https://doi.org/10.21122/1029-7448-2022-65-3-240-249

UDC 697.34:004.75

An Approach to Data Processing for the Smart District Heating System

A. V. Sednin[1], A. V. Zherelo[2]

[1]Belarusian National Technical University (Minsk, Republic of Belarus),
[2]Belarusian State University (Minsk, Republic of Belarus)

Abstract. The article deals with the district heating systems transition to intelligent systems by developing a united information system and obtaining a high level of controllability of the entire system. During the implementation of automated control systems of district heating, a number of information tasks of the lower level are being introduced, including the data collection for thermal and hydraulic modes of operation for monitoring, operational management and analysis of the effectiveness. One of the problems of intelligent systems is data collection and its further storage and processing. Methods for data collection for real energy facilities are considered and the usage of multi-level system with the allocation of the upper level in the cloud storage has been proposed. In addition to the currently implemented data collection scheme in automated control systems, a generalized method of data acquisition with the introduction of duplicate streams has been proposed to ensure their integrity. The paper presents the approaches to identifying the collected data, ensuring the stability of the collection process, reliability of data storage and their integrity. Role-based security model with a dedicated single certification authority helps to protect data. Approaches to further processing of the collected data are shown, differing in the way of parallel data processing. The next stage of development is global monitoring systems that will be aimed to prompt response at all levels. The accumulated data will allow bringing the operating systems to a new level through the use of tools such as forecasting and simulation modeling, which will allow creating digital twins of heat supply systems. The proposed data collection system will perform forecasting and modeling at a higher level, and, as a result, help in the formation of more balanced management decisions.

Keywords: district heating systems, intelligent systems, process control systems, storage, simulation, monitoring, digital twin, distributed data processing, network topology, cloud computing

For citation: Sednin A. V., Zherelo A. V. (2022) An Approach to Data Processing for the Smart District Heating System. *Energetika. Proc. CIS Higher Educ. Inst. and Power Eng. Assoc.* 65 (3), 240–249. https://doi.org/10.21122/1029-7448-2022-65-3-240-249

О подходе к обработке данных для интеллектуальных систем централизованного теплоснабжения

А. В. Седнин[1], А. В. Жерело[2]

[1]Белорусский национальный технический университет (Минск, Республика Беларусь),
[2]Белорусский государственный университет (Минск, Республика Беларусь)

Реферат. В статье рассматривается проблема перехода к интеллектуальным системам теплоснабжения за счет создания единого информационного пространства и достижения

Адрес для переписки	**Address for correspondence**
Седнин Алексей Владимирович	Sednin Alexei V.
Белорусский национальный технический университет	Belarusian National Technical University
пр. Независимости, 65,	65, Nezavistimosty Ave.,
220013, г. Минск, Республика Беларусь	220013, Minsk, Republic of Belarus
Тел.: +375 17 397-36-20	Tel.: +375 17 397-36-20
Sednin@bntu.by	Sednin@bntu.by

высокого уровня управляемости всей системы. В рамках внедрения проектов автоматизации технологических процессов решается ряд информационных задач нижнего уровня, в том числе по сбору данных о тепловых и гидравлических режимах работы объектов систем теплоснабжения для осуществления контроля, оперативного управления и анализа эффективности их функционирования. Одна из проблем интеллектуальных систем связана со сбором информации для ее дальнейшего хранения и обработки. Рассмотрены методы сбора информации на реальных энергетических объектах и предложено использовать многоуровневую систему с выделением верхнего уровня в облачное хранилище. Реализованная в настоящее время схема сбора данных в автоматизированных системах управления теплоснабжением может быть дополнена обобщенным методом аккумулирования данных с введением дублирующих потоков, позволяющих обеспечить их целостность. Предложены подходы к идентификации собираемых данных, обеспечению устойчивости процесса сбора, надежности хранения и целостности. Для защиты данных можно использовать ролевую модель безопасности с выделенным единым центром сертификации, а также параллельную обработку данных. Следующий этап развития – создание систем глобального мониторинга, деятельность которых направлена на оперативное реагирование на всех уровнях. Аккумулируемый массив данных позволит вывести эксплуатируемые системы на новый уровень за счет использования таких инструментов, как прогнозирование и имитационное моделирование, и создать цифровые двойники систем теплоснабжения. Дополнительное преимущество создаваемой системы сбора данных – возможность прогнозирования и моделирования на уровне выше отдельно взятой установки или предприятия и, как следствие, помощь в формировании более взвешенных управленческих решений.

Ключевые слова: системы централизованного теплоснабжения, интеллектуальные системы, автоматизированная система управления, хранение, имитационное моделирование, мониторинг, цифровой двойник, распределенная обработка данных, топология вычислительной сети, облачные вычисления

Для цитирования: Седнин, А. В. О подходе к обработке данных для интеллектуальных систем централизованного теплоснабжения / А. В. Седнин, А. В. Жерело // *Энергетика. Изв. высш. учеб. заведений и энерг. объединений СНГ.* 2022. Т. 65, № 3. С. 240–249. https://doi.org/10.21122/1029-7448-2022-65-3-240-249

Introduction

District heating systems (DHS) has a great potential to be one of the main parts of future low-carbon energy systems [1]. The implementation of smart energy systems is a key factor of fulfilling the objectives of future low-carbon strategies [2–3]. As it stressed in [4], the smart energy system concept represents a transition from single-sector development to a coherent energy systems and understanding of how to benefit from the integration.

Through the information network, different parts of DHS (thermal source, network of pipes, substation, and heat user) are connected together and integrated into a long-distance management controlled and intelligent system [5, 6]. The automated control systems for technological processes are being replaced by intelligent control systems with a multi-parameter industrial computer and controllers that provide control of technological processes based on big data analysis. Numerous sensors and devices need to be installed to secure the data collection [7]. The system obtains the real-time data from different parts and builds a running database, in order to realize the storage and analysis of all the information on a uniform platform of management in accordance with algorithms of processing and analyzing big data [8]. Based on the smart district heating system, the smart forecasting of load, the smart regulation of heat, the smart optimization of scheduling and the smart diagnosis of fault could be realized [5].

Digital technologies have been seen to be effective as an enabler of innovation across various economic sectors [9]. A wider implementation of information and communication technologies opens up for better network management based on real time measurement data and for the integration of new digital business processes. Digitalization or the wide implementation of digital technologies in energy systems are believed to make systems smarter, more efficient and reliable [10]. According to [11] for design smart thermal grids implementation of real-time operation monitoring shall ensure that all technological process is being performed according to the plan. Aggregating data and performing operational analytics are necessary.

According to [10] the digitalization in district heating systems is demanding a large number of different sensors, automated data storage and analyses systems. Data collection and storage is the first step of proposed in [10] basic structure of methodology of identifying efficiency potentials through digitalization of district heating.

Information structure and data collecting methods

This article aims to offer some opportunities for data collection and storage systems for smart district heating. The entire information that can be obtained from the object can be divided into two main classes:

- objective data – the information collecting from sensors and presented by personnel and representing some physical parameters of the investigated object, for example, temperature, pressure, etc.;
- expert data – information generated by personnel, which cannot be expressed in numerical form or in the case when it is formed on the basis of value judgments.

Taking into account the proposed ways of collecting data, the following methods can be introduced: automated, based on sensors measurements; manual, based on application, which helps personnel to enter the necessary data.

Globally, the current trend is the transition to complete automation of the data collection with the formation of the Internet of Things (IoT).

It is proposed to use a multi-level system, where at the lower level of installations, information is obtained from sensors and operators, and the high level is a cloud data storage, which is referred as DataLake [12].

Intermediate levels are aggregation levels for grouping an incoming data according to some criterion (for example, territorial) in order to ensure the possibility of a prompt response, preliminary processing, additional control and recovery of the general data array.

Fig. 1a shows generalized diagram of data collection with main data streams. Note, it is proposed to use actually two duplicate data streams, one of which is organized traditionally, i. e. hierarchically, and the second one is a generalized bus through which data is transmitted immediately to all levels. Implementing this scheme, the main load of data transmission is assigned precisely to the second stream, while hierarchical data transmission is used primarily to ensure data integrity.

Currently, within the development of automated control system of Minsk district heating system (DHS), the lower level of data collection has been implemented [13–14]. The implemented control system primarily aims to increase reliability and quality of operational control for different parts of DHS. Simultaneously, a number of information and technological tasks are being solved, including the data collection and archiving representing the information about thermal and hydraulic regimes of heat sources, heat pipelines, pumping stations. These data are necessary for monitoring, operational management and analysis of the DHS efficiency. The process of creating the information base for solving optimization problems is also underway.

In the compound software complex of DHS, it is necessary to select monitoring servers (usually two servers), which completely duplicate each other's functions and are interchangeable [15]. The exchange with data users external to the software and hardware complex is carried out through an intermediate server – a retranslation server. Fig. 1b shows the general scheme of data collection in the process control system. It should be noted that a bi-directional flow of information exchange is essentially organized between the monitoring servers, while the relay server can only receive information and this is limited due to the use of the internal data request protocol.

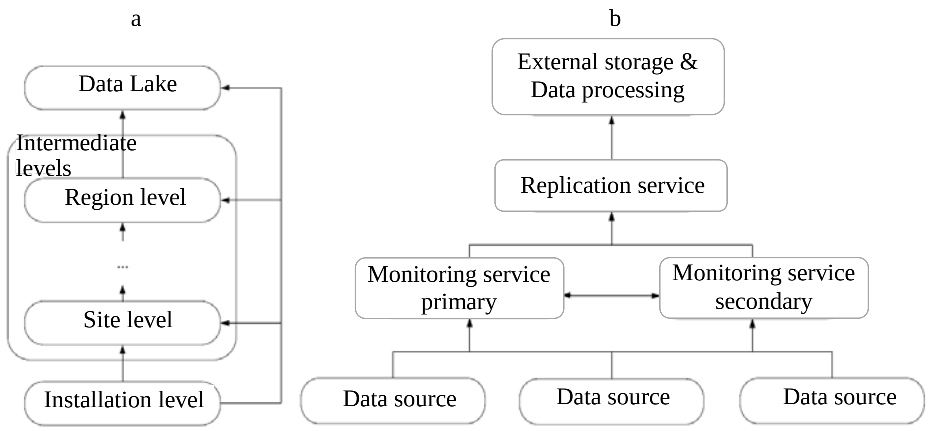

Fig. 1. Generalized diagram of data collection (a) and general scheme of data collection in the process control system (b)

Regardless of the methods of data collection, the main problem in their aggregation is the identification of data in the total volume. In this direction, it is proposed to use a hierarchical approach to identifying a chunk of data based on digital lines. In this paper, we propose the following general approach to the formation of an identifier:

- CC – country;
- DD – department;
- RRR – region;
- TT – type of installation;

- NN – number of installation;
- IIIIII – data source device number.

Then the identifier can be expressed as CCDDRRRTTNNIIIIII.

In the above identifier, each digit represents the number 0–9, since in the future it is supposed to use these identifiers in the form of linear bar-codes in order to simplify the maintenance of the information system being created (for example, search and identification of sensors/objects in case of need for repair or replacement).

The data volume that needs to be collected and processed is a classic BigData problem [12]. It is quite difficult to estimate directly the volume of data coming from one installation, because it depends on number of information sources and the time rate which the data are collected with. Two ways of collecting information can be proposed. In the first case, at the initial phase data is preliminary filtered and processed and then sent to the storage system. The second approach is based on the collection of the "raw" data with subsequent processing and using dedicated computing resources.

The first approach significantly reduces the requirements for data communication channels and computing resources used for following data storage and processing. However, at the stage of system deployment, this approach is rather difficult to use, since the quality of the collected data directly depends on the formed rules of the primary processing and, accordingly, in the process of the subsequent analysis, some important patterns will not be revealed. Therefore, at the stage of the initial implementation of the information collection system, it is better to use the second approach, although it is more demanding on resources.

Ensuring the sustainability of the data collection process

While collecting data, especially telemetry, it is necessary to highlight the formation of time series that correctly reflect the operation of the observed object. It is well known that gaps in the observed time series significantly complicate subsequent analysis and lead to filter out data or generate gaps. Such problems may occur due to the territorial distribution of the data collection system and the need to use the data communication channels of the third-party organizations, which may be either unstable or do not provide the required bandwidth.

To solve above the mentioned problems, it is useful to build a multi-level two-channel hierarchical data collection system using an intermediate storage. The key part of the proposed collection system is that data transmitted simultaneously through two logical channels. The first channel (I) transfers data to all intermediate data storage systems, and directly to the DataLake located in the cloud. The second channel (II) sequentially aggregates data to the enterprise-level storage, and then transfers it to the intermediate-level storage, for example, regional, etc. These channels are a technical implementation of the generalized data collection scheme shown in Fig. 1.

Transmitting data through the first channel could lead to possible loss due to the use of unstable communication lines, but this channel cannot be abandoned, since it provides information necessary for possible operational control. To eliminate the problem of completeness of incoming data, the second data transmission channel will be used, where local storage will be rewritten only after confirmation that data is transferred to the higher level.

Fig. 2 shows the approximate diagram of data transmission channels. In particular, only one intermediate level "Region" is used. The number of intermediate levels can be increased. Channels (II) are marked with double-headed arrows, supposed that they can be used in both directions. Since this channel, as mentioned above, will be used primarily to ensure data integrity, to organize data transmission, it will use session-type protocols based on the use of TCP as a transport layer protocol.

Channels (I) imply the immediate data transfer over the generalized bus at once to all points in the hierarchy. For this type of transmission, in order to reduce the load on transport networks and transmitting nodes, it is proposed to use protocols that do not establish connection and control data delivery, as well as that allow multicast broadcasting (for example, protocols based on the use of UDP). With this approach, partial data loss is possible. To eliminate this problem, it is also proposed to use channels (II). If the upstream density is high, it is technically possible to implement several independent buses for transmitting data to the cloud, uniting, for example, geographically localized resources (Fig. 2).

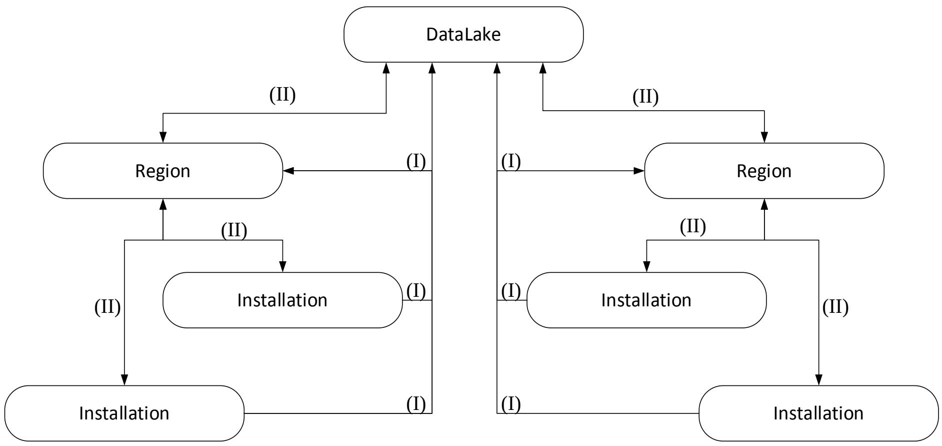

Fig. 2. An example of a data transmission channel organization scheme

Due to the importance of the processed data, it is necessary to ensure reliable storage at all system levels. At the level of an individual installation, like boilerhouse, it is possible to use hardware data backup systems, for example, RAID of various modifications [16], which is currently used in automated control system for Minsk DHS. In the case of large volumes, it is possible to use storage systems such as SAN [16].

However, at the upper levels, the total amount of data and the necessity of data processing requires the organization of storage reflection on distributed file systems, for example, Hadoop, Ceph, Gluster, etc. [17, 18]. Note that file systems of this type are already focused on ensuring reliable storage and allow setting the administrative factor of replication of stored data. Due to the peculiarities of the implementation of distributed file systems, access to stored data naturally reflects on the cloud infrastructure. The technical implementation of the access points to cloud resources can be realized in many ways, which increases the fault tolerance of the information collection process.

Data processing

Perhaps the processing of collected data is the main problem of the proposed system. The total amount of incoming data is large, so the dimension reduction could be possible, especially for solving the time restriction problems (for example, forecasting or control). Currently the dimension reduction is carried out according to the expert judgment, but, the amount of information that an expert can operate on is limited.

The purpose of collecting and processing data systems for industry is to accumulate the complete information for decision-making process. Even for full-accumulated information, there is a problem of insufficient computing resources to perform the required tasks such as searching, filtering data and calculating the simplest statistical indicators.

Here two approaches can be proposed. The first one involves the creation of a single computing center with significant computing resources in the form of the dedicated computing cluster. This approach is beneficial when it is necessary to carry out constant massive calculations (for example, as in meteorology). The essence of this approach is data transfer of to the point of computation.

The second approach is based on performing computations simultaneously with data storage, particularly for tasks related to primary data processing, since it does not require significant investments in computing infrastructure. It is more flexible due to the possibility to increase computing performance and storage space. For example, Hadoop framework could be used, which allows to solve problems both for storing large amounts of data, and for processing chunks of data directly at the nodes of storage.

The protection of communication channels is supposed to implement both at the physical level, by organizing isolated or dedicated communication channels, and at the presentation level, based on encryption of transmitted data. To organize data encryption and provide a unified network infrastructure the use of VPN with channel identification by a cryptographic key is more appropriate [19].

To protect data, a role-based security model with a dedicated single certification authority is proposed. Currently, in embed automated control system for Minsk DHS; such center operates in offline mode, while electronic keys are used for personal providing their authorization and authentication. However, if data processes in DataLake, due to the impossibility of determination of specific place for direct data processing, it is necessary to move to the online center. To consolidate information about users, their interconnections and rights, it is necessary to use a resource directory with standard interface, such as lightweight directory access protocol (LDAP).

Currently, a transition from reactive systems to proactive is required. Direct data collection and management system is not sufficient. The system should support operational and management decisions. Such systems demand significant resources, and some steps of building the system cannot be carried out in parallel due to the existing dependencies between them (for example, it is impossible to go to the modeling stage passing the data processing).

Fig. 3 shows the directions of development of the data processing.

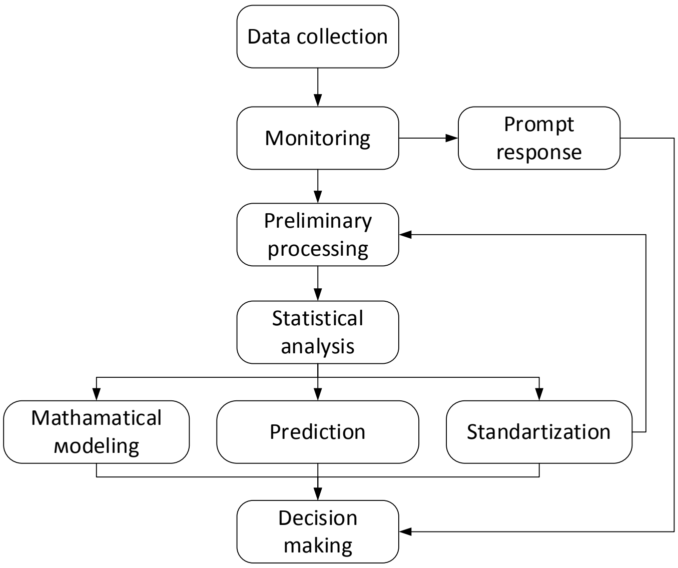

Fig. 3. Development of the data processing

At the moment, the data collection has been implemented and local monitoring tools allow prompt response to events reflected in the system at the local level (installation level). At the next stages, it is necessary to build up the monitoring system for the operational decision not only at the enterprise level, but also at the regional and global levels, as well as the introduction of primary data processing systems, for filtering, data validation and further analysis. As mentioned above, it is difficult to process data at the enterprise level due to the large amount and the variety of possible criteria, while a cloud solution allows processing with the involvement of additional computing resources on the temporary or permanent basis, for example, using MapReduce or Spark technologies [20].

After the preliminary data processing, statistical processing is possible, which helps to identify non-obvious patterns and relations in industrial process. The additional advantage of global data storage and processing is the possibility of deep comparative analysis of installations, industrial enterprises, regions, etc., which is impossible with the traditional approaches that do not stimulate the exchange of experience and the introduction of universal management solutions.

The proactive approach allows making actions to control the ongoing processes without expecting the occurrence of some event. At the moment, the elements of a proactive approach have been introduced into the implemented system, for example, there are preventive measures aimed at preventing emergencies. However, due to the complexity of the systems in operation, the competence and responsibility of personnel have a significant impact on reliability. The accumulated data storage will allow bringing the operating systems to a new level through the use of tools such as forecasting and modeling, and, as a result, the allocation of priority areas. For technological forecast it is advisable to use both classical approaches based on statistical approaches (for example, time series analysis) and relatively new methods for processing intelligent data (for example, using neural network approaches) [21], which will allow cross-checking the results obtained on different stages.

Mathematical and imitating modeling based on the mathematical formalization of physical processes is another important element of the proactive approach. The imitating modeling is aimed to produce the mathematical model that describes the functioning of the modeled system and its components as fully as possible, which is often called the digital twin of the system. They are continuously adapting to the current status of the system to ensure the ongoing accuracy with regard to the system status.

To describe a number of physical processes, without possibility to collect the information, it becomes necessary to switch to mathematical modeling. Regardless of the chosen modeling approach, it is not usually possible to conduct this type of research within enterprise due to the lack of necessary computing resources and the lack of personnel with the appropriate competencies, and, as a result, the joint work of specialists from various fields of science and industry is required.

CONCLUSIONS

1. District heating systems have a great potential and will remain one of the main parts of future low-carbon energy systems. The implementation of smart energy systems needs the real-time data from different parts and builds a running database, in order to realize the storage and analysis of all the information on a uniform platform of management in accordance with algorithms of processing and analyzing big data.

2. Based on current experience of data collection process while introducing automated control system of Minsk district heating system, a generalized method of data acquisition with the duplicate streams is proposed to ensure their integrity. The accumulated data will allow bringing the operating systems to a new level through the use of tools such as forecasting and simulation modeling, which will allow creating digital twins of heat supply systems.

3. For proposed data collection systems different aspects was discussed, such as data identification, storage volume, ensuring the sustainability, reliable storage and data integrity, data processing. The principles of development of data collection system were formed, focused on possible processing by using the existing technologies. The proposed approaches can be used in the formation of technical requirements for industry systems focused on the cloud technologies.

REFERENCES

1. Lund H., Østergaard P. A., Nielsen T. B., Werner S., Thorsen J. E., Gudmundsson O., Arabkoohsar A., Mathiesen B. V. (2021) Perspectives on Fourth and Fifth Generation District Heating. *Energy*, 227, 120520. https://doi.org/10.1016/j.energy.2021.120520.
2. Bezhan A. V. (2020) Performance Improvement of Heat Supply Systems Through the Implementation of Wind Power Plants. *Energetika. Izvestiya Vysshikh Uchebnykh Zavedenii i Energeticheskikh Ob'edinenii SNG = Energetika. Proceedings of CIS Higher Education Institutions and Power Engineering Associations,* 63 (3), 285–296. https://doi.org/10.21122/1029-7448-2020-63-3-285-296 (in Russian).
3. Sednin V. A., Bubyr T. V. (2019) Numerical Study of Complex Heat Transfer in Blown Impassable Channels of Heating Mains. *Energetika. Izvestiya Vysshikh Uchebnykh Zavedenii*

i Energeticheskikh Ob'edinenii SNG = Energetika. Proceedings of CIS Higher Education Institutions and Power Engineering Associations, 62 (1), 61–76. https://doi.org/10.21122/1029-7448-2019-62-1-61-76 (in Russian).

4. Lund H., Duic N., Østergaard P. A., Mathiesen B. V. (2018) Future District Heating Systems and Technologies: on the Role of Smart Energy Systems and 4th Generation District Heating. *Energy,* 165A, 614–619. https://doi.org/10.1016/j.energy.2018.09.115.
5. Gao L., Cui X., Ni J., Lei W., Huang T., Bai C., Yang J. (2017) Technologies in Smart District Heating System. *Energy Procedia,* 142, 1829–1834. https://doi.org/10.1016/j.egypro.2017.12.571.
6. Grzegórska A., Rybarczyk P., Lukoševičius V., Sobczak J., Rogala A. (2021) Smart Asset Management for District Heating Systems in the Baltic Sea Region. *Energies,* 14 (2), 314. https://doi.org/10.3390/en14020314.
7. Voropay N. I., Gubko M. V., Kovalev S. P., Massel' L. V., Novikov D. A., Raikov A. N., Senderov S. M., Stennikov V. A. (2019) Digital Energy Development Problems in Russia. *Problemy Upravleniya = Control Sciences,* (1), 2–14. https://doi.org/10.25728/pu.2019.1.1 (in Russian).
8. Kopaygorodsky A.N., Mamedov T.G. (2020) Architecture of the Intellectual Information System to Support Expert Decisions on Strategic Innovative Energy Development. *Information and Mathematical Technologies in Science and Management,* 4 (20) 168–176. https://doi.org/10.38028/ESI.2020.20.4.015.
9. Onile A. E., Machlev R., Petlenkov E., Levron Y., Belikov J. (2021) Uses of the Digital Twins Concept for Energy Services, Intelligent Recommendation Systems, and Demand Side Management: A Review. *Energy Reports,* 7, 997–1015. https://doi.org/10.1016/j.egyr.2021.01.090.
10. Schmidt D. (2021) Digitalization of District Heating and Cooling Systems. *Energy Reports,* 7 (4), 458–464. https://doi.org/10.1016/j.egyr.2021.08.082.
11. Lorenzen P., Janßen P., Winkel M., Klose D., Kernstock P., Schrage J., Schubert F. (2018) Design of a Smart Thermal Grid in the Wilhelmsburg District of Hamburg: Challenges and Approaches. *Energy Procedia,* 149, 499–508. https://doi.org/10.1016/j.egypro.2018.08.214.
12. Mikheev A. V. (2020) Big Data Analysis for Innovation Development Decision Making in Energy Sector. *Informatsionnye i Matematicheskie Tekhnologii v Nauke i Upravlenii = Information and Mathematical Technologies in Science and Management,* (4), 158–167. https://doi.org/10.38028/ESI.2020.20.4.014 (in Russian).
13. Sednin V. A., Sednin A. V. (2020) Reconstruction of Heat Sources of the Minsk Thermal Networks Branch of Minskenergo RUE. *Energeticheskaya Strategiya* [Power Engineering Strategy], (6), 16–22 (in Russian).
14. Sednin V. A., Sednin A. V. (2005) Automatic Control System of Regional Boiler House. *Energetika. Izvestiya Vysshikh Uchebnykh Zavedenii i Energeticheskikh Ob'edinenii SNG = Energetika. Proceedings of CIS Higher Education Institutions and Power Engineering Associations,* (4), 53–60 (in Russian).
15. Sednin V. A., Gutkovskii A. A. (2016) Modernization and Automation of the Heat Supply System. The Experience of Belarus. *Energosberezhenie* [Energy Saving], (8), 46–52 (in Russian).
16. Tate J., Beck P., Ibarra H. H., Kumaravel S., Miklas L. (2017) *Introduction to Storage Area Networks*. IBM Redbooks. Available at: https://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf.
17. Sajwan V., Yadav V., Haider M. (2015) The Hadoop Distributed File System: Architecture and Internals. *International Journal of Combined Research & Development (IJCRD),* 4 (3), 541–544.
18. Oh M., Park, S., Yoon J., Kim S., Lee Kang-won, Weil S., Yeom H., Jung M. (2018). Design of Global Data Deduplication for a Scale-Out Distributed Storage System. *IEEE 38th International Conference on Distributed Computing Systems (ICDCS),* 2–6 July 2018, 1063–1073. https://doi.org/10.1109/ICDCS.2018.00106.
19. Druk E. V., Levko I. V. (2019) Fundamentals of VPN-Networks Organization and Management for Special Purposes in Modern Information Systems. *T-Comm – Telekommunikatsii i Transport = T-Comm,* 13 (6), 19–29. https://doi.org/10.24411/2072-8735-2018-10276 (in Russian).
20. Buyya R., Broberg J., Goscinski A. (2011) *Cloud Computing: Principles and Paradigms*. Hoboken, John Wiley & Sons Inc. https://doi.org/10.1002/9780470940105.
21. Nikolenko S. (2018) *Deep Learning*. St. Petersburg, Peter Publ. 480 (in Russian).

Поступила 07.12.2021 Подписана в печать 08.02.2022 Опубликована онлайн 31.05.2022