

Article

Blockchain Behavioral Traffic Model as a Tool to Influence Service IT Security

Vasiliy Elagin ^{1,*}, Anastasia Spirkina ¹, Andrei Levakov ² and Ilya Belozertsev ¹

¹ Infocommunication Systems Department, The Bonch-Bruевич Saint-Petersburg State University of Telecommunications, 193232 Saint Petersburg, Russia; anastasia.4991@mail.ru (A.S.); ilya.belozertsev@outlook.com (I.B.)

² Center Macro-regional Branch of PJSC Rostelecom, 123298 Moscow, Russia; levakov1966@list.ru

* Correspondence: elagin.vas@gmail.com

Received: 25 March 2020; Accepted: 14 April 2020; Published: 15 April 2020



Abstract: The present article describes the behavioral model of blockchain services; their reliability is confirmed on the basis of experimental data. The authors identify the main technical characteristics and features associated with data transmission through the network. The authors determine the network scheme, working with blockchain transactions and the dependence of network characteristics on application parameters. They analyze the application of this model for the detection of the blockchain service and the possibility of the existing security mechanisms of this technology being evaded. Furthermore, the article offers recommendations for hiding the blockchain traffic profile to significantly complicate its identification in the data network.

Keywords: blockchain; distributed registry; security; data protection; network; decentralized systems

1. Introduction

Today, with the introduction of innovative technologies and various applications, blockchain technology is gaining an increasing importance in modern communication networks due to its technical capabilities.

Blockchain is a distributed database that includes an ever-growing organized list of records and data storage devices connected to different servers [1,2].

Blockchain functionality is obvious in all areas of data storage and authentication. This decentralized data system has the potential to eliminate corruption and can be helpful in tackling fraud. Anyone can access information posted on the internet by another person from anywhere in the world.

This stage of technological development contains both benefits and pitfalls, which are listed below [1,3].

Benefits include decentralization—network members are equal and data can be directly exchanged; reliability—any attempt at unauthorized changes will be rejected due to inconsistency with previous copies; compromise—data added to the system will be checked by other participants; transparency—any part of the transaction can be checked, and theoretically, it can be supplemented by new records an infinite number of times; and confidentiality—data are stored in an encrypted form and the user can track all transactions but cannot identify users that receive or send information.

The pitfalls include fraud and errors—blockchain data are irreversibly transmitted (wrong actions cannot be reversed), low transaction speed—a significant weakness of blockchain technology when it is used as a base for digital currencies, and the possibility of illegal operations.

2. Related Works

Due to the fact that blockchain has become popular due to cryptocurrencies, many studies are devoted to its analysis in terms of financial gain. For example, the article [4] presents the results of using blockchain technology, thanks to which we could be able to handle financial processes in a more efficient way than under the current system. This document discusses the problems with and possibilities of introducing blockchain technology in the banking and capital markets. In addition, many banks are experimenting with blockchain technology, betting on its ability to contribute to economic growth through free trade.

Today, blockchain research is the most relevant topic. Many scientists and tech companies analyze current problems and how to fix them. Therefore, in [5], the idea that blockchain is a highly secure storage medium and presents a technological quantum leap in maintaining data integrity is discussed. In this article, the authors show that blockchain as a medium with a high degree of protection represents a technological qualitative leap in maintaining data integrity. In addition, the immutability of the blockchain creates a fertile environment for the combination of AI and blockchain, which can affect areas such as the Internet of Things (IoT), financial markets, smart cities, supply chains, and other areas that will benefit society. In the article [6], the author applied blockchain-based smart contracts to bill resources. The results show that IoT had stable and reliable system performance with a certain data size and concurrency scale. In turn, in the article [7], the authors analyze the advantages and disadvantages of the technology and also discuss applications in the insurance sector, which can be easily extended to other areas. Currently, a number of use cases and prototype solutions have been developed in this sector. In particular, blockchain and smart contracts can be successfully used to speed up the processing of requirements and reduce operating costs. However, smart contracts are often attacked, which can lead to negative consequences. Blockchain technology offers to improve the exchange of information and data between objects. By creating trust without the need for a central authority, this is a technological breakthrough. However, the main aspects are security and ensuring the quality of service delivery. In the article [8], the authors examined the existence of vulnerabilities in blockchain technology and the consensus protocol based on PoW (Proof-of-work). Such vulnerabilities represent various security threats for standard functionality. The authors then examine the reliability of modern security and anonymity solutions, as well as analyzing existing solutions for maintaining confidentiality. However, blockchain technology has already been successfully applied to improve detection systems to protect against traditional cyber security attacks [9]. For example, Q. Xia et al. [10] report the implementation of blockchain technology for operations when working with the medical data of patients, ensuring minimal risks to data confidentiality.

In [11], the authors describe a decentralized personal data management system that provides users with ownership and control of their data. The authors propose a system based on blockchain technology in which users are not required to trust any third parties and are always aware of the data that is collected about them and how they are used. In this article [12], the authors examined the problem of ensuring transaction security in decentralized trading in intelligent systems without the use of trusted third parties. The authors conducted studies to analyze security and evaluate performance in the context of identified security and confidentiality requirements.

This study proposes an alternative approach to existing work to ensure data protection using blockchain technology. Our approach defines the network scheme for working with blockchain transactions and the dependence of network characteristics on application parameters. It analyzes the use of this model to discover the blockchain service and the possibility of discrediting the existing security mechanisms of this technology. In this case, it becomes possible to determine blockchain technology traffic based on behavioral analysis and to separate it from the traffic of other applications. We simulated network traffic to test various dependencies during the operation of the blockchain technology.

In this approach, it is proposed to rely on the characteristics of traffic behavior, including the lengths of packets and delays when transmitting packets with block data. Our approach has also been developed by taking into account widespread use by telecom operators and application developers.

In addition, the article offers recommendations for hiding the blockchain traffic profile so as to greatly complicate its identification in the data network.

3. Technical Aspects of Blockchain Technology

Security in blockchain technology is provided through a decentralized server that establishes peer-to-peer network connections. As a result, there is a new database that is managed independently without any center. After synchronizing with other network nodes, all transaction records are saved. The integrity and chronological order of transactions are preserved by using cryptographic rules [13]. Once nodes have been loaded, they perform a peer-to-peer detection in order to communicate with other valid nodes by means of TCP (Transmission Control Protocol) ports. The node is designed to maintain the network, maintain and update a copy of the blockchain, and process transactions [14].

The data exchange procedure within the blockchain technology includes a number of messages transmitted according to certain rules. An example of data exchange process is shown in Figure 1.

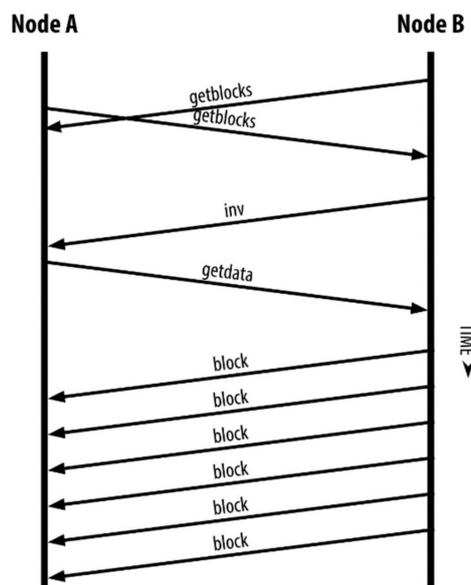


Figure 1. An illustrative example of node synchronization.

The main types of message used in the data exchange process [13] are version (to indicate the version of the node), verack (response to the version message), addr (to provide information about known nodes), getaddr (to request information about known active peers), getblocks (to return inv containing a list of blocks), inv (to distribute information about objects), getdata (to get the contents of an object), and block (response with transaction information from a block hash).

Figure 1 shows an example of the synchronization of nodes in a network. This graph shows the behavior of the protocol in the communication network.

The block is transmitted to the network by traditional TCP/IP technology. This block is a “container” that merges transactions for including them into the public registry. The block consists of a header containing metadata and a body from the list of transactions.

Figure 2 represents the algorithm of the blockchain technology operation with transactions, which adheres to the protocol in the communication network.

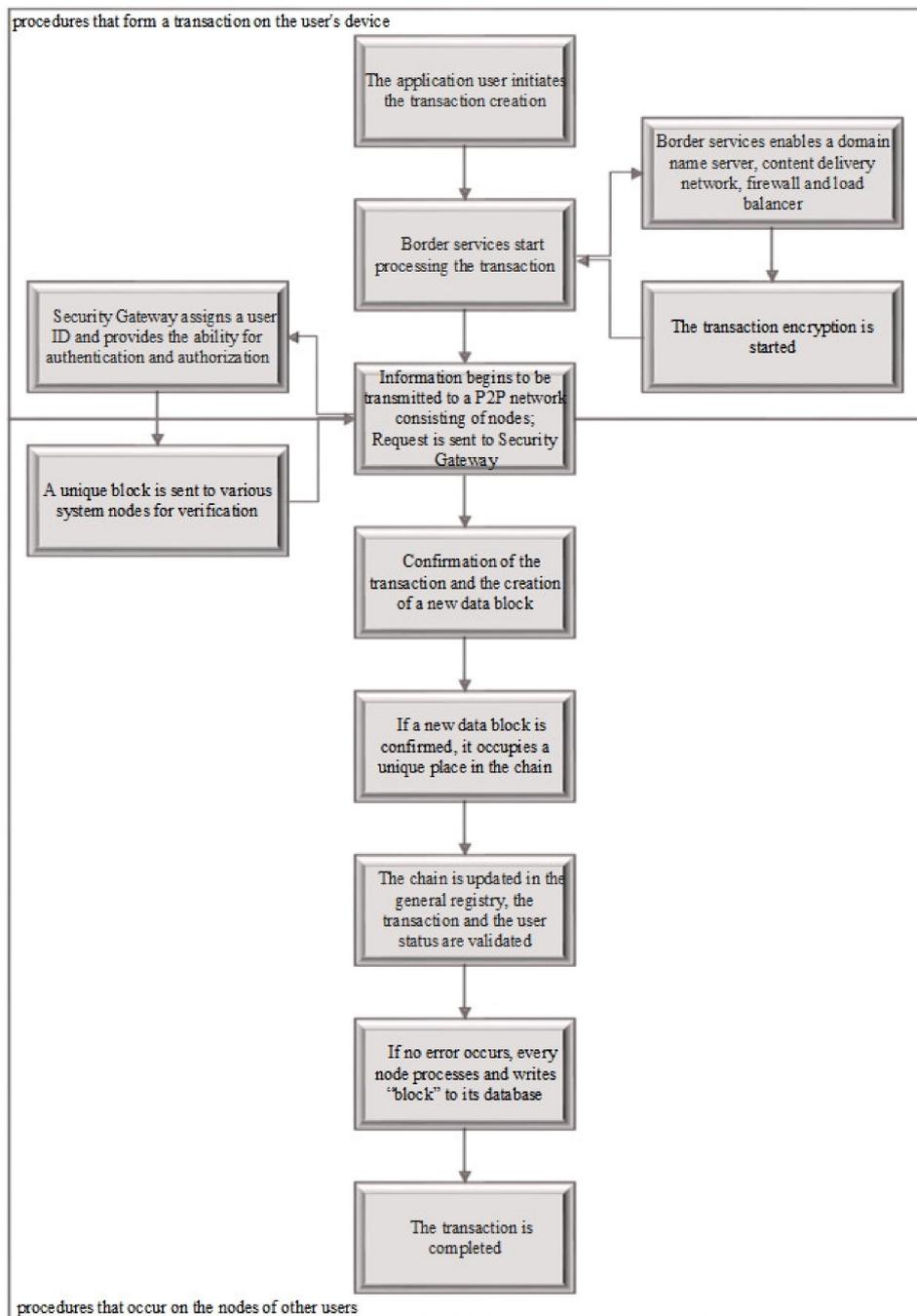


Figure 2. The algorithm of a blockchain technology operation with transactions.

The algorithm presented in Figure 2 allowed a phased analysis of the traffic received and the results on which the experiment presented in the study is based.

Blockchain technology uses cryptography to protect the identities of users, to ensure the security of transactions, and to guarantee that users change only those parts of blockchain to which they have private keys [15,16].

Blockchain cryptographic primitives are divided into two categories: primary and secondary primitives. The first category includes hash and standard digital signatures that are necessary to ensure protection against unauthorized access, public verification and reaching consensus. The second category is mainly used to increase the confidentiality and anonymity of transactions.

By means of the hash function, the data are protected against modifications until the intended user receives them, and the determination of the size of the encrypted information is prevented. Additionally, this function performs a description of all transaction changes, which definitively increases data security and eliminates the risk of previous data modifications, because one separate block cannot be modified without recalculating all of the subsequent ones, which requires many calculations; therein lies the security of blockchain technology. When one variant is found, the node sends the received block to other connected nodes for verification. If no error is detected, the block is considered to be added to the chain and the next block should include its hash.

A digital signature is a special document sign that allows the verification of the data authenticity in electronic documents and the confirmation of owner authenticity.

When the private key is used to sign transactions, the public key verifies the authenticity of other people's transactions. Since blockchain technology does not possess a central node that can authorize arbitrary transactions, the system security becomes decentralized, and the probability of successful technological intervention is minimized.

4. Network Attacks

Real networks make it possible to carry out various network attacks due to the incomplete security of user actions. A network attack is an action which is aimed at seizing control (elevation of privileges) over a remote/local computer system, its destabilization, denial of service, or receiving user data from the computer system.

Currently, the following attacks are distinguished: mailbombing, buffer overflow, the use of specialized programs, network intelligence, IP spoofing, man-in-the-middle, injection, denial of service attacks, and phishing attacks [15–18].

However, threats to distributed registries slightly differ from threats to typical computer networks. The main attacks, to which Blockchain systems are exposed, are as follows [15–18]:

- “Attack 51%”. Essentially, a hacker can type his/her own chain of blocks that will overtake the main one by controlling more than half of the confirming resources of the blockchain network. As a result, this chain becomes the main one. This allows including only one's own data in the chain. This attack can also be implemented at lower capacities, but, in this case, the potential success is sharply reduced. At the early stages of blockchain development, the network was vulnerable to this attack. Today, to attack the developed networks, the processing power would have to be many times higher than the power of all “supercomputers”.
- Sybil attack. A sybil attack implies a situation when one network node acquires several entities. A hacker may refuse to transmit and receive blocks by “disconnecting” users from the network or see all transactions by means of special programs. Sybil attacks are usually excluded by using a set of heuristic rules, contacting a trusted certification server or by restricting the number of accounts able to be created by one user in a given period of time.
- DDoS. DDoS is another type of hacker attack that is aimed at resending a huge number of similar requests. To prevent this attack, many blockchain systems use techniques based on limiting the block size, limiting the number of signature verifications, and blocking suspicious transactions or behavior.
- Cryptography hack. Algorithms for calculating the hash functions of SHA-256 and ECDSA standard are considered to be impossible to hack at current computing powers. However, new high-performance quantum computers will increase the risk of hacking these functions. In this case, the system hash function should be replaced with a more complex one.

Blockchain traffic is encrypted, but the analysis of the behavioral model allows the identification of data flows, with the possibility of further forecasting its impact on services.

5. Network Characteristics of Blockchain Technology

The network interaction process and traffic technical characteristics need to be identified to analyze the behavioral model [13,19]. It is also necessary to review the impact on the network due to the large number of transactions, because during the exchange process, the blockchain generates additional traffic to update the registries on all involved nodes, and there is an increased volume of service traffic that occurs when data is encrypted, which significantly reduces the proportion of useful traffic [20].

There are technologies that allow the capturing and analysis of traffic for subsequent attacks. Deep Packet Inspection (DPI) technology is one of them. Deep Packet Inspection is intended for deep packet analysis, which inspects not only headers of packets at different levels but also the data content.

DPI analysis depends on the following traffic identification instruments [3]:

- Clearly defined rules. The system administrator sets rules and policies by means of a full or partial activation of desired rulers and policies from kits provided by the software developer.
- Signature analysis. Signature analysis is an analysis where the system searches the packet structure and compares it with known cases. Simply put, this is a set of packet bytes intended for clearly defining which application or protocol the traffic belongs to and classifying it. DPI systems use scanning instruments based on the fact that all known protocols have their exact signatures. If a match is detected, the package is regarded as recognized and becomes available to other programs.
- Heuristic analysis. Heuristic analysis is a technology designed to detect traffic by concrete features (without guaranteed accuracy). This method is responsible for the incomplete but very close matching of packets with known signatures. The main advantage of this method is that it allows the detection of the desired traffic even before the signature is updated for it.
- Traffic behavior analysis. Traffic behavior analysis is regarded as one of the most promising analysis techniques in DPI traffic control systems, as it can define almost any traffic behavior model with high processing speed and a high accuracy of traffic identification. This technique increases the ability to identify applications with encrypted payloads.

To identify traffic by means of a behavioral analysis, a certain interval of time is tracked, during which traffic is transmitted. On the basis of the traffic behavior analysis performed during this interval and by comparison with the database, a decision is made.

The main characteristics for identifying a stream are the sequence of the sizes of transport layer segments, the sequence of the data chunk sizes, and size.

Deep Packet Inspection is designed to determine which application has generated or received data and take some action based on this data. General traffic identification can be reached by means of sharing with other analysis techniques.

What happened earlier on one device is now duplicated among all nodes, so the number of transmitted messages is higher (Figure 3). These data are transmitted in small portions but in a short period of time, which leads to an unexpected burst of transmitted messages.

Network load (ρ) will depend on:

$$\rho \sim F(n, \alpha_n, d, \lambda_n, \mu, k_i) \quad (1)$$

where:

n stands for the number of nodes in the blockchain network (units);

α_n stands for the intensity of formatted transactions (transactions per second);

k_i stands for the number of routers involved (units);

μ stands for the packet processing intensity of routers (packets per second);

λ_n stands for the intensity of packet formation (packets per second);

d stands for the block size (bytes).

Consequently, when forming and confirming a transaction, the blockchain provokes an avalanche of network load.

Blockchain technology affects the network load (Equation (1)), but network characteristics affect α_n , i.e., block formation, since delivering a packet with a lower delivery class can result in the retransmission of the packet and the changing of the order of transaction formation. This may lead to a leak of resources, which in return will increase the likelihood of packet loss and increase delays [3,20].

A detailed dependency analysis presented in Equation (1) will make it possible to estimate how each node loads the network and how a certain growth affects the network characteristics that are necessary for network quality [20–22].

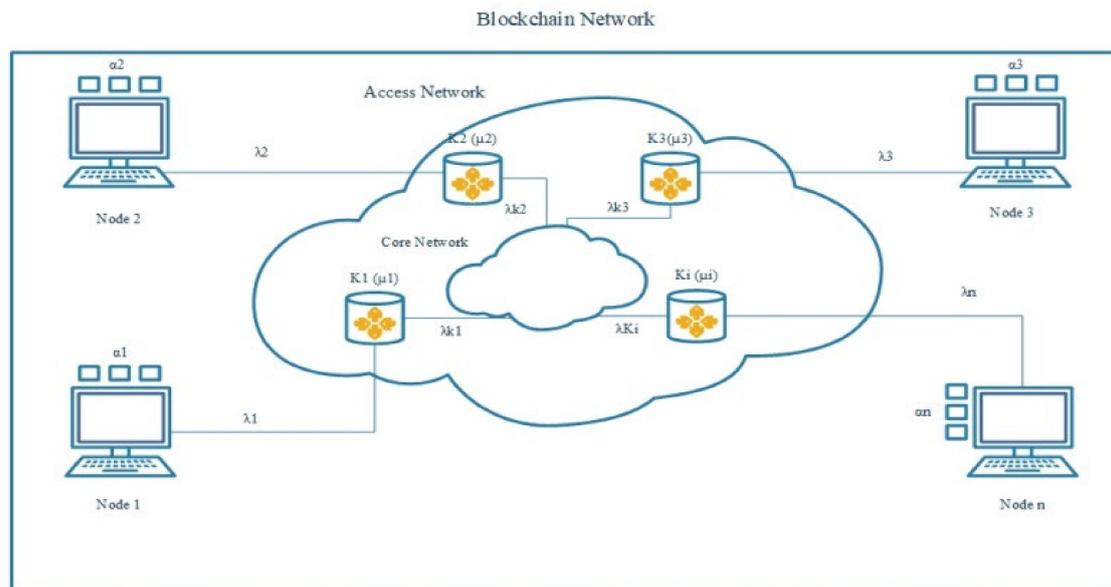


Figure 3. A blockchain traffic flow diagram.

6. Traffic Analysis of Blockchain Technology

To analyze the traffic behavior on the network, five virtual Linux Ubuntu 18.04 LTS -based clients were created. These clients were connected with each other and also with an external network, and were given minor and transaction initiator roles.

The generalized experiment design is presented in Figure 4.

In the course of experiments, virtual clients sent transactions to a similar client with a frequency of four transactions per second. The experiments were carried out at different times of the day with the same sequence of actions.

The data were obtained within 50 tests and processed by the mathematical tool of statistical analysis.

The results presented in diagrams of the packet traffic load (Figure 5, Figure 6, and Figure 7) reflect that, regardless of the time of day, traffic is transmitted by similar distributions with slight variation. It is a traffic stream with avalanche-like bursts of intensity.

In addition, diagrams that determine the distribution dependence of the packet number on the packet size and frequency distribution of time intervals between packets are of special interest.

The diagrams that determine the dependence of the packet number on the packet length (Figure 8) represent that most of the packets (85%) are between 65 and 250 bytes and include traffic transmitted during data synchronization between nodes. We can determine the traffic type (9%) transmitted within transactions with a length of over 1100 bytes.

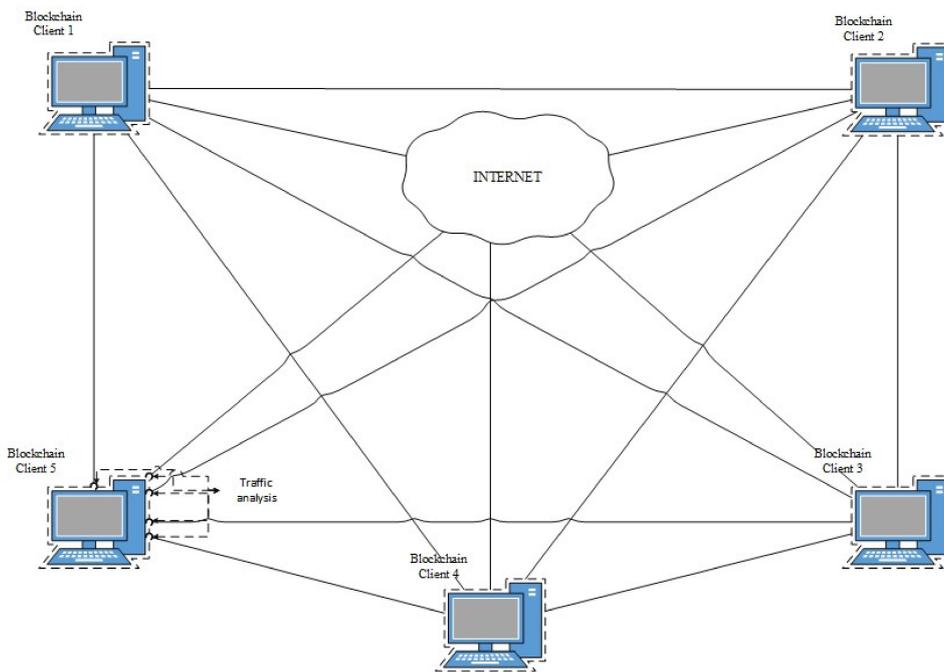


Figure 4. The generalized experiment design.

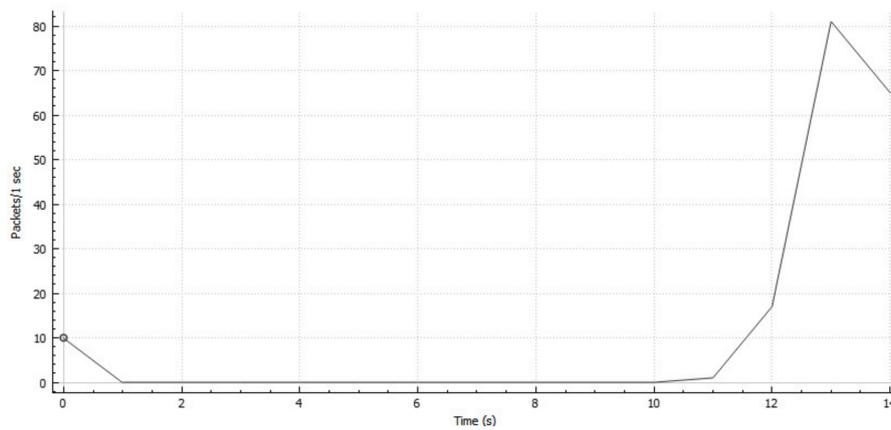


Figure 5. A diagram of the packet traffic load upon initial synchronization with the network.

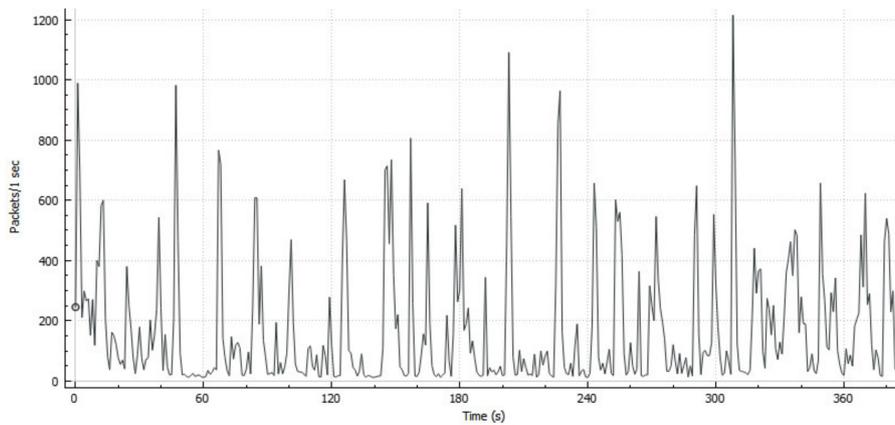


Figure 6. A diagram of the packet traffic load when synchronizing a node with neighbors and mining.

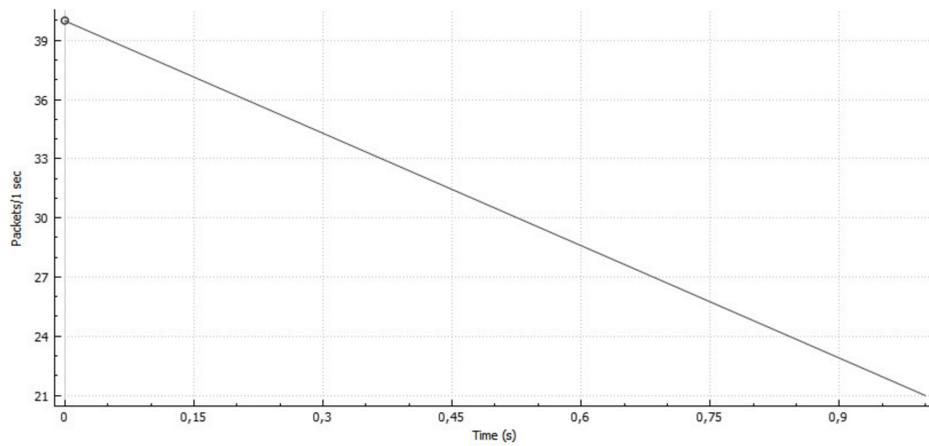


Figure 7. A diagram of the packet traffic load after synchronization and mining.

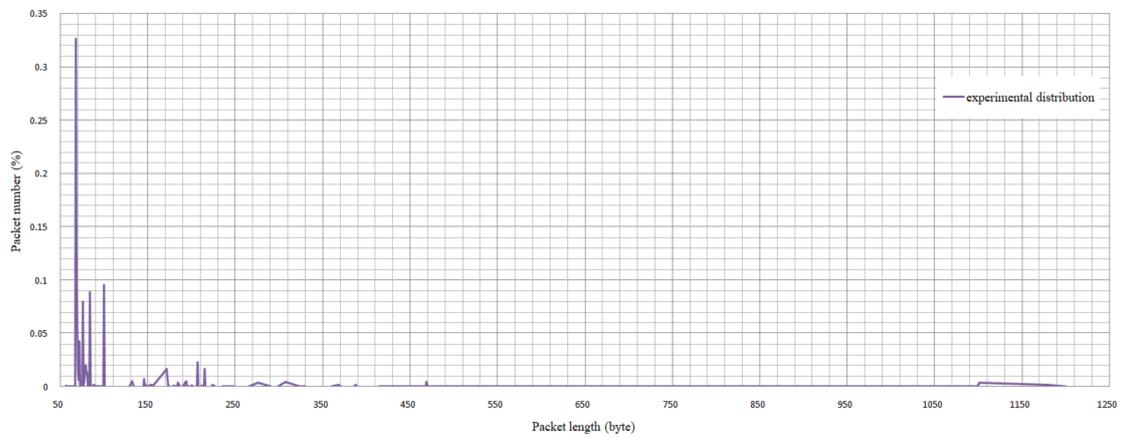


Figure 8. A diagram of the dependence of the packet number (%) on the packet length.

When analyzing the ratio of packet number to the time between packets, it can be noted that there is dependence as shown in Figure 9, but, in turn, the delay for most packets did not exceed 0.2 ms in any experiment.

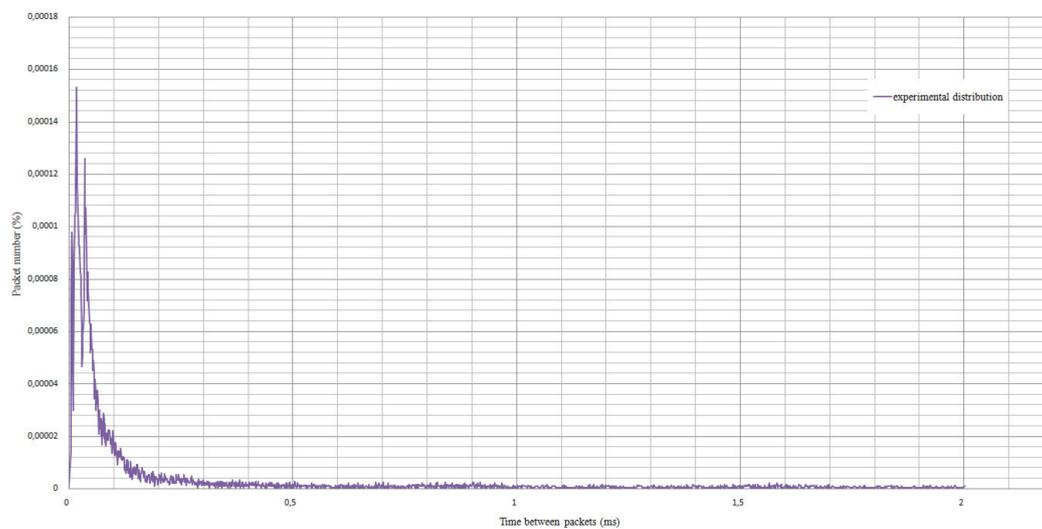


Figure 9. A diagram of the ratio of the packet number (%) to the time between packets (ms).

Subsequently, the study included a statistical analysis of the received data. The following are true for both dependencies:

- As the variation coefficient is over 70%, the totality is getting closer to the verge of heterogeneity.
- The As and Ex values are close to zero. Therefore, it can be assumed that this sample is close to the log-normal distribution.
- Testing the hypothesis by Pearson’s test shows that there is no reason to reject the hypothesis of log-normal distribution.
- Thereby, these dependencies can be represented in the context of the mathematical model of a logarithmic normal distribution with different weighting coefficients. The diagrams are presented in Figures 10 and 11.

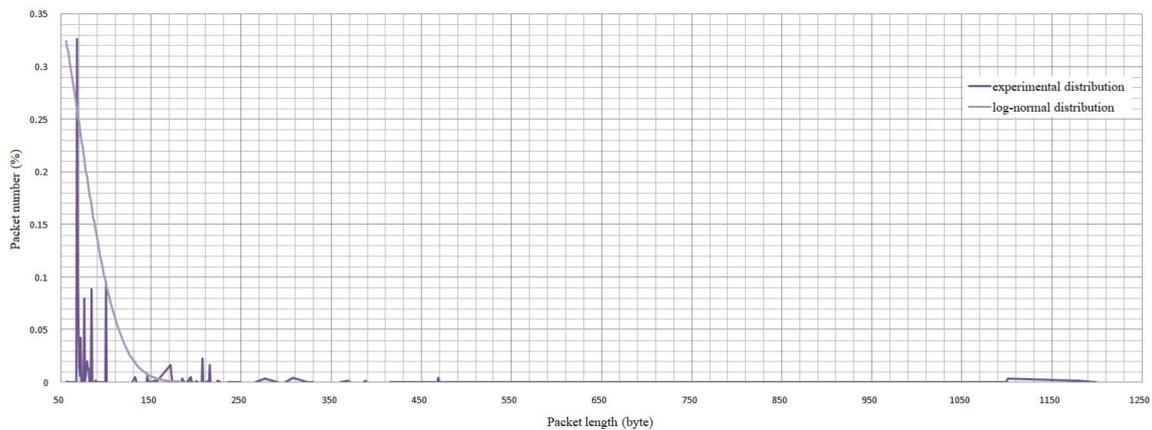


Figure 10. Comparison of the obtained results with the values obtained in the log-normal distribution.

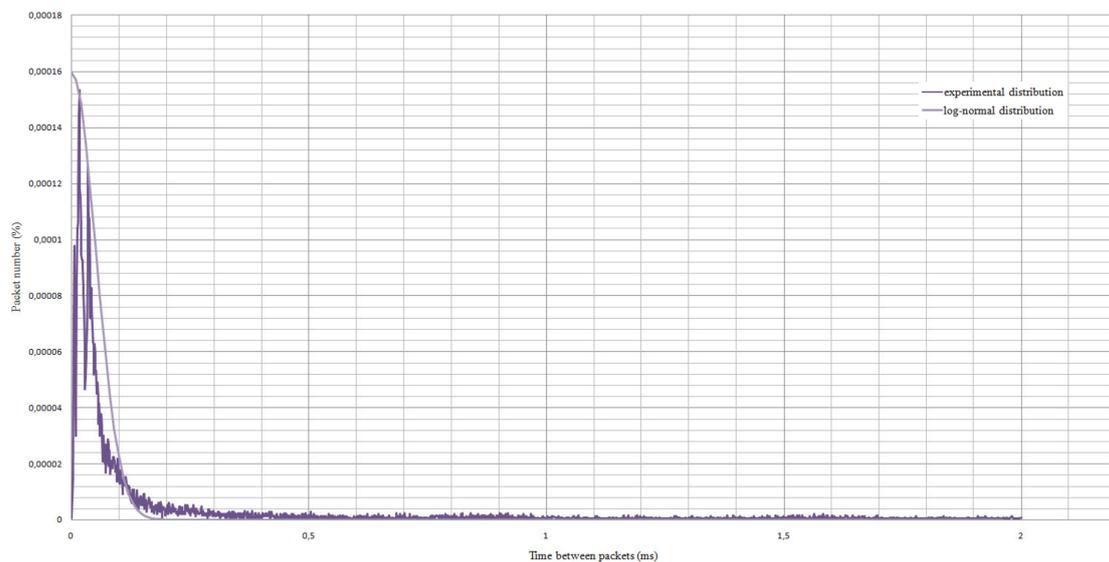


Figure 11. Comparison of the obtained results with the values obtained in the log-normal distribution.

Statistical analysis showed that the individual characteristics of blockchain traffic can be modeled by means of the closeness of the distribution to a normal one, allowing the use of analysis model data to identify blockchain traffic with its forecasting.

Therefore, to hide traffic, the characteristics, presented in the study, need to be dynamically changed. To change the distribution dependence of the number of packets on their size, the length of blocks can be artificially varied.

The main solutions for these kind of changes are software utilities or manual data transmission from the device. To change the dependence of the frequency distribution of time intervals between packets, artificial time shifts between packet transmissions, as well as changes in the packets' size and algorithm of transmission, should be considered.

7. Conclusions

Currently, the problem of ensuring data security is not completely resolved. In their study, the authors examined the possibility of identifying traffic to blockchain technology using behavioral analysis. The task of preventing the circumvention of the existing security mechanisms of this technology has not been completely achieved; however, network traffic was simulated to build dependencies during the operation of the blockchain technology. The authors identified the main network characteristics by which traffic identification can be made. The authors offer recommendations for hiding the blockchain traffic profile so as to greatly complicate its identification in the data network.

Despite the blockchain technology encryption, it can be identified in a flow with the help of deep packet analysis solutions that include behavioral analysis, which will further allow breaches and the use of services of this technology for nefarious purposes.

In this article, the authors analyzed the main characteristics of blockchain traffic and presented analysis models that allow the prediction of the on-network load, and the findings should help to ensure high-quality and safe data exchange on this part of the network in the future.

Under this experiment, traffic analysis was carried out within 3 min sessions, which turned out to be long enough to determine the behavior of blockchain technology traffic. The experiment showed that the distribution dependence of the number of packets on their size and the dependence of the frequency distribution of time intervals between packets are close to log-normal. Overall, it is recommended to hide blockchain technology traffic by means of various solutions. However, in the future, further work is to be done to develop recommendations for hiding traffic and preventing the interception of traffic from this technology, including by behavioral analysis.

Author Contributions: Conceptualization, V.E.; Formal analysis, A.S.; Methodology, A.L.; Software, I.B.; Writing—original draft, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by RFBR, project number 19-37-90050/19 (MOSCOW, RUSSIA).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mougayar, W. *The Business Blockchain*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2016.
2. Sun, Y.; Zhang, L.; Feng, G.; Yang, B.; Cao, B.; Imran, M.A. Blockchain-Enabled Wireless Internet of Things: Performance Analysis and Optimal Communication Node Deployment. *IEEE Internet Things J.* **2019**, *6*, 5791–5802. [[CrossRef](#)]
3. Goldstein, A.B.; Sokolov, N.A.; Elagin, V.S.; Onufrienko, A.V.; Belozertsev, I.A. Network Characteristics of Blockchain Technology of on Board Communication. In Proceedings of the 2019 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 20–21 March 2019; pp. 1–5.
4. Cocco, L.; Pinna, A.; Marchesi, M. Banking on Blockchain: Costs Savings Thanks to the Blockchain Technology. *Future Internet* **2017**, *9*, 25. [[CrossRef](#)]
5. Sgantzios, K.; Grigg, I. Artificial Intelligence Implementations on the Blockchain. Use Cases and Future Applications. *Future Internet* **2019**, *11*, 170. [[CrossRef](#)]
6. Li, Y. An Integrated Platform for the Internet of Things Based on an Open Source Ecosystem. *Future Internet* **2018**, *10*, 105. [[CrossRef](#)]
7. Gatteschi, V.; Lamberti, F.; Demartini, C.; Pranteda, C.; Santamaría, V. Blockchain and Smart Contracts for Insurance: Is the Technology Mature Enough? *Future Internet* **2018**, *10*, 20. [[CrossRef](#)]
8. Conti, M.; Sandeep Kumar, E.; Lal, C.; Ruj, S. A Survey on Security and Privacy Issues of Bitcoin. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 3416–3452. [[CrossRef](#)]

9. Meng, W.; Tischhauser, E.W.; Wang, Q.; Wang, Y.; Han, J. When intrusion detection meets blockchain technology: A review. *IEEE Access* **2018**, *6*, 10179–10188. [[CrossRef](#)]
10. Xia, Q.; Sifah, E.B.; Asamoah, K.O.; Gao, J.; Du, X.; Guizani, M. MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain. *IEEE Access* **2017**, *5*, 14757–14767. [[CrossRef](#)]
11. Zyskind, G.; Nathan, O.; Pentland, A. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In Proceedings of the 2015 IEEE Security and Privacy Workshops, San Jose, CA, USA, 21–22 May 2015; pp. 180–184.
12. Aitzhan, N.Z.; Svetinovic, D. Security and Privacy in Decentralized Energy Trading Through Multi-Signatures, Blockchain and Anonymous Messaging Streams. *IEEE Trans. Dependable Secur. Comput.* **2018**, *15*, 840–852. [[CrossRef](#)]
13. Antonopoulos, A.M. *Mastering Bitcoin*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2017.
14. Sharma, P.K.; Chen, M.; Park, J.H. A Software Defined Fog Node Based Distributed Blockchain Cloud Architecture for IoT. *IEEE Access* **2018**, *6*, 115–124. [[CrossRef](#)]
15. Buinevich, M.; Izrailov, K.; Vladyko, A. Testing of utilities for finding vulnerabilities in the machine code of telecommunication devices. In Proceedings of the 19th International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea, 19–22 February 2017; pp. 408–414.
16. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine Learning and Deep Learning Methods for Cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381. [[CrossRef](#)]
17. Salah, K.; Rehman, M.H.U.; Nizamuddin, N.; Al-Fuqaha, A. Blockchain for AI: Review and Open Research Challenges. *IEEE Access* **2019**, *7*, 10127–10149. [[CrossRef](#)]
18. Guan, Y.; Ge, X. Distributed Attack Detection and Secure Estimation of Networked Cyber-Physical Systems against False Data Injection Attacks and Jamming Attacks. *IEEE Trans. Signal Inf. Process. Over Netw.* **2018**, *4*, 48–59. [[CrossRef](#)]
19. Goldstein, A.B.; Zarubin, A.A.; Onufrienko, A.V.; Elagin, V.S.; Belozertsev, I.A. Synchronization of delay for OTT services in LTE. In Proceedings of the 2018 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Minsk, Belarus, 4–5 July 2018; pp. 1–4.
20. Elagin, V.S.; Belozertsev, I.A.; Goldshtein, B.S.; Onufrienko, A.V.; Vladyko, A.G. Models of QOE ensuring for OTT services. In Proceedings of the 2019 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 20–21 March 2019; pp. 1–4.
21. Makolkina, M.; Koucheryavy, A.; Paramonov, A. Investigation of Traffic Pattern for the Augmented Reality Applications. *Lect. Notes Comput. Sci.* **2017**, *10372*, 233–246.
22. Topór-Kamiński, T.; Krupanek, B.; Homa, J. Delays models of measurement and control data transmission network. *Stud. Comput. Intell.* **2013**, *440*, 257–278.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).