Computing, Telecommunication and Control, 2025, Vol. 18, No. 4, Pp. 87–101.
Информатика, телекоммуникации и управление. 2025. Том 18, № 4. С. 87–101.

Overview article
DOI: https://doi.org/10.18721/JCSTCS.18408
UDC 004.9:004.94

MITIGATING DATA GROWTH IN POW BLOCKCHAINS: STORAGE REDUCTION METHODS FOR SCALABILITY WITHOUT COMPROMISING DECENTRALIZATION

D.D. Razuvaev[1,2] ✉ , *S.M. Ustinov*[1]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;

[2] Lomonosov Moscow State University, Moscow, Russian Federation

✉ Razuvaev_DD@mail.ru

Abstract. The exponential growth of data volume in Proof-of-Work (PoW) blockchains threatens their decentralization. The article provides a systematic analysis of data growth mitigation methods (sharding, block pruning, off-chain storage etc.), identifying their key flaws: compromised auditability, increased synchronization complexity or centralization. To address the issue, a novel Periodic Aggregation with Dual Hash Anchoring (PADHA) method is proposed. Its key innovation is the synergy of data pruning and the controlled use of chameleon hash functions. The method enables linear reduction of stored history by creating final state aggregators and subsequent secure "cleansing" of blocks from past epochs. PADHA preserves cryptographic chain integrity and PoW support without trusted third parties. The method is designed for Fact-Oriented Blockchains that store final data states (facts), making it promising for registries, IoT and other applications where current information, not its change history, is critical.

Keywords: distributed registries, decentralized systems, blockchain, scalability, decentralization, block pruning

Citation: Razuvaev D.D., Ustinov S.M. Mitigating data growth in PoW blockchains: Storage reduction methods for scalability without compromising decentralization. Computing, Telecommunications and Control, 2025, Vol. 18, No. 4, Pp. 87–101. DOI: 10.18721/JCSTCS.18408

Обзорная статья
DOI: https://doi.org/10.18721/JCSTCS.18408
УДК 004.9:004.94

ОПТИМИЗАЦИЯ ХРАНЕНИЯ ДАННЫХ В PROOF-OF-WORK БЛОКЧЕЙН-СИСТЕМАХ: МЕТОДЫ СОКРАЩЕНИЯ ОБЪЕМА И ОБЕСПЕЧЕНИЕ МАСШТАБИРУЕМОСТИ В УСЛОВИЯХ ДЕЦЕНТРАЛИЗАЦИИ

Д.Д. Разуваев[1,2] ✉ , *С.М. Устинов*[1]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация;

[2] Московский государственный университет им. М.В. Ломоносова, Москва, Российская Федерация

✉ Razuvaev_DD@mail.ru

Аннотация. Экспоненциальный рост объема данных в блокчейнах на Proof-of-Work (PoW) угрожает их децентрализации. В статье системно анализируются методы борьбы с ростом данных (шардинг, обрезка блоков, офф-чейн хранение и др.) и выявляются их ключевые недостатки: нарушение аудируемости, усложнение синхронизации или централизации. Для решения проблемы предлагается новый метод периодической агрегации с двойным хеш-якорением (PADHA). Его ключевая инновация – синергия обрезки данных и контролируемого использования хеш-функций-хамелеонов. Метод обеспечивает линейное сокращение хранимой истории за счет создания агрегаторов финального состояния и последующего безопасного «очищения» блоков прошедших эпох. PADHA сохраняет криптографическую целостность цепочки и поддержку PoW, не требуя доверенных третьих сторон. Метод применим для факт-ориентированных блокчейнов (FOB), хранящих итоговые состояния данных (факты), что делает его перспективным для реестров, IoT и других приложений, где критична актуальная информация, а не история ее изменений.

Ключевые слова: распределенные реестры, децентрализованные системы, блокчейн, масштабируемость, децентрализация, обрезка блоков

Для цитирования: Razuvaev D.D., Ustinov S.M. Mitigating data growth in PoW blockchains: Storage reduction methods for scalability without compromising decentralization // Computing, Telecommunications and Control. 2025. Т. 18, № 4. С. 87–101. DOI: 10.18721/JCSTCS.18408

Introduction

Currently, blockchain technology is widely applied across various domains including finance [1, 4], intellectual property management [2], insurance, tourism [3], healthcare and biomedical systems [10, 21], government services and education [5, 9], as well as 5G networks [7, 8], to address challenges related to data integrity, transparency and disintermediation. Blockchain is a distributed digital registry based on the principles of decentralization, cryptography and consensus [11, 25]. Unlike traditional centrally managed databases, blockchain systems function as peer-to-peer networks of nodes (participants), each of which usually stores a complete copy of information. The data in such a system is organized as a sequence of blocks [22], where each block contains:

- a set of records (for example, transactions or any other data);
- hash identifier – a unique cryptographic signature calculated based on the block content;
- hash of the previous block, which ensures a cryptographic link between the chain elements.

This architecture guarantees the immutability of data (excluding the use of hash chameleons): any attempt to change information in existing block will disrupt communication with subsequent blocks,

which will be immediately detected by the network. To add new blocks, a consensus algorithm [24, 26] is used – a set of rules that allows nodes to coordinate the state of the system without trusting a central authority. The first implementation of this technology, which appeared in 2008, demonstrated the potential of blockchain technology and established itself as a tool for creating censorship- and fraud-resistant systems, as well as revealed fundamental limitations associated with the scalability of such systems.

A key feature of the Proof-of-Work (PoW) blockchain [27] is the mechanism for achieving consensus through computational tasks [28]. Nodes (miners) compete to solve a cryptographic puzzle that requires significant resources. The winner gets the right to add a block to the chain and a reward, and the other nodes check the correctness of the decision. This process, although it provides a high level of security, leads to an increase in data, since each new block increases the total volume of the chain, while deleting a block violates the integrity of the blockchain.

Over time, this creates "a paradox of centralization of decentralized systems": demands on computing power and storage become so high that participation in the network becomes available only to a limited number of specialized participants (who can commercialize their work by providing Software-as-a-Service solutions). For example, in January 2025, the size of the blockchain in the first implementation of the technology exceeded 600 GB, and in March 2025 it already stands at 612 GB[1], and the annual energy consumption of the network is comparable to that of entire countries [6]. These factors not only threaten the stability of the system, but also limit its application in areas where data processing speed and energy efficiency are critical, such as IoT, mobile devices and systems with data processing speed requirements. Fig. 1 shows a graph of the size of the data volume over a 10-year time scale, which clearly demonstrates the problem of data accumulation.

Thus, despite its revolutionary potential, PoW-based blockchain systems face a trilemma: scalability, decentralization and security cannot be simultaneously optimized within a classical architecture. Resolving this problem requires rethinking approaches to data storage and management, which is the focus of this study.

The exponential increase in data volume in PoW-based blockchain systems is a direct consequence of their architectural features. Each new block added to the chain not only expands the data history, but also requires all network participants to constantly verify and store a complete copy of the registry. This leads to a number of systemic contradictions:

• Accumulation of "historical load"

In classic PoW implementations [23] (for example, in bitcoin), the size of the chain increases by 1–4 MB daily (this value is individual for each implementation), which over 5 years of operation has created a load of several hundred gigabytes. For nodes with limited resources (for example, mobile devices), this makes participation in the network technically and economically impractical.

• Energy-computing imbalance

The PoW mechanism requires repeated recalculation of hashes to achieve consensus, which leads to duplication of computing operations on all nodes [18]. The growing volume of data exacerbates this problem: the validation of a long chain of blocks consumes more and more energy.

• Degradation of network synchronization

Increasing the download and verification time of the blockchain reduces the synchronization speed of new nodes. In high-bandwidth networks (for example, Ethereum before switching to Proof-of-Stake), full synchronization can take days, increasing the risks of chain splits and reducing resistance to attacks.

• Limitation of functionality of light clients

"Simplified" nodes (light clients) that do not store a complete copy of the blockchain are forced to rely on trusted third-party services to access the data. This violates the principle of decentralization and creates vulnerabilities such as censorship or information substitution.

[1] Protocol Labs. Filecoin: A decentralized storage network, 2017. Available: https://filecoin.io/filecoin.pdf (Accessed 10.05.2025)

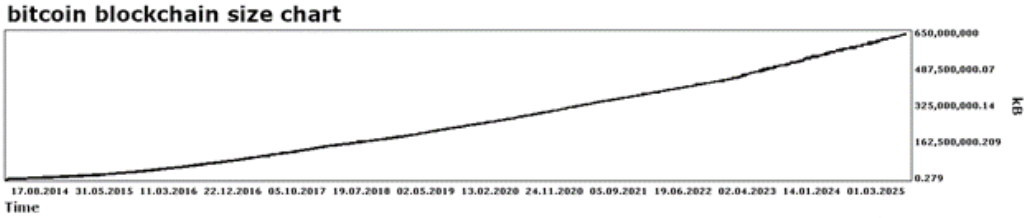

Fig. 1. Graph of the size of the Bitcoin blockchain over time

Attempts to solve these problems through traditional methods – sharding, block pruning or hybrid consensus – face fundamental limitations. For example, sharding, which divides the network into sub-chains, reduces the load on individual nodes, but increases the complexity of cross-shard operations and the risk of conducting attacks in isolated segments (shards) when conducting attacks of the "51 percent" class, where an attacker with more than 50% of the capacity (>50% of the total hashrate, which reflects how many cryptographic operations to find a block solution in the blockchain can be performed by devices in a unit of time), can single-handedly control a sub-chain or the entire block chain. Pruning, although it reduces the local amount of data, deprives the network of the total PoW power.

These contradictions actualize the search for alternative approaches that will preserve the advantages of PoW (for example, resistance to Sybil attacks), and eliminate the problem of endless growth of the stored amount of data. The key direction is the development of protocols that optimize information storage without compromising security, for example, through segmentation of data by relevance level or the introduction of mechanisms for "forgetting" outdated blocks.

Relevance of the topic

The relevance of this study is due to the systemic crisis of scalability faced by PoW blockchain solutions in the context of exponential growth of stored data. Despite the widespread introduction of technology into finance [1], intellectual property [2], logistics, healthcare [21] (including use in biomedical systems [10]), tourism and the public sector [3, 5, 9], blockchain technology has both advantages and fundamental limitations [13] – immutability data and full replication are becoming a barrier to sustainable development. Traditional optimization methods such as sharding or hybrid consensus demonstrate partial efficiency, but do not solve the key problem: inconsistencies between the growing volume of data and the requirements of decentralization[2] [12].

The scientific significance of this work lies in addressing several key gaps.

First, there is a lack of comprehensive analysis regarding how existing data reduction methods – such as multi-chain networks, off-chain storage and modifiable blockchains[3] [14–16] – affect the fundamental blockchain trilemma of security, decentralization and scalability.

Second, the theoretical framework lacks well-defined criteria for assessing "historical redundancy" of data; for instance, clear principles are needed to determine which blocks can be safely deleted or compressed without undermining the system's auditability.

Furthermore, a pragmatic paradox emerges: technologies originally designed to eliminate centralized intermediaries, such as cryptocurrencies [4], increasingly depend on centralized cloud storage solutions to archive old blocks [14], which fundamentally contradicts their decentralized ideology.

The practical relevance of this research is driven by pressing industry demands. Financial institutions seek to reduce the operational costs of maintaining full PoW network nodes while adhering to stringent regulatory standards like the European General Data Protection Regulation (GDPR).

[2] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.
[3] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.

Simultaneously, IoT [15] and 5G networks [7] require lightweight blockchain solutions capable of real-time data processing. Additionally, state registries – for land, education [9] and other public records – confront the challenge of long-term data preservation due to bandwidth constraints and ever-growing data volumes.

Therefore, this study conducts a systematic comparative analysis of data management methods in PoW systems, including:

- multichain architectures (AppChain)[4];
- off-chain storage with lazy loading [14];
- protocol modifications via hash chameleon functions [16].

The conducted analysis reveals the fundamental trade-offs inherent to each of the considered approaches. For instance, aggressive block pruning leads to the irreversible loss of the complete history and, consequently, to the ability to perform an independent chain audit. Hybrid consensus models, while aiming to reduce energy consumption, carry inherent risks of power centralization among a limited circle of stakeholders. Finally, modifiable blockchains based on chameleon hashes, while addressing data growth, introduce vulnerabilities related to the potential for targeted collision attacks and the centralization of editing control, which undermines the core principle of immutability. These identified limitations necessitate fundamentally new solutions based on the following design principles: dynamic data lifecycle management, enabling the deletion or archiving of obsolete information; segmentation of data based on its criticality to system operation; and minimizing reliance on external decentralized repositories (such as IPFS or Storj) to preserve the blockchain's self-sufficiency and security.

To overcome the noted shortcomings of existing methods, this article, based on their systematic analysis, proposes a new method of Periodic Aggregation with Dual Hash Anchoring (PADHA). Its goal is to resolve the contradiction between the need to preserve the key advantages of PoW, such as resistance to attacks and censorship, and the Web 3.0 requirements for high data processing speed and energy efficiency. The results of the study form the basis for the further development of data storage optimization methods[5] [19].

Methods

To combat the exponential increase in data volume in PoW systems, various approaches have been proposed that can be classified according to the level of impact on the blockchain architecture.

The effectiveness of data growth control methods in PoW systems directly depend on their ability to maintain a balance between volume optimization, decentralization, and security. However, most solutions upset this balance by introducing unacceptable trade-offs, especially in the context of PoW, where the integrity of the chain and the availability of the full history are the basis of consensus.

Traditional methods

Sharding

Bottom line: dividing the network into independent sub-chains (shards), each of which processes part of the transactions[6].

Example: Zilliqa, Ethereum before switching to Proof-of-Stake (PoS).

Advantages: reducing the load on individual nodes, increasing throughput.

Disadvantages: increased complexity of cross-shard operations; increased vulnerability of sub-chains to a sub-chain attack in cases where more than 50% of the hashrate of the sub-chain accumulates in the attacker; violation of the integrity of data auditing.

[4] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.
[5] Buterin V. Merkling in Ethereum, 2015. Available: https://blog.ethereum.org/2015/11/15/merkling-in-ethereum (Accessed 10.05.2025)
[6] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.

Key limitation: splitting into sub-chains complicates synchronization and increases the risk of attacks in small shards. This method complicates synchronization and increases the risk of attacks aimed at destabilizing the network.

Pruning

Bottom line: deleting outdated blocks while preserving only headers or critical data (for example, snapshots of the Unspent Transaction Output status) [12].

Example: Bitcoin Core ("pruned node" mode).

Advantages: reduction of the local data volume to 5–10% of the original.

Disadvantages: loss of the possibility of independent verification of the complete history; loss of historical validity; dependence on complete archive nodes, which leads to centralization.

Key limitation: deleting the initial blocks (including the genesis block) violates the integrity of the hash chain, which contradicts the PoW principle. This method may disrupt the stability of the network.

Hybrid consensus models (*PoW/PoS*)

Bottom line: a combination of PoW for creating blocks and PoS for validation[7].

Example: Decred [16].

Advantages: reduction of energy consumption and performance requirements due to partial abandonment of mining.

Disadvantages: conflicts between consensus mechanisms; vulnerability to "nothing-at-stake" attacks; risk of centralization of power among stakeholders.

Key limitation: the combination of PoW and PoS creates contradictions between miners and stakeholders, and also leads to the centralization of power among large stakeholders and a potential decrease in the security of the blockchain.

Innovative methods

Multi-chain networks

Bottom line: creating a hierarchy of blockchains, where the main chain (Layer 1) coordinates the work of sidechains (Layer 2)[8].

Example: Polkadot (based on Nominated PoS) [17], Cosmos (based on Tendermint BFT), Rootstock RSK (PoW).

Advantages: isolation of application data; scalability due to parallelism.

Disadvantages: the difficulty of synchronization between circuits.

Key limitation: the hierarchy of blockchains complicates auditing and synchronization [20]. Synchronization problems directly reduce the stability of blockchain systems.

Off-chain storage with lazy loading

Bottom line: transferring old blocks to external decentralized storage (for example, InterPlanetary File System[9]) with on-demand download [14].

Example: Arweave (permanent storage), Filecoin (archiving on request)

Advantages: reducing the load on the nodes; maintaining access to the full history.

Disadvantages: delays in requesting archived data; vulnerability to storage failures; dependence on the stability of third-party networks.

Key limitation: transferring data to IPFS/Filecoin introduces delays and risks of information loss, and does not solve the problem of blockchain bloat, shifting responsibility for the preservation of archived data to a distributed solution. The self-sufficiency of the blockchain is being violated.

Redactable blockchains

Bottom line: using hash chameleon functions to edit or delete blocks without violating the integrity of the chain [16].

[7] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.
[8] Huobi Research Institute Report. Game of Thrones in Blockchain: Multi-Chain Networks Battle for Supremacy, 2022.
[9] Protocol Labs. Filecoin: A decentralized storage network, 2017. Available: https://filecoin.io/filecoin.pdf (Accessed 10.05.2025)

Example: Hyperledger Fabric (primarily in enterprise contexts)[10].

Advantages: dynamic data management.

Disadvantages: centralization of control (modifiers); vulnerability to collision attacks.

Key limitation: editing blocks using hash chameleon functions undermines the basic principle of block immutability. For systems that prioritize the principle of immutability of information, this method is not optimal. The risk of collision attacks harms the data integrity paradigm in the system.

Merkle Patricia Trie

Bottom line: optimizing the storage of network status through tree-like hash structures that allow you to remove duplicates [8].

Example: Ethereum State Trie

Advantages: reduction of data volume by 30–50%; acceleration of transaction search.

Disadvantages: the complexity of the implementation for PoW networks; the risk of data loss in case of failures.

Key limitation: this method is applicable to the classical implementation of the PoW blockchain only to a limited extent, for working with the state tree stored inside the blockchain. At the same time, the method solves the problem of inflating the PoW of the blockchain system by using pruning method, with all the conclusions drawn from this.

Summarizing the review and analysis of existing solutions, addressing the inherent limitations of existing methods is challenging. However, a viable path forward involves introducing specific architectural constraints. These constraints narrow the method's applicability to a particular class of tasks – specifically, systems that only need to store final state data rather than complete transactional histories. While this represents a focused application domain, it remains broad enough to cover numerous practical use cases, such as registries, sensor data logging and state tracking. The significant gain achieved – dramatic data volume reduction without breaking cryptographic chain integrity – justifies this targeted approach.

Proposed Method

An analysis of existing traditional and innovative methods for combating data growth in PoW blockchain systems has revealed fundamental problems affecting the scalability, security and/or decentralization of solutions based on them. Models of sharding, block pruning and hybrid consensus models, while reducing the load on nodes, violate the principles of data integrity, auditability and participant equality. Architectures with application-specific sidechains and off-chain storage, in turn, introduce dependency of the main chain on third-party networks and complicate synchronization or lead to forced centralization of solutions. Furthermore, modifiable blockchains based on chameleon hashes, despite offering dynamic data management, also contain inherent risks.

For an unambiguous definition of the proposed method and its applicability boundaries, it is necessary to introduce the following terminology:

• Atomic Fact – an immutable, self-contained unit of information representing an arbitrary set of data (an assertion), which does not require references to other facts for its interpretation or integrity.

• Fact Block – an operationally atomic structural unit of a distributed ledger, whose sole content is an ordered set of atomic facts, supplemented by a unified service header (timestamp, hash). The acceptance or rejection of a fact block by the system follows an "all-or-nothing" principle.

• Fact-Oriented Blockchain (FOB) – a distributed ledger implemented as a chronologically and cryptographically linked chain of fact blocks, which guarantees the immutability and order of atomic facts.

This method is specifically applicable to FOB, which imposes a key architectural constraint: such a system can only store final data (facts), but cannot store procedural or transactional records describing the process of changes.

[10] Hyperledger Fabric. A Blockchain Platform for the Enterprise, 2021. Available: https://hyperledger-fabric.readthedocs.io/en/release-2.5/ (Accessed 10.05.2025)

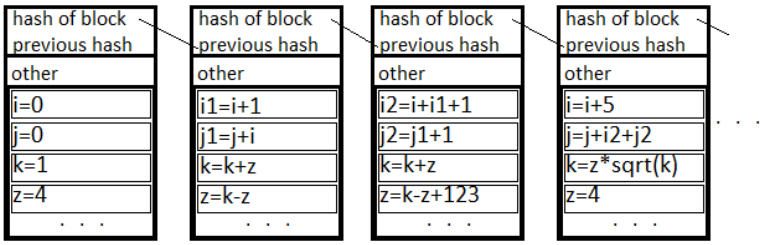

Fig. 2. Blockchain with transactions

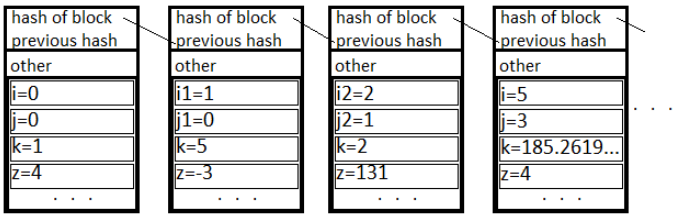

Fig. 3. Fact-Oriented Blockchain

The main difference between a classical blockchain model and an FOB lies in the nature, connectivity and interpretation of the data stored in the blocks (Figs. 2, 3).

In a classical blockchain (Fig. 2), blocks typically contain transactions or change commands (e.g., i = i + 5). The information in subsequent blocks is semantically and logically dependent on the information in previous blocks. To obtain the current state of the system (for example, the final value of *j*), it is necessary to replay the entire transaction history from the very beginning (the genesis block). The deletion or corruption of an intermediate block makes it impossible to calculate or verify any subsequent data.

In FOB (Fig. 3), this limitation is eliminated due to a different organization of data. Each atomic fact within a block represents a final value or statement (for example, i = 3), which is semantically complete and does not require referencing other facts for its interpretation. Let us define that a system based on FOB interprets data by treating the last value recorded in the chain for a given entity (for example, variable i) as its current state. Thus, a new fact i = 5 in a later block does not reference the previous value but semantically overwrites it for the system. A fact-block serves for the operationally atomic batch transfer of such independent facts.

Thus, a situation is achieved where:

- information (an atomic fact) in any block is self-sufficient for interpretation;
- understanding the current state of the system does not require reproducing the full history of changes; it is sufficient to read the latest facts for the entities of interest;
- integrity of a block is ensured cryptographically, and the semantic value of each fact is contained within itself and its position in the chain, which determines the relevance of the value.

To solve the problem of uncontrolled “bloating”, the use of a combination of two method concepts is proposed: the pruning method and the editable blockchains method.

The classic implementation of the data pruning method involves removing transactions that are technically unnecessary for the target user. This results in the loss of the entire chain’s validity and reduces the weight of the blockchain (Fig. 4).

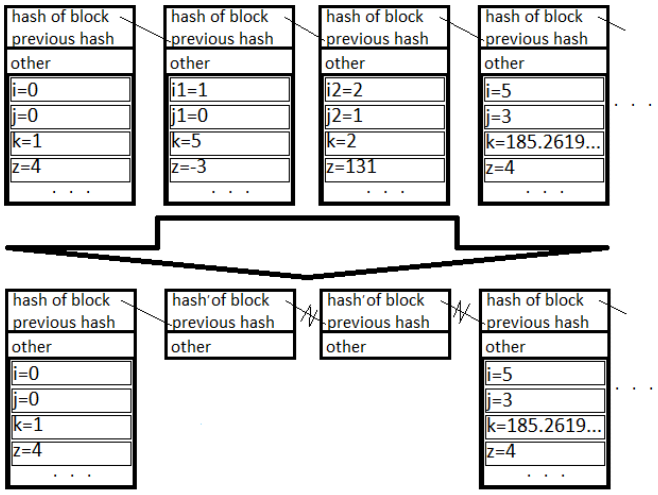

Fig. 4. Pruning method for an FOB: in blocks 2 and 3,
the facts have been removed and the hash chain is broken

The second method proposes using a chameleon hash function instead of a regular hash function. A chameleon hash function allows generating collisions using a special key. Fig. 5 demonstrates the method of using an editable blockchain based on chameleon hash functions.

The PADHA method is designed for FOB and combines the advantages of data pruning and the use of chameleon hash functions, enabling a radical reduction in the volume of stored data without losing the cryptographic integrity of the chain. The core idea of the method is the periodic creation of special aggregator blocks, which contain the final (current) values of all entities for a certain period, and the subsequent "clearing" (replacing the bodies with empty ones) of ordinary blocks in that period while preserving the hash chain thanks to chameleon hash functions.

Architectural features and prerequisites

As noted, an FOB operates with atomic facts, each of which is a self-sufficient statement (e.g., "i = 5"). The system interprets the most recent (in chain order) value for a given entity (i) as the current one. This property enables state aggregation: instead of storing all intermediate changes, it is sufficient to store only the latest values at the moment of aggregation.

In the PADHA method, each block contains two cryptographic hashes: a regular one (e.g., SHA-256) and a chameleon hash. The link between blocks is considered valid until an aggregator block is created if at least one of the two hashes matches (i.e., either the regular hash or the chameleon hash), or the link between blocks is valid if both hashes match when the aggregator block has not yet been formed (the epoch is not yet completed). Such link validation rules allow, after cleaning the block bodies, the chain to remain valid via chameleon hashes, while the regular hashes may become invalid due to changes in the block contents. They also protect against malicious attempts to modify blocks during an unclosed epoch.

Method description

1. Normal operation mode (within an epoch)

The blockchain is divided into epochs of a fixed length (for example, 1000 blocks). During an epoch, all blocks are created in the normal mode: each block contains atomic facts (changes in entity values) and has a header with two hashes: hash_std (standard hash) and hash_cham (chameleon hash). Both hashes are calculated based on the block's content and the corresponding hash of the previous block. Thus, at the beginning of an epoch, the chain is valid according to both hashes (Fig. 6).

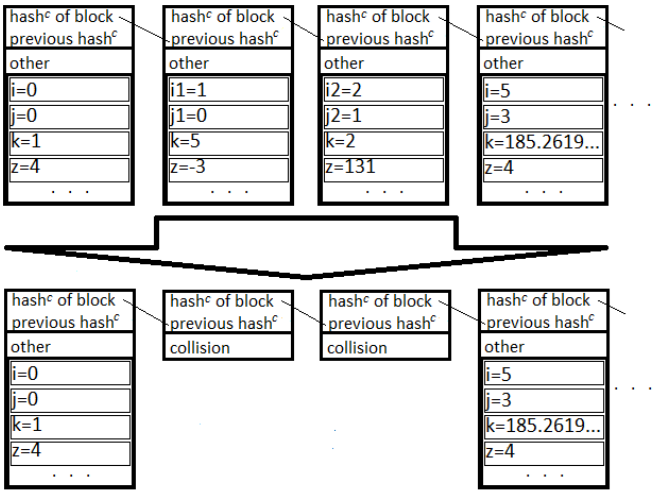

Fig. 5. Chameleon hash method: blocks 2 and 3 have been cleared,
the hash chain remains intact because collisions have been inserted

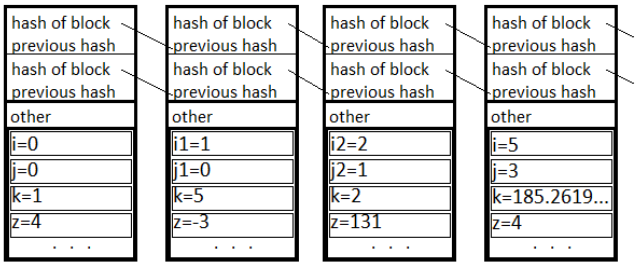

Fig. 6. PADHA – operation within an epoch with a length of 5

2. Creating an aggregator block

At the end of each epoch (every 1000th block), a special aggregator block is created. This block does not contain ordinary facts; instead, it contains:

• The final values of all entities that were changed during the epoch (essentially, the last assignments for each variable).

• The aggregator may also contain values of entities that were not changed during the epoch but are relevant at the time of aggregation (i.e., a complete snapshot of the system state). However, considering that an FOB can contain a vast number of entities, it is more practical to include only those entities that were changed in the given epoch. In this case, to obtain the current value of any entity, it will be necessary to find the last aggregator in which it was changed or read the value from the current epoch (if it has not yet been aggregated).

The aggregator also has two hashes that reference the corresponding hashes of the previous (999^{th}) block of the epoch (Fig. 7).

3. Cleaning of epoch blocks

After the aggregator is created and accepted by the network, the process of cleaning the blocks of this epoch (from the 1^{st} to the 999^{th}) begins. Cleaning consists of replacing the body of each block with a collision (essentially clearing all atomic facts in the block) in such a way that the block's chameleon hash remains unchanged. This is achieved due to the property of the chameleon hash function:

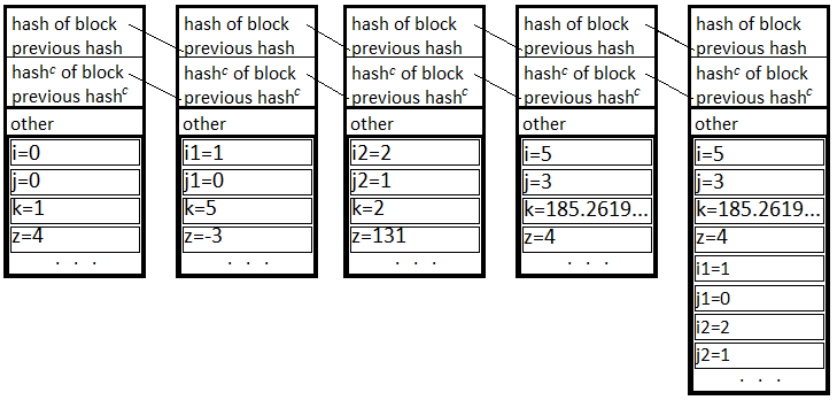

Fig. 7. PADHA – generation of an epoch aggregator block with a length of 5

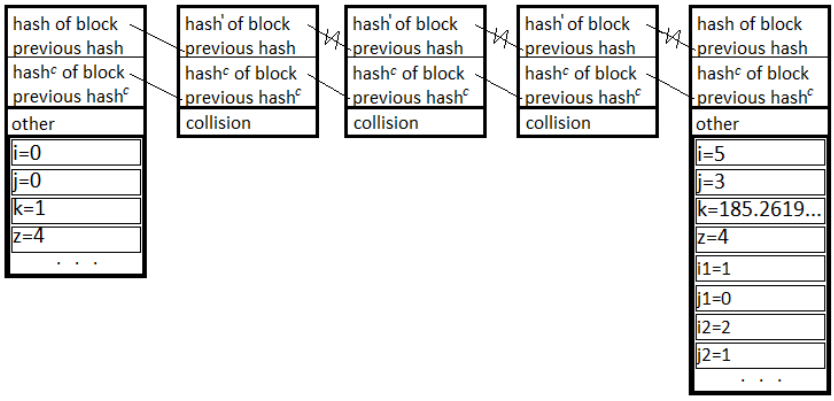

Fig. 8. PADHA – data deletion, chameleon hash correction, completion of an epoch with a length of 5

knowing a special key (which is publicly available in the system), one can find a new body such that the chameleon hash matches the original one. In this case, the regular hash of the block will change.

After cleanup, the connection between blocks within an epoch remains valid only via hash chameleons. The standard hashes no longer form a continuous chain. However, since we allow validation of blocks up to the aggregator using either of the two hashes, the chain remains valid.

4. Transition to the next epoch

After the epoch blocks are cleared, the new blocks (starting from block 1001) reference the aggregator (block 1000) and subsequent blocks accordingly via both hashes. Thus, the aggregator becomes the new reference point for the next epoch. It is important to note that the aggregator itself is not cleared (it remains full), as it contains the summary information necessary for state recovery.

5. Multiple aggregation and history compression

The process repeats every epoch. After creating the aggregator for epoch 2 (block 2000) and cleaning blocks 1001–1999, we are left with a chain consisting of:

- genesis block (full);
- cleaned blocks of epoch 1 (1–999);
- aggregator of epoch 1 (block 1000, full);
- cleaned blocks of epoch 2 (1001–1999);
- aggregator of epoch 2 (block 2000, full), and so on.

6. Recursive aggregation

To further reduce the data volume, recursive aggregation can be applied: after creating the aggregator for epoch 2, which also includes information from aggregator 1, the aggregator of epoch 1 can also be cleaned (replaced with an empty one while preserving its chameleon hash), because the final state of all entities by the end of epoch 2 will be contained in the aggregator of epoch 2.

Advantages of the PADHA method:

• Significant reduction of data volume: The bodies of blocks, except for the last aggregator, can be cleared. This provides a linear (or even logarithmic with recursive aggregation) reduction in the volume of stored history.

• Preservation of cryptographic integrity: Thanks to the use of chameleon hash functions, the block chain remains cryptographically linked, and chain validation is possible (via chameleon hashes).

• Support for PoW: Since block headers (including the chameleon hash) remain unchanged, the proof of work performed for each block remains valid. Mining new blocks is not disrupted.

• Decentralization: The method does not require trusted third parties or centralized archives. All nodes can independently perform the clearing, as the chameleon hash key is public. The chameleon key is public for all participants because, prior to epoch closure, blocks at the protocol level must maintain validity through both chameleon hash and standard hash links, and validation is performed against both hashes. Consequently, any unauthorized modification of a chameleon hash and its corresponding block body will disrupt the standard hash chain and be rejected by the network. After an epoch is closed, the blocks are "cleansed" (their bodies are emptied) at the protocol level, meaning they cannot legitimately contain data. Should data appear in such blocks, the network will also reject them.

• A malicious actor cannot create an alternative history without performing PoW for all blocks.

Limitations of the PADHA method:

• Requirement for a fact-oriented model: The method is only applicable to blockchains where data is represented as atomic facts, and the current state is determined by the last value. It is not suitable for transactional models where history is important.

Conclusion

In response to the limitations of existing methods, this paper proposes a novel method of PADHA, designed for FOBs. The exponential growth of data volume in PoW blockchain systems represents a fundamental challenge to their long-term sustainability and decentralization. The systematic analysis conducted in the article confirmed that traditional optimization methods – sharding, pruning of outdated blocks and hybrid consensus models – do not eliminate the key contradiction between the need to reduce "historical load" and preserve the basic properties of PoW: immutability, full chain auditability and distributed consensus. Each of these approaches introduces unacceptable compromises, whether it is the complication of synchronization and increased risks of attacks (sharding), the loss of the possibility for independent audit of the full history (pruning) or the conflict of consensus mechanisms (hybrid models). The PADHA method key innovation lies in the synergy of pruning principles and the controlled use of chameleon hash functions, ensures linear reduction in the volume of stored data through the periodic creation of aggregator blocks containing the system's final states, followed by cryptographically secure "cleansing" of the bodies of blocks from past epochs. This achieves the preservation of cryptographic chain integrity through dual hash anchoring and full support for the PoW mechanism, which distinguishes this approach from classical pruning or editable blockchains. However, the method's application necessitates introducing specific architectural constraints, narrowing its applicability to systems designed to store only final state data (facts) rather than complete transactional histories. While this represents a focused application domain, it remains broad enough to cover numerous practical use cases, such as registries, sensor data logging and state tracking. The significant gain achieved – dramatic data volume reduction without breaking cryptographic chain integrity – justifies this targeted approach.

Compared to existing methods, the proposed PADHA approach offers distinct qualitative advantages. Unlike sharding, it maintains a single coherent chain without cross-shard complexity. In contrast to classical pruning, it preserves cryptographic chain integrity via chameleon hashes. And unlike editable blockchains, it employs a controlled, protocol-level use of chameleon functions that does not undermine immutability for data. This enables the creation of scalable and energy-efficient PoW systems suitable for deployment in resource-constrained environments (IoT, mobile devices), without compromising their decentralization and security. The directions for further research are: the development of cryptographic proofs of the aggregators' correctness to minimize trust, the creation of adaptive protocols for selecting compression parameters, and an in-depth analysis of the method's resilience to new attack vectors.

Thus, the presented PADHA method offers a concrete path to overcoming the key limitation of PoW blockchains, demonstrating that data storage optimization is achievable not through the abandonment of fundamental principles, but through their adaptive evolution and the application of modern cryptographic primitives. This opens prospects for a new generation of decentralized systems that combine the robustness of PoW with the scalability demands of the Web 3.0 environment.

REFERENCES

1. **Yusupova D.R., Tazetdinova L.R.** The advantages of using blockchain technology in the insurance business. *Forum molodykh uchenykh* [*Forum of Young Scientists*], 2019, Vol. 33, No. 5, Pp. 1437–1442.

2. **Shomakhov A.R.** Use of blockchain technologies in the field of intellectual property rights. *Voprosy studencheskoi nauki* [*Student Science Issues*], 2020, Vol. 51, No. 11, Pp. 134–137.

3. **Marchenko K.Iu., Gimadeeva A.S.** Vnedrenie tekhnologii blokchein na mezhdunarodnyi rynok turistskikh uslug [Implementation of blockchain technology in the international tourism services market]. *Vestnik sovremennykh issledovanii* [*Bulletin of modern research*], 2018, Vol. 21, No. 6.4, Pp. 231–234.

4. **Kozin A.D., Makarova L.N., Buter A.P.** Realizatsiia tekhnologii blokchein na primere kriptovaliut [Implementation of blockchain technology using cryptocurrencies]. *Evraziiskii Soiuz Uchenykh* [*Eurasian Union of Scientists*], 2020, Vol. 81, No. 12, Pp. 10–17.

5. **Agamalian N.Kh.** Blokchein v auditorskoi deiatel'nosti [Blockchain in auditing activities]. *E-Scio*, 2020, Vol. 46, No. 7, Pp. 449–455.

6. **Rîndaşu S.-M.** Blockchain in accounting: Trick or treat? *Quality – Access to Success*, 2019, Vol. 20, Pp. 143–147.

7. **Bezzateev S.V., Fedorov I.R.** Blockchain technology in 5G networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, Vol. 20, No. 4, Pp. 472–484. DOI: 10.17586/2226-1494-2020-20-4-472-484

8. **Zheng Z., Xie S., Dai H., Chen X., Wang H.** An overview of blockchain technology: Architecture, consensus, and future trends. *2017 IEEE International Congress on Big Data* (*BigData Congress*), 2017, Pp. 557–564. DOI: 10.1109/BigDataCongress.2017.85

9. **Turginbayeva A.N., Tarabella A., Tanassoglo Y.** Use of blockchain technology in the highest education in Republic of Kazakhstan. *Bulletin of the Karaganda University*, 2018, Vol. 91, No. 3, Pp. 114–121.

10. **Kuo T.-T., Kim H.-E., Onho-Machado L.** Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association* (*JAMIA*), 2017, Vol. 24, No. 6, Pp. 1211–1220. DOI: 10.1093/jamia/ocx068

11. **Toilybayev A.Ye., Aitimov M.Zh., Bimuratkyzy Zh., Aitimova U.Zh.** Principles of the technology blockchain. *The Bulletin of Kazakh Academy of Transport and Communications named after M. Tynyshpayev*, 2017, Vol. 103, No. 4, Pp. 249–257.

12. **Sokolova T.N., Voloshin I.P., Petrunin I.A.** Pros and cons of the blockchain technology. *Ekonomicheskaia bezopasnost' i kachestvo* [*Economic Safety and Quality*], 2019, Vol. 34, No. 1, Pp. 49–52.

13. **Zharkova Yu.S.** Blockchain digital technologies: Advantages and disadvantages. *Zametki uchenogo* [*Scientist's Notes*], 2020, No. 10, Pp. 232–235.

14. **Dubovitskaya A., Xu Z., Ryu S., Schumacher M., Wang F.** Secure and trustable electronic medical records sharing using blockchain. *American Medical Informatics Association Annual Symposium Proceedings*, 2017, Pp. 650–659.

15. **Shafagh H., Burkhalter L., Hithnawi A., Duquennoy S.** Towards blockchain-based auditable storage and sharing of IoT data. *CCSW'17: Proceedings of the 2017 on Cloud Computing Security Workshop*, 2017, Pp. 45–50. DOI: 10.1145/3140649.3140656

16. **Ateniese G., Magri B., Venturi D., Andrade E.** Redactable blockchain – or – rewriting history in bitcoin and friends. *2017 IEEE European Symposium on Security and Privacy* (*EuroS&P*), 2017, Pp. 111–126. DOI: 10.1109/EuroSP.2017.37

17. **Kwon J., Buchman E.** Cosmos: A network of distributed ledgers, 2019. Available: https://resources.cryptocompare.com/asset-management/15/1662453504718.pdf (Accessed 10.05.2025).

18. **Kiayias A., Russell A., David B., Oliynykov R.** Ouroboros: A provably secure Proof-of-Stake blockchain protocol. *Advances in Cryptology – CRYPTO 2017*, 2017, Vol. 10401, Pp. 357–388. DOI: 10.1007/978-3-319-63688-7_12

19. **Reshi I.A., Sholla S.** IBF network: enhancing network privacy with IoT, blockchain, and fog computing on different consensus mechanisms. *Cluster Computing*, 2025, Vol. 28, Art. no. 208. DOI: 10.1007/s10586-024-05026-w

20. **Poon J., Buterin V.** Plasma: Scalable autonomous smart contracts, 2017. Available: https://plasma.io/plasma.pdf (Accessed 10.05.2025)

21. **Sharma A., Chauhan R., Gupta S., Kapruwan A.** Blockchain revolution in healthcare: A comprehensive survey. In: *Challenges in Information, Communication and Computing Technology* (eds. V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila), 2025, Pp. 218–223. DOI: 10.1201/9781003559085-38

22. **Maadallah Y., El Bouzekri El Idrissi Y., Baddi Y.** Enhancing IoT security through blockchain an in-depth analysis of the proof-of-work consensus mechanism. *EDPACS – Electronic Data Processing Audit, Control and Security*, 2025, Vol. 70, No. 5, Pp. 1–44. DOI: 10.1080/07366981.2025.2454095

23. **Ghorbian M., Ghobaei-Arani M.** Key concepts and principles of blockchain technology. *arXiv:2501.11707*, 2025. DOI: 10.48550/arXiv.2501.11707

24. **Bathini P.K., Manideep D.** Applications of artificial intelligence to blockchain consensus mechanisms: Using machine learning to make decentralized networks faster and more scalable. *2025 Global Conference in Emerging Technology* (*GINOTECH*), 2025, Pp. 1–6. DOI: 10.1109/GINOTECH63460.2025.11076838

25. **Soundararajan G., Tyagi A.K.** Blockchain technology: An introduction. In: *Blockchain Technology in the Automotive Industry* (eds. G. Yasin, A.K. Tyagi, T.A. Nguyen), 2025, Pp. 3–36. DOI: 10.1201/9781003450306

26. **Marwaha M., Bedi R.K., Gupta S.K.** An analysis of blockchain ecosystems: Understanding types, consensus models and security. *2025 12th International Conference on Computing for Sustainable Global Development* (*INDIACom*), 2025, Pp. 1–6. DOI: 10.23919/INDIACom66777.2025.11115579

27. **Niu G.** A Blockchain-based secure and privacy-preserving healthcare data management framework with SHA-256 and PoW consensus. *Informatica*, 2025, Vol. 49, No. 20, Pp. 149–162. DOI: 10.31449/inf.v49i20.8392

28. **Yu S., Qiao Y., Yang F., Bo J.** DPoW: A decentralized proof-of-work consensus mechanism for blockchain system. *Computer Networks*, 2025, Vol. 270, Art. no. 111490. DOI: 10.1016/j.comnet.2025.111490

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Daniil D. Razuvaev
Разуваев Даниил Дмитриевич
E-mail: Razuvaev_DD@mail.ru

Sergey M. Ustinov
Устинов Сергей Михайлович
E-mail: usm50@yandex.ru
ORCID: https://orcid.org/0000-0003-4088-4798

Submitted: 30.06.2025; Approved: 25.11.2025; Accepted: 19.12.2025.
Поступила: 30.06.2025; Одобрена: 25.11.2025; Принята: 19.12.2025.