SAINT PETERSBURG STATE UNIVERSITY

As a manuscript

KIYAMOV

Jasur Utkirovich

ON THE PROBLEMS OF OPTIMIZATION AND SECURITY FOR A MULTILEVEL VIRTUAL NETWORK

1.2.2. Mathematical modeling,
numerical methods and software packages

THESIS

For a degree of
candidate for technical sciences

Translation from Russian

Scientific supervisor:
Doctor of Physical and
Mathematical Sciences, Professor,
Bogdanov Alexander Vladimirovich

Saint Petersburg
2023

Content

INTRODUCTION

The progress of information technology is inextricably linked with the widespread use of computationally intensive applications in various fields of activity. These applications, which require significant computing resources, play a key role in both scientific research and advanced industrial sectors, including aircraft, shipbuilding, biotechnology, pharmaceuticals, genetics, and others.

In scientific research, the use of computationally intensive applications has become an integral part of the work of scientists. They are used to model complex physical and biological processes, analyze large amounts of data, predict experimental results, and develop new theories. These applications allow us to expand the boundaries of our knowledge and deepen our understanding of fundamental scientific principles.

In industry, computationally intensive applications are widely used in advanced industries that require a high degree of accuracy, complex modeling and optimization. In the aircraft industry, they are used to develop new aircraft, predict and analyze aerodynamic characteristics. In shipbuilding, they help to optimize the design of ships, improve their maneuverability and efficiency. In biotechnology and pharmaceuticals, computational applications play a critical role in the development of new drugs and the modeling of biological processes. In genetics, they help analyze genomes, investigate genetic diseases, and conduct personalized medicine.

The use of computationally intensive applications in these cutting-edge industries is driving progress, streamlining processes, and achieving new results. They help accelerate and improve scientific research, improve the quality and efficiency of industrial development, and also promote innovation and the development of important sectors of the economy.

In the context of significant changes in the technological platform and the transition to a new era, it is necessary to develop new approaches to building information systems of the future. In fact, most of these systems are distributed, and as their importance grows,

security problems arise. Recently, blockchain technology has been presented as one of the possibilities for solving these problems. Given the importance of blockchain in the context of 6G mobile networks, it is necessary to explore the various possibilities of this technology and the challenges associated with its implementation. At the moment, there is no comprehensive overview that covers the key aspects of the role of blockchain in 6G. It is all the more important to create such an overview, considering both the technical features of 6G and the applications and use cases envisioned by this new era.

It is important to study the key ideas, applications, requirements and core technologies that are shaping the 6G ecosystem in order to understand the trends driving future 6G applications and define clear requirements for this new network, including key enabling technologies. In the context of this study, blockchain is one of the key technologies that is attracting particular attention.

Among the many aspects of blockchain applications, it is important to highlight a high-level understanding of its role in 6G trends and applications, given the observed trends and anticipated requirements, it is necessary to understand what the blockchain has to offer for 6G. This creates the basis for the most detailed study of the use of the blockchain in the 6G ecosystem.

Today it is clear that large distributed systems can best be built on the basis of cloud computing. Cloud computing is a computing model where resources such as computing power, data storage, networks, and software are provided as services over the Internet to a remote user. They provide resource provisioning availability, dynamic scalability, and virtually limitless possibilities for solving various problems. Cloud computing technology has a number of benefits, including high performance, lower costs, high availability, and easy scalability.

However, in its practical implementation, a number of still unresolved scientific problems arise that prevent the full use of all the potential advantages of this approach. In the practical development of these technologies, it is necessary to solve a number of technical problems, of which we highlight:

First, in the process of creating a universal cloud system, there is a need to work in a heterogeneous environment and ensure that users can access their individual

applications without sacrificing performance. This means that the system must be able to support various platforms, programming languages and programming environments so that users can work with applications designed for their specific requirements. At the same time, it is necessary to ensure high performance and responsiveness of the system so that users can perform their tasks efficiently.

Secondly, the security and reliability of storing individual data in cloud environments and organizing access for multiple users is a significant problem. Security is one of the important challenges of cloud computing as it affects the entire system. It is important to secure the APIs used to manage resources, virtual machines, and services. These interfaces must provide user authentication and authorization, as well as data encryption to protect against unauthorized access. In addition, the system should provide convenient and consistent authorized access to resources, accounting for resource usage, and protection against unauthorized use of data and resources.

Thirdly, the practical use of a heterogeneous cloud environment in various areas requires the development of a universal system for launching individual applications. This system should provide the ability to launch and run applications in a heterogeneous cloud environment, where different users can work with their own applications without loss of performance. Such a system must be flexible and scalable to support a variety of user requirements and make efficient use of computing resources. It must also provide resource management and performance monitoring to ensure that applications perform optimally in a heterogeneous environment.

Purpose of the study.

This study is focused on the study of methods of theoretical analysis and experimental research related to the organization of a user access system to a distributed computing environment based on cloud computing technology. The main focus is on the following aspects:

1. User authorization methods based on the one-stop shop principle. The possibility of creating a single authorization mechanism that allows users to access various resources in a cloud environment without the need for re-

authentication is being explored. This improves the usability of the system and reduces the risk of multiple authentication vulnerabilities.

2. Methods for building an open source cloud infrastructure. The possibility of creating a cloud infrastructure based on open standards and open source software is being explored. This reduces dependency on specific cloud service providers and allows users to customize and modify the infrastructure to suit their needs.
3. Creation of an efficient distributed computing environment based on cloud computing technology. Methods are being studied to optimize the use of resources and improve performance in a distributed computing environment. Various resource scheduling algorithms, mechanisms for load balancing and optimization of computations are considered.
4. Methods of integration and consolidation of software products in a distributed computing environment. Approaches are being explored that allow integrating and combining various software products in a cloud environment in order to ensure their interaction and collaboration. The issues of compatibility, standardization and interfaces for effective interaction of software components are considered.

During the study, both theoretical analysis and experimentation are carried out to better understand these methods and their applicability in real-world scenarios. This improves the practical applicability and efficiency of the developed approaches and creates a basis for the development of more advanced and reliable distributed computing systems in the cloud.

Subject of study.

This study is focused on the study of methods of theoretical analysis and experimental research related to the organization of a user access system to a distributed computing environment based on cloud computing technology. The main focus is on the following aspects:

1. User authorization methods based on the one-stop shop principle. The possibility of creating a single authorization mechanism that allows users to

access various resources in a cloud environment without the need for re-authentication is being explored. This improves the usability of the system and reduces the risk of multiple authentication vulnerabilities.

2. Methods for building an open source cloud infrastructure. The possibility of creating a cloud infrastructure based on open standards and open source software is being explored. This reduces dependency on specific cloud service providers and allows users to customize and modify the infrastructure to suit their needs.
3. Creation of an efficient distributed computing environment based on cloud computing technology. Methods are being studied to optimize the use of resources and improve performance in a distributed computing environment. Various resource scheduling algorithms, mechanisms for load balancing and optimization of computations are considered.
4. Methods of integration and consolidation of software products in a distributed computing environment. Approaches are being explored that allow integrating and combining various software products in a cloud environment in order to ensure their interaction and collaboration. The issues of compatibility, standardization and interfaces for effective interaction of software components are considered.

During the study, both theoretical analysis and experimentation are carried out to better understand these methods and their applicability in real-world scenarios. This improves the practical applicability and efficiency of the developed approaches and creates a basis for the development of more advanced and reliable distributed computing systems in the cloud.

Research methods.

In this research work, modern methods are used, based on the principles of parallel and distributed information processing, data transmission in computer systems, as well as on the principles of protection of computer systems. The work also uses modern software design technologies.

To analyze and ensure the reliability of information systems, the study is based on the theory of information systems reliability, the theory of random processes and flows. These methods make it possible to analyze various aspects of the reliability of information systems, including assessing the probability of failures and failures, assessing the reliability of the system as a whole, as well as determining the optimal strategies for detecting and recovering failures.

The use of modern principles and technologies makes it possible to conduct a study that contributes to the achievement of the set goals of the work. This includes improving the efficiency and performance of information systems, ensuring their reliability and security, as well as developing optimal strategies and methods for managing data and resources.

Scientific novelty of the work.

1. As part of this research work, a new methodology for organizing a computing system using a multi-level virtual blockchain network was developed. This technique is aimed at improving the efficiency of the system by applying the principles and mechanisms of blockchain technology. A multi-level virtual blockchain network allows you to effectively distribute computing resources and ensure data security, creating a reliable and fault-tolerant environment for computing tasks. This approach is an innovative way of organizing computing systems, which can significantly increase their efficiency and reliability.

2. As a result of this study, a new methodology was developed based on existing methods, which is aimed at increasing the degree of data protection and computing in a virtualized blockchain environment. This technique includes a multi-level protection system that ensures the reliability and security of transactions within the blockchain network.

The reliability of scientific results and conclusions is confirmed by the results of testing algorithms and software, as well as the practical use of the developed algorithmic and software methods and tools on the operating software and hardware complex of the faculty of PM-PU of St. Petersburg State University. In addition, the reliability of

scientific results and conclusions is confirmed by approbation of research results at a number of scientific conferences.

Scientific provisions submitted for defense.

1. A new methodology and a set of programs based on this methodology have been developed to create an operating environment for a multi-level virtual blockchain network. This approach can significantly increase the overall performance of heterogeneous hardware and software systems. On average, performance is improved by an order of magnitude by adapting the architecture of each individual virtual machine to a specific user application. Thus, a new methodology and a set of programs provide more efficient use of computing resources and optimal functioning of a multi-level virtual blockchain network. This represents significant progress in improving the performance of heterogeneous appliances and adapting virtual machines to specific user applications.
2. A methodology has been developed to create a cloud computing system that contributes to an increase in overall performance. This is achieved by virtualizing not only processors, but also memory and communication networks. An important feature of this technique is dynamic balancing and management of process migration, not data. This allows you to optimize the use of system resources, ensure efficient load balancing and improve performance in the cloud environment. Thus, this technique represents an important step in the field of cloud computing optimization and system performance improvement.
3. A technique has been developed that allows increasing the degree of data and resource security by introducing a multi-level protection system. This technique contributes to a higher degree of reliability of the system as a whole. By applying various levels of protection, including authentication, authorization, encryption, and other security measures, this technique provides effective protection of data and resources from unauthorized access.

Approbation of work. The main results of the work were reported and discussed at national and international scientific and technical conferences.

Publications

1. Bogdanov, A. V. Digitalization of healthcare: what can be done now / A. V. Bogdanov, N. M. Zalutskaya, N. L. Shchegoleva, N. R. Zaynalov, J. U Kiyamov, A. G. Dick // INFORMATION SOCIETY. — 2022. — № 5. — P. 58–70.

2. Bogdanov, A. A Multilayer Approach to the Security of Blockchain Networks of the Future / A. Bogdanov, A. Degtyarev, N. Shchegoleva, V. Korkhov, V. Khvatov, N. Zaynalov, J. Kiyamov, A. Dik // Computational Science and Its Applications — ICCSA 2022 Workshops. — 2022. — P. 205–216. («Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)»; Vol. 13377).

3. Shchegoleva, N. New Technologies for Storing and Transferring Personal Data / N. Shchegoleva, N. Zalutskaya, A. Dambaeva, J. Kiyamov, A. Dik ; O. Gervasi, B. Murgante, S. Misra, A. M. A. C. Rocha, C. Garau (Eds.) // Computational Science and Its Applications — ICCSA 2022. — Cham : Springer Nature, 2022. — Vol. 13380. — P. 680–692. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 13380 LNCS).

4. Bogdanov, A. Comparative analysis and applicability determination for several dlt solutions / A. Bogdanov, V. Korkhov, N. Shchegoleva, V. Khvatov, N. Zaynalov, J. Kiyamov, A. Faradzhov, A. Dik // Proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education". — 2021. — Vol. 3041. — P. 592–596. — (CEUR Workshop Proceedings).

5. Degtyarev, A. Risk Model of Application of Lifting Methods / A. Degtyarev, A. Bogdanov, N. Shchegoleva, A. Dik, V. Khvatov, N. Zaynalov, J. Kiyamov, A. Faradzhov // Proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education". — 2021. — Vol. 3041. — P. 369–374. — (CEUR Workshop Proceedings).

6. Degtyarev, A. Solving the Problems of Byzantine Generals Using Blockchain Technology / A. Degtyarev, A. Bogdanov, N. Shchegoleva, V. Korkhov, A. Dik, J. Kiyamov, A. Faradzhov, N. Zaynalov // Proceedings of the 9th International Conference "Distributed Computing and Grid Technologies in Science and Education". — 2021. — Vol. 3041. — P. 573–578. — (CEUR Workshop Proceedings).

7. Bogdanov, A. Testing and Comparative Analysis of the F-BFT-based DLT Solution / A. Bogdanov, N. Shchegoleva, V. Korkhov, V. Khvatov, N. Zaynalov, J. Kiyamov, A. Dik, A. Faradzhov ; O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. M. Rocha, E. Tarantino, C. M. Torre (Eds.) // Computational Science and Its Applications — ICCSA 2021. — Cham : Springer Nature, 2022. — P. 31–41. — (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12952 LNCS).

8. Bogdanov, A. Protection of Personal Data Using Anonymization /, A. Bogdanov, A. Degtyarev, N. Shchegoleva, V. Korkhov, V. Khvatov, N. Zaynalov, J. Kiyamov, A. Dik, A. Faradzhov ; O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. M. Rocha, E. Tarantino, C. M. Torre (Eds.). — Computational Science and Its Applications — ICCSA 2021. — Cham : Springer Nature, 2021. — P. 447–459. — (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12956 LNCS).

9. Zaynalov, N. Hiding short message text in the uzbek language / N. Zaynalov, U. Narzullaev, A. Muhamadiev, O. Mavlonov, J. Kiyamov, D. Qilichev // 2020 International Conference on Information Science and Communications Technologies, (ICISCT). — Institute of Electrical and Electronics Engineers Inc., 2020. — 9351521.

10. Zaynalov, N., Information Security Issues for Travel Companies / N. Zaynalov, A. Mukhamadiev, B. Ugli, O. Mavlonov, J. Kiyamov, Q Dusmurod. ноя 2019, International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities (ICISCT 2019). — Institute of Electrical and Electronics Engineers Inc., 2019. — 9011896).

I-CHAPTER. MODERN APPROACH TO SHORTWAVE DATA EXCHANGE TECHNOLOGIES IN DISTRIBUTED REGISTRIES

1.1. Using next generation networks as a distributed network

How blockchain can improve the technical aspects of 6G: Next-generation mobile networks will require significant improvements in existing technical aspects in the current generation, as well as new technical building blocks. For each technical aspect considered, key issues are identified and blockchain data security is explored by providing decentralized and cryptographic database protection, where each block of information is confirmed and verified by a network of participants. In addition, efforts have been made to fairly present both the pros and cons of using blockchain for the technical aspects under consideration.

6G mobile systems must be non-volatile on both the infrastructure side and the device side to ensure uninterrupted connectivity anywhere in the world. The development of energy harvesting capabilities will extend the life cycle of both network infrastructure devices and end devices such as IoT devices [11, 40].

Advances in sensor technologies and their direct integration with mobile networks, including low power communication capabilities, will lead to the creation of advanced 6G networks that combine communication, discovery, control, localization and computing in one system [63]. Through this integration, the 6G network will be able to provide sensing and localization services, as well as advanced communication and computing functions. [55].

Typically, IoT devices consume more power for communication than for recognition and data processing, but the development of ultra-low power communication mechanisms and efficient power harvesting mechanisms can lead to zero power or energy efficient IoT devices. [59].

Advances in wireless communication technologies, including coding schemes and antenna technologies, will expand the use of available spectrum and increase the amount of information transmitted over existing wireless channels, allowing more information bits to be transmitted more reliably. [31,36].

Blockchain has gone beyond its original use in cryptocurrencies and is increasingly becoming a promising technology in other industries such as supply chain management, healthcare, smart manufacturing, education, and other types of business and commerce. [3, 68]. Blockchain has huge potential to transform the way peer-to-peer transactions, log management, record keeping, decentralized negotiation, trade agreements, auditing and compliance, dispute resolution, access control, and secure automation in various sectors.

Following the general trend that we have already seen in the development of 5G, it is safe to assume that 6G will include more and more software, virtualized, intelligent and programmable systems. [12, 70]. However, it is interesting to note the following: on the one hand, concepts such as programming, virtualization and programmability of next generation mobile networks will lead to huge benefits, such as flexible and open network management and orchestration, multiservices and microservices, virtual networking on demand.

On the other hand, these concepts tend to exacerbate issues such as security vulnerabilities, confidential information leaks, unauthorized access to user data, disputed soft spectrum sharing, spoofed or corrupted software network functions and management APIs, illegal resource usage and failure to provide differential security for differentiated services [65, 69]. Blockchain can play an important role in solving these problems. For example, in 6G networks, open source software solutions may be available to implement various network functions at different levels, including core, transport, and access. This can reduce security vulnerabilities, reduce the risk of confidential information leakage and unauthorized access to user data, and provide a more differentiated level of security for different types of services. [56, 66].

The use of blockchain in an open environment can guarantee the correct functioning, versioning, integrity and overall security and trust management of software available for deployment in 6G networks (Figure 1). This can help improve the technical

aspects of 6G mobile networks and enhance the application and implementation of 6G in various fields, applications and use cases [15].

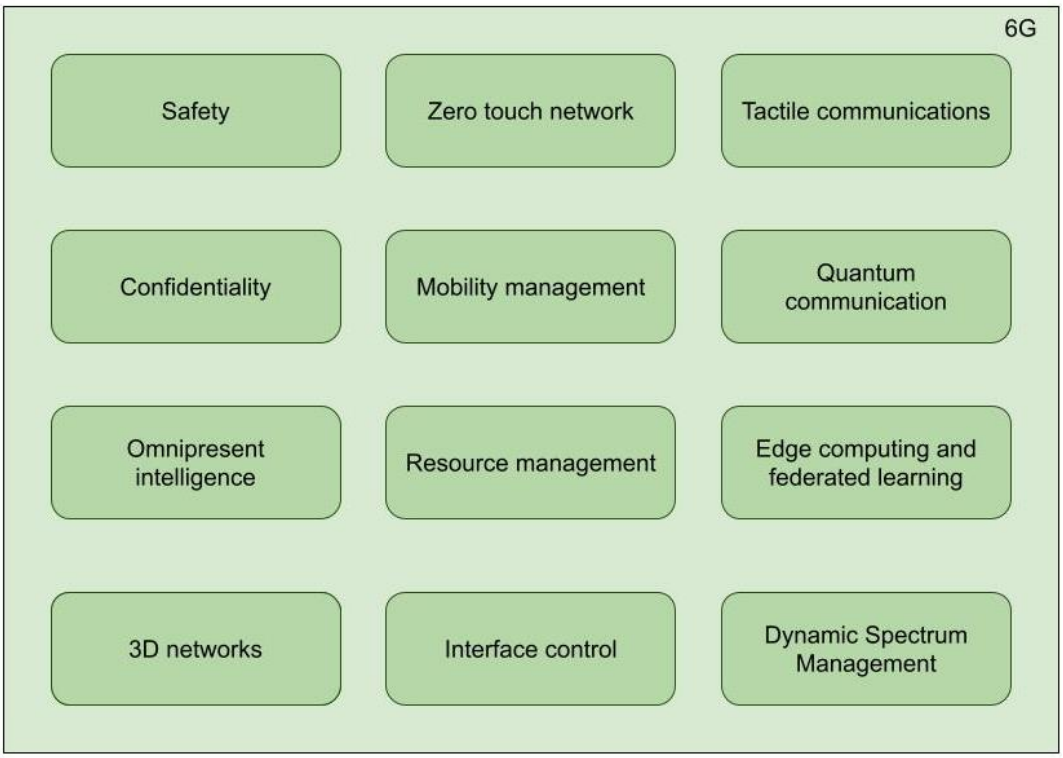

Figure 1 - Using Blockchain to Improve the Technical Aspects of the Universal Communication System Envisioned by 6G

Some of the hurdles that exist in the creation of the 3D network paradigm are as follows:

- The blockchain network can provide an effective solution for data management and security in heterogeneous networks where there is an interoperability issue. Blockchain allows decentralized versioning and data integrity management, as well as ensuring transparency and reliability in the process of data transfer between different devices and systems.
- Adding height as a new dimension may introduce new vulnerabilities because attackers can use it to gain access to security keys in a distributed network.

Thus, the blockchain is proposed to be used to ensure the secure and decentralized management of heterogeneous networks that are located on the ground, in the air and in

space. Through the use of well-defined smart contracts and blockchain technology, it is possible to create a decentralized solution for security key management, authentication, authorization and auditability, as well as interoperability between different types of networks. This will ensure a high level of security and protection against cyber attacks in integrated heterogeneous networks.

1.2. Spectrum management with dynamic resource allocation

The radio frequency spectrum is a scarce and irreplaceable resource for mobile communications, and its overall management (distribution, discovery, and sharing) must be done efficiently in 6G networks.

The traditional fixed spectrum access (FSA) policy approach to static spectrum allocation, which is used by communications regulators, may allow for legitimate spectrum use, but may also lead to under-utilization of allocated spectrum [37]. In recent years, various Dynamic Spectrum Management (DSM) mechanisms have emerged, such as Dynamic Spectrum Access (DSA), Licensed Shared Spectrum (LSS), and Spectrum Access System (SAS) [22]. In general, the CRS system includes two types of users: a primary user and a secondary user. The main user is an authorized entity with the exclusive right to the allocated spectrum. It may not be able to fully use the spectrum all the time. The secondary user is the one who gets access to the unused spectrum shared by the primary user. ACD allows secondary user devices to constantly evaluate the RF spectrum environment and automatically adjust operating frequencies to accommodate bandwidth conditions [17].

The emergence of new mobile operators in the market leads to increased competition and the need to use spectrum more efficiently to provide high quality services. In this context, dynamic spectrum allocation to operators can be one way to optimize resource utilization and provide a satisfactory level of service to end users.

Main problems:

- The traditional (centralized) way of allocating and managing spectrum presents many challenges. These are high administrative costs, security vulnerabilities, privacy leaks, and difficulties in providing transparent user network access.
- Analogue data transmission methods for users have their drawbacks, such as limited access and the possibility of DoS attacks, which can cause network congestion and reduce network performance.
- In the case of private network sharing, the main issue is the management of the priority access license and the general authorized access of users in order to ensure their coexistence through relative spectrum sharing [10]. The development of a dedicated SAS access system will be critical to 6G as private networks are expected to continue to exist. However, the creation of such a system will be a complex task requiring significant efforts in development and implementation.

Blockchain has become a promising technology for dynamic distribution of access and general management, including for conducting auctions and settling payments. The FCC has already recognized the potential of blockchain for effective spectrum management [51, 64]. The use of blockchain technology can significantly improve spectrum monitoring and simplify the audit process, making it more transparent and efficient to implement sharing rules between local mobile operators [58].

The use of distribution by participation of several organizations with common spectrum access rights differs from the traditional method. Here, users can exchange spectrum rights to suit their needs. Blockchain can be used to create decentralized trading platforms with open access to the spectrum. However, blockchain-based open trade in spectrum resources can lead to confidential issues, such as trading patterns being exposed or bids made in auction-based spectrum allocation being leaked. Various privacy protection mechanisms based on blockchain and smart contracts can be used to protect merchants from such privacy threats.

To implement this approach, you can use the dual auction scheme on the blockchain platform using smart contracts. To ensure user privacy during trading,

differential privacy is used in addition to symmetric encryption. Integer linear programming is used to determine the winner in the double auction scheme. To solve the problems of privacy, administrative overhead and single point of failure associated with the traditional broadband broadcasting service, a distributed model using blockchain has been proposed. The scheme, originally proposed by the FCC, is seen as a potential candidate for high spectral efficiency requirements for 6G as a new consensus algorithm that is not only lightweight, but also integrates with the spectrum allocation process, i.e. the system reaches consensus by evaluating the strategy spectrum distribution. In addition, the use of the ring signature method protects privacy by ensuring that there is no connection between the real identity of users and their aliases.

With the use of blockchain, it is possible to implement a two-level hierarchical architecture of a resource management system, which can be supplemented with artificial intelligence. In particular, it demonstrates the use of blockchain to improve the function of regulating and accounting for financial settlements. AI is supposed to enable usage pattern recognition through deep reinforcement learning and intelligent decision making. In addition, the use of a hierarchical blockchain reduces administrative costs, reduces computational and storage overhead, and minimizes the complexity of consensus algorithms. However, blockchain overheads and methods for obtaining training data for AI models with full privacy are still difficult.

Blockchain can provide fast financial agreements for real-time sharing between different infrastructure operators and between mobile operators [60]. In particular, spectrum tokenization in the form of a new virtual coin. Thus, their use makes it possible to abandon financial (central) exchanges and speed up financial calculations. to maximize profits, minimize losses and ensure break-even trading.

The use of blockchain helps human-to-human (H2H) communication users to interact with mobile networks to meet the huge connectivity needs for M2M communication. The blockchain-backed architecture provides privacy for sharing costs, provides optimized incentives for H2H users through the use of smart contracts, and optimizes registry allocation for M2M devices by developing consensus. In general, this

technique can be useful for dynamic spectrum management in the future convergence of heterogeneous networks with 3D networks.

In the context of mobile operators, blockchain has been proposed as a service to automate the secure management and use of spectrum between mobile operators and local operators. This service can provide:

- choice of spectrum owners;
- public access to the secondary market;
- P2P direct payments for sharing and dynamic agreement creation using smart contracts. [42].

Security has always been an important feature of telecommunications and networks. Starting with the lack of a security mechanism and restrictions in the 1G network, authentication, anonymity and encryption-based security have been introduced in 2G [30]. In 3G, 2G security features have been enhanced with the introduction of authentication and key agreement, two-way authentication, and 3GPP air interface security and network user authentication. With 4G Enhanced Packet System-AKA (EPS AKA) trust mechanism and handover, key management is introduced. 5G includes many advanced security features such as enhanced mutual authentication capabilities, Hidden Subscription ID, Extensible Authentication Protocol (EAP-AKA) [28], and 5G identification mechanism based on an integrated elliptic curve encryption scheme. [26]. In future 6G networks, many security issues may arise due to the extremely large network of heterogeneous networks and the expected extremely high reliability requirements. To thwart more sophisticated adversaries, advanced and intelligent security mechanisms will be required [47, 48]. Some of the security issues have been identified in alleged 6G networks:

- Confidentiality and integrity will be challenging as the future 6G infrastructure could create huge threat surfaces with wireless connectivity and huge amounts of data generated on the network.

- Continuous service availability will be another challenge as a wider threat surface and greater connectivity will increase the risk of distributed denial of service (DDoS) attacks [30].
- Authentication and access control mechanisms should be expanded to match the diversification of 6G networks, which are resource intensive and can create data vulnerabilities in related services. In particular, the centralized way of access control creates serious problems in the design of future networks.
- Auditing will be another challenging aspect of security due to the requirements for evaluating a huge amount of big data (eg: managing network segments between multiple nodes).
- With the advent of AI in 6G, AI/ML-based security attacks can also occur in 6G networks. On the contrary, AI can be used as a tool to detect, predict and mitigate security attacks. Thus, the deployment of proactive security mechanisms and the detection of zero-knowledge attacks (ZKP) will be vital security issues in 6G [28].

Blockchain is becoming an intriguing solution for the security, accountability, surveillance and management of mobile networks. Figure 2 shows the possible security attacks that could occur in next generation mobile networks and how blockchain can mitigate them. With the use of blockchain technologies, 6G networks can withstand eavesdropping threats due to the properties of immutability, transparency, non-repudiation and distributed access.

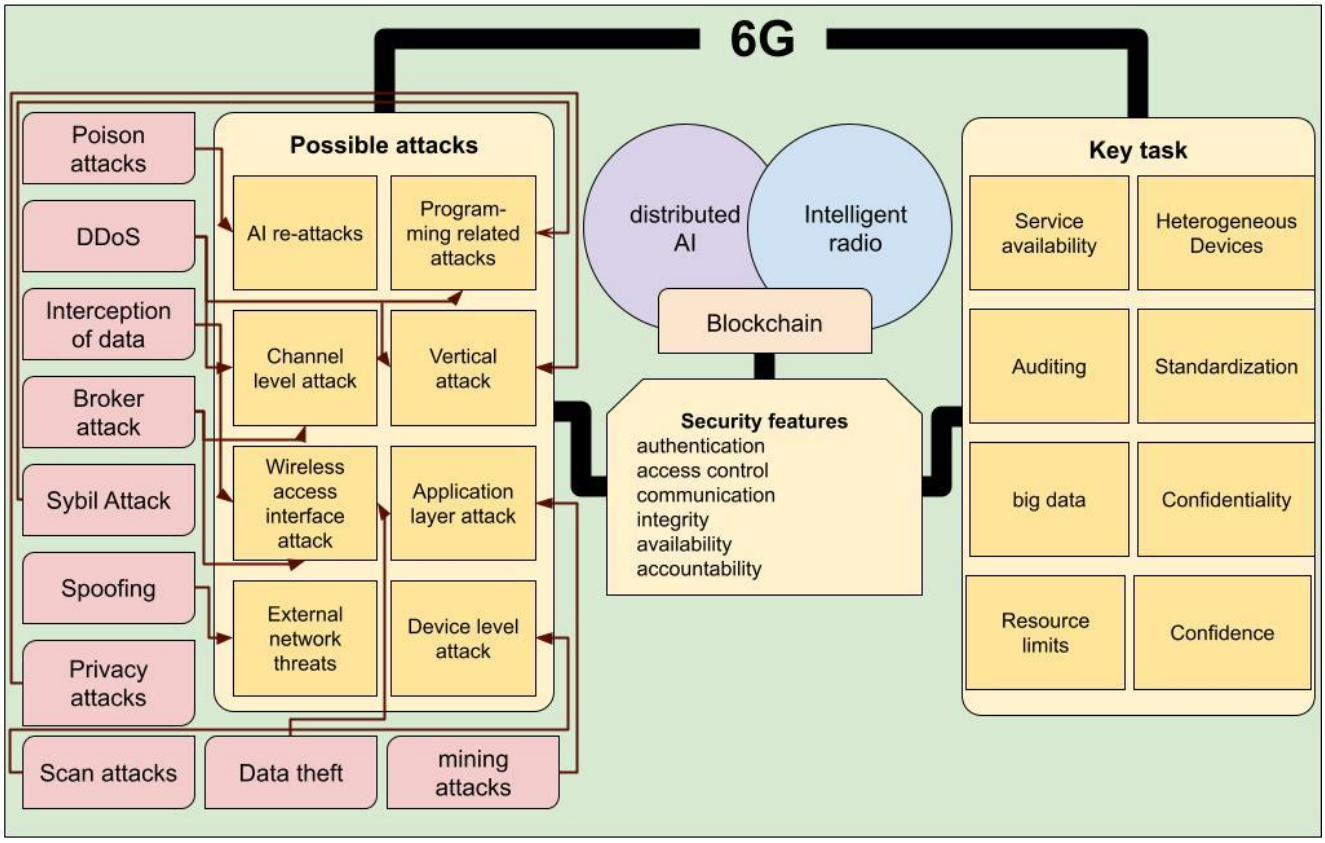

Figure 2 - Possible attacks and security issues in 6G networks, as well as security features that may accompany communications

Blockchain creates a decentralized means of trust between untrusted users, as it can withstand Sibyl attacks [9,35,66]. Note that in a Sybil attack, the attacker creates a large number of fake identities in a peer-to-peer based system and gains more exposure due to the scale of the fake identities. However, through the use of efficient consensus protocols such as P-BFT, the blockchain ensures that only the real users are in control of the system. [52,64]. This approach also reduces the chance of data modification and link layer attacks, since only participating nodes can view or add new transactions. By deploying well-designed smart contracts on top of the blockchain, key security aspects such as authentication, access control, and accountability can be addressed.

In addition, as noted in [47], the use of the blockchain allows AI/ML models to be securely trained based on immutable data stored in the blockchain, usually in a heterogeneous network scenario, to build trust by transparently storing related data. For AI/ML based systems, it is important to have trusted execution to ensure security in the decision making process. Collaborative optimization of ML models can occur by securely

exchanging model parameters, without requiring trust between participants. Smart contracts can be used as a distributed mechanism for security in AI/ML functions [71]. For example, in industrial IoT systems with a trained deep learning model, the developed smart contract can be used to detect network anomalies and malicious traffic. This enables network traffic analysis in a secure distributed environment where AI/ML models can be trained securely based on immutable data stored on the blockchain.

Security will be paramount to building trust between potential hosts. Thus, in Figure 2 shows possible attacks and security issues in 6G networks, as well as the security features offered by the blockchain. Two key components of the 6G network, namely distributed AI and smart radio, are identified as being most closely related to the applicability of blockchain for security.

Attacks at different levels in next-generation 6G mobile networks, including the edge layer, application layer, device layer, wireless radio interfaces, external network, and various verticals such as autonomous cars, industrial control, etc. Each layer has its own unique vulnerabilities and risks for security, which can be used by attackers to attack the system. Therefore, it is important to provide full protection at all levels and make 6G mobile networks more resilient to possible attacks.

Blockchain can be used to provide authentication, authorization and key management in communication networks [53,67,69]. It is a common communication channel that can establish trust between stakeholders in the 6G network and facilitate their cooperation on a common platform, even when unexpected network failures occur. [27].

Based on the above works, it can be concluded that blockchain can be used as a security technology in 6G networks to solve security functions such as authentication, access control, communication integrity, availability and accountability. In addition, blockchain-based solutions can be formed to protect 6G networks from multiple attacks, as shown in Fig. 2. However, the inclusion of such blockchain-based security solutions in 6G networks will continue to be problematic due to the increase in device heterogeneity, the convergence of heterogeneous networks towards 3D networks, the growth of big data, and the additional computing resources needed for blockchain.

Difficulties in maintaining privacy while collecting, combining and processing data are possible in many 6G applications, such as healthcare, environmental protection, smart city. Hence, privacy will be an important requirement for 6G networks. In addition, it is necessary to ensure that such networks comply with GOST R 59407-2021 privacy laws.

1. Key issues:
   - Unauthorized access to virtualized network resources can pose a high risk to the privacy of connected users, so secure access control is one of the key concerns when maintaining privacy in rich datasets in 6G networks.
   - Emergence of large-scale sensor systems with cameras and sensors continuously collecting data from the environment, which can lead to data breaches. For example, crowd surveillance drones that collect information about large gatherings can pose a threat to a person's privacy.
   - Due to the huge amount of data in centralized locations such as data aggregation servers, data providers and service providers, there may be a risk of confidential information leakage [61]
   - Since AI is a key technology enabling 6G, when collecting individual personal data for training models, the risk of data disclosure can lead to privacy breaches.
2. Role of blockchain.

Privacy is another technical aspect that can be implemented using blockchain in 6G networks [24]. The presence of a common communication channel in the form of a blockchain can allow the identification of network users by pseudonyms instead of direct identification of a person [34]. Privacy preservation performance is measured by user and resource level in relation to time complexity and memory consumption. This can provide greater privacy and transparency during data collection and air traffic control, while allowing faster data downloads with higher bandwidth and lower inter-application latency.

In 6G, the concept of hyperintelligence provides for intelligent and autonomous networks that will facilitate the performance of key operational functions [24]. In

addition, network services and applications will be highly flexible, customizable and adaptive to provide the best user experience. This cognitive network at its core aims to integrate AI as a native element into future networks.

The main challenges of ubiquitous use of intelligence in 6G networks can be summarized as follows:

- Complexity and optimization for AI/ML. The diversity and complexity of 6G networks will lead to challenges in terms of human control and system optimization tasks. Autonomous and self-managed control and optimization schemes are crucial. However, the complexity of such intelligent circuits and how to optimize them are not easy issues. It is necessary to manage the overall complexity and optimize the implementation of AI / ML.
- The distributed nature of intelligence in 6G requires the integrated work of disparate parties for an intelligent infrastructure. This is also due to the multi-connected and multi-domain structure of the software and service architecture, inherited from 5G networks and expanded further [16]. However, the reliability of AI/ML agents is a non-trivial attribute and may degrade intelligence performance in 6G.
- Secure data exchange for comprehensive analytics. The centralized storage and management function in 6G AI applications may be subject to data security threats (such as theft or tampering due to server compromise), resulting in a breach of privacy and disruption of the network. This is an important task for large-scale and multi-user network systems such as 6G.
- Large-Scale AI/ML Applications: Omnipresent intelligence for different use cases and applications will lead to a more decentralized architecture for storing and sharing data in 6G. This drive is also linked to cooperative and federated services and infrastructure flexibly linked together for intelligent immersion environments and objects. An important issue is how to develop and integrate technologies that provide this scalability for ubiquitous intelligence.

Since the ubiquitous use of AI/ML will be implemented in a distributed and large-scale 6G system for various technical aspects including network management, distributed AI/ML methods are expected to provide rapid control and analytics of the extremely large amount of generated data in the 6G network.

Blockchain can provide important features for ubiquitous intelligence in 6G networks. One such aspect is AI management and data sharing. Efficient, reliable, and secure management and sharing of learning outcomes is hampered by heterogeneity and mistrust between nodes. Each edge node has the same number of neural cells and the same number of layers in their neural network. In real scenarios, this is not always possible. Another important role of the blockchain is to support contextual awareness and autonomy. 6G hyperintelligent networks must be context-aware for network management as well as for providing intelligent digital services. This is possible with superscale sensing features that will support such context-aware operation. In addition, they must be self-managed and autonomous in order to meet performance KPIs.

Blockchain will enable reliable execution through smart contracts and data sharing to understand the context and therefore work autonomously. However, the complexity of blockchain-based solutions for this purpose can lead to an encumbrance.

Superscale sensing is another important opportunity, as the integration of advanced sensor technologies with mobile networks, combined with low power communication capabilities, will lead to ubiquitous smart sensing and localized services [8]. Blockchain will also serve as a decentralized and trustworthy data layer for super-scale sensing and analysis of big data in 6G. This is the most important useful feature of the blockchain, since sensors are the basis of intelligent services and networks. A security architecture with the ability to detect and track objects across the blockchain by deploying artificial intelligence algorithms on edge servers, while the 5G network provides low latency and high communication reliability to serve these time-sensitive applications. For 6G networks, the expected services require very strict and stable QoE/QoS performance. This is only possible thanks to a very efficient and intelligent resource management structure in network access.

However, resource constraints and efficiency requirements make this type of operation difficult. Large-scale smart surface control is a particular control challenge in 6G. The use of programmable intelligent surfaces to cover artificial structures allows for intelligent control of the transmission medium in wireless systems [39]. Such smart surfaces will play an important role in 6G networks and contribute to the deterioration of the radio signal quality. In a blockchain-based 6G network, blockchain provides a key opportunity to facilitate federated control across disparate smart surfaces. Moreover, intelligent surface management functions from different administrative domains (for example, different network operators) can exchange synchronization and reconfiguration information using the blockchain.

Blockchain can serve as a means of implementing pervasive smart grids due to secure data exchange, immutability and decentralized architecture. The decentralized paradigm also provides a more trusted intelligence stratum for the entire network, from smart devices to cloud services. It can facilitate very large-scale federated AI/ML applications for intelligent and autonomous digital connected services. Thus, in the future of 6G, a decentralized and secure method of data exchange without the so-called trusted third party or intermediary, that is, through the blockchain, is promising. However, the overhead, scalability, and interoperability issues of the blockchain are especially evident on the way to achieving the ubiquitous smart goal of 6G. For example, energy consumption due to blockchain features such as consensus and cryptographic operations is an issue that limits the usefulness of blockchains.

This phenomenon demonstrates the critical need to analyze the energy and environmental impacts of the widespread adoption of blockchain in 6G, and therefore research is needed to minimize it. In addition, the scalability aspect needs to be improved to realize super-scale intelligent solutions. New attacks are also emerging that attempt to reduce the reliability and accuracy of AI/ML solutions using blockchains, and the research community needs to listen more closely in the future. [61].

Because 6G networks will be an ultra-dense network of networks and will use a higher frequency spectrum (GHz/THz band), 2D cell boundaries will shrink to a few meters. This is a drastic reduction in the number of base stations to cover and the waiting

area for handovers for mobile users will increase exponentially. In addition, due to non-optimized handover mechanisms for various base stations with overlap and low coverage (such as ABS Picocell, Femtocell and UAV), the probability of handover failure, probability effect and radio link failure will increase significantly [31]. Subsequently, this will negatively affect both throughput and latency. On the other hand, the emergence of 3D networks will support 3D mobility, which will require new methods of mobility management, since handover will also occur in the vertical direction [62]. As 6G seeks to provide ultra-reliable, low latency communications, effective mobility management is essential.

Here are some of the challenges associated with mobility management in next generation networks:

- Distributed mobility management raises many security issues such as pseudodata attack, session hijacking, DoS attack, man-in-the-middle attack, and malicious attacks. These issues make it difficult to use a DMM for mobility management.
- To ensure secure data transfer to protect the past session with key access, there is a threat of a compromised privilege node. In addition to forward secrecy, concurrent fulfillment of other requirements (e.g., mutual authentication, key agreement, traceability, and reliability) is a significant challenge [17].
- Using a centralized authentication server for mobility management introduces both performance and security issues.
- Roaming fraud, due to inefficient and latency-prone communication between the home network and the guest network, results in large economic losses. Such roaming fraud could become a major problem in future 6G networks due to network densification and the popularity of local/private networks.

The combination of blockchain and smart contracts can establish trust between mobile users and mobile networks in a distributed and anonymous manner. The various uses of blockchain for efficient mobility management in wireless networks are discussed as follows: distributed mobility management aims to overcome the problems (e.g. single

point of failure and sub-optimal routing) associated with traditional centralized mobility management. Distributed mobility management suffers from security issues such as session hijacking, DoS attack, and mobile anchor and access router attacks. These problems are mainly due to the fact that security is implemented centrally. While their scheme has shown promise, however, high storage redundancy due to multiple distributed ledgers is challenging.

Handover authentication solution for multicast communication scenario in networks where vehicle mobility needs to be managed: aggregate message authentication code, one-time password and blockchain, can reduce transmission congestion. The proposed protocol uses the Elliptic Curve Diffie-Hellman algorithm to manage session keys. The protocol is used in a blockchain with database management and the mobility access router stores key information about vehicle authentication. In addition to meeting security requirements such as session key confidentiality and immunity to retaliation and sybil attacks, the proposed scheme is effective in the context of handoff delay and computational and communication overhead. Similarly, the use of blockchain is possible for authentication during handover while protecting user privacy in SDN-based 5G networks. As 6G will use SDN and NVF to effectively manage networks, blockchain is expected to play an important role in mobility management.

Anonymous mutual key authentication has strong traceability and perfect forward secrecy using blockchain hash functions. The user prepares for the transfer phase. First, he chooses a pseudo-identifier, calculates various parameters to prove the legitimacy of his identity, and passes the authentication parameters to the access point. The access point checks the relevance of the information to prevent replay attacks and checks the legitimacy of the authentication parameters. The authors compared the proposed mechanism in terms of authentication message sizes and storage costs and concluded that both parameters are reduced in the proposed scheme. Using Burrow-Abadi-Needham (BAN) logic to test the security properties offered by anonymous mutual authentication. In addition, model checking tools are used to formally verify the proposed authentication mechanism.

With the same intention of providing forward secrecy and satisfying lightness requirements for resource-constrained mobile devices, as well as protecting user privacy and tamper resistance, the proposed scheme uses a blockchain-secured hash function and a user certificate to perform in-transit authentication. When a user registers with an authentication server, the latter generates a unique certificate and uploads it to the blockchain. When a user moves to a new access point, the user is authenticated by cross-validation with the original certificate available on the blockchain. This scheme provides conditional privacy based on the pseudonymous function of the blockchain, immunity to various attacks (such as replay, man-in-the-middle attacks, and passive listening), and perfect forward secrecy of the session key.

Zero Touch Network and Service Management (SMS): The SMS initiative represents a new paradigm for achieving automation of E2E networks and services. The ultimate goal of an SMS is to implement 100% autonomous networks capable of self-configuration, self-optimization, self-monitoring, self-healing and self-scaling without any human intervention. Thus, SMS leads to flexible and fast service delivery, ensuring economic sustainability in diversified networks [36, 44]. The SMS architecture is formed using intent-based interfaces, feedback operations, and AI/ML techniques to provide full automation of management operations. The complex 6G ecosystem is expected to consist of multiple operators and multiple service providers (in a virtual network). Thus, SMS will play an important role for 6G networks in achieving E2E automation of network resources and services that relate to different parts of the mobile network, spanning multiple operators and service providers.

Blockchain and AI governance can support cross-domain security and trust management mechanisms in the proposed SMS architecture, helping to automate the service life cycle in multi-user and multi-party environments. The implementation of distributed security in the blockchain is carried out by the trust of smart contracts between trusted and untrusted parties. This approach can be further combined with distributed artificial intelligence to orchestrate the cognitive network, resulting in the automation of network processes. Thus, blockchain can become a promising technology to increase trust in zero-touch network management. However, it is still difficult to maintain a proper

balance between data privacy and trust in a blockchain-based automated system to work with guaranteed information security.

1.3. Grid Resource Sharing

A way to solve complex problems that require a large number of workers (computing resources, as well as storage and data transfer resources) working in parallel on different aspects of the task. The general term "grid systems" covers the various approaches and specific technologies used to solve such problems.

The basis of Grid systems is to provide stable operation of a set of services based on widely accepted open standards and software, called "middleware" or "middleware" in English. This provides reliable and unified access to geographically distributed information and computing resources, including individual computers, clusters, supercomputer centers and information storages [72].

The creation of grid systems has become possible due to significant advances in several areas:

- Increasing the performance of mass-produced microprocessors. Modern personal computers have achieved performance comparable to the supercomputers that were relevant ten years ago.
- Emergence of fast network connections. Today, the main data transmission lines have a throughput of several gigabits per second.
- Globalization of information exchange via the Internet and the Web.
- Development of methods of metacomputing, a scientific discipline that deals with the organization of massive and distributed computing processes.

Grid infrastructure is based on the provision of resources for general use and the use of publicly available resources. An important concept in this concept is the virtual organization (VO).

The idea of the grid arose in response to the need for access to large information and computing resources that are dynamically allocated to solve complex problems in various fields such as science, industry, administration and commerce. The creation of a

grid environment includes the distribution of computing resources to different geographically located sites on which special software is installed. This software is responsible for distributing tasks between sites, receiving and returning results to the user, managing user access rights to resources, and monitoring resources.

The public resources of the grid system include computing nodes, data storage and transmission nodes, data, and application software. Computing resources provide the user with processing power and can be represented as clusters or individual workstations. Any computing system, with the appropriate software, can be a potential computing resource of a grid system [73].

Storage resources also use software that implements a unified management and data transfer interface. The physical architecture of storage resources is not fundamental to a grid system, whether it be a workstation hard drive or hundreds of terabytes of mass storage. The main criterion for storage resources is their volume, which is currently measured in terabytes (TB).

Information resources and catalogs are a special kind of data storage resources used to store metadata and information about other resources of the Grid system. Information resources allow you to store large amounts of information about the state of the grid system in a structured manner and efficiently perform resource search tasks. The network resource plays the role of a link between the distributed resources of the grid system.

The concepts of "service state" and "stateless service" are key concepts in the theory of loosely coupled services. In the context of loose coupling, the benefits come from the fact that the client can use any service capable of fulfilling its request. If the client is limited to a single service, the benefits of loose coupling are reduced. For example, for simple requests, such as using a calculator or getting information about a stock price, the client requests the information and receives it, after which the transaction is considered completed, and the client does not have a particular need to contact the same service again. In this case, the connection between the client and the service is weak.

However, for more complex requests that require multiple steps, the service can store information (state) about the previous steps in its local memory. This allows the service to use the stored information on subsequent requests from the client. In this case,

the service has a stateful service, and the client must access the same service in the next step. This can lead to delays or denial of service, especially if multiple clients are using the same service, or if the service crashes between steps.

A more correct approach to developing services is based on "stateless services". This is especially important for multi-step query processing. At each intermediate step, the service must provide the client with sufficient state information so that another service with the appropriate properties can identify and continue serving. The client must pass state information to any service in the next step. The selected service must be able to accept and process state information provided by the client, whether it handled the request in the previous step or another service handled it.

In the context of Figure 3, the client makes a request that requires three processing steps and may involve multiple services. Each service may be able to process any part or all of the request. It is important that state information be passed between the services and the client to ensure that the request is continuously processed.

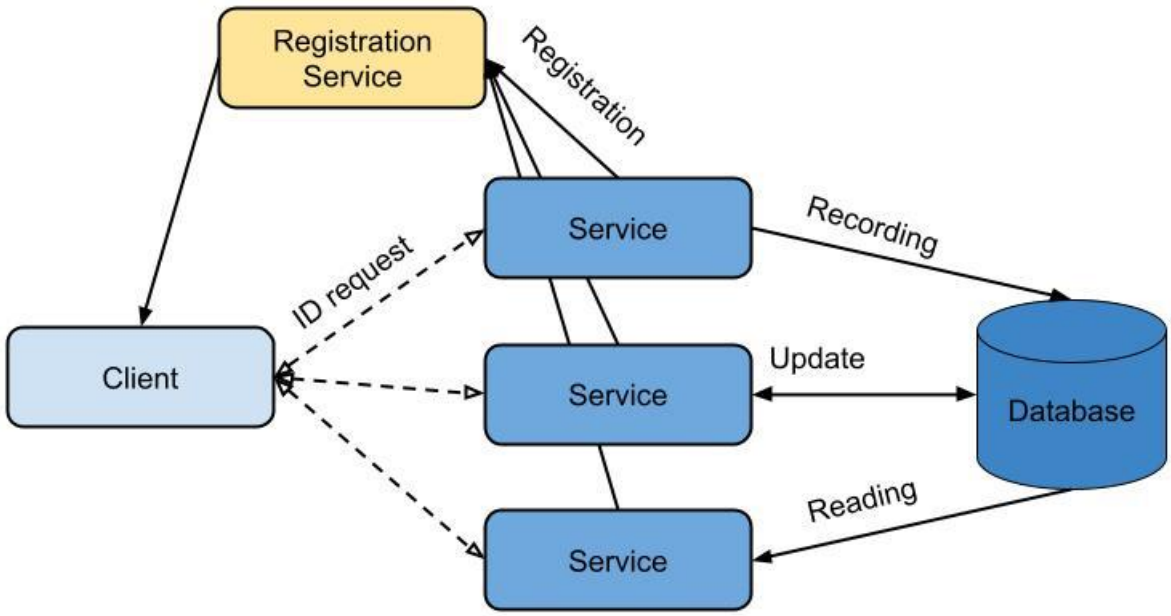

Figure 3 - Multi-step interaction between the client and services

In the approach described above, the service that handles the first step stores the processing details of the current request in the database and returns the information to the client along with the request ID. The client can then ask for confirmation from the user before passing this identifier to another service, which uses it to look up state information in the database and initiates the second processing step. This service updates the database

and returns additional information to the client. Finally, the client passes the transaction identifier to the third service along with the request to complete processing [72].

Most non-trivial applications require access to state information, and the question is not whether states should exist, but where they should be stored. In the approach described above, the processing state of a request is separated from the service performing the processing, providing a loose coupling between the services and the client. To reduce the amount of state information passed between the client and services, the essential details of request processing are stored in a database. All participating services should be able to access the database and get the information they need based on a client/request ID that is easily passed from the client to the services.

Note that the operation of grid systems is based on middleware that provides controlled access to resources. At the initial stage of development, grid systems were built on the basis of specially developed public components or closed (proprietary) technologies. Although various public and commercial solutions have shown success in their respective fields of application, each with its own advantages and limitations, they have had limited potential as the basis for the next generation of grid systems, which must be scalable and interoperable to meet the needs of large-scale scientific and industrial projects.

Figure 4 shows a diagram of a simple service-oriented grid where services are used to both virtualize resources and provide other grid functionality.

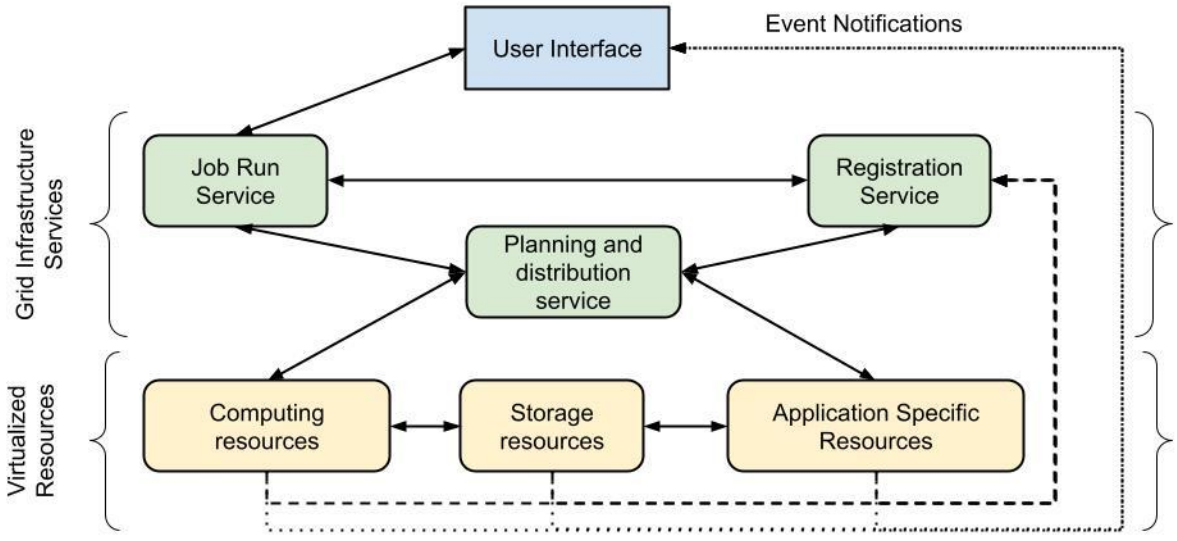

Figure 4 - Simplified Service-Oriented Grid Diagram

The diagram shows a single console for both running jobs in the grid environment and managing grid resources. The user interface (console) software calls the registration service to obtain information about existing grid resources. The user then contacts the services "representing" (virtualizing) each resource through the console to request periodic retrieval of resource performance data and notification of significant changes in their state (for example, if the resource becomes unavailable or heavily loaded). The user sends a request to run a job to the startup service, which forwards the request to the job distribution service (often referred to as the "scheduler"). The distribution service contacts the service representing the application and requests information about the resource requirements to complete the job. The distribution service then queries the registration service for information about all eligible resources in the grid and contacts them directly to make sure they are available. If suitable resources are available, the scheduler selects the best available set of resources and passes information about them to the application service with a request to start execution. Otherwise, the scheduler queues the job and executes it when the required resources become available. When the job ends, the application service reports the result to the scheduler, which notifies the job launch service. The job launch service, in turn, notifies the user. Note that this example is greatly simplified for clarity: the functioning of a real industrial-grade grid is much more complex than shown in the diagram. The main result of his work should be a high degree of automation and optimization of the use of resources within the grid environment.

Conclusions on the 1 chapter

The identified tasks of the distributed ledger require the following approaches to solutions:

1. On 6G networks, open source software may be available to implement various network functions at the core, transport, or access level;
2. The two-level hierarchical architecture of the blockchain can be considered as the main resource management system;
3. Using the concept of a grid as a response to emerging needs for large information and computing resources dynamically allocated for solving cumbersome tasks;
4. General principles of grid technology: add users and transfer work resources;
5. By efficiently allocating resources, grid technology significantly reduces the waiting time for access to them.

In the next chapter, we will look at data storage methods and possible security risks.

II-CHAPTER. ACCESS TO DATA DIFFERENTIATION BY CLASSIFIERS

2.1. Comprehensive personal data security models

All established management processes are based on balanced management systems, the level of which is measured by maturity models. Another direction of building complex personal data security systems is the creation of frameworks aimed at building a personal data management system in an organization.

Two different approaches prevail: vertical - a cross-cutting approach for organizing management processes, and horizontal - focusing on a particular technological ecosystem (for example, the Internet of Things (IoT), mobile communications, social networks).

ISO/IEC 29100:2011 (as well as its updated version ISO/IEC 29100:2020) raises a number of issues specific to the processing of personal data:

- defines general privacy terminology;
- defines the participants and their roles in the processing of personally identifiable information;
- describes privacy protection considerations based on organizational and technical aspects;
- contains references to known privacy practices.

In addition to the specified standard, ISO / IEC has released a number of documents describing approaches to the organization and technical architecture of personal data security issues [74]:

- ISO/IEC 27403. IoT security and privacy. Guidelines for IoT-domestics. The standard is still in draft status;
- ISO/IEC 27550. Privacy engineering for system life cycle processes. The standard describes an approach to privacy management through the life cycle of automated systems. Still in draft status;

- ISO/IEC 27556, User-centric framework for the handling of PII based on privacy preferences. Forms a personal data management system on the part of the user;
- ISO/IEC 27561, Privacy operationalization model and method for engineering (POMME). An operational model for managing the security of private data. Preparing for release;
- ISO/IEC 27570. Privacy guidelines for smart cities. Addresses personal data security issues for smart cities;
- ISO/IEC 29134, Guidelines for privacy impact assessment. Assesses the impact/damage resulting from violations of the confidentiality of personal data;
- ISO 31700. Privacy by design for consumer goods and services. Ensures the security of personal data when selling consumer goods and services.

Taken together, the ISO/IEC initiatives cover a wide range of issues related to the organization of work and the technical architecture for various systems. At the same time, many of the standards are still in development, and their multiplicity and inconsistencies make it difficult to apply them as a single set of guidelines.

The US National Institute of Standards and Technology (NIST) has also attempted to develop a comprehensive personal data security model [74]. Among the published documents and guidelines, the following should be noted:

- NIST.SP.800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Guidelines for protecting the privacy of personal data. Contains a wide range of measures, including organizational ones, for managing personal data, preparing them for depersonalization, documenting the procedures and actions of personnel, the process of publishing personal data, as well as actions as a result of leaks;
- NIST.SP.800-53. Security and Privacy Controls for Federal Information Systems and Organizations. Organization of control over the processing of personal data for government organizations;

- NIST.SP.800-37. Risk Management Framework for Information Systems and Organizations A System Life Cycle Approach for Security and Privacy. Describes the process of risk management in organizations, taking into account the security of personal data;
- NIST.IR.8053. De-Identification of Personal Information. Description of depersonalization methods, approaches to calculating risks and general organization of the depersonalization process;
- NIST PRIVACY FRAMEWORK 1.0 (PRAM): A tool for improving privacy through enterprise risk management. General system (framework) for managing security, personal data in organizations, taking into account organizational procedures, instructions for the risk prioritization system, a set of instructions and checklists for known personal data security issues.

The provided NIST documents make it possible to consider the problem of working with personal data in the context of the life cycle in a single organizational and technical field.

Security of private data through system building. Ensuring the security of personal data through a system of consistent implementation of technical and organizational procedures in the practice of the organization (Privacy By Design) is an important component of the overall approach to building reliable procedures for processing personal data, including data depersonalization [57].

Although there is no single organization behind this concept, most standards and technology companies recognize this approach as the most important approach to personal data security. [13, 41,46].

This concept is often opposed to the concept of “Privacy By Default”, however, in reality, these approaches complement each other. The idea of Privacy By Design is a top-level concept that declares the absence of a static solution to the issue of personal data security and proposes instead to use a dynamic approach in which there is a constant adjustment to external threats and challenges, and the system is built by the organization step by step.

The main principles of this approach:

- A proactive approach takes precedence over reactive actions. Organizations should provide preliminary data security management actions, not solve problems that have already arisen;
- privacy as default setting. This means that confidentiality should be built into every system that handles personal data. No action is required from an individual, in their absence, his data remains safe;
- confidentiality is built into the design and architecture of IT systems, as well as the organization's business practices;
- Ensuring full functionality while minimizing risks. The principle declares the existence of a possible compromise that benefits all parties;
- end-to-end security - protection of the full life cycle of data, including their transformation, transfer and destruction;
- visibility and transparency of the personal data processing policy;
- respect for user privacy - user-centric.

The analysis carried out allows us to generalize some of the provisions of existing practices and draw a number of significant conclusions:

- The need for data depersonalization is directly related to the data exchange processes that arise in modern integration processes. Models that focus on static data storage situations without regard to data sharing or publishing do not cover the entire lifecycle;
- The issue of depersonalization of personal data is closely related to the overall management of personal data security;
- Although this area of anonymization is still in the development stage (new works appear regularly), there is an established consensus regarding the management of anonymization based on risks: all current models are based on risk assessment in the process of anonymization;
- The contextual risks associated with organizational and technical measures within organizations are important for conducting depersonalization processes.

Ignoring such measures leads to greater availability of third-party information resources, which significantly increases the likelihood of attacks using additional resources and more stringent requirements for anonymized data.;

- The current domestic system of working with personal data in terms of depersonalization does not take into account the changed information landscape, is focused on a system of prohibitions and is overly state-centered. Ignoring the use of big data by businesses and non-profit organizations only increases the risk of data breaches and, ultimately, leads to a technological backlog.

2.2. General Approach to Risk Calculation

Data is depersonalized for the purpose of efficiently and securely exchanging information between several identified parties or publishing it to an indefinite circle of persons in order to ensure transparency and trust in decisions made, research of data sets by independent data analysts, and for a number of other purposes (figure 5). At the same time, data security (in terms of confidentiality of personal data) and information efficiency (usefulness) of the data obtained as a result of depersonalization are in conflict.

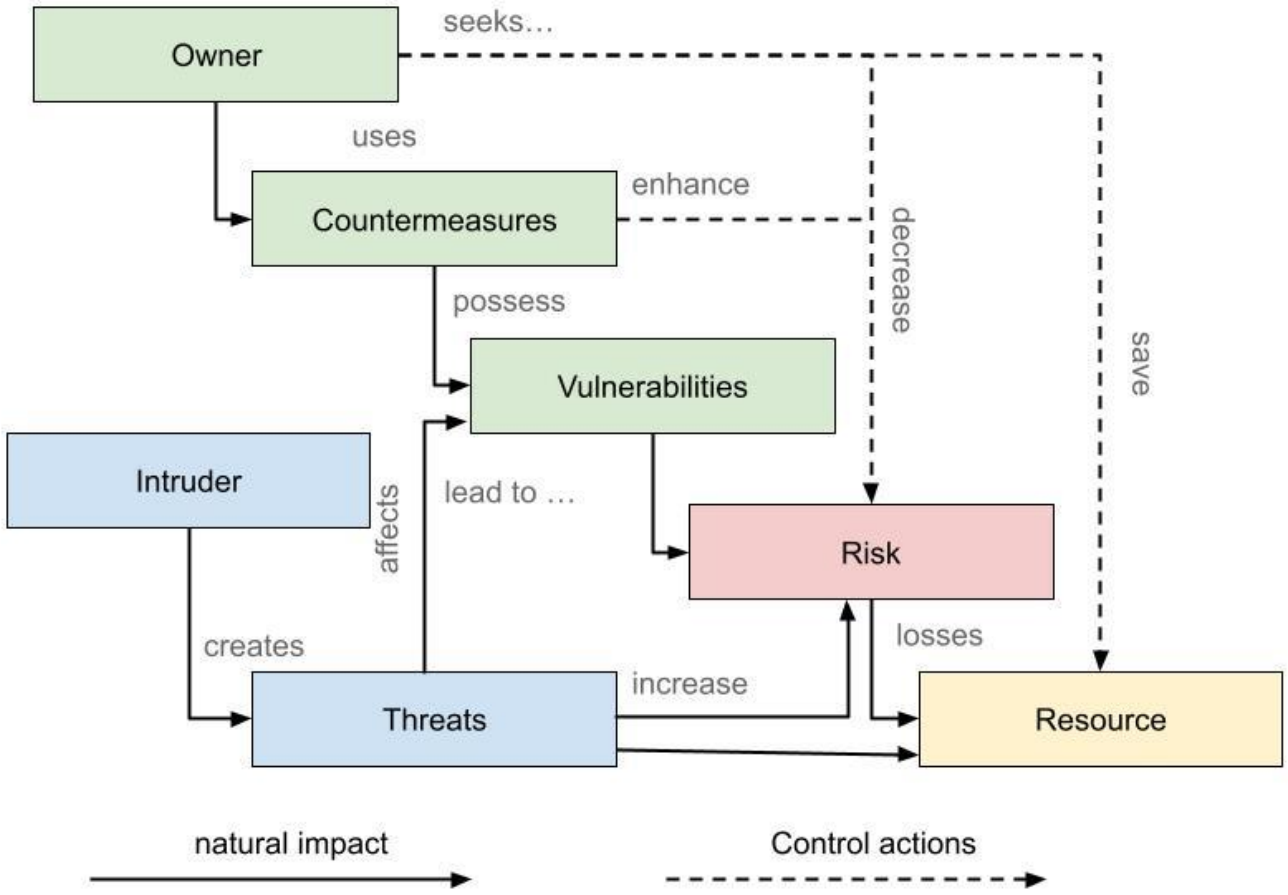

Figure 5 - Risk assessment process adopted in information security

The security of personal data differs from the classical concept of information security [6, 45], the focus of which is unauthorized activity (which actually leads to loss of confidentiality, integrity or availability). When processing personal data for the purpose of depersonalization, we are talking about planned activities, which, however, may create negative consequences for the private life of individuals. Threats to the confidentiality of personal data arise both as a result of authorized data processing and as a result of unauthorized access:

- as a result of unauthorized access to data - emotional suffering of individuals, economic losses from identity theft, and physical or psychological harm due to harassment;
- as a result of the decline in the quality of information processing within the framework of planned processing - a decrease in the quality of services provided, failures in the operation of systems, incorrectly made or delayed decisions, which may also affect the health of individuals or entire populations.

The key provisions of the proposed risk model for the depersonalization procedure take into account confidentiality issues arising from both unauthorized access and scheduled data processing:

- the risk of publishing anonymized data is a measure of the extent to which the subject is threatened by potential circumstances or events and is a function of the probability ***P*** of the occurrence of such events and the adverse consequences of a violation of the confidentiality of personal data - damage (***I***, Impact) [77]:

$$R = P * I \quad (1)$$

- building completely anonymized datasets that retain useful information is an idealized hypothesis that is not feasible in practical solutions. The proposed approach is instead based on the probability of conducting re-identification attacks and setting an acceptable risk.;
- the process by which risks are identified is known as risk assessment;
- risk assessment requires a risk model that identifies risk factors and their relationships:
  - risk factorization;
  - data usage scenarios;
  - quantitative risk thresholds;
  - the required level of usefulness of the received depersonalized data;
  - the procedure for building a risk model for a specific depersonalization procedure, including the possibility of risk reassessment when using various depersonalization methods.

Damage may include:

- Compliance costs related to privacy concerns for individuals;
- direct costs of fines and legal fees;
- loss of business, refusal to use products or services;
- reputational costs leading to loss of trust from users;
- decreased performance or inability to achieve the mission of the organization.

Thus, the risk assessment proposed in this model contains damage as a parameter, in the future, risk assessment is understood as an assessment of the probability of occurrence of events $P \in [0;1]$;

general factorization formula re-identification risk probability:

$$P_{re-id} = P_{context} * P_{data} \quad (2)$$

Here: $P_{re\text{-}id}$ – general risk, $P_{context}$ - contextual risk defined by a set of organizational and technical risks, P_{data} – data risks.

- data risks are associated with a specific data set and should be calculated taking into account the characteristics of the data set - in particular, the allocation of direct identifiers and quasi-identifiers. Direct identifiers and their substitution are the smallest problem in solving the reidentification risk problem. Quasi-identifiers, their correct identification and application of depersonalization methods, taking into account their structure and type, is a much more difficult task, taking into account the probabilistic approach to risk assessment.

Carrying out depersonalization taking into account the risk model requires a balance between the usefulness of the data obtained as a result of depersonalization and the acceptable level of risk. Risk thresholds are set according to use cases. A general idea of the orders of risk can be drawn from the data in the table 1 [5, 18]:

Table 1 - General risk threshold values

Impact on PD privacy	Acceptable vice of risk	Equivalence class size for aggregated data
Low	0.1	10
Average	0.075	15
High	0.05	20

The acceptable level of risk is determined taking into account the following parameters:

- likelihood of a threat to confidential data ($P_{context} * P_{data}$) ;

- potential damage I (assessed in terms of “low”, “medium”, “high” or through points) [14, 33]. The damage is also determined taking into account the volume of published data and the sensitivity of the attributes;
- publishing scenario - affects the recommended equivalence classes and therefore data risk;
- the business model of the organization publishing the data. This parameter describes risk tolerance taking into account the characteristics of the business (industry), the value of assets (data containing private characteristics), management preferences.

Within the framework of the model under consideration, a number of quantified metrics are used:

- *Risk level.* It is the product of the damage and the probability of the risk of re-identification. Probability, in turn, is factorized by the probability of contextual risks and the probability of data. The latter is a metric tied to the data - their structure and content - and reflects the ability of an external agent ("intruder") to restore the connection between the subject of personal data and its characteristics (attributes). The assessment of the likelihood of re-identification risk for data depends on whether the attacker knows any data about individuals included in the set under consideration;
- *Data utility level.* In the process of depersonalization, there is inevitably a loss of information associated with the use of depersonalization methods. Estimates for large sets are conveniently based on internal metrics tied to the properties of the set. It is also an assessment of data quality;
- *Reversibility level.* Reversibility allows you to maintain a connection between the original and anonymized data set, can be evaluated in the range [0-1]. Reversibility can be assessed based on two options:
  - inclusion in the impersonal set of prepared unambiguous data identifiers (these can be special identifiers stored during the conversion process, hash values of real identifiers or specially formulated aliases). In this case:

$$K_{revers} = 1 \quad (3)$$

– the absence in the set of attributes that uniquely link the original and impersonal set. In this case, the reversibility coefficient is equal to the probability of re-identification for the data set:

$$K_{revers} = P_{data} \quad (4)$$

- *Variability of the depersonalization method* (*variability*)**.** A number of anonymization methods allow the use of various parameters that affect the way the anonymized set is formed:
  – in the case of deterministic depersonalization methods with a finite set of values accepted by the parameters *N* (generalization, suppression) coefficient is determined by: (обобщение, подавление) коэффициент определяется:

$$K_{var} = \frac{N}{2} \quad (5)$$

  – in the case of continuous changes in parameters, the coefficient is determined by the dispersion of values as a result of the transformation:

$$K_{var} = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad (6)$$

- *Variability (flexibility).* Evaluates the possibility of making additions (distortions) to the anonymized data array. For some reason, an already anonymized data set may be subject to additional changes. Because depersonalization methods target individual attributes, applying repeated changes to previously modified columns should be supported by criteria that ensure that the underlying distributions for that attribute remain unchanged:
  – to control the correctness of the changes made, it is recommended to use null hypothesis tests (similarity hypothesis, equality of variances):

$$H_0 = \sigma_1^2 - \sigma_2^2 \quad (7)$$

– This approach recommends using *F*-test. Let a sample *X* of *m* random values of the initially impersonal set be compared with a sample *Y* of a set with *n* random values that has undergone a repeated value. Fisher function:

$$F = \frac{\sigma_X^2}{\sigma_Y^2} = \frac{\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (8)$$

– with such a transformation $F \approx 1$. There are several limitations to this test: it is assumed that the values of the attributes in both sets are normally distributed, the sample is limited.

- *Durability impersonal set to attacks*. Determined by the probability of success of re-identification attacks;
- *Compatibility various impersonal sets (when matching attributes).* Two sets can be compared provided that the equivalence classes for the compared attributes are the same, in this case it is recommended to compare using the *l*-diversity metrice;
- *The parametric volume* is determined by the required amount of additional (service) information for the method. In the simplest cases, it represents the number of parameters for setting up anonymization methods, as well as auxiliary data for logging experiments, including the storage of intermediate versions of anonymized sets.

Conclusions on the 2 chapter

Carrying out depersonalization taking into account the risk model requires a balance between the usefulness of the data obtained as a result of depersonalization and the acceptable level of risk. Risk thresholds are set according to use cases.

Within the framework of the model under consideration, a number of quantified metrics are used:

- *Risk level.* It is the product of the damage and the probability of the risk of re-identification. Probability, in turn, is factorized by the probability of contextual risks and the probability of data. The latter is a metric tied to the data - their structure and content - and reflects the ability of an external agent ("intruder") to restore the connection between the subject of personal data and its characteristics (attributes). The assessment of the likelihood of re-identification risk for data depends on whether the attacker knows any data about individuals included in the set under consideration.;
- *Data utility level.* In the process of depersonalization, there is inevitably a loss of information associated with the use of depersonalization methods. Estimates for large sets are conveniently based on internal metrics tied to the properties of the set. It is also an assessment of data quality;
- *Reversibility level.* Reversibility allows you to maintain the relationship between the original and anonymized dataset.

With that said, in the next chapter, we will look at layered access to transaction processing.

III-CHAPTER. A LAYERED APPROACH TO SECURITY IN A DISTRIBUTED LEDGER

3.1. Stages of forming transaction processing on vertical scalability

Storing and using data at the hardware level requires a solution for distributing data across communication nodes with vertical scalability. Vertical scalability is actually the ability to increase the efficiency of currently used hardware or software by adding additional resources to it, by adding fast processors to the server to increase its speed [54]. This means adding additional hardware resources to an existing machine by using more processors, memory, etc.

In fact, virtual server deployment is a fundamental prerequisite for vertical deployment. This technology allows us to mutually use the same functions as physical servers. In fact, when a user logs into the server's console (figure 6), it's virtually impossible to tell the difference between a physical server and a virtual one until you look at the drivers [19, 20].

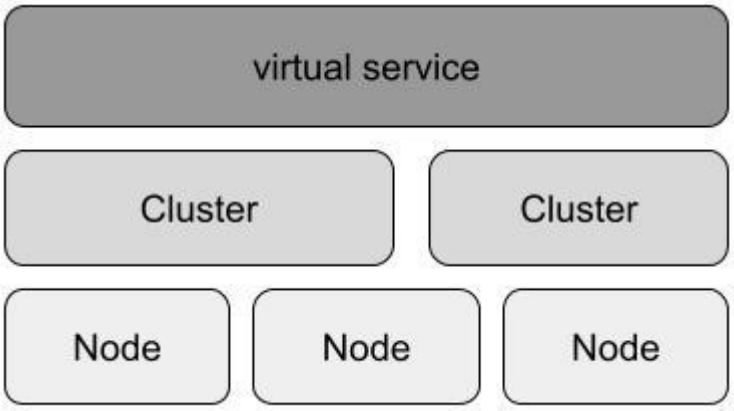

Figure 6 - Virtual Server Architecture

The difference between physical servers and virtual servers is that virtual servers are not installed on physical hardware. They are rather installed on a thing called a hypervisor. The hypervisor allows you to run more than one virtual server on the same physical hardware. Using virtualization makes it easier to dynamically scale an

application as clients interact with a virtual server on the application delivery controller, and then it interacts with virtual servers located on physical servers inside the data center [2,3].

In a grid system, ensuring interoperability between different platforms, languages and software environments is a key condition for efficient operation. For this, common protocols are used that regulate the interaction between elements of a distributed system and the structure of the transmitted information [72,73].

The general structure of the global grid is described as a protocol stack. In this model, each layer of the stack is dedicated to a specific set of tasks and provides services to the higher layers. The upper levels of the stack are closer to the user and work with abstract objects, while the lower levels are closely related to the physical implementation of the grid resources [1]. This protocol stack structure is similar to the Open Systems Interconnection Reference Model (OSI) for network communications and network protocol development.

The grid protocol stack includes the following layers (Figure 7):

1. *Hardware layer (Fabric Layer):* contains the protocols that are used by the corresponding services to work with resources directly.
2. *Connectivity Layer:* includes protocols that provide communication between components of the base layer, as well as authentication protocols.
3. *Resource Layer:* is the core of a multi-layer system, where protocols interact with resources through a unified interface, regardless of the architectural features of a particular resource.
4. *Collective Layer:* responsible for coordinating the use of available resources.
5. *Application Layer:* Describes user applications running in a virtual organization environment. Applications use protocols defined at lower levels of the stack.

Such a structure of protocols makes it possible to effectively organize interaction and ensure compatibility between various components of the grid system, using common principles and standards.

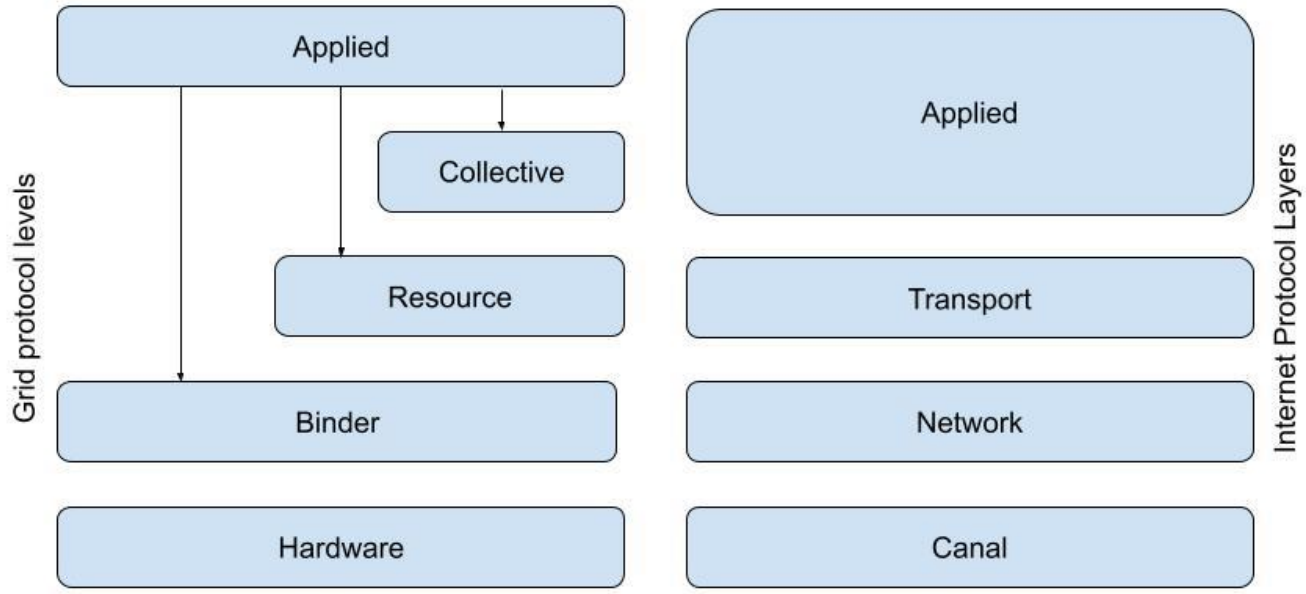

Figure 7 - Protocol stacks of the grid system and network model

The number of servers inside the data center can be changed without any security, acceleration and availability issues of the respective application as these features are centered on the application delivery controller which can be automated to add/remove servers inside the data center without any issues. As each virtual server is added, its software accesses the underlying hardware, reducing downtime [10].

Most optimization algorithms require synchronization of local peer-to-peer access to WAN information. This is a separate problem, known as the aggregation problem [64], and refers to a set of functions that provide access to such components of a distributed system as network size, load average and uptime, and so on.

The problem solving was installed on the DGT platform, which considers the problem of fault tolerance with a multi-level data processing approach. This approach is based on the deployment of a virtual network for solving problems or when developing applications. Servers or nodes can also be located on different physical networks (figure 8). Node clusters are part of a larger network division - segments, which can be of two types: public and private. In a separate network, only one public segment is possible, joining nodes can freely interact with other segment nodes. The network can have several private segments, the main difference of which from the public segment is the controlled topology.

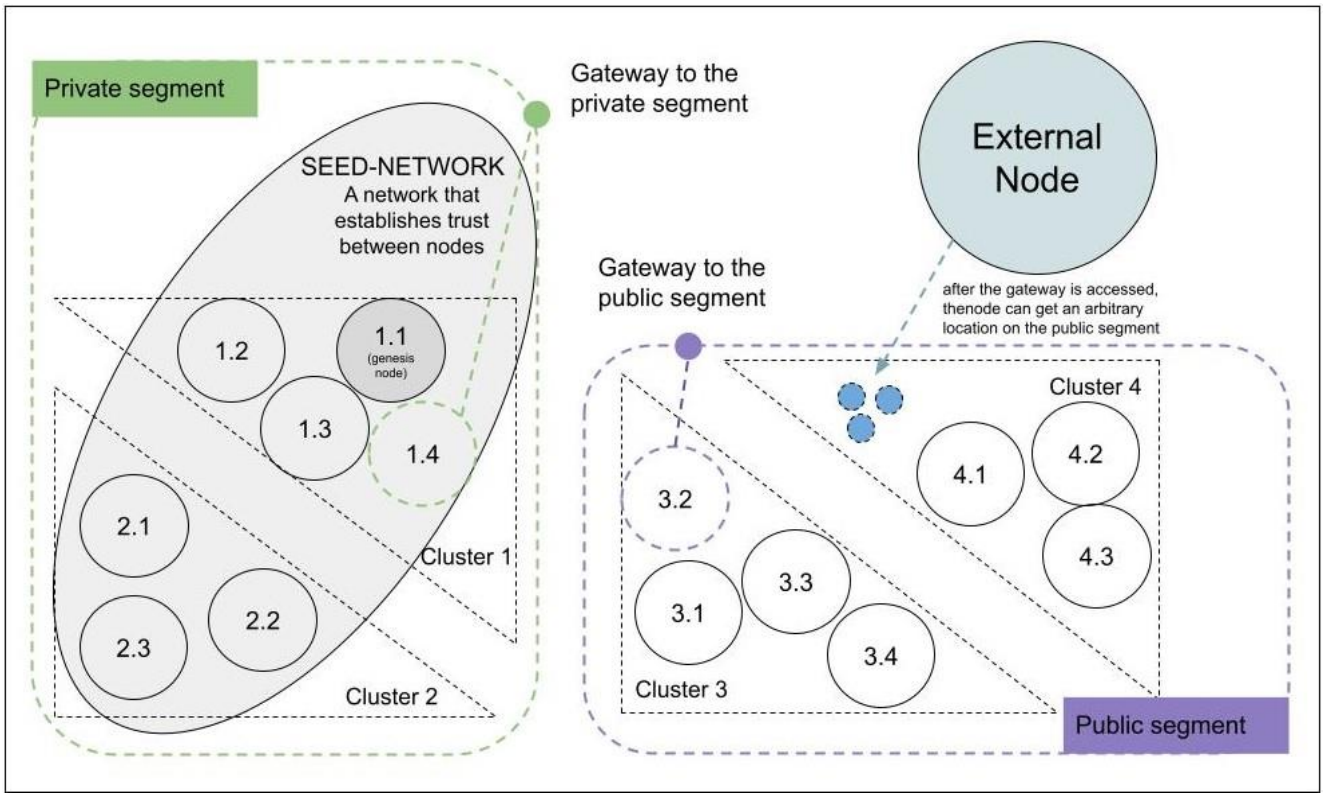

Figure 8 -DGT Network Topology and Node Attaching

The initial implementation of the network, also called the “static core”, which is a group of nodes / clusters that form special trust relationships (public keys of such nodes are known in advance and are registered at the time of kernel deployment). Joining other nodes requires the processing of node certificates for private shards and/or dynamic joining in the case of public shards.

Additionally, you can consider data segmentation on the DGT platform, taking into account the cluster topology. A cluster is a group of nodes that carry out the primary round of voting. Clusters can form complex structures.

The size of the cluster, the attachment of nodes to it, is determined by the topology processor, which is a separate family of transactions. Within each cluster, a variable leader is determined, which changes after several rounds of voting. If at the request of the leader he does not answer within a certain time, a procedure for selecting a new leader takes place. Voting is initially carried out in the cluster, then an arbitrator outside the cluster is selected according to a special algorithm. P-BFT prevents attacks such as “double spending” due to the difference in “voting” times within the cluster and the characteristic DAG (state) synchronization time, which is carried out through permalinks.

Clustering provides the following benefits:

- Formation of "topologically close" groups of nodes, which improves trust between nodes (reduces the risks of attacks) and improves performance;
- Allows for “sharding” of the network, including the formation of private DAG branches;
- Increases horizontal network scalability.

DGT is positioned as a platform for distributed and decentralized computing, where the system processes data regardless of the specific application task. To solve a specific task, it is required to set up a family of transactions, as well as add an application client part [7,19]. In fact, the DGT software is the Nth number of typical nodes - Node, which provide interaction with other nodes, data validation and insertion of new data into the storage (registry), also called DAG or State. It is aimed at supporting consortium-based networks. This means that a node can join the network if certain conditions are met. In the simplest terms, this could be checking the host for a certificate. Depending on the implementation of the anchor mechanism, the degree of openness of the network varies - from completely open (public) to completely closed (private).

Nodes are combined into groups and are called clusters. The initial interaction is carried out through connections between nodes with one dedicated node in the cluster - Leader. The leader collects data from transaction checks at each node. Such checks are called “votes”. If the number of votes exceeds a certain threshold, then the transaction is considered approved in the cluster and awaits arbitration, an additional check performed outside the cluster. Within a cluster, nodes interact with each other via dedicated channels, also called permalinks.

Following Sawtooth, DGT is a multi-transactional system in which multiple families of transactions can be addressed. Each family is processed by a separate transaction processor. Transaction families complement the technology of smart contracts, and also allow you to set the boundaries of the availability of different types of transactions for different network segments.

But this approach cannot provide complete system protection, as server components or the server itself may fail.

Initially, a set of experiments was carried out with a two-level processing system on the DGT platform. During the experiments, the fault tolerance of the system was tested, which successfully coped with more than 1000 transactions with stable 24 nodes. The results of the experiments showed that the throughput averaged 0.009 seconds per transaction (Figure 9).

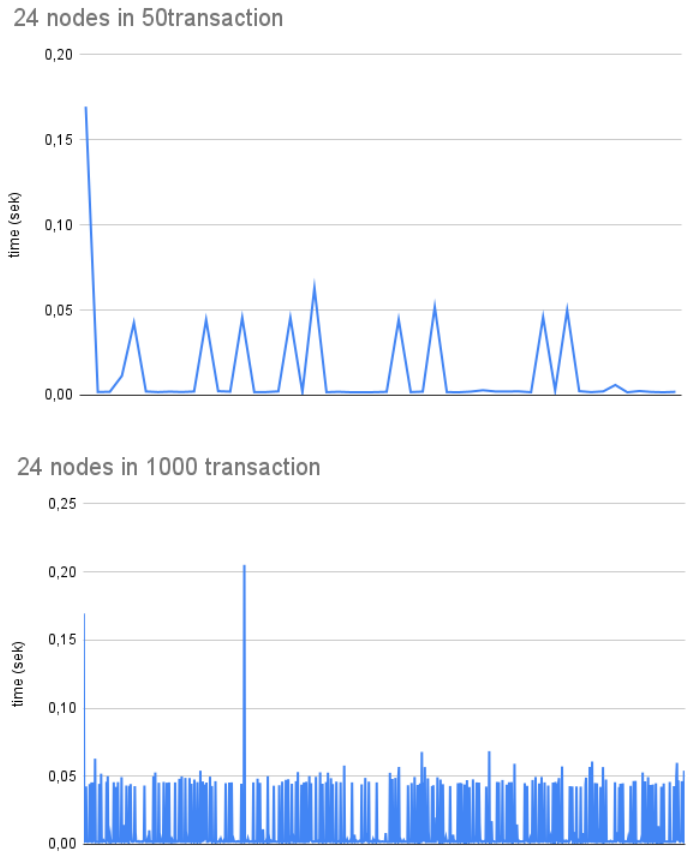

Figure 9 - Transaction Processing Graph with 24 Nodes Deployed

Figure 10, 11 simulated a forced failure of 6 nodes out of 24 when processing 1000 transactions, in this scenario, the leader of the node began processing rounds of voting among 18 nodes, and continued to process data. In this position, processing took an average of 0.0077 seconds (Figure 10).

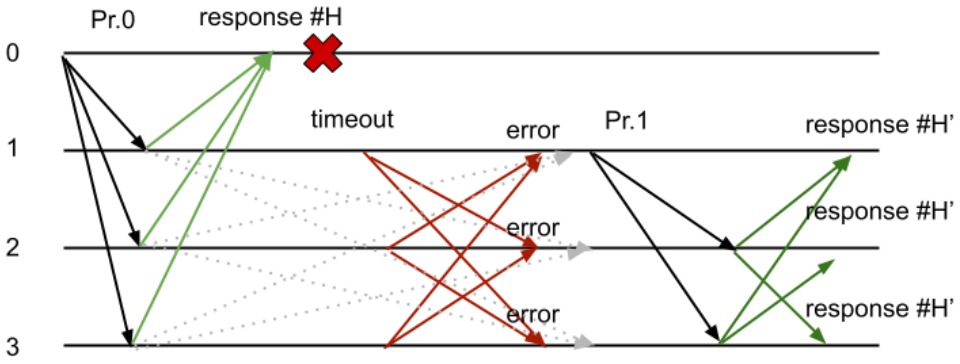

Figure 10 - If one node fails, move on to the next to process transactions

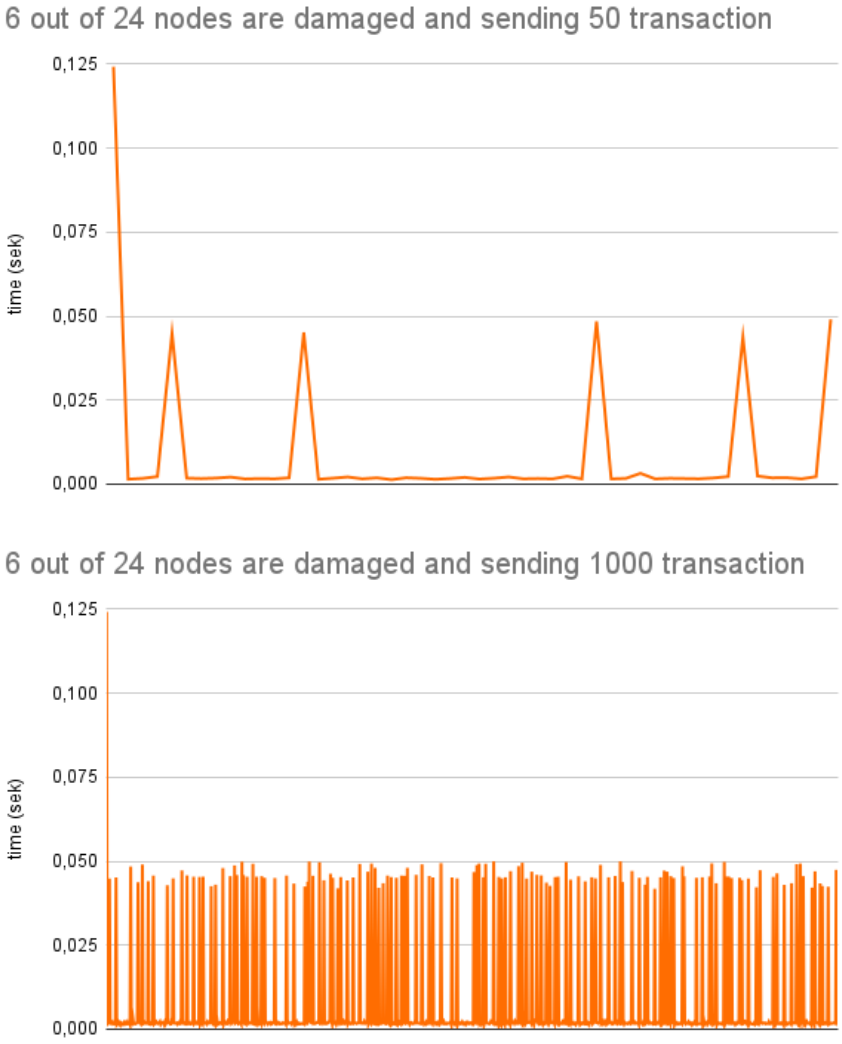

Figure 11 - Transaction Processing Graph with 24 Nodes Deployed, with 6 Nodes Forced to Fail.

In the third scenario, the failover solution was tested on 12 nodes out of a total of 24. Despite such a large-scale failure, the PBFT consensus algorithm continued to work stably, processing one transaction in an average of 0.0073 seconds (Figure 12).

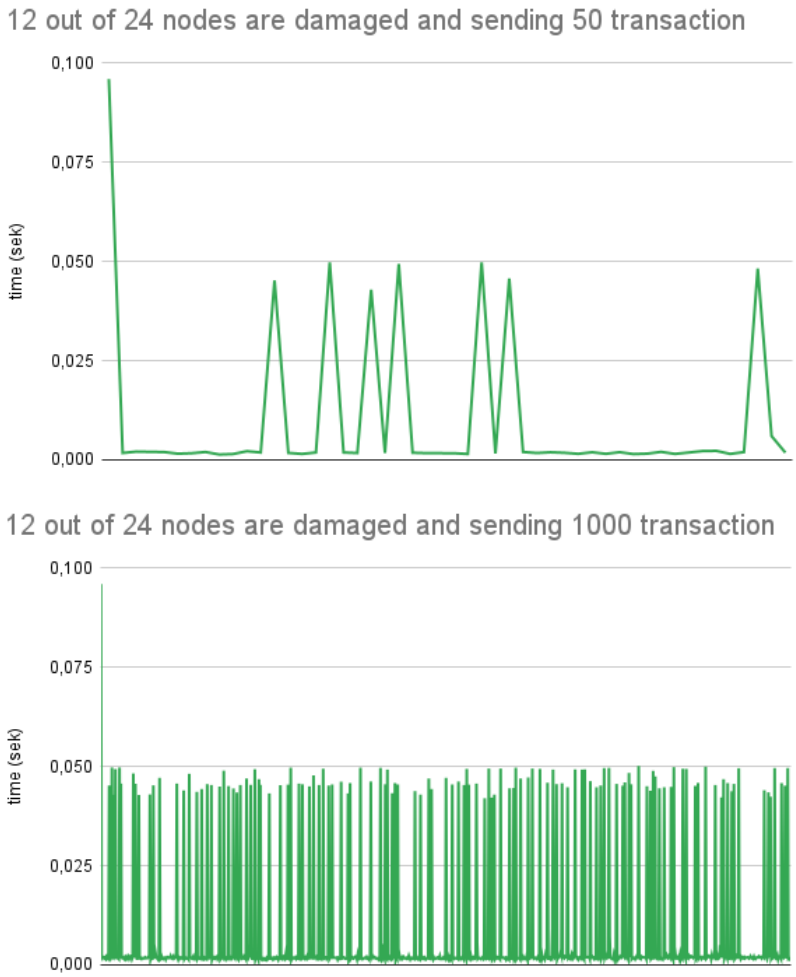

Figure 12 - Graph of transaction processing with 24 nodes deployed, with 18 nodes forced to fail over

Each node is a set of services that interact with each other and are responsible for organizing the network, storing data, and processing transactions. Even a single node delivers a significant service that supports client applications via APIs. At the same time, a number of network capabilities of the platform can be used only if there are several nodes. Therefore, thanks to this technology, we eliminate the main drawback of a distributed system - data loss / failure, which allows us to reduce the risks associated with the failure of a part of the system.

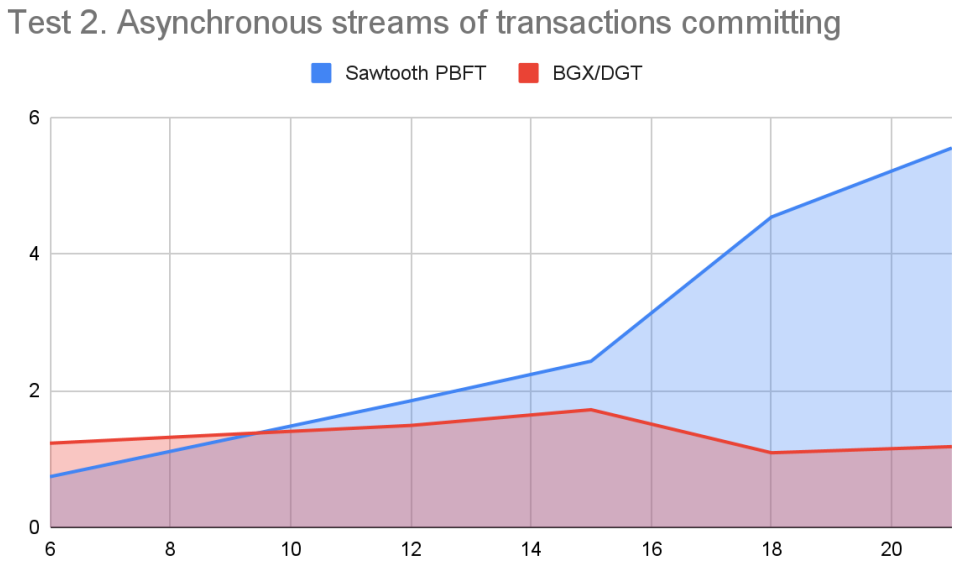

Figure 13 - Comparisons of Initial BGX/DGT Transactions with Sawtooth PBFT Platform for Upgraded Throughput

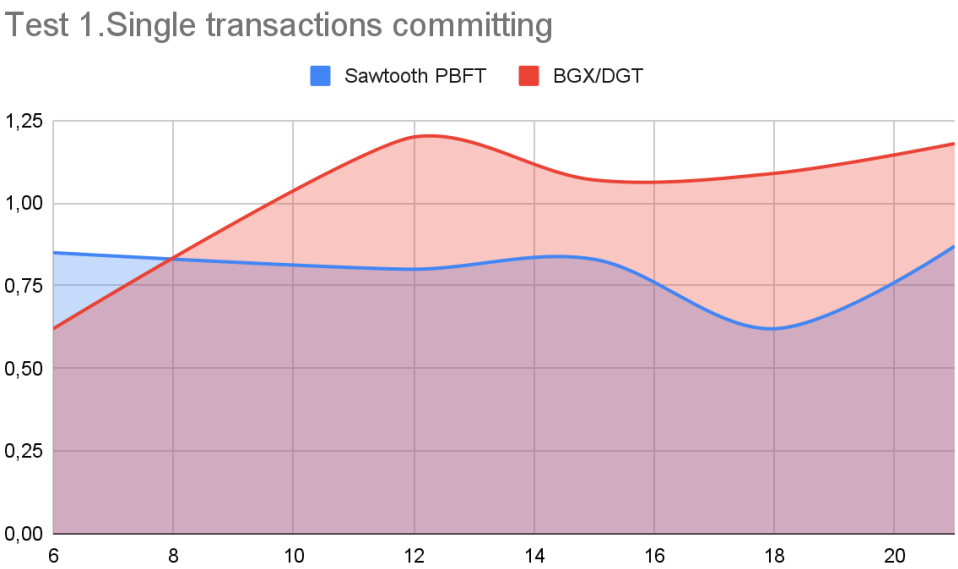

Figure 14 - Comparisons of BGX/DGT Asynchronous Transaction Flows with Sawtooth PBFT Platform for Upgraded Throughput

Note in figure 13, 14, the time it takes to send a block commit event from the identifier is considered negligible for the subscriber, since both of them are on the same virtual machine and as such there are no network delays. The transaction scheduler can be either serial or parallel. When using sequential transaction execution is blocked until the execution of the previous transaction is completed, with parallel transaction

scheduling, the subsequent transaction, which has no dependencies on the current transaction, can be executed in parallel with the current transaction sent for execution.

3.2. Formation of a new level of verification in the P-BFT consensus, access to demarcation of transactions

The PBFT (Practical Byzantine Fault Tolerance) consensus can provide a mechanism for exchanging replicas of files with each other so that each copy remains consistent even in the event of corruption. Alternatively, in Bitcoin, ordering occurs through a process called mining, where competing computers compete to solve a cryptographic puzzle that determines the order on which all processes are subsequently built.

Each node must communicate with other nodes to secure the network, which means that as the number of nodes increases, there is a huge network overhead. The PBFT consensus works great with small groups. The PBFT model is vulnerable to Sybil attacks, in which a large number of nodes can be exploited by one party on the network, compromising security. The threat can be mitigated by increasing the size of the network, but the PBFT mechanism does not support large networks for the above reason. So by using it in combination with another consensus mechanism, it can be optimized [75].

A public key infrastructure is used to generate cryptographic certificates that are tied to organizations, network components, and end users or client applications. As a result, data access control can be managed across the wider network and at the channel level. This approach helps address issues where privacy is paramount.

A blockchain network is a technical infrastructure that provides accounting services and smart contracts (which are packaged as part of a “chaincode”) to applications. First of all, smart contracts are used to generate transactions, which are subsequently distributed among all peers in the network, where they are invariably written to their copy of the ledger. Application users can be end users using client applications or blockchain network administrators.

Transaction processing initially went through 2 levels, at the first level, N nodes were deployed in the cluster and voting was held. At the end of the voting, the results were transferred to the leader of the node for the next, 2-stage voting, at the 3rd level, the notary node confirmed the transactions using the RAFT algorithm and added them to the new block chain (Figure 15).

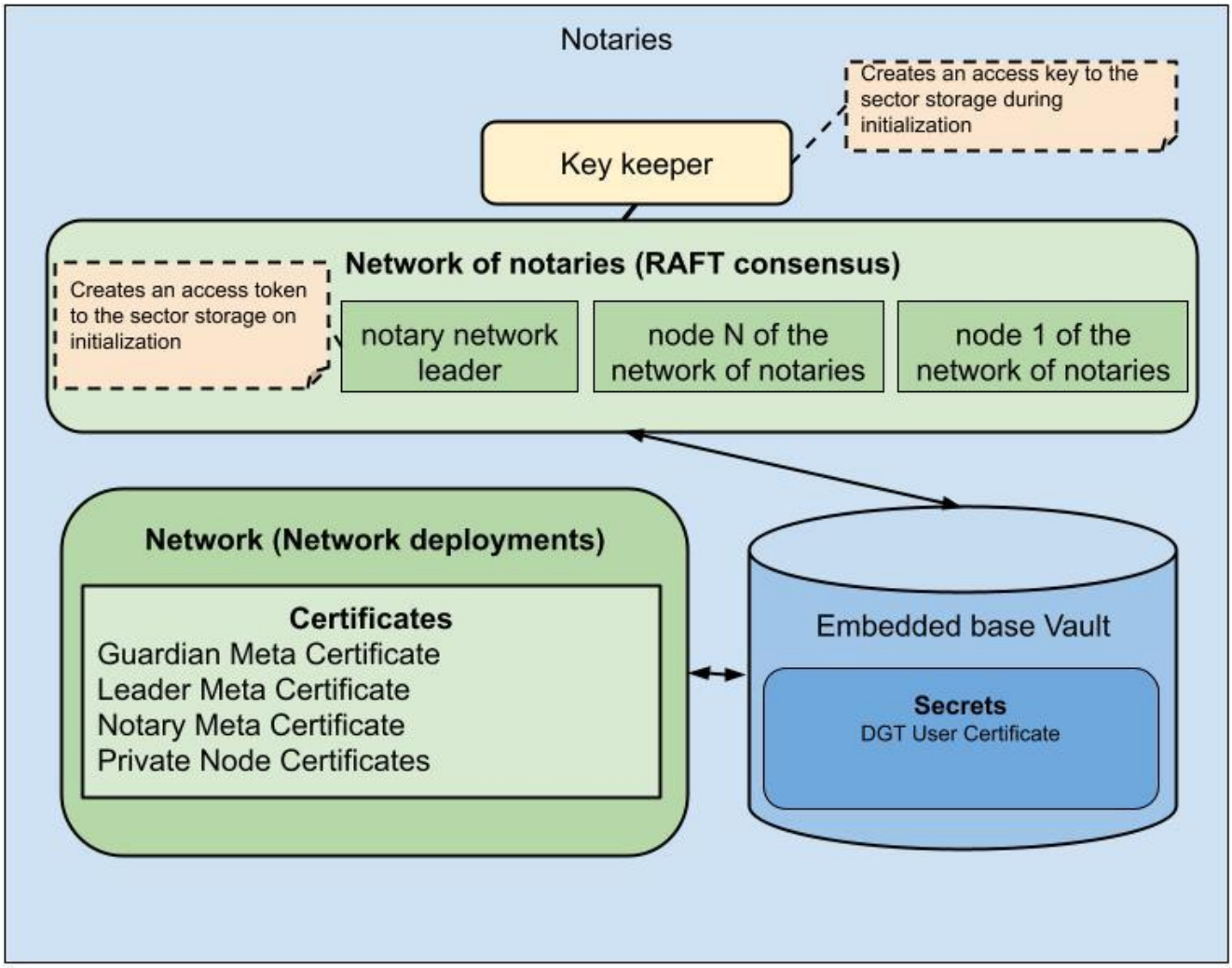

Figure 15 - Confirmation of data using a notary node

The Raft algorithm consists of two phases: leader election and transaction log replication. If the leader fails, the execution of the algorithm is repeated from the first phase. At any given time, a node is in one of three states: leader, candidate, or follower.

The Raft algorithm is numbered by consecutive integers. Each stage begins with an election in which one or more candidates try to become the leader. If the candidate wins the election, then he remains the leader until the end of the term. The state transition

is shown in figure 15 [3]. All nodes start with a slave state. If the follower does not hear from the leader within a certain period of time, then he becomes a candidate. The candidate then requests the votes of other nodes to become the leader. Other nodes will respond to the voting request. If the candidate receives the votes of the majority of the nodes, he becomes the leader. This process is called leader election. In particular, if a follower receives a heartbeat within the minimum election timeout from the current leader, it does not cast its vote for the candidate. This helps to maximize the uptime of the leader and avoid frequent failures due to some isolated nodes. In normal Raft operation, there is exactly one leader and all other nodes are followers. The leader periodically sends all his followers to maintain authority. All transactions during this period go through the leader. Each transaction is added as an entry to the node's registry.

There are several timeout settings in the Raft algorithm. One of them controls the electoral process. Election timeout is the amount of time a follower needs to wait to become a candidate. The follower enters the candidate state when the election time reaches zero. The time counter is reset to a random value when the slave receives a pulse from the leader [76]. Random election timers in Raft help reduce the chance of multiple subscribers moving to candidates at the same time.

We define the packet loss probability as p and suppose that p is a constant value for this network. Let's denote the timeout value for each round of elections as E_t, which is initially uniformly selected from the range $[a, b]$. The interval between two frequencies is τ. Discrete and integer time scales are adopted. Thus, if the slave cannot consistently receive $K = [E_t/\tau]$ cycles, it assumes that there is no viable leader and goes into candidate state to start the election. notice, that $K \in \{K_1, K_2, \ldots, K_r\}$, K – uniformly chosen from a set $\{K_1, K_2, \ldots, K_r\}$, where $K_1 = [a/\tau]$ and $K_r = [b/\tau]$. In the following analysis, K denotes the maximum number of clock pulses for an elective counter before timeout.

Let $g(n)$ — stochastic process representing the state of a stage $\{1, 2,...,r\}$ given node at time n. Let $b(n)$ — stochastic process representing the left steps of the selection time counter for a node at a point in time n. Once independence between $g(n)$ and $b(n)$ is assumed, we can model this as a two-dimensional process $\{g(n), b(n)\}$.

Let's take a short notation: $P\{i, k_i - 1|i, k_i\} = P\{g(n + 1) = i, b(n + 1) = k_i - 1|g(n) = i, b(n) = k_i\}$. In this Markov chain, the only non-zero one-step transition probabilities are

$$P\{i, k_i - 1|i, k_i\} = p \tag{9}$$

$$P\{i, K_i|i, k_i\} = (1 - p)/r \tag{10}$$

$$P\{i, 0|i, 0\} = 1 \tag{11}$$

Where $i, j = 1, 2, \ldots, r$ and $k_i \in \{1,\ldots,K_i\}$.

Equation (9) is based on the fact that the slave receives no clock from the current leader and its election time counter is decremented by 1. Equation (10) shows the fact that the slave receives a heartbeat and resets the election time counter. Equation (11) shows that once the selection time counter reaches zero, the slave transitions to the candidate state.

Denote $\{i, 0\}$ as an absorbing state. Because the $i = 1, 2,\ldots,r$, available r absorbing states. Let us designate other states, except for the state $\{i, 0\}$ in the state space $\{g(n), b(n)\}$, as transition states. There are t transition states, where $t = \sum_{i=1}^{r} K_i$. We order the states so that the first t states are transitional, and the last r states are absorbing. The transition matrix has the following canonical form:

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \tag{12}$$

where Q — size matrix $t \times t$, R — nonzero size matrix $t \times r$, 0 — zero size matrix $r \times t$, I — identity matrix of size $r \times r$. In particular, the input q_{ij} in Q is defined as the probability of transition from the transition state S_i into a transition state S_j, and the input r_{mn} in R is defined as the probability of transition from the transition state S_m into an absorbing state S_n.

When $K_r - K_1 < K_1$ or $b - a < \tau$, the selection timeout value has only one value. Thus, $r = 1$ and $t = K$, and then the only nonzero one-step transition probabilities in (9)–(11) can be simplified as follows:

$$P\{k - 1|\, k\} = p \tag{13}$$

$$P\{K|k\} = 1 - p \tag{14}$$

$$P\{0|0\} = 1 \tag{15}$$

where $k \in \{1,...,K\}$. So the transition matrix $\boldsymbol{P}$ becomes a matrix $(K + 1) \times (K + 1)$ as

$$P = \begin{pmatrix} 1-p & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \tag{16}$$

At step n, the transition matrix is P^n, and the element $p_{ij}^{(n)}$matrix P^n there is a possibility of being able to S_j out of state S_i.

For simplicity, let's assume that the election timeout counter has a fixed value K in the next analysis. It is easy to extend the analytical results to the case when the election timeout is a random value $r > 1$.

Conclusions on the 3 chapter

P-BFT (Practical Byzantine Fault Tolerance) and RAFT (Replicated State Machine Protocol) Consensus Layered Network is a system that uses a combination of two consensus protocols to provide high fault tolerance and data validity in a distributed environment.

P-BFT is a consensus protocol that provides fault tolerance in the face of malicious attacks on the network. It allows distributed nodes to reach agreement on the state of the system using an algorithm that processes individual blocks of data in series. The P-BFT protocol provides reliable data delivery and protection against duplication of data blocks.

RAFT is another consensus protocol that is also used to ensure data validity in a distributed environment. It is based on a replicated state model, where each node in the system holds a complete copy of the data. Replicas use the voting protocol to determine the correct order of operations and confirm that consensus has been reached.

A multi-layer network using a combination of P-BFT and RAFT protocols usually consists of several layers. Each layer uses separate instances of the consensus protocols that process blocks of data at that layer. This allows you to achieve higher fault tolerance and data reliability, since the system can detect and correct errors at an earlier stage.

Layer 1 networks can use the P-BFT protocol to ensure high transaction processing speed, and layer 2 networks can use the RAFT protocol to ensure high data reliability and protection against malicious attacks. This allows you to achieve maximum fault tolerance and data reliability in a distributed environment.

IV-CHAPTER. RESEARCH RESULTS

4.1. Analysis of multi-level transaction processing access

A multilayer network based on a combination of P-BFT and RAFT protocols consists of several layers. At the top level, P-BFT is used to select a leader who will coordinate the network at the next level. At the next level, RAFT is used, which allows the leader of this level to coordinate the work of nodes at lower levels. Each level can have its own leader, who is chosen by voting.

Thus, the combination of P-BFT and RAFT protocols makes it possible to achieve a high degree of security and fault tolerance in large networks at a lower cost of communication and message processing.

Multilayer networks based on a combination of P-BFT and RAFT protocols showed the following results:

1. High performance: A multi-layer network based on P-BFT and RAFT protocols has shown high performance in transaction processing and consensus building in a distributed environment. The researchers noted that the use of a combination of protocols allows achieving high performance and fault tolerance in various conditions.
2. Data Reliability: Multi-layer networks based on a combination of P-BFT and RAFT protocols also provide high data reliability in a distributed environment. Studies have shown that systems using this combination of protocols can detect and correct errors earlier, which improves data reliability and reduces the likelihood of problems.
3. Reducing network load: Using a combination of P-BFT and RAFT protocols can reduce network load in distributed systems. This is due to the fact that protocols can use different methods of data transfer and transaction processing, which can reduce the amount of data transferred and speed up the processing.

4. Resilience to malicious attacks: Layered networks based on a combination of P-BFT and RAFT protocols are highly resistant to malicious attacks. Research has shown that systems using this combination of protocols can effectively defend against Sybil and DDoS attacks, which improves data security in distributed systems.

The overall research results show that the combination of P-BFT and RAFT protocols can be an effective tool for providing high fault tolerance, data reliability and security in distributed systems.

The numerical method of a multilayer network based on a combination of P-BFT and RAFT protocols can be implemented using algorithms that provide high transaction processing speed and data reliability in a distributed environment. It does the following steps:

1. Creation of a multilayer network based on a combination of P-BFT and RAFT protocols with several nodes and levels.
2. Separation of nodes into groups that can work independently of each other and process transactions in parallel.
3. Implementation of a consensus mechanism based on the P-BFT and RAFT protocols, which allows achieving unity among nodes and maintaining data reliability.
4. Optimization of the transaction processing process, for example, by using parallel computing and load distribution between nodes
5. Implementation of protection mechanisms against various types of attacks, including DoS attacks and "transaction manipulation" attacks, to ensure the security and reliability of the network.
6. Implementation of mechanisms for monitoring and managing the network, including tools for analyzing and monitoring performance, mechanisms for detecting and correcting network errors and problems, as well as mechanisms for updating and scaling the network.
7. Ensuring that the network is sufficiently scalable and flexible to adapt to changing user needs and changing conditions in a distributed environment.

8. Implementation of mechanisms for storing and managing data in the network, including mechanisms for backing up, restoring and synchronizing data between nodes.
9. Development and implementation of interfaces and applications that allow users to interact with the network and use it to exchange data and conduct transactions.
10. Testing and optimizing the network, including testing for strength, performance, reliability and security, as well as finding and eliminating errors and vulnerabilities.

Optimized Byzantine Fault Tolerance Algorithm for Blockchain Consortium based on a combination of P-BFT and RAFT protocols is a solution to ensure high fault tolerance of a blockchain system in a distributed environment. It is based on a combination of the P-BFT and RAFT protocols, which provide robust consensus building and failure tolerance.

The algorithm works as follows: when a new block appears in the blockchain, network nodes send their votes for this block. If a quorum of votes is reached, the highest numbered block received from the leader is elected as the new current block. If a quorum has not been reached, then the RAFT protocol is used to elect a new leader, who will then continue with the block selection process.

This approach to the fault tolerance of the blockchain consortium based on a combination of the P-BFT and RAFT protocols allows for high transaction processing speed and consensus building, subject to the conditions for reliable network operation. It also provides the fault tolerance of the blockchain system in a distributed environment and failure tolerance.

4.2. Combined transaction processing approach

A mathematical model of a multilayer network built on a combination of P-BFT and RAFT protocols can be represented as a system of equations describing the processes of messaging between network nodes and consensus decision making.

One of the possible mathematical models of such a network can be built on the basis of the following equations:

1. Equations for messaging between network nodes:

- Block request message from node *i* to node *j*: $M_{ij} = (i, j, n, H)$, where M_{ij} - message, *i* and *j* - node identifiers, *n* - block number, *H* - block hash.
- Block request response message from node *j* to node *i*: $M_{ji} = (j, i, n, H, D)$, where *D* – data of block.
- Voice request message from node *i* to node *j* $M_{ij} = (i, j, n, V)$, where *V* – vote of node *i* for block number *n*.
- Voice response message from node *j* to node *i*: $M_{ji} = (j, i, n, V)$, where *V* – vote of node *j* for block number *n*.
- Consensus request message from node *i* to nodes $j_1, j_2, \dots, j_k$: $M_{ij} = (i, j_1, j_2, \dots, j_k, n)$, where *n* – number of block.
- Node consensus response message $j_1, j_2, \dots, j_k$ to node *i*: $M_{ij} = (j_1, j_2, \dots, j_k, i, V)$, where *V* - decision to approve the block with the number *n*.

2. Equations for consensus decision making:

- The condition for achieving a quorum on the votes of the nodes: $|V| > f$, where $|V|$ - number of votes, *f* - the number of errors the system can make.
- The condition for achieving a quorum on responses to block requests: $|D| > f$, where $|D|$ - number of blocks received from other nodes, *f* - number of errors.
- Condition for achieving a quorum on responses to consensus requests: $|V| > f$, where $|V|$ - number of votes received from other nodes, *f* - number of errors;
- The function of choosing a leader responsible for collecting and processing votes.

3. Leader selection algorithm in the RAFT protocol:

- Each node starts in the "suspect" state;
- A node may transition to the "viewing" state if it receives a message with a higher proposed leader number;

- A node transitions to the "confirming" state if it receives messages from a majority of nodes in the "browsing" state;
- A node transitions to the "leader" state if it receives an acknowledgment from a majority of nodes in the "acknowledging" state;
- If the node does not receive confirmation for some time, it starts a new round of elections, increasing the number of the proposed leader.

4. Block formation algorithm:

- The node collects transactions in a pool of transactions;
- The node adds a block header that includes the number of the previous block and the hash of the previous block;
- The node calculates the hash of the current block based on its header and transactions;
- The node adds the hash of the current block to the header and generates the final block.

5. Algorithm for block verification:

- The node checks that the number of the previous block in the header matches the number of the previous block in the block chain;
- The node checks that the hash of the previous block in the header matches the hash of the previous block in the block chain;
- The node checks that the hash of the current block in the header matches the hash of the current block calculated from the header and transactions;
- The node checks that all transactions in the block are correct;
- If the verification succeeds, the node accepts the block and adds it to the block chain.

Decision selection function based on a combination of P-BFT and RAFT protocols: if a quorum of votes has been reached, then the block with the highest number received from the leader is selected; if a quorum has not been reached, then the RAFT algorithm is used to select a leader and then make a decision.

A decision function based on a combination of P-BFT and RAFT protocols has the following algorithm:

1. The leader node proposes a new block to be added to the blockchain.
2. Nodes in the network vote for a block: each node sends a vote request to other nodes in the network.
3. If a quorum of votes has been reached, i.e., the number of votes exceeds the number of errors f, then the block with the highest number received from the leader is selected. If the quorum has not been reached, then go to step 4.
4. The RAFT algorithm is used to select a new leader. Nodes conduct leader elections using the RAFT protocol.
5. The new leader proposes a decision block.
6. This block can be accepted as the correct solution if the majority of participants support it.

Thus, the decision selection function combines the advantages of the two negotiation protocols and allows you to choose the right decision, even if not all participants agree with it. If a quorum has been reached, then the block with the highest number proposed by the leader is selected. If the quorum has not been reached, then a new leader is selected using the RAFT protocol, after which this leader proposes a new block for decision making.

This mathematical model allows one to describe the processes occurring in a multilayer network built on a combination of the P-BFT and RAFT protocols, and allows one to determine the conditions necessary for making a consensus decision. It also demonstrates how these protocols can work together to provide a robust and efficient system for reaching consensus in a distributed environment.

An important aspect of a multilayer network based on a combination of P-BFT and RAFT protocols is the support for security certificates. Security certificates are used to provide security in the leader selection and decision making process. Security certificates can be used to authenticate network participants, thereby providing protection against possible message spoofing or spoofing attacks. In addition, certificates can be used to

secure the leader election process by ensuring that only trusted participants can participate in the leader election process.

In a layered network based on a combination of the P-BFT and RAFT protocols, certificates can be used to authenticate the members of each layer of the network, as well as to secure the leader election and decision making process at each layer. In addition, certificates can be used to ensure the confidentiality and integrity of data transferred between network participants.

Support for security certificates is an important aspect of a multi-layer network based on a combination of P-BFT and RAFT protocols, ensuring the security of the leader election and decision-making process at each layer of the network.

Conclusions on the 4 chapter

A multi-layer network based on a combination of P-BFT and RAFT protocols may include the following steps:

1. Setting up a multi-layer network: depending on the specific task, it is necessary to select the optimal network configuration, including the number of layers, the number of nodes at each layer, and the parameters of the P-BFT and RAFT protocols.
2. Distribution of nodes by tiers: nodes must be tiered so that each tier contains a certain number of nodes necessary to achieve the required performance and reliability
3. Network initialization: when starting the network, it is necessary to initialize each node and establish a connection between the nodes.
4. Transaction processing: P-BFT and RAFT protocols are used for transaction processing. In addition, each level processes transactions with a certain level of complexity. Transactions are processed using the P-BFT protocol, which provides high processing speed and transaction reliability.
5. Data Synchronization: In order to ensure data consistency across all layers of the network, the RAFT protocol is used, which provides data synchronization and error detection. At the same time, each level has its own independent transaction log, which is synchronized with the logs at other levels.
6. Error Handling: When network errors occur, such as node failure or data conflicts, the RAFT protocol provides automatic network recovery and conflict resolution.
7. Monitoring and analysis: to optimize the performance and reliability of the network, it is necessary to constantly monitor its condition and analyze the results of work

CONCLUSION

Based on the results of the research, the following conclusions can be drawn:

1. The creation of virtual machines that implement the operating environment of the blockchain system is an important step in the development of heterogeneous software and hardware systems.
2. The developed methodology for launching applications in multi-level virtual environments really makes it possible to increase the overall performance of such complexes.
3. The developed approach to building the operating environment of the user subsystem provides secure user access to resource-intensive applications in a heterogeneous distributed cloud computing environment.
4. The study of methods for improving the reliability of authentication and authorization and the developed methodology for their application in a heterogeneous cloud environment are important components for ensuring the security of such systems.

Thus, the analysis and development of blockchain systems significantly speeds up the data processing process.

Bibliography

1. Bogdanov, A. Risk model of application of lifting methods [Electronic resourse] / A. Bogdanov, A. Degtyarev, N. Shchegoleva [et al.] // The 9th International Conference «Distributed Computing and Grid Technologies in Science and Education» on. — 2021. — Vol. 3041. P. 369–374. — URL: https://doi.org/10.54546/mlit.2021.77.69.001 (date of access: 24.06.2023).

2. Bogdanov, A. Solving the problems of byzantine generals [Electronic resourse] / A. Bogdanov, A. Degtyarev, N. Shchegoleva [et al.] // The 9th International Conference «Distributed Computing and Grid Technologies in Science and Education» on. — 2021. — Vol. 3041. P. 573–578. — Режим доступа: https://doi.org/10.54546/mlit.2021.72.42.001 (date of access: 24.06.2023).

3. Bogdanov, A. A Multilayer Approach to the Security of Blockchain Networks of the Future Bogdanov / A. Bogdanov, A. Degtyarev, N. Shchegoleva [et al.] // ICCSA 2022 Workshops. — [s. l.] : Springer, 2022. — Vol. 13377. — P. 205–216.

4. Bogdanov, A. Protection of Personal Data Using Anonymization / A. Bogdanov, A. Degtyarev, N. Shchegoleva [et al.] // Computational Science and Its Applications. ICCSA 2021. — [s. l.] : Springer, 2021. — Vol. 12956. — P. 447–459.

5. Bogdanov, A. Testing and Comparative Analysis of the F-BFT-based DLT Solution / A. Bogdanov, N. Shchegoleva, V. Korkhov [et al.] // International Conference on Computational Science and Its Applications. — [s. l.] : Springer, 2021. — Vol. 12952. — P. 31–41.

6. Bogdanov, A. Digitalization of health care: what can be done now / A. Bogdanov, N. Zalutskaya, N. Schegoleva [et al.] // Information Society Journal. — 2022. — P. 58–70

7. Bogdanov, A. Comparative analysis and applicability determination for several dlt solutions [Electronic resourse] /A. Bogdanov, V. Korkhov, N. Shchegoleva [et al.] // The 9th International Conference «Distributed Computing and Grid Technologies in Science and Education» on. — 2021. — URL: https://doi.org/10.54546/mlit.2021.13.56.001 (date of access: 24.06.2023).

8. Imoize, A. L. 6g enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap / A. L. Imoize, O. Adedeji, N. Tandiya, S. Shetty — Sensors, 2021. — Vol. 21. — I. 5.

9. Musleh, A. S. Blockchain Applications in Smart Grid–Review and Frameworks / A. S. Musleh, G. Yao, S. M. Muyeen // IEEE., — 2019. — Vol. 7. — P. 86746–86757.

10. Aijaz, A. Private 5G: The Future of Industrial Wireless / IEEE Industrial Electronics Magazine. — 2020. — Vol. 14. - №4 — P. 136–145.

11. Anshuman Kalla. A survey on the use of blockchain for future 6G: Technical aspects, use cases, challenges and research directions [Electronic resourse] / Anshuman Kalla Chamitha de Alwis, Pawani Porambage, Gürkan Gür, Madhusanka Liyanage. — Journal of Industrial Information Integration. — 2022. — P. 100404. — URL: url: https://doi.org/10.1016/j.jii.2022.100404 (date of access: 24.06.2023).

12. Aazhang, B. Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence / Matti Latva-aho, Kari Leppänen (eds.) — Finland : University of Oulu, 2019. — 36 p.

13. Burtyka, Ph. Batch Symmetric Fully Homomorphic / “Proceedings of ISP RAS” journal. — 2014. — Vol. 26. — I. 5. — P. 99–116.

14. Graham, C. Anonymisation: managing data protection risk code of practice / C. Graham. — [s. l.] : Information Commissioner’s Office, 2012. — P. 108.

15. Benzaid, C. AI-driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions / C. Benzaid, T. Taleb. — IEEE Network. — 2020. —Vol. 34. — № 2 — P. 186–194.

16. Gaber, C. Liability-aware security management for 5G / C. Gaber, J. S. Vilchez, G. Gür [et al.] // 2020 IEEE 3rd 5G World Forum (5GWF). — IEEE, 2020. — P. 133–138.

17. Hu, C. A novel blockchain-based anonymous handover authentication scheme in mobile networks / C. Hu, D. Zheng, R. Guo, A. Wu, L. Wang, S. Gao. — International Journal of Network Security. — 2020. — Vol. 22. — №5 — P. 874–884.

18. Children’s Hospital of Eastern Ontario Research Institute Pan-Canadian De-Identification Guidelines for Personal Health Information. — Canada : Office of the Privacy Commissioner of Canada, 2007. — 87 p.

19. DGT ONE PAGER [Electronic resourse] //DGT. — URL: https://dgt.world/docs/DGT_OnePager.pdf. (date of access: 10.05.2023)

20. DGT The Blockchain Handbook [Electronic resourse] //DGT. —URL: https://dgt.world/docs/DGT_BLOCKCHAIN_ABC.pdf (date of access: 10.05.2023)

21. ENISA Data Pseudonymization: Advanced Techniques & Use Case / Athena Bourka (ENISA), Prokopios Drogkaris (ENISA) (eds.). — [s. l.] : ENISA, 2021. — 53 p.

22. Hu, F. Full spectrum sharing in cognitive radio networks toward 5G: A survey / F. Hu, B. Chen, K. Zhu // IEEE Access. — 2021. — V. 6. — P. 15754–15776.

23. Fraser, R. Tools for de-identification of personal health information / R. Fraser, D. Willison. — Canada : Pan Canadian Health Information Privacy (HIP), 2009. — 40 p.

24. Dik, G. Challenges of IoT Identification and Multi-Level Protection in Integrated Data Transmission Networks Based on 5G/6G Technologies / G. Dik, A. Bogdanov, N. Shchegoleva [et al.] — Computers. — 2022. — Vol. 11. — № 12. — 178 p.

25. Goverment of Canada Personal Information Protection and Electronic Documents Act. — Canada: Minister of Justice, 2023.

26. GSMA Securing the 5g era [Electronic resourse] / GSMA. — URL: https://www.gsma.com/security/securing-the-5g-era/. (date of access: 03.05.2021)

27. Zhang, H. Blockchain-based trust management for internet of vehicles / H. Zhang, J. Liu, H. Zhao [et al.] // IEEE Transactions on Emerging Topics in Computing. — 2021.- Vol. 9. — № 3 — P. 1397–1409.

28. Haverinen, J. Extensible authentication protocol method for 3rd generation authentication and key agreement / J. Haverinen, H. Arkko. — USA : RFC Editor, 2006. — 79 p.

29. “Best Practice” Guidelines for Managing the Disclosure of De-Identified Health Information / Health System Use Technical Advisory Committee Data De-

Identification Working Group. — Ottawa : Canadian Institute for Health Information, 2010. — 53 p.

30. Ahmad, I. Security for 5G and beyond / I. Ahmad, S. Shahabuddin, T. Kumar [et. al.] // IEEE Communications Surveys & Tutorials. —2019. — Vol. 21. — № 4. — P. 3682–3722.

31. Shayea, I. Key Challenges, Drivers and Solutions for Mobility Management in 5G Networks: A Survey / I. Shayea, M. Ergen, M. H. Azmi [et al] // IEEE Access. — 2020. — Vol. 8. — P. 172534–172552.

32. De-identification Guidelines for Structured Data / Information and Privacy Commissioner of Ontario. — Toronto, Ontario : [s.n.], 2016.

33. ISO 17432:2004 Health informatics. Messages and communication. Web access to DICOM persistent objects. - [s.l.] : Standardinform, 2010.

34. Demertzis, K. Anomaly detection via blockchained deep learning smart contracts in industry 4.0 / K. Demertzis, L. Iliadis, N. Tziritas [et al.]. — Neural Computing and Applications. — 2020. — Vol. 32. — P. 17361–17378.

35. Gai, K. Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks / K. Gai, Y. Wu, L. Zhu [et al.] // IEEE Internet of Things Journal. —2019. — Vol. 6. — P. 7992–8004.

36. Leppänen, K. Key drivers and research challenges for 6G ubiquitous wireless intelligence / K. Leppänen, M. Latva-Aho. — Finland : Oulu, 2019.

37. Liang, Y.-C. Dynamic Spectrum Management / Y.-C. Liang. — [s .l.] : Springer Nature, 2020. — P. 21–27.

38. Crosby, M. Blockchain technology: Beyond bitcoin / M. Crosby, P. Pattanayak, S. Verma [et al.]. — [s.l.] : Applied Innovation, 2016. — Vol. 2. — 16 p.

39. Renzo, M. Di. Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead / M. Di Renzo, A. Zappone, M. Debbah [et al.] // IEEE Journal on Selected Areas in Communication. — 2020. — Vol. 38. — I. 11. — P. 2450–2525.

40. Chowdhury, M. Z. 6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions / M. Z. Chowdhury,

M. Shahjalal, S. Ahmed [et al.]. // IEEE Open Journal of the Communications Society — 2020. — Vol. 1. — P. 957–975.

41. Fredrikson, M. Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing / M. Fredrikson, E. Lantz, S. Jha // 23rd USENIX Security Symposium. — San Diego, CA : Security Symposium, 2014. — P. 17–32.

42. Weerasinghe, N. A novel blockchain-as-a-service (baas) platform for local 5g operators / N. Weerasinghe, T. Hewa, M. Liyanage [et al.] // IEEE Open Journal of the Communications Society. — 2021. — Vol. 2. — P. 576–601.

43. Zaynalov, N. Information Security Issues For Travel Companies / N. Zaynalov, A. Muhamadiev, U.bekburodov [et al.] // 2019 International Conference on Information Science and Communications Technologies (ICISCT). — [s.l.] : IEEE, 2019. — P. 1–4.

44. Zaynalov, N. Hiding short message text in the uzbek language / N. Zaynalov, U. Narzullayev, A. Muhamadiev [et al.] // 2020 International Conference on Information Science and Communications Technologies (ICISCT). — Tashkent, Uzbekistan : IEEE, 2020. — P. 1–6.

45. NISTIR 8062 An Introduction to Privacy Engineering and Risk Management / S. Brooks, M. Garcia, N. Lefkovitz [et al.]. — [s. l.] : U.S. Department of Commerce, 2017.

46. Privacy Enhancing Technologies – A Review of Tools and Techniques / Office of the Privacy Commissioner of Canada. — Canada : Office of the Privacy Commissioner of Canada, 2017. — 11 p.

47. Porambage, P. The Roadmap to 6G Security and Privacy / P. Porambage, G. Gur, D. P. M. Osorio [et al.] // IEEE Open Journal of the Communications Society. — 2021. — Vol. 2. — P. 1094–1122.

48. Porambage, P. 6G Security Challenges and Potential Solutions / P. Porambage, G. Gür, D. P. M. Osorio [et al.] // 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). — [s.l.] : IEEE, 2021. — P. 622–627.

49. Advisory Guidelines on enforcement of the data protection provisions / Personal Data Protect Comission, Singapore. — [s.l.] : PDPC, 2016. — 52 p.

50. General Data Protection Regulation: [regulation of the European Parliament and of the Council]. — [s. l.] : Intersoft Consulting, 2018. — P. 1–5.

51. Hu, S. Blockchain and artificial intelligence for dynamic resource sharing in 6g and beyond / S. Hu, Y.-C. Liang, Z. Xiong [et al.] // IEEE Wireless Communications. — 2021. — Vol. 28. — № 4 — P. 145–151.

52. Kim, S. A Survey of Scalability Solutions on Blockchain / S. Kim, Y. Kwon, S. Cho // 2018 International Conference on Information and Communication Technology Convergence (ICTC). — Jeju, Korea (South) : IEEE, 2018. — P. 1204–1207.

53. Kiyomoto, S. On blockchain-based authorization architecture for beyond-5G mobile services / S. Kiyomoto, A. Basu, M. S. Rahman [et al.] // 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST). — Cambridge, UK : IEEE, 2018. — P. 136–141.

54. Tanwar, S. Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward / S. Tanwar, Q. Bhatia, P. Patel [et al.] // IEEE Access. — 2019. — Vol. 8. — P. 474–488.

55. A survey on mobile edge networks: Convergence of computing, caching and communications / S. Wang, X. Zhang, Y. Zhang [et al.] // IEEE Access. — Vol. 5. — P. 6757–6779.

56. 6G The Next Hyper — Connected Experience for All / Samsung Research. — [s. l.] : Samsung Research, 2020. — 46 p.

57. Shchegoleva, N. New Technologies for Storing and Transferring Personal Data / N. Shchegoleva, N. Zalutskaya, A Dambaeva [et al.] // Computational Science and Its Applications — ICCSA 2022 Workshops. — 2022. — V. 13380. — P. 680–692.

58. Ariyarathna, T. Dynamic Spectrum Access via Smart Contracts on Blockchain / T. Ariyarathna, P. Harankahadeniya, S. Isthikar [et al.] // WCNC. — Marrakesh, Morocco : 2019 IEEE Wireless Communications and Networking Conference (WCNC), 2019. — P. 1–6.

59. Higashino, T. Context Recognition of Humans and Objects by Distributed Zero-Energy IoT Devices / T. Higashino, A. Uchiyama, S. Saruwatari [et al.] // 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). — Dallas, TX, USA : IEEE, 2019. — P. 1787–1796.

60. Maksymyuk, T. Blockchain-empowered framework for decentralized network management in 6g / T. Maksymyuk, J. Gazda, M. Volosin [et al.] // IEEE Communications Magazine. — 2020. — Vol. 58. — № 9 — P. 86–92.

61. Privacy-aware blockchain innovation for 6g: Challenges and opportunities / T. Nguyen, N. Tran, L. Loven [et al.] // IEEE. — 2020. — P. 1–5.

62. Miao, W. Unlocking the potential of 5g and beyond networks to support massive access of ground and air devices / W. Miao, C. Luo, G. Min [et al.] // IEEE Transactions on Network Science and Engineering. — 2021. — Vol. 8. — № 4 — P. 2825–2836.

63. Saad, W. A vision of 6G wireless systems:Applications, trends, technologies, and open research problems / W. Saad, M. Bennis, M. Chen // IEEE Network. — 2020. — Vol. 34. — № 3 — P. 134–142.

64. Yang, W. A Survey on Blockchain-Based Internet Service Architecture: Requirements, Challenges, Trends, and Future / W. Yang, E. Aghasian, S. Garg [et al.] // IEEE Access. — 2019. — Vol. 7. — P. 75845–75872.

65. Li, X. Network Slicing for 5G: Challenges and Opportunities / X. Li, M. Samaka, H. A. Chan [et al.] // IEEE Internet Computing. — 2017. — Vol. 21. — P. 20–27.

66. Liang, X. Integrating blockchain for data sharing and collaboration in mobile healthcare applications / X. Liang, J. Zhao, S. Shetty [et al.] // 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). — [s.l.] : IEEE, 2017.

67. Ling, X. Blockchain radio access network (B-RAN): Towards decentralized secure radio access paradigm / X. Ling, J. Wang, T. Bouchoucha [et al.] // IEEE Access. — 2019. — Vol.7. - P. 9714–9723.

68. . Liu Y. Blockchain and Machine Learning for Communications and Networking Systems /F. R. Yu, X. Li [et al.] // IEEE Communications Surveys & Tutorials. — 2020. — Vol. 22. — № 2 — P. 1392–1431.

69. Haddad, Z. Blockchain-based Authentication for 5G Networks / Z. Haddad, M. M. Fouda, M. Mahmoud, M. Abdallah // 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). — [s.l.] : IEEE, 2020. — P. 189–194.

70. Zhang, Z. 6G wireless networks: Vision, requirements, architecture, and key technologies / Y. Xiao, Z. Ma [et al.] // IEEE Vehicular Technology Magazine. — 2019. — Vol. 14. — № 3 — P. 28–41.

71. Towards artificial intelligence enabled 6G: State of the art, challenges, and opportunities / D. Zhu, S. Zhang // Computer Networks. — 2020. — Vol. 183. — P. 107556.

72. Naing, Y. M. Development of a system for launching resource-intensive applications in a cloudy heterogeneous environment : abstract of Ph.D. diss.... cand. tech. : Sciences: 05.13. 15 / Y. M. Naing. — ETY “LETI” : 2013. — 125 p.

73. Demichev, A. P. Introduction to Grid Technology / A. P. Demichev, V. A. Ilyin, A. P. Kryukov // Preprint NIINP MSU-2007-11/832. — 2007. — 87 p.

74. Paraskevov, A. V. Comparative analysis of legal regulation of personal data protection in russia and abroad [Electronic resourse]/ A. V. Paraskevov, A. V. Levchenko, Y. A. Cuhol // Scientific journal KubSAU. — 2015. — № 110. — URL: https://cyberleninka.ru/article/n/sravnitelnyy-analiz-pravovogo-regulirovaniya-zaschity-personalnyh-dannyh-v-rossii-i-za-rubezhom (date of access: 23.06.2023).

75. Romanenko, K. S. Features of the development of information systems on the hyperledger blockchain platform / K. S. Romanenko, E. A. Ishukova // Perspektiva— 2021. — 2022. — P. 51–55.

76. Huang, D. Performance analysis of the raft consensus algorithm for private blockchains / D. Huang, X. Ma, S. Zhang //IEEE Transactions on Systems, Man, and Cybernetics: Systems. — 2019. — Vol. 50. — №. 1. — P. 172–181.

77. Certificate of state registration of the computer program No. 2023618607 Russian Federation. Data Privacy Framework: No. 2023617657: App. 04/26/2023 : publ. April 26, 2023 / A. G. Dik, J. U. Kiyamov, V. V. Khvatov, N. L. Shchegoleva; applicant Limited Liability Company "ASSOCIATION OF DIGITAL TECHNOLOGIES AND INTELLIGENT SYSTEMS". – EDN CKPZHW.

APPLICATIONS A

Approach for determining quasi-identifiers by the index of the upper access level.

A big data platform typically consists of an infrastructure platform, a structured and unstructured data storage platform, and a data processing platform. Therefore, ensuring the protection of a big data platform is a very laborious process: it is necessary to ensure the security of processing in distributed software systems, the protection of information in databases using various DBMS; data and transaction logs must be protected; key management must be provided for access control and key tracking. In addition, to ensure the proper security context and functioning of the data at each stage, it is important to ensure the legitimacy of the origin of the data, and to ensure their availability, it is necessary to provide measures to counteract DoS attacks.

Securing big data system controls. Big data system management tools provide ample opportunities for implementing security mechanisms that allow real-time monitoring of the state of components, managing access control rules, identifying data sources, etc. value to offenders.

International standards and approaches to depersonalization

European Union

In the European Union, the basic document regulating relations related to PD is the General Data Protection Regulation (GDPR) [50].

The GDPR uses the terms “pseudonymization” and “anonymous data”. Their difference is revealed through the provisions of paragraph 26 of the Preamble and paragraph 5 of Art. 4. Since “personal data subjected to pseudonymization” can be “associated with an individual through the use of additional information”, then pseudonymized data is considered as PD. Anonymous data is not classified as PD, and the relevant requirements do not apply to them, since they “do not refer to an identified or identifiable natural person” or are provided in such a way that “the data subject is not

identified". But to further distinguish between pseudonymized and anonymous data, the GDPR requires "paying attention to all objective factors, such as the costs of identification and the amount of time required for identification, taking into account the technologies and technological developments available at the time of processing."

The purposes of using pseudonymization may be "to reduce risks for data subjects" and to provide "data protection", along with other technical and organizational data protection measures. At the same time, the measures applied should be correlated with the assessment of the degree of risk.

When assessing the degree of risk, it is necessary to take into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing, as well as the risk of varying probability and severity for the rights and freedoms of natural persons. Taking into account the degree of risk, the controller and the processor must implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk, including, but not limited to, where necessary:

- pseudonymization and encryption of personal data;
- the ability to ensure the continued confidentiality, integrity, availability, and resiliency of processing systems and services;
- the ability to restore availability and access to personal data in a timely manner in the event of a physical or technical incident;
- the process of regular testing, evaluating the effectiveness of technical and organizational measures to ensure the security of processing.

An assessment of the appropriate level of security should also take into account the possibility of accidental or unlawful destruction, loss, alteration, unauthorized disclosure or access to personal data transmitted, stored or otherwise processed.

The European Union Cybersecurity Agency (ENISA) is responsible for developing pseudonymization provisions. Thus, one of the newest acts of ENISA is "Pseudonymization of ENISA data: Best practices and use cases. Technical analysis of cybersecurity measures in the field of data protection and privacy" (January 2021) [21].

Canada

Canada has a wide regulatory framework regarding data anonymization, especially in the medical field. Canada's founding act on PD is the Personal Information Protection and Electronic Documents Act (PIPEDA), 2000 [25]). Special acts include:

- Pan-Canadian De-Identification Guidelines for Personal Health Information, 2007 [18],
- tools for depersonalization of personal health information (Tools for De-Identification of Personal Health Information (Authored by: Ross Fraser and Don Willison, 2009)) [23],
- 'Best Practice' Guidelines for Managing the Disclosure of De-Identified Health Information (Prepared by the: Health System Use Technical Advisory Committee Data De-Identification Working Group, October 2010) [29],
- Guidance document on Public Release of Clinical Information (March 12, 2019) [18],
- De-identification Guidelines for Structured Data (June 2016) [32].

In these documents, despite some difference in terminology, related, among other things, to the different time of adoption of acts, the depersonalization process is considered in connection with two aspects - reducing the risk of re-identification, on the one hand, and preserving the usefulness of the data, on the other, - in order to find a balance between them. Therefore, the task of depersonalization is to reduce the level of risk of re-identification to the minimum possible level, for which the procedure for assessing the degree of risk of re-identification is used for each specific case.

It is expected that the owner of the data set will not disclose the data set if the level of risk of re-identification is higher than the set threshold. The threshold value, in turn, is determined by three parameters:

1) the probability of attempts to re-identify;
2) the consequences of successful re-identification;
3) the impact of depersonalization on the useful properties of data.

Particular attention in the process of calculating risks in relation to publicly available data sources is given to quasi-identifiers, the presence of which in the data set represents a relatively low level of risk of re-identification:

- region of residence (without the presence of other quasi-identifiers);
- gender (no other quasi-identifiers present);
- year of birth (no other quasi-identifiers present);
- combination of gender and region of residence.

The remaining quasi-identifiers and their combinations are considered as having a sufficiently high level of de-identification risk.

The conceptual scheme of depersonalization is as follows:

1. Determination of the method of dissemination of anonymized data, the circle of persons having access to anonymized data, the most likely persons who will try to re-depersonalize;
2. Determination of quasi-identifiers in the anonymized data array and determination of the presence in open sources of data, which, together with the quasi-identifiers available in the anonymized array, can lead to re-identification;
3. Assessing the risk of re-identification by applying a heuristic approach to the available quasi-identifiers and then applying various anonymization methods to these quasi-identifiers;
4. Depending on the results of applying the above steps, the last steps in the process will be (a) defining a reidentification risk threshold and (b) data cleansing, i.e. extracting personal identifiers and encoding, deleting or randomizing them.

The threshold value of the acceptable risk level determines the minimum level of anonymization to which the PD array must be subjected in order for the resulting array to be recognized as anonymized.

Risk assessment can be carried out on the basis of quantitative (percentage, numerical value from zero to one) or qualitative (“low”, “medium”, “high”) approaches.

The quantitative approach is based on empirical measurement and is therefore more accurate, less subjective and tends to preserve the useful properties of the data to a greater extent. So, since direct identifiers significantly affect the possibility of re-identification, the risk in relation to them is estimated as 100%, or 1.0. Direct identifiers must be anonymized on a mandatory basis so that the level of risk of re-identification is below the threshold.

The re-identification risk for indirect identifiers is assessed at the data subject level. For example, by calculating the cell sizes, which is determined by the number of data subjects in the array that have the same indirect identifier values. The recommended reidentification risk threshold of 0.09 is achieved with a cell size of 11.

The Structured Data Anonymization Guide [32] provides a methodology for calculating contextual risks.

It is necessary to determine the likelihood of three different reidentification attacks or threats:

- deliberate internal attack,
- unintentional re-identification of an individual in a data set by acquaintances,
- data leak.

When assessing contextual risk, the highest of these probabilities should be used.

The likelihood that the recipient of the data will attempt to de-anonymize depends on two factors:

- the degree of control established in the data exchange agreement regarding the confidentiality and security of data,
- the recipient's motives and capabilities to carry out a re-identification attack.

Both of these factors imply a qualitative assessment, resulting in values in the "low", "medium", or "high" range.

To assess the first factor - the degree of control established in the agreement - it is necessary to compare the measures taken with the measures proposed by regulations and recommendations.

The assessment of the second factor - the motives of the recipient - can be carried out taking into account the following points:

- the presence of incidents during the work of the recipient;
- reasons for these incidents;
- whether the recipient has the technical knowledge and/or financial resources to attempt re-identification;
- whether the recipient has access to other private databases or datasets that may be associated with the re-identification data;
- the level of privacy and security controls in the data sharing agreement.

The guide in question proposes the following table for estimating the likelihood of an attack on datasets:

Table A1. Boundary values for the probabilities of reidentification attacks

Privacy and Security Control	The motives and capabilities of the recipient	Probability of a re-identification attack
High	Low	0,05
	Medium	0,1
	High	0,2
Medium	Low	0,2
	Medium	0,3
	High	0,4
Low	Low	0,4
	Medium	0,5

	High	0,6

In addition to a deliberate attack attempt, the recipient of the data may also inadvertently re-identify one or more individuals. The probability of such an "attack" occurring is equal to the probability that a random recipient knows someone from the data set. To calculate it, you can use the following equation:

$$P = 1 - (1 - p)^m \quad (17)$$

where *P* is the percentage of people in the population who have the condition or characteristic in question in the dataset, and m is the number of people a person knows on average. The p-value should be determined based on the latest population statistics. On the other hand, the value of m may vary depending on what kind of relationship with a person is required in order to have information about him regarding the condition or characteristic discussed in the data set.

The third attack to consider is data leakage. The probability of such an attack is equal to the probability of a data security breach on the recipient's objects. To calculate this value, publicly available data on the prevalence of data breaches in the recipient's respective industry should be used.

The total value of the risk of re-identification is calculated by the formula:

$$P_{general} = P_{data\ risk} * P_{context\ risk} \quad (18)$$

Depending on the calculated risk indicator and the need to preserve the usefulness of the data array, various depersonalization methods and their combinations are used.

Russian approach to data depersonalization

The basics of the current regulation of the processing of personal data are determined by the Federal Law of July 27, 2006 N 152-FZ "On Personal Data" [4]. This law contains the main provisions on the concept and types of PD, the principles and conditions for processing PD, the rights of PD subjects and the obligations of PD operators, measures to ensure the security of PD, and also lays the foundations for by-law regulation of this area.

The last update of the Federal Law on Personal Data took place in March 2021, when the Federal Law of December 30, 2020 N 519-FZ "On Amendments to the Federal Law" On Personal Data" came into force. The main innovation was the adoption of a new category of personal data - “personal data authorized by the subject of personal data for dissemination”.

However, the Federal Law on PD practically does not address the issue of data depersonalization, limiting itself to indicating the definition of this term and cases of using the depersonalization procedure:

- depersonalization of processed PD upon reaching the goals of processing or in case of loss of the need to achieve these goals, unless otherwise provided by federal law;
- to process personal data for statistical or other research purposes;
- for the purposes of the Federal Law "On conducting an experiment to establish special regulation in order to create the necessary conditions for the development and implementation of artificial intelligence technologies in the subject of the Russian Federation - the city of federal significance Moscow and amending Articles 6 and 10 of the Federal Law "On Personal Data".

With regard to the direct regulation of PD depersonalization procedures, the following legal acts can be distinguished:

1. Decree of the Government of the Russian Federation of March 21, 2012 N 211 "On approval of the list of measures aimed at ensuring the fulfillment of the obligations provided for by the Federal Law "On Personal Data" and the regulatory legal acts adopted in accordance with it, by operators that are state or municipal bodies" [49].

   Among such measures related to depersonalization, the act prescribes the following measures to be taken in state and municipal bodies:

   – approve the rules for working with depersonalized data in case of depersonalization of personal data;

- approve the list of positions of employees of the state or municipal body responsible for carrying out measures to depersonalize processed personal data, in case of depersonalization of personal data;
- in cases established by regulatory legal acts of the Russian Federation, in accordance with the requirements and methods established by the authorized body for the protection of the rights of personal data subjects, they depersonalize personal data processed in personal data information systems, including those created and operating as part of the implementation of federal target programs.

2. Order of the RKN dated September 5, 2013 N 996 "On approval of requirements and methods for depersonalization of personal data".

The order was adopted in accordance with subparagraph “3” of paragraph 1 of Decree of the Government of the Russian Federation of March 21, 2012 N 211 “On approval of the list of measures aimed at ensuring the fulfillment of obligations stipulated by the Federal Law “On Personal Data” [49] and adopted in accordance with them by regulatory legal acts, operators that are state or municipal bodies" and contains a description:

- properties of anonymized data;
- requirements for the properties of received depersonalized data;
- characteristics of PD depersonalization methods;
- requirements for depersonalization methods;
- requirements for the properties of the depersonalization method;
- the most promising and practical depersonalization methods.

Comparison of approaches to depersonalization

Initially, the Organization for Economic Cooperation and Development (OECD) drew attention to the problem of personal data protection at the international level, which adopted in 1980 the Directive on the Protection of Privacy and International Exchanges of Personal Data. These principles were further detailed in the Council of Europe Convention on the Protection of Persons with regard to Automatic Processing of Personal

Data (1981), in the Directive of the European Community on the Protection of Citizens with regard to the Processing of Personal Information of July 27, 1990, in the Directive of the European Union and Parliament 95/46/EC of 24 October 1995 on the protection of the rights of individuals with regard to the processing of personal data and the free movement of such data and Directive 97/66/EC of 12/15/1997 on the processing of personal data, protection, privacy in telecommunications sector.

In the noted policy documents, the main principles for organizing the processing of personal data and ensuring the right of citizens to the protection of personal data were defined:

- personal data should be collected only for certain purposes and in strict accordance with the law;
- the data must comply with the requirements, be accurate, complete and up to date;
- the purposes for which personal data are collected and processed must be determined and approved before the commencement of activities and used only for these purposes;
- Mechanisms should be introduced in personal data accounting systems to prevent the loss or misuse of personal data;
- the activities of organizations (both public and private) that have databases containing personal data should be open;
- data holders should be subject to control in order to ensure compliance with these principles, for these purposes the creation of an independent supervisory body should be provided as an important element of protecting the identity of the automated processing of personal information.

Consider the regulation of personal data protection rules at the federal and regional levels in various countries and the existence of an authority to monitor compliance with personal data protection requirements. There are three types of legal regulation systems: decentralized, centralized and mixed.

1. Decentralized system:

- lack of a unified approach to the protection of personal data within the framework of industry legislation;
- regulation of personal data protection is carried out through specialized regulations of complex branches of legislation or at different levels of government;
- advisory acts play a significant role;
- lack of a unified supervisory body.

2. Centralized system:
   - direct effect of international norms harmonizing the national laws of states (Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Directive 95/46/EC, Directive 2002/58/EC);
   - the existence of national sectoral laws containing generally binding rules regarding the protection of personal data;
   - regulation of the processing of personal data through the establishment of a single supervisory agency ("mega-regulator"). Examples: EU countries, Israel, Mexico, Hong Kong, Switzerland, Singapore.
3. A mixed system of legal regulation implies the presence of one or more features that make it possible to attribute the system of legal regulation of the protection of personal data of the state to a centralized or decentralized system.

From a practical point of view, of great interest is the Canadian Personal Information Protection Act, which provides for real mechanisms for protecting personal data and exercising the right to access information about oneself.

The main conclusions as a result of the comparative analysis:

- the considered systems are generally similar in terms of addressing personal data protection issues. The Russian approach is more conservative, broadly interpreting the composition of personal data and ignoring a number of security measures;
- all considered approaches support the rights of the subject of personal data on consent to the processing of personal data, as well as the right to be forgotten;

- the composition of personal data is interpreted differently in different legal systems. Thus, technical parameters (online identifiers) such as IP or cookies are personal data in the GDPR and do not apply to those in the Russian Federation;
- The GDPR and other Western laws specifically highlight the issue of data leakage: organizations are required to report such cases, while there are no such requirements in Russian legislation;
- Unlike the Russian legal system, most Western systems have the ability to transfer / transfer data and liability, which makes publication issues and related depersonalization actions more addressable;
- liability for violations under the laws of the Russian Federation is still significantly lower than in the GDPR or PIPEDA (Canada). Western counterparts require significantly higher fines, in addition, they are extra-territorial in nature;
- the use of methods such as pseudonymization, in which a number of direct identifiers are replaced by an abstract identifier, is still not clear in the Russian legal field. For the purposes of the GDPR, personal data using a pseudonym that can be attributed to a natural person after using additional information must be considered as information about a natural person that can be identified;
- in the Russian Federation, in contrast to the compared approaches, there is no categorization according to the level of data sensitivity, which makes it difficult to assess the damage suffered by individuals.

APPLICATIONS B

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023618607

Data Privacy Framework (Система поддержки конфиденциальных данных)

Правообладатель: ***Общество с ограниченной ответственностью "АССОЦИАЦИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ И ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ" (RU)***

Авторы: ***Дик Александр Геннадьевич (RU), Киямов Жасур Уткирович (UZ), Хватов Валерий Владимирович (RU), Щеголева Надежда Львовна (RU)***

Заявка № **2023617657**
Дата поступления **26 апреля 2023 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***26 апреля 2023 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 68b80077e14e40f0a94edbd24145d5c7
Владелец Зубов Юрий Сергеевич
Действителен с 2.03.2022 по 26.05.2023

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

RU2023618607

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2023618607
Дата регистрации: **26.04.2023**
Номер и дата поступления заявки:
2023617657 26.04.2023
Дата публикации и номер бюллетеня:
26.04.2023 Бюл. № 5
Контактные реквизиты:
n.shchegoleva@spbu.ru

Автор(ы):
Дик Александр Геннадьевич (RU),
Киямов Жасур Уткирович (UZ),
Хватов Валерий Владимирович (RU),
Щеголева Надежда Львовна (RU)
Правообладатель(и):
Общество с ограниченной ответственностью "АССОЦИАЦИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ И ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ" (RU)

Название программы для ЭВМ:
Data Privacy Framework (Система поддержки конфиденциальных данных)

Реферат:
Программный комплекс предназначен для анализа защиты персональных данных. Для оценки эффективности защиты обезличенного набора данных реализованы следующие программные модули: основанные на деобезличивании хеш-функцией с солью по заданным начальным условиям (атака методом перебора или со словарем на применяемые хеш-функции), применении метрики k-Anonymity для оценки эффективности обезличенного набора данных (поиск минимального значения количества повторяющихся строк атрибутов), использовании метода Мондриана для определения степени приватности в обезличенном наборе данных (разбиение поступающей информации на прямоугольные области с одинаковым значением метрики анонимизации и последующей обработкой до достижения заданного уровня обезличивания в системах обработки больших данных).

Язык программирования: Python

Объем программы для ЭВМ: 1,8 Мб