Blockchain Systems Review and Analysis for Information Security of Big Data

Igor Zakharov
GeoAnalytics
C-CORE
Ottawa, ON, Canada
igor.zakharov@c-core.ca

Jonathan Anderson
Electrical and Computer Engineering
Memorial University
St. John’s, NL, Canada
jonathan.anderson@mun.ca

Garrett Parsons
GeoAnalytics
C-CORE
Ottawa, ON, Canada
garrett.parsons@c-core.ca

Michael Henschel
GeoAnalytics
C-CORE
Ottawa, ON, Canada
michael.henschel@c-core.ca

***Abstract.* Data quantities are rapidly increasing in many industry sectors due to the development of new sensors, mobile and cloud technologies, advancements in IoT and AI, and growth of social and entertainment media. Many applications (e.g. in finance, healthcare, government data) have strict information security requirements for simultaneous access, record updating, and validation in an immutable manner, which can be achieved with distributed ledger technology (DLT). In this paper we review and analyze fifty-eight currently available blockchain (BC) systems and their components in context of the DLT for big data storage, provenance tracking, replication, and sharing. We also elaborate the key BC system components and architecture for major information security concerns.**

***Keywords:* information security, DLT, blockchain, big data**

I. Introduction

Data volumes are rapidly increasing across different industries and governments due to the development of new sensors, mobile and cloud technologies, advancements in IoT and AI, and growth of social and entertainment media. The amount of data created, captured, copied, and consumed within an organization can reach volumes on the order of 100s of terabytes to multi-petabytes [1] with millions and even billions of records approaching to big data scales (beyond computing power of modern data centers).

Information security measures to protect data from internal and external cyber threats include access control (authentication and authorization), data integrity maintenance, encryption and digital signatures, and monitoring [2, 3]. Beyond these traditional measures, centralized digital ledger recordkeeping enables verifiable transaction logs to be created and maintained by a central authority to simplify auditing mechanism [4].

In contrast to the centralized ledger, the distributed ledger is a decentralized database which is synchronized and accessible across different users (nodes) on a network. In multiple applications, such as finance, legal services, medicine, earth observation etc., the distributed ledger technology (DLT) can be used to meet requirements in secured transactions, simultaneous data access, validation, record updating and storing in an immutable manner [5, 6]. The immutability helps DLTs access a different point in the consistency-availability-protection (CAP) trade-off space [7, 8] than the traditional solutions [9]. In addition to the key mechanisms, such as, database, consensus algorithm, peer-to-peer (P2P) network and transactions logging, the main DLT features also include [10]:

- immutability (new data can only be appended, but not changed or deleted), and
- cryptography.

Distributed ledgers enable to form and maintain consensus about the existence, status and evolution of a set of shared facts [11]. The major differences between distributed ledgers and traditional distributed databases (distributed across sites in a network) is the use of an adversarial threat model and a very different view of authority.

Blockchains (BCs) can be considered to be a subset of distributed ledgers that share the same adversarial threat model (assuming that not all nodes are trustworthy) over a P2P network and have additional characteristics, such as [10], linked blocks which chronologically stored transactional data.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In general, the implementation of a custom, private permissioned blockchain from scratch is a difficult task [12] and therefore a comprehensive survey of currently available BCs is essential to achieve advanced performance. The goal of our paper is to review and analyze the currently available BC systems in context of the DLT component configuration for big data management. The novelty of the paper also lies in the elaboration of possibilities for BC implementation for information security of big data.

II. Blockchain

A. BC Surveys

Numerous publications provided comprehensive and systematic literature reviews of BC technology analyzing its concept, architecture, components and implementation focusing on its:

- evolution and architecture [13, 14],
- performance evaluation [15],
- identity management [16],
- post-quantum cryptography [17],
- use by cryptocurrencies [18],
- industrial IoT applications [19],
- healthcare application [20],
- data management [21],
- engineering and manufacturing applications [22],
- supply chain management applications [23],
- connection with cloud computing [24],
- connection with communication networks [25],
- and other numerous applications [26–30].

A survey on BCs for several smart applications (city, healthcare, transportation and grid), which can generate big data [32], discussed approaches, opportunities, challenges and future directions for secure big data acquisition, data storage, data analytics, and data privacy preservation. BC architecture for massive data storage was recently analyzed and implemented [31] to improve scalability and performance. Various aspects of BC-enabled cyber-physical systems, including security, privacy, immutability, fault tolerance, interoperability, data provenance, atomicity, automation, data/service sharing, and trust were reviewed in [33]. BCs can be categorized as permissionless and permissioned (requiring authorization). Based on the usage and ownership the BCs can be divided [29] into public, private and consortium (to record cross-organizational transactions).

B. Block Structure

The block structure includes a block header and block body [29]. The block header specifies the metadata, including various fields:

- hash (ID) of previous block (to connect its previous block called a parent block),
- hash of current block,
- timestamp (creation time of the block),
- Nonce (relates to consensus mechanism for validation),
- Merkle root (to store the transactions for efficient data verification).

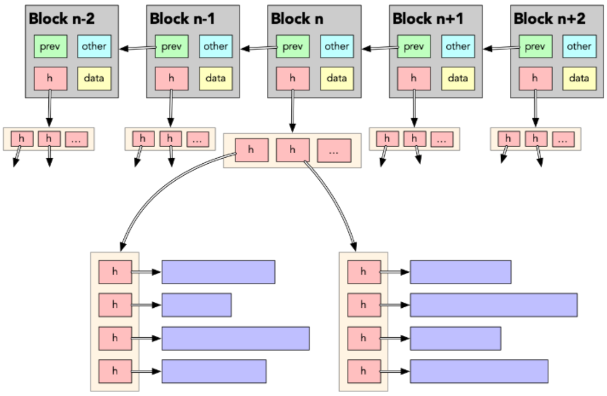

Fig. 1. A typical BC structure

The block's header may also contain other information, for example, block version (software/protocol version), nBits (target threshold of a valid block hash) [34] and confirmation [25]. The block body (also called as block data) stores transactions (work process resulting in a state change) which are assembled using cryptographic functions. All performed transactions (e.g. transfers of money) are hashed and hash values are structured into a Merkle DAG. The linked blocks form a BC (Fig. 1).

C. BC Architecture

The three main components which enable BC technology are a decentralized P2P network, distributed consensus and cryptographically secure algorithms [30]. Work [11] highlights importance of ledger and validity rules (when transactions are considered valid and how the ledger gets updated). The basic BC architecture distributed over P2P network is shown in Fig. 2.

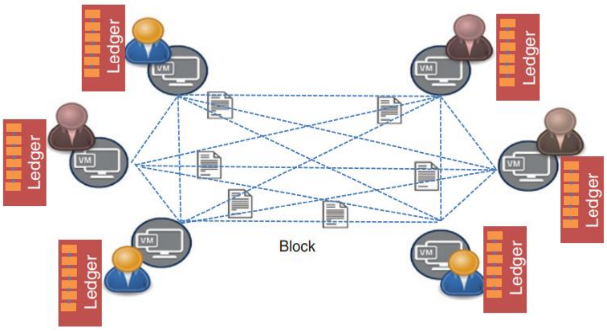

Fig. 2. Blockchain architecure. Adapted from [30]

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A smart contract is a computer program deployed using cryptographically signed transactions in the distributed ledger (BC network), in which parties agree to be bound by the program's output.

D. Consensus Mechanisms

Consensus algorithms enable agreement among decentralized nodes before a block is included into the blockchain. Since, BC networks are designed to function with no trusted central node [29], there are several major consensus mechanisms: Proof of Work (PoW), Proof of Stake (PoS), Practical Byzantine Fault Tolerance (PBFT), and Delegated Proof of Stake (DPoS). However, BFT still requires either a central authority to identify a BFT node (i.e., permissioning) or PoW to maintain overall consensus. Xie et al. [30] review more consensus mechanisms: Proof of Activity (PoA), Proof of Elapsed Time (PoET), Proof of Luck (PoL), and Proof of Space (PoSpace). Other mechanisms were comprehensively analyzed in different works (e.g. [35]). These include: Multi-signature (majority rule), delayed Proof of Work (DPoW), Proof of Importance, Delegated Proof of Stake (DPoS), Delegated Byzantine Fault Tolerance (DBFT), Partioned Consensus, Smilo BFT+, Stellar Consensus Protocol, Proof of Stake Velocity (PoSV), Proof of Burn (PoB), Proof of History (PoH), Proof of Importance (PoI), Proof of Believability (PoBelievability), Federated Byzantine Agreement (FBA), Combined DPoS+BFT, Proof of Authority (PoAuthority), and Raft.

Some consensus mechanisms have multiple in-service implementations. For example, PoS has several implementations [25], such as, Chain of Activity, Casper, Algorand, and Tendermint.

III. Information Security with BC

BC is not automatically more secure than other DLT systems. Although DLT makes it difficult to manipulate and attack by a single node, the major BC vulnerabilities are related to [30]:

- fork (intentional or unintentional condition in which nodes in the network have diverging views),
- stale or orphaned blocks,
- different attacks (e.g. 51%, DNS, DDoS, selfish mining, consensus delay, double-spending).

Some of these vulnerabilities are specific to cryptocurrency applications and associated with public permissionless BCs. In permissioned BC, multiple validation nodes are trusted to maintain the consensus. To avoid certain nodes accepting inconsistent messages, consensus mechanisms (e.g. PBFT, PoAuthority, DPoS+BFT) are used; decisions are encoded using digital signatures with algorithms such as Elliptic Curve Cryptography and SHA3-256 [30].

IV. BC Systems for Big Data

A. Implementations of BC and Other DLT Systems

The key aspects described in the above sections (block structure and BC architecture) have to be implemented and customized based on the type of application and considering the data to be stored, secured and managed with the BC. The block components, algorithms and protocols have to follow the latest standards and meet the defined requirements. The BC implementation has to take into account vulnerabilities and the performance of the hash function considering the data size and file system.

TABLE I. Implementations of BC and Other DLT Systems

Platform/Framework	**Platform/Framework**
1. Aion	30. Komodo
2. ArcBlock	31. Lisk
3. Ardor	32. MedRec
4. BigchainDB	33. MediChain
5. Binded	34. MultiChain
6. Bitcoin	35. ModelChain
7. BlocHIE	36. Neblio
8. Burst	37. NEM
9. Cardano	38. NEO
10. Chain core	39. Nxt
11. Corda	40. OmniPHR
12. Credits	41. OpenChain
13. Democracy Earth	42. Oracle
14. Elements	43. Qtum
15. Enigma	44. Quorum
16. EOS	45. Paperchain
17. Ethereum	46. Parity
18. Exonum	47. Po.et
19. Follow My Vote	48. Propy
20. Graphene	49. Stellar
21. Guardtime	50. Straitis
22. Hydrochain	51. Tezos
23. Hyperledger Fabric	52. Ubitquity
24. Hyperledger Indy	53. Ujomusic
25. Hyperledger Burrow	54. Verisart
26. Hyperledger Sawtooth	55. Wanchain
27. ICON	56. Warranteer
28. IOTA	57. Waves
29. KSI Blockchain	58. Zilliqa

Multiple resources (e.g. [35, 36]) analyzed available BC and other DLT frameworks and platforms. Their application areas are varying between finance, government, IoT, legal and other industries. A comprehensive list of BC and DLT systems identified on July 15, 2021 from internet and research publications is provided in Table I. This list includes both frameworks (e.g. Ethereum, Hyperledger, Qtum) and the implementations of BCs for specific applications. These implementations were analyzed in

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

context of their applicability and benefits to big data applications (details are presented at the conference). The analysis also considered different factors affecting the implementation, for example, programming language, support documentation, and availability (commercial or open source).

B. BC Systems for Big Data

The BC system can be implemented for increasing level of information security of large and growing data volumes in data centers and cloud environments. The largest cloud service providers (e.g. AWS, Azure, IBM, Oracle) have already developed environments with blockchain as a service (BaaS). In a cloud environment, the usage of BC enables multiple nodes to participate in maintaining transparent and immutable provenance information for tracking data transactions and detect malicious activities [30]. The API and web user interface can be used to monitor BC activity and visualize provenance records. The BC system applications for sharing and enabling integrity of large data files such as medical [37] and Earth Observation [38] images provide examples of the required definition and design activities.

An important problem in BC is scalability, which coincides with growing data volumes. An increasing number of transactions (and block validations) from an increasing number of users, leads to communication overheads that limit the network scalability [35]. BC systems for massive data storage can potentially enable scalability with acceptable performance [25]. To address issue of redundant massive data storage, a secure data storage and recovery scheme in the BC-based network was proposed [39] by improving the real-time monitoring, and supporting the dynamic storage, and update of distributed data.

The main development activities to achieve the best performance and security capabilities of BC system are currently focused on definition and design of its technological elements, including block structure and information, hash function, encryption algorithm, consensus mechanism and integration in file system. The architecture of the BC system and its components for big data were analyzed (presented at the conference with more details).

V. Conclusion

This paper reviewed and analyzed current concepts, applications and implementations of BC systems. In total, 58 BC platforms and frameworks were identified. The main elements of a BC system to enable information security of big data were analyzed and defined.

References

[1] NASA OCIO, “The Big Data Wave,” IT Talk, vol. 8, no. 3, Sep. 2018. Accessed: Jul. 15, 2021. [Online]. Available: https://www.nasa.gov/sites/default/files/atoms/files/365378_it_talk_design_-_july_2018_0.pdf

[2] M. Stamp, Information security: principles and practice, 2nd ed. Hoboken, NJ: Wiley, 2011.

[3] M. Copeland and M. Jacobs, Cyber Security on Azure: An IT Professional’s Guide to Microsoft Azure Security. Berkeley, CA: Apress, 2021. doi: 10.1007/978-1-4842-6531-4.

[4] B. Camci, S. Bayar, and M. G. Ulkar, “A simple auditing mechanism for financial reports in e-Ledger project,” in 2015 9th International Conference on Application of Information and Communication Technologies (AICT), Rostov on Don, Russia, Oct. 2015, pp. 244–248. doi: 10.1109/ICAICT. 2015.7338555.

[5] World Bank, Distributed Ledger Technology and Secured Transactions. World Bank, Washington, DC, 2020. doi: 10.1596/34007.

[6] M. M. Akhtar, D. R. Rizvi, M. A. Ahad, S. S. Kanhere, M. Amjad, and G. Coviello, “Efficient Data Communication Using Distributed Ledger Technology and IOTA-Enabled Internet of Things for a Future Machine-to-Machine Economy,” Sensors, vol. 21, no. 13, p. 4354, Jun. 2021, doi: 10.3390/s21134354.

[7] E. A. Brewer, “Towards robust distributed systems (abstract),” in Proceedings of the nineteenth annual ACM symposium on Principles of distributed computing - PODC ’00, Portland, Oregon, United States, 2000, p. 7. doi: 10.1145/343477.343502.

[8] S. Gilbert and N. Lynch, “Perspectives on the CAP Theorem,” Computer, vol. 45, no. 2, pp. 30–36, Feb. 2012, doi: 10.1109/MC.2011.389.

[9] M. Litoiu et al., “How do I choose the right NoSQL solution? A comprehensive theoretical and experimental survey,” Big Data Inf. Anal., vol. 1, no. 2/3, pp. 185–216, Sep. 2016, doi: 10.3934/bdia.2016004.

[10] M. Lange, S. C. Leiter, R. Alt, “Defining and Delimitating Distributed Ledger Technology: Results of a Structured Literature Analysis,” in Business Process Management: Blockchain and Central and Eastern Europe Forum, vol. 361, C. Di Ciccio, R. Gabryelczyk, L. García-Bañuelos, T. Hernaus, R. Hull, M. Indihar Štemberger, A. Kő, and M. Staples, Eds. Cham: Springer International Publishing, 2019, pp. 43–54. doi: 10.1007/978-3-030-30429-4_4.

[11] G. Hileman and M. Rauchs, “2017 Global Blockchain Benchmarking Study,” SSRN Electron. J., 2017, doi: 10.2139/ssrn.3040224.

[12] F. Knirsch, A. Unterweger, and D. Engel, “Implementing a blockchain from scratch: why, how, and what we learned,” EURASIP J. Inf. Secur., vol. 2019, no. 1, p. 2, Dec. 2019, doi: 10.1186/s13635-019-0085-3.

[13] M. N. M. Bhutta et al., “A Survey on Blockchain Technology: Evolution, Architecture and Security,” IEEE Access, vol. 9, pp. 61048–61073, 2021, doi: 10.1109/ACCESS.2021.3072849.

[14] W. Yang, E. Aghasian, S. Garg, D. Herbert, L. Disiuta, and B. Kang, “A Survey on Blockchain-Based Internet Service Architecture: Requirements, Challenges, Trends, and Future,” IEEE Access, vol. 7, pp. 75845–75872, 2019, doi: 10.1109/ACCESS.2019.2917562.

[15] C. Fan, S. Ghaemi, H. Khazaei, and P. Musilek, “Performance Evaluation of Blockchain Systems: A Systematic Survey,” IEEE Access, vol. 8, pp. 126927–126950, 2020, doi: 10.1109/ACCESS.2020.3006078.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

[16] T. Rathee, P. Singh, “A systematic literature mapping on secure identity management using blockchain technology,” J. King Saud Univ. Comput. Inf. Sci., p. S1319157821000690, Mar. 2021, doi: 10.1016/j.jksuci.2021.03.005.

[17] T. M. Fernandez-Carames and P. Fraga-Lamas, “Towards Post-Quantum Blockchain: A Review on Blockchain Cryptography Resistant to Quantum Computing Attacks,” IEEE Access, vol. 8, pp. 21091–21116, 2020, doi: 10.1109/ACCESS.2020.2968985.

[18] M. H. ur Rehman, K. Salah, E. Damiani, and D. Svetinovic, “Trust in Blockchain Cryptocurrency Ecosystem,” IEEE Trans. Eng. Manag., vol. 67, no. 4, pp. 1196–1212, Nov. 2020, doi: 10.1109/TEM.2019.2948861.

[19] T. Alladi, V. Chamola, R. M. Parizi, and K.-K. R. Choo, “Blockchain Applications for Industry 4.0 and Industrial IoT: A Review,” IEEE Access, vol. 7, pp. 176935–176951, 2019, doi: 10.1109/ACCESS.2019.2956748.

[20] S. Namasudra and G. C. Deka, Eds., Applications of Blockchain in Healthcare, vol. 83. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-15-9547-9.

[21] H.-Y. Paik, X. Xu, H. M. N. D. Bandara, S. U. Lee, S. K. Lo, “Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance,” IEEE Access, vol. 7, pp. 186091–186107, 2019, doi: 10.1109/ACCESS.2019.2961404.

[22] J. E. Kasten, “Engineering and Manufacturing on the Blockchain: A Systematic Review,” IEEE Eng. Manag. Rev., vol. 48, no. 1, pp. 31–47, Mar. 2020, doi: 10.1109/EMR.2020.2964224.

[23] S. E. Chang and Y. Chen, “When Blockchain Meets Supply Chain: A Systematic Literature Review on Current Development and Potential Applications,” IEEE Access, vol. 8, pp. 62478–62494, 2020, doi: 10.1109/ACCESS.2020.2983601.

[24] K. Gai, J. Guo, L. Zhu, and S. Yu, “Blockchain Meets Cloud Computing: A Survey,” IEEE Commun. Surv. Tutor., vol. 22, no. 3, pp. 2009–2030, 2020, doi: 10.1109/COMST.2020.2989392.

[25] M. H. Rehmani, Blockchain Systems and Communication Networks: From Concepts to Implementation. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-71788-9.

[26] D. Cagigas, J. Clifton, D. Diaz-Fuentes, and M. Fernandez-Gutierrez, “Blockchain for Public Services: A Systematic Literature Review,” IEEE Access, vol. 9, pp. 13904–13921, 2021, doi: 10.1109/ACCESS.2021.3052019.

[27] C. Shen and F. Pena-Mora, “Blockchain for Cities–A Systematic Literature Review,” IEEE Access, vol. 6, pp. 76787–76819, 2018, doi: 10.1109/ACCESS.2018.2880744.

[28] T. Ali Syed, A. Alzahrani, S. Jan, M. S. Siddiqui, A. Nadeem, and T. Alghamdi, “A Comparative Analysis of Blockchain Architecture and its Applications: Problems and Recommendations,” IEEE Access, vol. 7, pp. 176838–176869, 2019, doi: 10.1109/ACCESS.2019.2957660.

[29] J. Xie et al., “A Survey of Blockchain Technology Applied to Smart Cities: Research Issues and Challenges,” IEEE Commun. Surv. Tutor., vol. 21, no. 3, pp. 2794–2830, 2019, doi: 10.1109/COMST.2019.2899617.

[30] S. S. Shetty, C. A. Kamhoua, and L. L. Njilla, Blockchain for Distributed Systems Security. 2019. Accessed: Jul. 14, 2021. [Online]. Available: http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781119519591

[31] X. Chen, K. Zhang, X. Liang, W. Qiu, Z. Zhang, and D. Tu, “HyperBSA: A High-Performance Consortium Blockchain Storage Architecture for Massive Data,” IEEE Access, vol. 8, pp. 178402–178413, 2020, doi: 10.1109/ACCESS.2020.3027610.

[32] N. Deepa et al., “A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions,” ArXiv200900858 Cs, Feb. 2021, Accessed: Jul. 18, 2021. [Online]. Available: http://arxiv.org/abs/2009.00858

[33] W. Zhao, C. Jiang, H. Gao, S. Yang, and X. Luo, “Blockchain-Enabled Cyber–Physical Systems: A Review,” IEEE Internet Things J., vol. 8, no. 6, pp. 4023–4034, Mar. 2021, doi: 10.1109/JIOT.2020.3014864.

[34] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” in 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA, Jun. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.

[35] Ismail and Materwala, “Article A Review of Blockchain Architecture and Consensus Protocols: Use Cases, Challenges, and Solutions,” Symmetry, vol. 11, no. 10, p. 1198, Sep. 2019, doi: 10.3390/sym11101198.

[36] TechnoDuet, “A Comprehensive List of Blockchain Platforms,” 2021. technoduet.com (accessed Jul. 15, 2021).

[37] M. Sultana, A. Hossain, F. Laila, K. A. Taher, M. N. Islam, “Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology,” BMC Med. Inform. Decis. Mak., vol. 20, no. 1, pp. 256, Dec. 2020, doi: 10.1186/s12911-020-01275-y.

[38] A. Burzykowska, M. Iapaolo, A. Priit, and A. Sisask, “EO Data Provenance with KSI Blockchain,” ESA, Issue Brief. ESA Blockchain / Distributed Ledgers and EO Community of Practice (CoP), 2020.

[39] W. Liang, Y. Fan, K.-C. Li, D. Zhang, and J.-L. Gaudiot, “Secure Data Storage and Recovery in Industrial Blockchain Network Environments,” IEEE Trans. Ind. Inform., vol. 16, no. 10, pp. 6543–6552, Oct. 2020, doi: 10.1109/TII.2020.2966069.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.