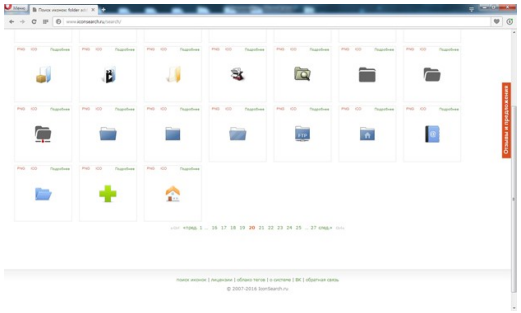

Figure 8: Testing

The application gives the following result (packages containing this file). But this search method only works for data transmitted in an unencrypted and immutable form (without compression).

Analysis of the implemented system

We will use the hierarchy analysis method to compare the developed DLP system with those discussed earlier[13].

This method is a mathematical tool that allows you to apply a systematic approach to multidimensional solution problems. This method allows you to interactively determine which option to solve the problem meets the requirements defined for its solution.[14]

This method is used in the following order:

- Initially, the purpose and options for achieving it are determined;
- A problem model is created in the form of a hierarchy, defining criteria for determining alternative quality;
- The priority of each criterion and each element of the hierarchy is distributed by pairwise comparison;
- By comparing criteria priorities, Global options priorities are determined;
- The correctness of the conclusions drawn differs;
- Based on the study, the preferred option is indicated.[15]

In this case, the purpose of the comparison is to select the most appropriate DLP system to prevent leaks in the organization's information system. To do this, we compare three of the four information leakage protection systems discussed earlier, namely Zecurion DLP, Solar Dozor, Trend Misgo Data Loss Prevention, Symantec Data Loss Prevention, and a system designed according to ten independent characteristics (three leak channels, six ways to prevent information leakage).[16]

The data required by the expert to determine the functionality of each system is obtained from the manufacturer's websites for this comparison, based on the documentation and test results of the program demos. To determine priorities, pair comparison matrices are developed, which are shown in Table 1. Expert data is formed within the framework of the above data (characteristics of the DLP systems to be provided and characteristics of the protection object). All comparisons were made on the significance scale from 1 to 9 (1 is the same value, 3 is a slight advantage, etc., inverse values - if the object in question is lower than the actual characteristic).[17]

Table 1:Pair comparison matrix.

	Channel 1	**Channel 2**	**Channel 3**	**Method 1**	**Method2**	**Method 3**	**Method 4**	**Method 5**	**Method 6**	**Cert.**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Channel 1	1	1/3	1/5	1	1	1	1	1	1	1/3
Channel 2	3	1	1/3	1	1	1	1	1	1	1/3
Channel 3	5	3	1	1	1	1	1	1	1	1/3
Method 1	1	1	1	1	1/9	1/7	1/7	1/7	1/7	1/3
Method 2	1	1	1	9	1	5	3	1	3	1/3
Method 3	1	1	1	7	1/5	1	1	1/5	1/7	1/3
Method 4	1	1	1	7	1/3	1	1	1/5	1/7	1/3
Method 5	1	1	1	7	1	5	5	1	1/3	1/3
Method 6	1	1	1	7	1/3	7	7	3	1	1/3

For each of the N matrices, a normalized vector of local priorities with the following components is defined(1):

$$\sqrt[n]{\prod_{i=1}^{n} a_{ij}} = a_j \qquad (1)$$

where N is the dimension of the Matrix element AJ is the i – th row of the Matrix.

Thus, The Matrix N is compared with the vector A.[18]

Component rationing is performed by dividing each component of Vector a by the sum of all components of this vector(2):

$$b_j = \frac{a_j}{\sum_j a_i} \quad (2)$$

Next, priorities are calculated to compare alternatives for all criteria. The result of the calculation is presented in Table 2.

Table 2: priorities for comparing alternatives by all criteria.

	Channel 1	Channel 2	Channel 3	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Zecurion DLP	0,08	0,33	0,26	0,33	0,14	0,08	0,08	0,33	0,09
Solar Dozor	0,46	0,33	0,1	0,33	0,72	0,46	0,46	0,33	0,45
Symantec Data Loss Prevent	0,46	0,33	0,64	0,33	0,14	0,46	0,46	0,33	0,45
Trend Micro DataLoss Preventio	0,33	0,1	0,33	0,33	0,1	0,33	0,33	0,1	0,33
Developed DLP	0,08	0,33	0,26	0,33	0,14	0,08	0,08	0,33	0,09

The priority vector obtained to compare the importance of criteria with each other is shown in Table 3.

Table 3: Criteria significance priorities.

Channel 1	Channel 2	Channel 3	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,06	0,07	0,1	0,02	0,13	0,057	0,059	0,116	0,139

By multiplying one matrix by another, we obtain the final vector of the priorities for the alternative (A - 0,17; B - 0,36; C - 0,46, D - 0,32; E - 0,49).

Based on the results of the calculations, we obtain the values of the overall rating of alternatives:

A - 0,17; B - 0,36; C - 0,46, D - 0,32; E - 0,49

Thus, the appraiser is a DLP system designed as the most suitable alternative for the expert.[19]

The third chapter is devoted to the development of an additional subsystem that prevents the leakage of personal information, in particular in the analysis of network packets in cross-border communication networks.

3. CONCLUSION

Each of the companies that develop a data leak protection system (DLP) usually offers a similar system structure, which differs only in detail. The main modules of such a system are:

- control modules for each channel where information leakage is possible;
- agency modules installed in end-user workplaces;
- control link with control panel for System Administrator.

Controlling modules, analyzing all information passing through channels outside the perimeter of the organization's information system, identify the data to be protected, perform its classification and distribution, and transmit this information to the DLP server for decision-making. Such modules can be installed for output and input information.

Control modules for detecting data stored on network resources perform special detection processes that may differ in the way confidential information is detected. This can be scanning traffic or running individual iTunes modules on servers, as well as on workstations.

Control modules on workstations act in accordance with a previously defined security policy, analyzing user actions with protected information, and transmit detected events to the DLP management server.

Agent programs on workstations and servers monitor compliance with the rules for processing confidential information.

The control server analyzes all the above modules, generates reports based on the results of their work using the control con

Thus, DLP provides effective protection against intentional unauthorized dissemination of information, both by employees who have any right to access the system, and by third parties.

In the course of this research work, issues related to ensuring the information security of the enterprise by developing the implementation of the DLP system were considered.

The analysis of information security threats allowed us to conclude that the number of risks associated with information calculators is increasing. According to statistics, the main channels of information leakage are internet connection channels and removable information stores.

As a result of the comparison of software and hardware, it was concluded that in order to achieve the security of confidential information and reduce the likelihood of information leakage, it is necessary to use the DLP system.

As a result of the comparison of existing DLP on the market according to a number of formulated criteria, it was found that all considered systems have a low level of protection when transmitting data over HTTP, FTP and other protocols, as a result of which it was decided to independently develop a DLP system.

After the development was carried out, it was compared with other DLP systems using the Saati method, as a result of which it was found that the implementation of this system is optimal for a number of selected parameters.

Thus, as a result of the implementation of the selected software product, the enterprise will have the following advantages::

- Ensuring the ability to identify emerging information security incidents and automate response procedures;
- Reducing the labor intensity of operations performed by an employee of the Information Security Department to identify and block information security incidents, channels of information leakage;
- More clearly understand the processes of ensuring information protection at the enterprise;
- Simplify the preparation of reports on the state of the information protection system;
- Facilitate the search for unscrupulous employees of the organization when trying to distribute confidential information without permission;
- General improvement of the information protection system and increasing the level of Information Security.

Therefore, the goal set for this research work has been achieved and the formulated tasks have been fulfilled.

REFRENCES:

[1] Serdyuk V.A. Organization and technologies of information protection. Detection and prevention of information attacks in automated systems of enterprises. - M.: Economics of zhogary mektebi, 2011. - 576 p.

[2] Skiba V.Yu., Kurbatov V.A. Guide to protection from internal threats to information security. - St. Petersburg, 2011 - 320 p.

[3] Chefranova A. O. ViPNet information protection system. Course of lectures. - M.: DMK-Press, 2015. – 392 p.

[4] Shangin V. F. Information security. - M.: DMK-Press, 2014– - 702 p.

[5] Shangin V. F. Complex protection of information in corporate systems. - M.: Forum, Infra-M, 2010. - 592 p.

[6] Dunaev VV Base data. Written SQL. - M .: БХВ-Петербург, 2016 - 288 б.

[7] Carvin B. SQL database programming. Typical errors and their elimination. - M .: Reed Group, 2013. - 336 p.

[8] Tashenova, Z., Nurlybaeva, E., Tulegulov, A., Abdugulova, Z. Sql-attack research and protection. Journal of Theoretical and Applied Information Technology. 2021, 99(19), 4536–4545 p.

[9] Michael J. Practical guidance on data manipulation in SQL. - M .: ЛОРИ, 2013. - 458 б.

[10] Markin A.V. Query construction and programming in SQL. Textbook. - M .: Диалог-Мифи, 2014. – 384 p.

[11] Yegorov M. Identification and operation of SQL injections in applications / / Complex and information security. URL: https://npo-echelon.ru/doc/echelon-sql. pdf (accessed: 12.02.2018).

[12] Martishin SA Designing and implementing databases in MySQL DBMS using MySQLWorkbench. Textbook. - M .: Forum, Infra-M, 2015. - 160 p.

[13] AirJones. SQL functions. Programmer's Handbook.- M .: Dialectics / Williams, 2014. – 556p.

[14] Graber M. Understanding SQL. - M .: Lori, 2012. - 125 p.

[15] Zhukov Yu.V. Basics of web hacking: attack and protection. - SPB .: Peter, 2012 - 208 p.

[16] Astakhova LV Theory of information security and methodology of protection of information. Chelyabinsk, 2006. - 361 p.

[17] Galatenko VA Fundamentals of information security. - SPB .: Peter, 2006. - 205 p.

[18] Joseph, J. Bambara SQL Server® Developer's Guide / Joseph J. Bambara, Paul R. Allen. - Moscow: Mir, 2016. - 235 p.

[19] Opel, Andrew J. SQL. Complete Guide / Opel Andrew J.-M.: Dialectics / Williams, 2016. - 902 p.