

Article

BMIT: A Blockchain-Based Medical Insurance Transaction System

Jun Fei and Li Ling *

Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China;
jfei22@m.fudan.edu.cn

* Correspondence: lingli@fudan.edu.cn

Abstract

The Blockchain-Based Medical Insurance Transaction System (BMIT) developed in this study addresses key issues in traditional medical insurance—information silos, data tampering, and privacy breaches—through innovative blockchain architectural design and technical infrastructure reconstruction. Built on a consortium blockchain architecture with FISCO BCOS (Financial Blockchain Shenzhen Consortium Blockchain Open Source Platform) as the underlying platform, the system leverages FISCO BCOS's distributed ledger, granular access control, and efficient consensus algorithms to enable multi-stakeholder on-chain collaboration. Four node roles and data protocols are defined: hospitals (on-chain data providers) generate 3D coordinate hashes of medical data via an algorithmically enhanced Bloom Filter for on-chain certification; patients control data access via blockchain private keys and unique parameters; insurance companies verify eligibility/claims using on-chain Bloom filters; the blockchain network stores encrypted key data (public keys, Bloom filter coordinates, and timestamps) to ensure immutability and traceability. A 3D-enhanced Bloom filter—tailored for on-chain use with user-specific hash functions and key control—stores only 3D coordinates (not raw data), cutting storage costs for 100 records to 1.27 KB and reducing the error rate to near zero (1.77% lower than traditional schemes for 10,000 entries). Three core smart contracts (identity registration, medical information certification, and automated verification) enable the automation of on-chain processes. Performance tests conducted on a 4-node consortium chain indicate a transaction throughput of 736 TPS (Transactions Per Second) and a per-operation latency of 181.7 ms, which meets the requirements of large-scale commercial applications. BMIT's three-layer design ("underlying blockchain + enhanced Bloom filter + smart contracts") delivers a balanced, efficient blockchain medical insurance prototype, offering a reusable technical framework for industry digital transformation.

Keywords: blockchain technology; Blockchain-Based Medical Insurance Transaction System; medical insurance system; Bloom filter



Academic Editors: Andrea Prati and George Drosatos

Received: 30 June 2025

Revised: 27 August 2025

Accepted: 15 October 2025

Published: 17 October 2025

Citation: Fei, J.; Ling, L. BMIT: A Blockchain-Based Medical Insurance Transaction System. *Appl. Sci.* **2025**, *15*, 11143. <https://doi.org/10.3390/app152011143>

Copyright: © 2025 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Against the backdrop of the global digitalization wave, the traditional medical insurance system is facing increasingly severe structural challenges. In terms of information, there is a prominent "information silo" phenomenon among medical institutions, insurance agencies, and insured individuals. The electronic medical record systems of Grade A Class III hospitals, prescription management platforms of chain pharmacies, and underwriting databases of insurance agencies vary in data formats and storage standards due to being developed by different providers, leading to the formation of "data silos". For example,

when chronic disease patients apply for insurance, they need to visit various institutions (e.g., hospitals, medical examination centers, and pharmacies) multiple times to collect their past medical records, medication lists, and health examination reports. Compiling these materials alone can take 1–2 weeks. As insurance underwriters cannot access cross-institutional data in real time, they can only rely on manual review of paper documents submitted by insured individuals, leading to an underwriting cycle of 5–7 working days and an underwriting error rate exceeding 12% due to incomplete information [1,2].

In terms of trust mechanisms, the traditional insurance system heavily relies on the credit endorsement of centralized institutions, but the security vulnerabilities of centralized databases have emerged as a critical industry concern [3]. In 2024, a large insurance company was subjected to a hacker attack on its internal system, resulting in the leaking of sensitive information (e.g., names, ID numbers, and medical histories) of millions of policyholders, which directly triggered a class-action lawsuit. In the same year, Wuxi Hongqiao Hospital in Jiangsu province forged CT reports by copying imaging data of other patients, involving a sum of over CNY 50 million. In October, the Chongqing Municipal Public Security Bureau cracked the “3.10” special case, with the involved amount reaching as high as CNY 330 million, exposing significant latent risks in the traditional system in data security and fraud prevention [4].

As a distributed ledger technology, blockchain offers a revolutionary solution for reconstructing the medical insurance ecosystem, thanks to its core features such as the decentralized consensus mechanisms and encryption algorithms. Specifically, its decentralized nature has the potential to break the traditional centralized data storage model, enabling multiple stakeholders (including medical institutions, insurance agencies, and patients) to jointly maintain an immutable distributed ledger, thereby ensuring the consistency and authenticity of medical data. For example, each medical record and medication history of insured individuals can be stored on the chain in real time, and each participant can only view data within their authorized scope without the ability to tamper with it, fundamentally solving the problems of information asymmetry and data fraud [5–7].

To address the aforementioned issues, this paper proposes a Blockchain-based Medical Insurance Transaction System (BMIT) that reconstructs the medical insurance ecosystem through innovations in technological architecture, with the core goals of “efficient collaboration, security, trustworthiness, and privacy protection.” The specific implementation paths are as follows:

- Selection of underlying architecture: The FISCO BCOS consortium blockchain is adopted as the basic platform. Leveraging its distributed ledger, permission management, and optimized PBFT consensus mechanism, it supports multiple entities such as hospitals, patients, and insurance companies to collaboratively maintain data, ensuring that transactions are traceable and immutable.
- Design of roles and processes: The functional boundaries of the four parties are clarified. Hospitals are responsible for collecting medical data and generating 3D coordinate hash values for on-chain storage through an improved Bloom filter. Patients control data access permissions through private keys and unique verification parameters (keys). Insurance companies verify insurance eligibility and claim conditions based on on-chain Bloom filters. The blockchain network stores encrypted core data (such as user public keys, Bloom filter coordinates, timestamps, etc.), achieving “data availability without visibility.”
- Innovation in privacy protection: To address the false positive problem of traditional Bloom filters, they are upgraded to a 3D structure, storing only the 3D coordinates of medical data instead of original information. Combined with user-specific hash functions and key control, this not only reduces storage costs (only 1.27 KB for 100 records)

but also lowers the error rate to nearly 0 (a 1.77% reduction compared to the traditional scheme when processing 10,000 data entries).

- Support from smart contracts: Registration contracts, medical information data contracts, etc., are developed to automatically handle identity authentication, data on-chain, and verification processes. In a 4-node consortium chain test, the system achieved a transaction processing capacity of 736 TPS, with an overall process calculation overhead of approximately 181.7 ms, meeting the needs of large-scale applications.

Through the architectural design of “consortium blockchain + improved Bloom filter + smart contracts,” this study constructs a medical insurance transaction system that balances efficiency, security, and privacy, providing an implementable technical solution for the digital transformation of the industry.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the relevant literature, summarizing the research landscape of prior studies. Section 3 details the architecture of the BMIT model and the design of its operational workflow. Section 4 discusses the advantages of the blockchain platform adopted in this study and explains the rationale for selecting FISCO BCOS. Section 5 presents the improvements and applications of the Bloom filter in the BMIT system. Section 6 describes the design of key smart contracts in the BMIT system. Section 7 introduces simulation tests conducted on the enhanced Bloom filter, smart contracts, and the primary computational costs of the BMIT system. Finally, Section 8 summarizes the key findings, draws conclusions, and suggests directions for future research.

2. Related Work

Early researchers typically used blockchain directly to store all medical data as follows: Antwi M S, Adnane A, et al. introduced blockchain for medical data storage. They built a permissioned chain based on Hyperledger Fabric, supported the storage of unstructured data such as medical images, and ensured immutability through PBFT consensus [8]. Ismail L and Zeadally S introduced blockchain technology to verify medical insurance orders. They proposed a classification of 12 fraud scenarios, utilized a permissioned chain with PBFT consensus mechanism, and dynamically allocated verifiers via smart contracts (e.g., drug claims were co-verified by pharmaceutical companies and patients). NS3 (Network Simulator 3) simulations showed that performance declined linearly as the number of nodes scaled [9]. Wei Liu, Qinyong Yu, et al. adopted a multi-layer architecture including a cloud platform layer, network layer, and core layer, integrating multi-source data from 120 emergency command centers, public security agencies, etc. They achieved traceability of medical data and uniqueness verification of invoices through medical chains, third-party liability chains, and invoice chains [10]. Gokay Saldamli, Vamshi Reddy, et al. innovatively built a system based on BigchainDB, combining the Neo4j graph database to analyze patient–doctor–drug relationship networks for duplicate claim detection [11]. Anuj S S Mishra implemented multi-institutional permission management using Attribute-Based Signatures (MA-ABS), allowing patients to authorize data access and triggering claim payments automatically via smart contracts [12].

With the growing awareness of patient privacy protection, researchers gradually introduced cutting-edge cryptographic technologies to safeguard patient privacy. Houyu Zheng, Lin You, et al. integrated zk-SNARK, homomorphic encryption, and the Schnorr protocol, ensuring medical data authenticity through zero-knowledge proof in the data verification phase and hiding transaction amounts and patient information during the transaction privacy phase [13]. Lijing Zhou, Licheng Wang, et al. adopted (t, n) threshold secret sharing protocol and homomorphic encryption to support server-side operations on encrypted data [14]. Md Al Amin, Rushabh Shah, et al. designed a three-stage

multi-signature claim process (submission, approval, and confirmation), combined with Optimism Layer 2 to reduce transaction costs. However, the multi-signature mechanism increases process complexity, potentially prolonging claim processing time [15].

As blockchain technology continues to develop and be applied, more researchers have realized the preciousness of on-chain storage space and begun to introduce IPFS for separate storage of on-chain and off-chain data: Purshottam Purswani combined Ethereum smart contracts with IPFS to automatically trigger payouts based on parameter thresholds (e.g., blood pressure indicators), shortening claim time to within 5 min [16]. Aysha Alnuaimi, Amna Alshehhi, et al. used a private Ethereum chain and IPFS to store prescription files, achieving process automation through dual contracts (registration contract; approval contract), with smart contract approval time reaching the second level [17]. Namitha T S, Bipin Kumar Rai, et al. constructed a three-layer architecture with a blockchain layer, IPFS storage layer, and application layer, integrating Aadhaar identity verification to adapt to Indian scenarios, expecting to reduce administrative costs by USD 375 billion [18]. Xueping Liang, Juan Zhao, et al. based on Hyperledger Fabric and IPFS, used Merkle trees to improve data integrity verification efficiency, with 10,000 records verified in approximately 2500 ms [19]. Abrar Ahmed Raza, Bhavna Arora, et al. adopted PoS consensus to reduce energy consumption, combined with IPFS to store medical images, achieving 92% fraud detection accuracy [20]. However, the node admission mechanism of PoS (Proof of Stake) consensus may lead to power centralization, and applications of zero-knowledge proof only protect claim condition certification without covering end-to-end privacy protection.

In recent years, with the development of AI technology, many researchers have attempted to introduce AI algorithm models. Khyati Kapadiya, Fenil Ramoliya, et al. used ensemble learning (Bagging; Stacking) to improve the F1 score of fraud identification to 0.9862, with the blockchain storing claim records [21]. However, the high computational complexity of ensemble models may affect real-time detection efficiency, and the threshold setting criteria for dynamic early warning mechanisms are not specified. Shravan Kumar Joginipalli combined smart contracts with AI algorithms to detect abnormal transactions, reducing claim time by 37% and achieving 92% accuracy [22]. However, the cross-chain data synchronization efficiency of the hybrid chain architecture (private chain + consortium chain) was not tested, and security verification for data access from IoT devices was absent. Kruthika Alnavar and Dr. C. Narendra Babu analyzed medical data through Random Forest and SVM, with smart contracts automatically verifying claims, achieving 88% classification accuracy [23]. Wael El-Samad, Mirna Atieh, et al. adopted a dual-chain collaborative architecture (private chain + consortium chain), combined with IoT device data to trigger smart contracts, and machine learning to improve fraud detection accuracy [24]. However, the consensus consistency mechanism between the dual chains was not elaborated, and the compliance challenges of real-time fiat currency settlement (such as different national regulatory policies) were not mentioned. Khyati Kapadiya and Usha Patel proposed a four-layer architecture of blockchain and AI (user layer, data generation layer, data analysis layer, and blockchain layer), combining algorithms like LightGBM to identify fraud patterns [25], but did not provide specific algorithm parameter optimization processes, and the solutions to data standardization challenges remained at the theoretical level. Elhence A, Goyal A, et al. utilized ridge regression to dynamically calculate insurance premiums, employed random forest to assess risk levels, and implemented automated claim settlement through smart contracts, leading to a significant reduction in transaction costs [26]. However, this framework relies on the integrity of historical data, demonstrates insufficient adaptability to new fraud patterns, and does not elaborate on the interpretability solutions for machine learning models.

Some researchers have also attempted to develop consensus mechanisms more suitable for medical insurance. For example, Alhasan B, Qatawneh M, et al. proposed a consortium chain consensus algorithm based on FIFO (First-In-First-Out) and result length, which achieves a verification time of 785 ms for 10,000 transactions and a 100% tampering detection rate [27]. However, this consensus algorithm relies on preselected verification nodes, potentially posing risks of node collusion. Additionally, the lightweight design does not clarify how to balance decentralization and security.

With recent advancements in distributed digital identity (DID) technology, scholars have begun exploring its applications in the medical insurance industry. Meriem Guerar, Mauro Migliardi, Enrico Russo, and colleagues proposed the SSI-MedRx system, which leverages blockchain technology and Self-Sovereign Identity (SSI) to ensure cross-border interoperability, protect patient privacy, and prevent various types of medical fraud [28]. Gunjan Kumar Chouhan, Shubhangi Singh, Aanya Jaduvansy, and colleagues focused on the challenges in Electronic Health Record (EHR) management, highlighting that traditional cloud-based EHR systems suffer from centralized control, privacy threats, limited interoperability, and vulnerability to data breaches. To address these issues, they proposed a patient-centric blockchain-based EHR management system, integrating Ethereum smart contracts, Decentralized Identifiers (DIDs), and the InterPlanetary File System (IPFS) [29].

Table 1 summarizes the main methods used in several major categories, their key findings, and the limitations of previous studies.

Table 1. The main methods, key results, and limitations of previous studies in Related Work.

Methods	Key Outcomes	Limitations
Directly storing all medical data ^a	Support the storage of various unstructured data and introduce blockchain to ensure immutability	Excessive on-chain data storage and storing plaintext cannot protect user privacy
Introduce cutting-edge cryptographic technologies ^b	zk-SNARK, Homomorphic Encryption, Schnorr Protocol, and Three-Phase Multi-Signature Claim Process	There has been significant progress in terms of privacy protection, but the issue of excessive on-chain data remains
Introduce IPFS ^c	Integrate with the IPFS system to achieve collaborative storage of on-chain and off-chain data, reducing the burden on the blockchain.	The problem of excessive on-chain data has been solved, but the IPFS system cannot guarantee data persistence
Introduce AI algorithm model ^d	Enhance the accuracy of fraud detection through ensemble learning, machine learning, etc.	It is highly dependent on datasets, mainly providing fraud identification, and fails to leverage the advantages of blockchain
Improve the consensus mechanism to be more suitable for medical insurance ^e	Propose a consortium blockchain consensus algorithm based on FIFO and result length	This consensus algorithm relies on pre-selected verification nodes, which may face the risk of node collusion. Moreover, its lightweight design fails to clarify how to balance decentralization and security
Introduce distributed digital identity ^f	Decentralized Identifiers (DIDs), SSI-MedRx System	The introduction of DID enables permission control, but it lacks encryption methods to protect privacy and does not process on-chain data

a: References [8–12]; b: References [13–15]; c: References [16–20]; d: References [21–26]; e: Reference [27]; f: References [28,29].

3. Design of Blockchain-Based Medical Insurance Transaction Model

3.1. Design Goal

The design objectives of BMIT focus on reconstructing medical insurance transaction processes through blockchain technology to address core pain points in traditional models, while incorporating innovative applications of Bloom filters and smart contracts, thereby achieving efficient, secure, and transparent medical data management and automated insurance transactions. The specific design goals are as follows:

- **Fraud Prevention and Transparency Enhancement:** Transaction data should be recorded on the blockchain in chronological order, leveraging blockchain and cryptographic technologies to ensure data transparency, immutability, traceability, verifiability, and censorship resistance. This guarantees the tamper-proof nature of health data and transaction records, preventing insurance companies or hospitals from falsifying data;
- **Privacy Protection and Data Minimization:** Bloom filters store hashes of health information rather than raw sensitive data (e.g., hypertension diagnoses) on the blockchain, retaining only the capability for existence verification. This reduces the risk of privacy leakage by minimizing the exposure of sensitive information;
- **Efficient Querying:** Leveraging the efficient querying features of Bloom filters—specifically low storage requirements and fast verification—smart contracts facilitate rapid result queries, support high-concurrency transaction processing, and meet the needs of large-scale user bases.

In the subsequent sections of this paper, we will demonstrate, through detailed explanations of the design of each system component, how BMIT achieves the design objectives established in this section.

3.2. Why Blockchain Is Needed

Why choose a blockchain network? The core of the answer lies in its tamper-proof nature. This feature is not merely a simple technical gimmick but reshapes the trust foundation of data interaction from the underlying logic, bringing revolutionary changes to numerous fields in the digital era.

The tamper-proof property of blockchain stems from its unique technical architecture. Each data record is packaged into a “block” and closely linked to the previous block through an encryption algorithm, forming a chain-like structure. When an attempt is made to modify the information in a specific block, it is necessary not only to tamper with the block itself but also to synchronously modify the encryption verification information of all subsequent blocks—which is almost an impossible task in a blockchain network with distributed storage. Since each node in the network holds a complete copy of the ledger, any malicious operation at a single point will be identified and rejected by other nodes, ultimately leading to the failure of the modification. This “collective supervision” mechanism ensures that once data is written, it is akin to being carved into rock, making it difficult to tamper with or delete privately. The value of this feature is particularly prominent in scenarios with extremely high requirements for data authenticity. For example, in financial transactions, the tamper-proof nature of blockchain can eliminate issues such as false transactions and accounting fraud. Every transfer record is clearly traceable and non-repudiable, significantly reducing the risk of financial fraud. In supply chain management, full-process data from raw material procurement to finished product circulation is permanently recorded. Once a quality problem arises, it can be quickly traced back to the specific link, avoiding responsibility shirking. In intellectual property protection, creators can record their work information on the blockchain; the timestamp and tamper-proof

records serve as the most powerful proof of ownership, effectively addressing the pain point of “easy forgery of evidence” in infringement disputes.

More profoundly, the tamper-proof nature is redefining how “trust” is generated. In the past, people often relied on intermediary institutions (e.g., banks, notary offices) to guarantee data authenticity, which not only increased process costs but also carried the risk of intermediaries making errors or breaking promises. However, blockchain has established a “code as trust” mechanism through technology itself, enabling unfamiliar entities to collaborate directly based on tamper-proof data without expending extra time and resources on verifying information authenticity. This “disintermediated” trust model is gradually being implemented in fields such as government affairs, healthcare, and law—electronic contracts cannot be tampered with after being uploaded to the chain; government affairs data avoids human operational errors after being on-chain; and medical records ensure the authenticity and integrity of patient information after being on-chain.

It is this ability to fully guarantee data authenticity at the technical level that makes blockchain networks a “trusted infrastructure” in the digital era. It addresses the inherent issue of “easy data manipulation” in traditional centralized systems and provides more secure, efficient, and low-cost collaboration possibilities for various industries. Choosing a blockchain network essentially means choosing a more reliable, transparent, and credible mode of data interaction—this is the key to coping with the complex challenges of the digital era and the fundamental reason why this system selects blockchain as its cornerstone.

3.3. System Model

To complete the entire process of medical insurance purchase, the BMIT model mainly involves four roles: hospitals, patients (clients), the blockchain network, and insurance companies. As shown in Figure 1, each time a patient visits a hospital for treatment or medical examination, medical data is generated. The hospital processes the generated data to obtain a hash value and stores this hash value in the Bloom filter on the blockchain. When the patient purchases medical insurance from an insurance company, the insurance company constructs query terms, generates a hash value using the same hashing rules, and then verifies the existence of a matching item through the Bloom filter. If a match is found, the smart contract is triggered to reject the transaction; otherwise, the policy is allowed to be issued.

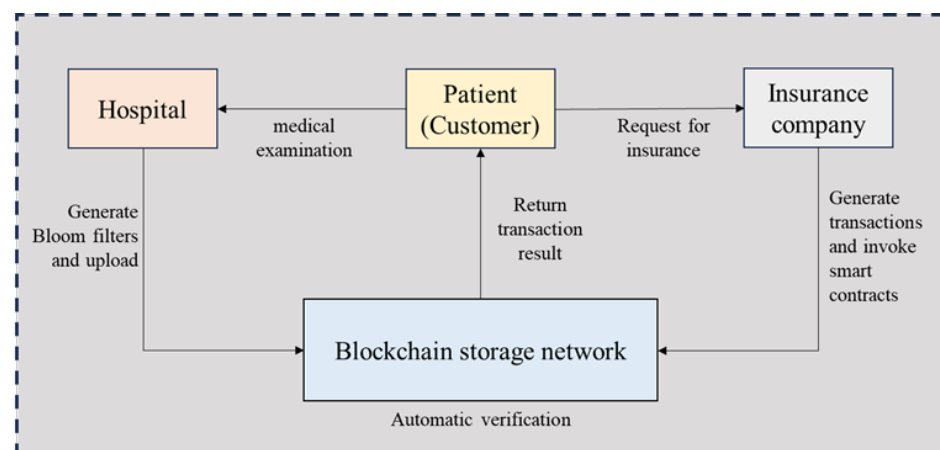


Figure 1. System model of the BMIT.

In the BMIT model, insurance companies cannot directly access users’ medical data. Instead, they can only construct terms and use Bloom filters to verify whether users have a specific disease specified in the terms. However, if traditional Bloom filters are used directly, false positives may occur—i.e., when the judgment result indicates “existence,” the disease

may not actually exist, whereas a “non-existence” judgment is absolutely reliable. We have improved the Bloom filter to eliminate false positives. Based on the improved Bloom filter, insurance companies can clearly determine whether users have specific diseases without accessing their complete medical data.

Meanwhile, since the data stored on the blockchain is a Bloom filter constructed through a specific method—with the key to this construction method held by users—even if network attackers obtain the data stored in the blockchain, they cannot convert it into users’ medical information. This significantly enhances the security of users’ private data.

The main functions of each role in the BMIT model are as follows:

- **Hospitals:** Responsible for collecting patients’ health data (e.g., hypertension; diabetes), processing it through a hash function, storing it in a Bloom filter, and uploading the data digest to the blockchain;
- **Patients (Customers):** Entities assuming different roles when interacting with different parties, holding a unique private key and verification parameter;
 - As patients: After collaborating with hospitals to collect health data, they provide the unique verification parameter to assist hospitals in data processing, then use the unique private key to sign off, facilitating hospitals to upload the processed data to the blockchain;
 - As customers of insurance companies: Provide the unique verification parameter to help insurance companies validate eligibility criteria.
- **Blockchain Network:** Serving as the core layer, it stores the set of 3D coordinates processed by the improved Bloom filter and provides an environment for data verification and transaction execution through smart contracts;
- **Insurance Companies:** Initiate query requests via smart contracts, using the Bloom filter to determine whether patients’ health conditions meet the criteria for insurance application/claim settlement.

The workflow of the BMIT is described as follows:

1. **Registration of Roles in the Blockchain Network:** First, patients, hospitals, and insurance companies register in the blockchain network. The blockchain network assigns each role a unique public key address pb and a unique private key pr . The public key serves as the role’s unique account, while the private key is stored independently by each role. For users, a unique verification parameter key_h is generated during registration, encrypted with the user’s public key, and stored on the blockchain network to prevent malicious tampering. Since only the user’s private key can decrypt the ciphertext, only the user holding the private key can access the relevant key.
2. **Interaction Between Patients and Hospitals:** Patients (customers) undergo health checks at hospitals in daily life, and each check generates a medical data term m (e.g., “hypertension”). The process is as follows:
 - (a) The hospital and the patient (customer) jointly retrieve the unique verification parameter key_h stored in the blockchain network;
 - (b) The term m is processed through a specialized Bloom filter based on key_h , generating a string bf containing the 3D coordinate set of the health check term;
 - (c) The patient (customer) uses their unique private key pr_p and a timestamp t to perform an SM2 digital signature on bf , obtaining the signature $sign_p$;
 - (d) The hospital uses its unique private key pr_h and timestamp t to sign bf again, generating the second signature $sign_h$;
 - (e) The hospital uploads bf , the two signatures $sign_p$ and $sign_h$, and the public keys pb_p and pb_h to the blockchain network;

- (f) The blockchain network first checks the timestamp t : if t is earlier than the stored timestamp, the request is rejected as invalid or overdue; if t is later, the network verifies the signatures using the public keys. Upon successful verification, bf and t are saved under the patient's unique blockchain account.

3. Interaction Between Patients and Insurance Companies:

- Insurance Purchase:** When a patient (customer) applies for insurance, the insurance company conducts the following:
 - Collaborates with the user via a smart contract to retrieve key_h and the 3D coordinate set string bf from the blockchain;
 - Generates health data parameter terms v based on the policy requirements, using the user's specialized Bloom filter to verify compliance. For example, if a query for "hypertension" in bf returns a match, the insurance application is rejected; otherwise, an electronic policy is issued.
- Claims Settlement:** When a patient (customer) files a claim, they do not need to manually provide medical records. The insurance company verifies the Bloom filter stored on the blockchain: if the required medical condition exists in bf , the claim is approved; otherwise, it is rejected.

The complete workflow of the BMIT system is illustrated in Figure 2. This workflow integrates SM2 digital signatures, Bloom filters, and the immutability of blockchain to enable secure, privacy-preserving medical insurance transactions.

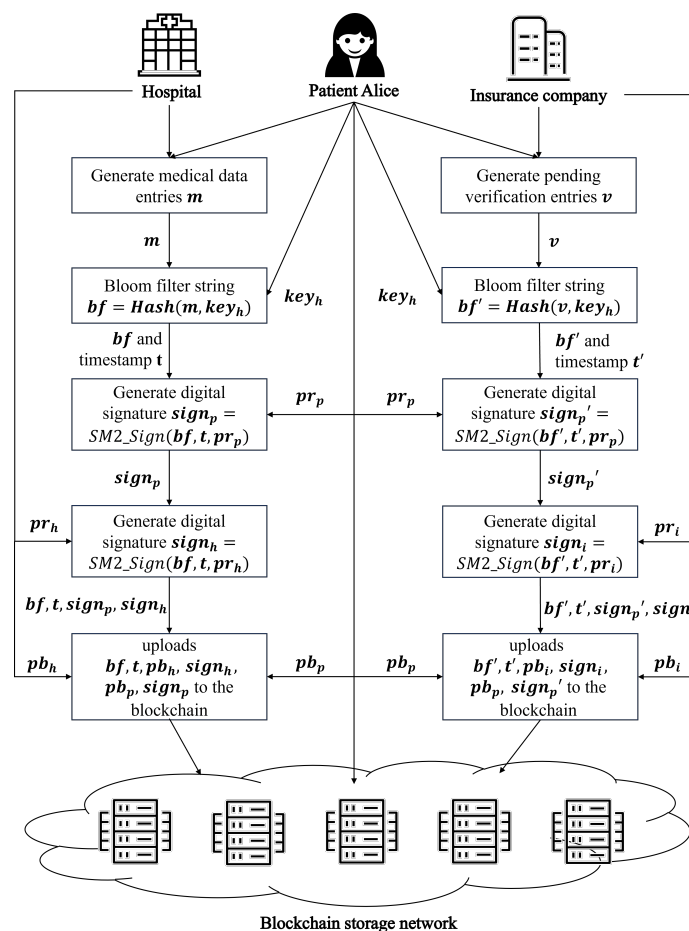


Figure 2. The workflow of the BMIT.

3.4. BMIT Data Storage Solution

As a peer-to-peer (P2P) network-based distributed ledger system, the blockchain achieves ledger consistency through redundant storage of transaction history across all network nodes—each node must maintain a complete record of all transactions since the genesis block (Bitcoin full nodes now store 430 GB of data with an annual growth rate of 19%; Ethereum state data exceeds 1.2 TB, with gas-consuming transactions accounting for 37%). This full-replica mechanism grants the system core advantages in resisting single points of failure and data tampering (MIT experimental data shows that altering a Bitcoin transaction requires 6.7×10^{24} hash computations, costing over USD 250 billion). However, it also introduces three critical engineering challenges:

- **Rigid Storage Cost Growth:** For Amazon EC2 nodes, storing 1 TB of blockchain data annually costs USD 276. A network with tens of thousands of nodes would incur total storage expenses exceeding USD hundreds of millions.
- **Node Synchronization Efficiency Degradation:** Bitcoin’s median block propagation delay reaches 12.6 s, while new nodes require 72 h to complete initial block synchronization (under 10 Mbps bandwidth conditions).
- **State Verification Performance Bottlenecks:** Ethereum’s global state tree contains 140 million account nodes, with average query latency increasing from 8 ms in 2016 to 53 ms currently.

To address this challenge, when designing blockchain systems, we must adhere to the ‘minimum necessary data on-chain’ paradigm, carefully selecting the data to be stored on-chain to avoid imposing excessive burdens on the blockchain network. Based on the aforementioned system requirements, the following core datasets will be stored on-chain:

- **Patient pb_p :** A unique public key issued to the patient upon registration in the blockchain system. It serves as the primary key for on-chain data storage, enabling data retrieval. Additionally, as a public key, it validates SM2 digital signatures uploaded by users (e.g., verifying whether the signature of user-uploaded data is generated by the legitimate user);
- **Bloom Filter String bf :** This core data component stores 3D coordinate values of medical terms processed by the 3D Bloom filter, deliberately avoiding storage of the complete filter structure. The design leverages inherent sparsity in real-world applications, where individual patients typically trigger fewer than 100 active markers in the filter space. By storing only populated coordinates, the system achieves an over 99% storage reduction while preserving essential verification capabilities;
- **Timestamp t :** A timestamp (e.g., Unix epoch time with millisecond precision) recorded when the user uploads the Bloom filter string at the hospital, enabling chronological identification of data versions;
- **Encrypted key_h :** The ciphertext of a user’s unique verification parameters, generated by encrypting the verification parameters directly with the user’s public key. Only the holder of the corresponding private key can decrypt this ciphertext to retrieve the original verification parameters. Patients subsequently provide the decrypted verification parameters to hospitals or insurance providers, enabling these entities to generate the corresponding Bloom filter string from the patient’s medical data;
- **Hospital pb_h :** A unique public key issued to the hospital upon blockchain registration. Combined with the timestamp t , it forms a composite identifier for medical examination records, facilitating future data audits.

Figure 3 illustrates the on-chain data schema and memory layout of the Bloom filter string.

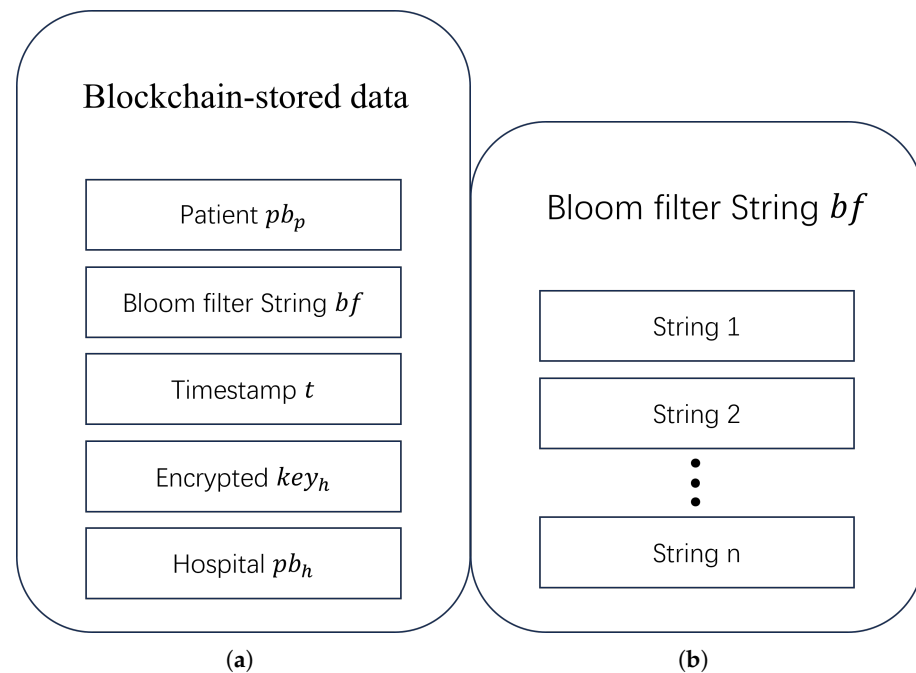


Figure 3. (a) Schematic diagram of data components stored in the blockchain. (b) Schematic diagram of the memory layout of the Bloom filter string.

4. Blockchain Network Platform

Since the official launch of the Bitcoin network in 2009, blockchain technology—its underlying core—has embarked on a long evolutionary journey from marginal innovation to mainstream exploration. Initially, its application scenarios were strictly confined to the field of digital currencies. Bitcoin, relying on its decentralized ledger system, immutable transaction records, and peer-to-peer value transfer capabilities, achieved global financial transactions without intermediaries for the first time. At this stage, blockchain was more like a “dedicated tool” serving the sole purpose of currency.

However, with the deepening of technical understanding and the expansion of industry demands, the potential of blockchain has gradually been explored. Developers have found that its core features—such as “distributed accounting,” “consensus mechanisms,” and “smart contracts”—can completely break away from the scope of currency and be applied to broader social and economic scenarios. Starting with the launch of the Ethereum network in 2015, which introduced smart contract functionality, blockchain technology has officially entered the stage of “multi-domain penetration”: the financial sector uses it to build cross-border payment systems to reduce settlement costs; the supply chain sector uses it to realize commodity traceability to improve process transparency; the government affairs sector uses it to store electronic certificates to ensure data authenticity. Nowadays, blockchain has evolved from a single carrier for currency applications into one of the important infrastructures supporting the development of the digital economy.

4.1. Classification of Blockchain Platforms

To adapt to the differentiated needs of various scenarios—such as the demand for “high performance” in financial transactions, the emphasis on “high security” for government data, and the pursuit of “low energy consumption” in IoT scenarios—the blockchain network architecture has been continuously iterated, forming a diversified system. In terms of the degree of decentralization, there are public blockchains that pursue “complete decentralization,” such as Bitcoin and Ethereum; private blockchains oriented to internal scenarios of enterprises or institutions, which are only open to authorized nodes; and

consortium blockchains (e.g., Hyperledger Fabric) that balance openness and controllability. In terms of technical characteristics, there are Layer 2 solutions focused on improving transaction throughput (e.g., Ethereum's Rollup) and vertical domain chains customized for specific industries (e.g., energy blockchains and medical blockchains).

At present, mainstream blockchain network platforms that have been verified by the market and recognized by the industry can be mainly divided into the following categories:

- **Public Chain Platforms:** Represented by Ethereum, they support the deployment of smart contracts and the development of decentralized applications (DApps) and are the most active blockchain networks in the global developer ecosystem. As the earliest public chain, Bitcoin remains the cryptocurrency network with the highest market value, with its security and degree of decentralization widely recognized. Additionally, emerging public chains such as Solana and Avalanche have risen rapidly in decentralized finance (DeFi) and non-fungible token (NFT) fields, relying on their high TPS (transactions per second) performance;
- **Consortium Chain Platforms:** Managed jointly by multiple pre-selected organizations, nodes require authorization to join, and read/write permissions are customized according to alliance rules, making them suitable for inter-institutional collaboration (e.g., finance and supply chains). For example, Hyperledger Fabric—led by the Linux Foundation—is specifically designed for enterprise-level applications, supporting permission management and flexible consensus mechanisms, and is widely used in scenarios such as supply chains and cross-border trade. R3 Corda focuses on the financial services sector and has become the preferred solution for inter-bank settlement, insurance claims, and other scenarios through its unique “peer-to-peer transaction privacy protection” technology. China's consortium chain platforms, such as Ant Chain and WeBank's FISCO BCOS, have implemented numerous practical applications in fields such as government services and digital copyright;
- **Private Blockchain Platforms:** controlled by a single organization, node access and permissions are highly centralized, making them suitable for scenarios such as internal enterprise audits and government management (e.g., digital currency systems of some central banks).

These mainstream platforms, each leveraging their unique technical advantages, serve as infrastructure in different scenarios.

4.2. Advantages of Consortium Blockchains

As a key branch of blockchain architecture that balances “controllability” and “collaborativeness,” consortium blockchains demonstrate significant advantages in specific scenarios compared to public and private blockchains, particularly adapting to enterprise-level and inter-institutional collaboration needs. Their core advantages are as follows:

- **In terms of permissions:** The core feature of public blockchains is “full decentralization,” where any node can freely join or exit the network, and data is visible to all participants. While this ensures extreme decentralization and transparency, it increases the risk of intrusion by malicious nodes from unknown entities and fails to meet institutions' requirements for data privacy (e.g., corporate business data; sensitive government information). Private blockchains, by contrast, are completely closed, with all nodes controlled by a single institution—resulting in high data privacy and centralized permissions. However, this sacrifices blockchain's core value of “multi-node collaboration,” making them more akin to traditional distributed databases. Consortium blockchains, by contrast, require nodes to be authorized by consortium institutions (e.g., an inter-bank consortium blockchain only allows member banks to access it). Unauthorized nodes cannot participate in ledger maintenance or view sensitive data. Meanwhile, consortium

nodes jointly maintain the ledger, avoiding the “single-entity control risk” of private blockchains and addressing the “privacy leakage risks” of public blockchains through permission management. This makes them particularly suitable for multi-institutional collaboration scenarios (e.g., upstream and downstream enterprises in supply chains; cross-departmental government collaboration);

- **In terms of performance and efficiency:** To ensure decentralization and security, public blockchains typically adopt inefficient consensus mechanisms such as “Proof of Work (PoW)” (e.g., Bitcoin supports only 7 transactions per second). Additionally, all nodes must synchronize full data, leading to slow transaction confirmation and high costs (e.g., soaring transaction fees during Ethereum network congestion). Consortium blockchains significantly improve performance by optimizing consensus mechanisms (e.g., adopting “Practical Byzantine Fault Tolerance (PBFT)” or “RAFT”). Since the number of nodes is limited and trusted (consortium institutions are usually known and reputable), there is no need for computing power competition to verify transactions. Transaction confirmation time can be shortened to seconds or even milliseconds, with transactions per second (TPS) reaching thousands (e.g., Hyperledger Fabric can achieve tens of thousands of TPS). Furthermore, consortium blockchains can flexibly design data storage strategies (e.g., synchronizing only key transaction information instead of the full ledger), further reducing node burdens and meeting the needs of high-frequency transaction scenarios (e.g., cross-border payments; real-time updates of supply chain logistics information);
- **In terms of operation and maintenance:** Public blockchains rely on “miners” or “validator nodes,” and their incentive mechanisms (e.g., Bitcoin mining rewards) result in high operational costs (e.g., electricity and hardware investments). Moreover, the threshold for ordinary users to participate in ledger maintenance is extremely high (requiring meeting computing power or token staking requirements). Consortium blockchains have a limited number of nodes (usually dozens to hundreds) with operational costs jointly borne by consortium institutions, without reliance on external incentives. Hardware and network investments for nodes can be flexibly configured according to consortium needs, significantly lowering the technical threshold for multi-institutional collaboration (e.g., small and medium-sized enterprises joining industry consortium blockchains do not need to build complex infrastructure themselves);
- **In terms of rules:** Public blockchain rules (e.g., consensus mechanisms; smart contract logic) are difficult to modify once deployed and require “hard forks” for upgrades—demanding consensus from all network nodes, resulting in extremely poor flexibility. Consortium blockchains can flexibly customize technical rules based on the joint needs of consortium institutions: for example, adjusting consensus mechanisms to adapt to different trust environments (e.g., using PBFT for low-trust scenarios and RAFT for high-trust scenarios); designing differentiated data permissions (e.g., core transaction data visible only to key consortium nodes, while basic logistics information is public to all nodes); and even supporting dynamic upgrades of smart contracts (without forking). This customization capability allows them to precisely match industry scenarios (e.g., medical consortium blockchains need to protect patient privacy while enabling hospitals to share diagnostic data).

In conclusion, public blockchains excel in “decentralization and absolute transparency,” making them suitable for transactions between strangers without trusted intermediaries (e.g., cryptocurrencies). Private blockchains excel in “extreme privacy and efficiency,” suitable for internal management of a single institution (e.g., enterprise internal audits). Consortium blockchains, however, through the combined advantages of “authorized access, efficient collaboration, privacy protection, and regulatory compliance,” serve as an

intermediate form connecting “centralization” and “full decentralization.” They perfectly adapt to multi-institutional, multi-entity collaboration scenarios and act as the core carrier for blockchain technology to transition from “conceptual experimentation” to “industrial implementation” (e.g., large-scale applications in cross-border trade, digital government, and supply chain finance mostly rely on consortium blockchain architectures).

When designing the system discussed in this paper, fully considering practical needs is crucial to ensuring its stable, efficient, and secure operation. Below, we analyze the system’s practical requirements from several key dimensions:

1. The precise division of access permissions is the primary link in ensuring system information security. Given that the system processes data potentially involving hospital diagnosis records, patients’ personal health information, and insurance companies’ business data—all of which are highly sensitive and private—any leakage could cause incalculable losses to relevant parties. Therefore, access permissions must be strictly and meticulously divided. Specifically, only hospitals, patients, and insurance companies that have formally registered and completed identity verification are granted system access. This mechanism effectively blocks unauthorized access at the source, minimizing the risk of privacy leaks. For example, hospitals must submit relevant qualification documents during registration to confirm their legitimate medical service qualifications; patients must complete real-name registration and identity verification to protect their legal rights to health information; insurance companies must provide corresponding business license documents to ensure their access complies with industry norms and legal regulations;
2. Second, the system’s operational efficiency is critical to meeting large-scale application scenarios. With the accelerating digitalization of the medical industry and the continuous expansion of insurance services, the system will inevitably face massive transaction processing demands in the future, placing high requirements on its transactions per second (TPS). While public blockchains offer decentralization and transparency, their consensus mechanisms often result in slow processing speeds when handling large-scale transactions, failing to meet high-concurrency business needs. Consortium blockchains, by contrast—managed jointly by pre-selected nodes—maintain a certain degree of decentralization while significantly improving transaction efficiency through optimized consensus mechanisms, achieving higher TPS. This enables consortium blockchains to better handle potential large-scale transaction scenarios, ensuring stable operation even under high loads. Therefore, considering the system’s efficiency requirements, consortium blockchains are more suitable than public blockchains for this system.

4.3. FISCO BCOS

FISCO BCOS, as China’s leading open-source consortium blockchain platform, was jointly developed and open-sourced by WeBank and multiple institutions. Its core mission is to provide a secure, efficient, and user-friendly blockchain infrastructure for multi-organization collaboration scenarios. Unlike public blockchains designed for open environments, FISCO BCOS focuses on the core needs of consortium chains—grounded in “controlled node access and governance,” it is better suited for data sharing and business coordination among specific organizations or institutional groups. Its key advantages span five dimensions:

- **Security:** It fully adopts national cryptographic algorithms such as SM2 and SM3, strictly complies with national cryptographic standards, and provides high-strength encryption protection for the entire process of data transmission, storage, and transac-

tions. It is particularly suitable for fields with extremely high security requirements, such as healthcare and finance;

- **Performance:** Through technological innovations such as the optimized PBFT consensus mechanism and parallel computing, it can support high-concurrency transaction processing, effectively meeting the efficiency needs of large-scale business scenarios;
- **Flexibility:** It provides rich smart contract interfaces and development tools, facilitating developers to quickly build and deploy applications according to actual business needs. Additionally, it supports dynamic node expansion and flexible permission configuration, enabling adaptation to collaboration modes of organizations of different scales;
- **Cost advantage:** As a fully open-source platform (with code hosted on GitHub), enterprises do not need to pay authorization fees, and the supporting full-link toolchain significantly reduces the development threshold;
- **Ecological and implementation capabilities:** It has one of the most active consortium blockchain ecosystems in China. A community of over 100,000 developers provides detailed documents, tutorials, and offline training support. To date, it has been implemented in over 300 cases across fields such as finance (supply chain finance; cross-border payment), government affairs (electronic certificates; judicial evidence storage), and public welfare (medical data sharing; charity traceability), verifying its adaptability in actual business.

Based on the above advantages, we have decided to adopt FISCO BCOS as the blockchain network platform for this research.

5. Bloom Filter in BMIT

5.1. The Traditional Bloom Filter

The Bloom filter was proposed by Burton Howard Bloom in 1970. It is essentially a long bit array combined with a series of random hash functions. This probabilistic data structure is primarily used to determine whether an element belongs to a set. It offers significant advantages: fast operation (high time efficiency) and low memory consumption (high space efficiency). However, it has a certain false positive rate. Crucially, the Bloom filter's errors are exclusively false positives, meaning it will only misclassify elements not in the set as being present. It will never produce false negatives (misclassifying elements in the set as absent).

The traditional Bloom filter, whose structure is roughly illustrated in Figure 4, is built using a bit array and multiple hash functions. During initialization, it creates an m -bit array, with all bits initially set to 0.

- **Adding Elements:** To add an element to the Bloom filter, k distinct hash functions are applied to it. Each hash function produces an index value between 0 and $m-1$, and the corresponding bits at these index positions in the bit array are set to 1. For example, if the first hash function yields index 10 and the second yields index 25, bits 10 and 25 in the array are set to 1. This process uses multiple hash functions to map the element onto the bit array from different perspectives, creating a distributed representation;
- **Querying Elements:** To check if an element is present, the same k hash functions are applied to it, yielding k index values. The Bloom filter then checks if all the bits at these index positions in the array are set to 1:
 - If all corresponding bits are 1, the Bloom filter considers the element possibly present in the set;
 - If any bit is 0, it can definitively conclude the element is not in the set.

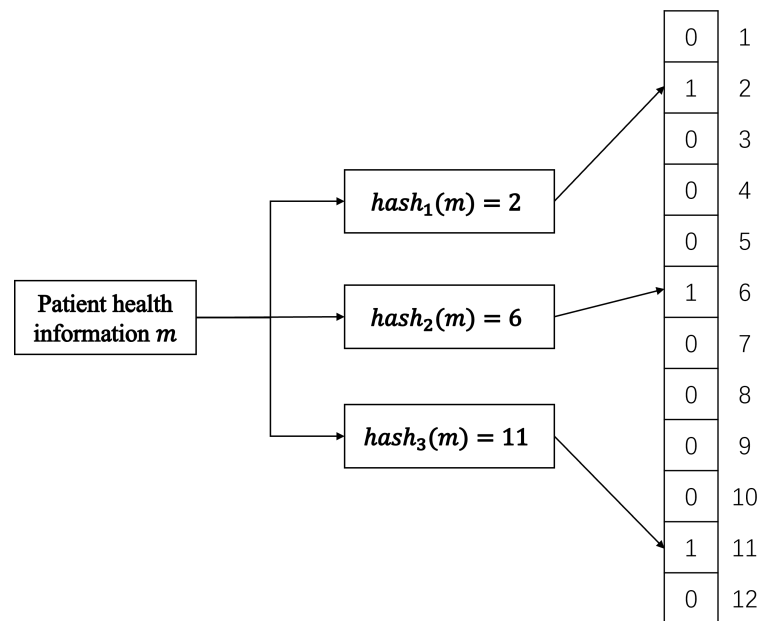


Figure 4. Traditional bloom filter basic structure diagram.

This operational mode introduces the possibility of false positives: when all checked bits are 1, the element may not actually be in the set. This occurs because different elements can set the same bits to 1 due to hash collisions. However, the Bloom filter never produces false negatives: if any checked bit is 0, the element is guaranteed to be absent from the set.

5.2. Improved Bloom Filter for BMIT

To reduce the false positive rate, improvements to the traditional Bloom filter are necessary. Based on the principle of the Bloom filter described earlier, its false positives fundamentally stem from hash collisions between multiple elements. Although it uses multiple hash functions to reduce the probability of complete collisions between any two elements, compressing all information into a single one-dimensional bit array inevitably leads to partial index collisions among multiple elements. Cumulative collisions can potentially cover all positions corresponding to a specific element, resulting in a false positive.

The traditional solution to this problem is to increase the length of the bit array and the number of hash functions according to actual needs. While this preserves the Bloom filter's advantage of fast retrieval, it significantly increases computational overhead and storage requirements. For blockchain systems, maintaining a large Bloom filter incurs relatively significant overhead. Now, let us consider the specific requirements for the Bloom filter in this study:

1. Since we are maintaining medical information for individual patients, a single patient cannot have an excessive number of disease terms. Therefore, an extremely large indexing space is unnecessary;
2. Given the involvement of insurance claims, insurance companies naturally desire the lowest possible false positive rate, prioritizing accuracy over retrieval speed.

Based on these two points, we propose the following improvements to the Bloom filter (as shown in Figure 5):

1. **From Bit Array to Multidimensional Array:** We replace the one-dimensional bit array with a multidimensional array structure. Considering practical needs, we specifically adopt a three-dimensional array. This structure significantly reduces the mutual influence between different elements, thereby greatly lowering the probability of false positives during retrieval;

2. **Storing Array Coordinates on the Blockchain:** Instead of storing the entire Bloom filter array, we store only the coordinates within the three-dimensional array on the blockchain. This approach is motivated by the observation that most patients in reality do not suffer from a large number of diseases. Storing coordinates drastically reduces storage requirements compared to storing the entire multidimensional Bloom filter space;
3. **User-Specific Hash Functions with Key Control:** Each user is assigned dedicated hash functions. The user retains the associated key, and only possession of this key allows generation and retrieval of that specific user's Bloom filter. This measure transfers control of the data to the user via the key, significantly enhancing user privacy and security. Furthermore, since only coordinate values are stored on the blockchain, even if hackers obtain these values, they cannot decipher them, thus protecting user privacy.

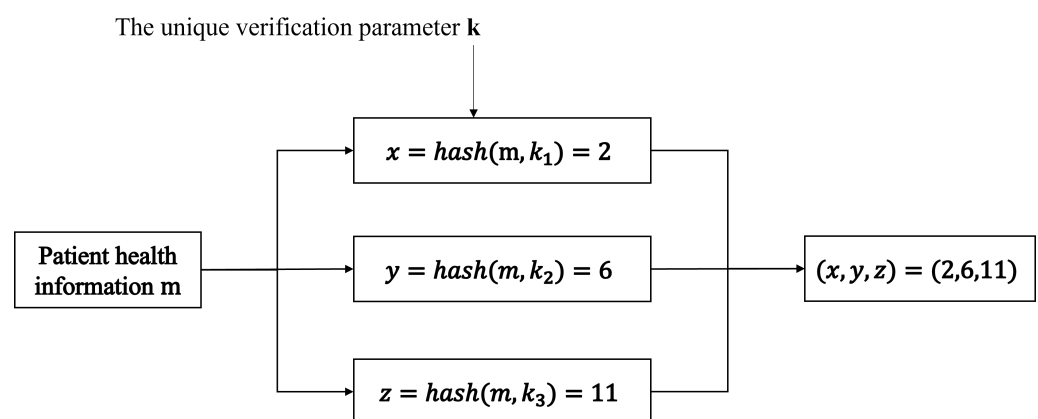


Figure 5. Improved Structure Diagram of the BMIT Bloom Filter.

5.3. Operation of the Improved Bloom Filter in the BMIT System

In the BMIT system, the Bloom filter primarily undertakes the following three tasks:

1. User Registration:

- **k-value generation:** During the registration process, the user is assigned three large random prime numbers. These primes collectively serve as the user's Unique Verification Parameter k . For practical computational efficiency, these primes are randomly selected from a pre-generated prime number database rather than being generated on-the-fly for each use;
- **Integrity Verification:** To ensure the integrity and authenticity of the assigned k value throughout its lifecycle, a digital signature is generated using the user's private key. This signature corresponds specifically to the k value;
- **Verification Process:** Whenever the validity of a user-provided k value needs to be confirmed (e.g., by a hospital or insurer), the verifier performs the following steps:
 - (a) Compute the hash value of the provided k ;
 - (b) Verify this hash against the digital signature stored on the blockchain using the user's public key;
 - (c) A successful verification confirms that the k -value is correct and untampered with.

2. Adding Elements (Medical Term Creation):

- When a medical term (e.g., a diagnosis) is generated for the user at a hospital, the user provides their k value to the hospital;

- The hospital first verifies the authenticity of the provided k using the digital signature stored on the blockchain, as described in step 1;
- Upon successful verification, the hospital generates a new element for the user's Bloom filter, represented as a three-dimensional coordinate;
- **Coordinate Generation:** The coordinate is algorithmically generated based on the user's verified k value and the specific medical term being recorded;
- **Storage:** The hospital then uploads this newly generated three-dimensional coordinate to the blockchain system, where it is stored as part of the user's Bloom filter representation.

3. Retrieving Elements (Insurance Application):

- When a user applies for insurance with a company, they provide their k value to the insurer;
- The insurer verifies the authenticity of the provided k using the digital signature stored on the blockchain, following the same process as the hospital;
- Upon successful verification, the insurer generates a set of three-dimensional coordinates corresponding to the medical terms (e.g., pre-existing conditions; specific diagnoses) relevant to the insurance policy being applied for;
- **Coordinate Generation:** These query coordinates are algorithmically generated based on the verified user k and the specific medical terms the insurer needs to check;
- **Comparison and Decision:**
 - (a) The insurer retrieves the user's stored Bloom filter coordinates from the blockchain;
 - (b) The insurer then compares the generated query coordinates against the retrieved coordinates (user's actual record);
 - (c) Based on the comparison results (e.g., presence or absence of matches for specific query coordinates), the insurer determines whether the user qualifies for the policy or under what terms. This allows confirmation of relevant medical history without accessing raw medical data.

5.4. The Hash Function of the Improved Bloom Filter in the BMIT System

In this system design, a conventional polynomial hash algorithm is selected as the hash function. The polynomial hash algorithm is a string hashing technique based on polynomial evaluation, which treats a string as polynomial coefficients and generates a hash code by computing the polynomial's value at a specified point. The formula is expressed as follows:

$$\text{hash}(s) = (s_0p^{n-1} + s_1p^{n-2} + \dots + s_{n-1}p^0) \bmod M \quad (1)$$

where:

- s is the input string;
- s_i denotes the ASCII value of the i -th character in the string;
- p represents the chosen prime base;
- n is the string length;
- M is the modulus.

Within this Bloom filter design, three distinct prime numbers (p_1 , p_2 , and p_3) are employed as moduli to compute three separate hash values. These values map the original information into a three-dimensional space of size M^3 . Considering practical requirements and collision probability, M is set to 1000, creating an element space of 10^9 (1,000,000,000). This space magnitude is sufficient to meet real-world application demands.

6. Smart Contract Design

6.1. Registry Contract

The core objective of this contract is to provide standardized and secure identity registration services for patients, hospitals, and insurance companies and to establish an identity management system for entities involved in medical-related affairs through this mechanism. Specifically, the contract will automatically create a structured registry and be responsible for the full-lifecycle maintenance of the registry to ensure the accuracy and timeliness of information.

This registry not only records key information in detail, including the SM2 public key address, the specific role corresponding to the address (clearly classified into three categories: patient, hospital, or insurance company), the precise timestamp of the registration application, and the current status (e.g., normal, canceled), but also generates and associates a unique UID (user identifier) for each registered user. The introduction of the UID is of great significance: in subsequent information query operations, the system can directly retrieve data via the UID, avoiding additional resource consumption caused by direct invocation of the SM2 (ShangMi 2) public key address. This significantly improves query response speed and overall system efficiency while reducing direct exposure of the public key address.

Through continuous dynamic maintenance of this public key address registry—including real-time information updates and timely cleaning of abnormal data—a solid foundation for identity verification in all subsequent medical-related transaction scenarios is provided. When transactions are conducted, the system will automatically and strictly verify the public key addresses of participating parties. If an unregistered illegal entity attempts to initiate a request, the request will be immediately intercepted and rejected.

This mechanism realizes strict control over access conditions from the source, ensuring that only legally registered entities can participate in related transactions and effectively eliminating interference from illegal entities. The ultimate goal of this series of processes is to effectively protect privacy and security during medical data interaction, prevent unauthorized access and information leakage, and build a reliable security barrier for digital transactions in the medical industry.

6.2. Medical Information Data Contract

As the core component of the medical information management system, this contract is primarily responsible for recording and maintaining key information, such as patients' current medical records and unique verification parameters. It functions as a highly secure and orderly digital archive, ensuring that information related to patients' health and rights is accurately stored and efficiently managed. The core data it stores includes several key elements: the patient's public key address pb_p , which serves as the patient's digital identity in the system and the basis for information ownership and permission management; the improved Bloom filter string bf , whose optimized design effectively enhances information matching accuracy while ensuring query efficiency, providing reliable support for subsequent information verification; the timestamp t , which accurately records the generation or update time of information, offering a critical basis for judging timeliness and traceability; the ciphertext key_h of the unique verification parameter, whose encryption ensures the security of this core parameter during storage and prevents information leakage from unauthorized access; and the hospital public key address pb_h , which clarifies the identity of the medical institution uploading medical information, facilitating liability tracing and information source verification.

Additionally, the system uses the UID as the primary key. This design not only enables quick localization of specific patients' information—greatly improving query con-

venience and response speed—but also reduces unnecessary resource consumption caused by complex query logic, optimizing overall system performance. Through fine-grained maintenance of this key information in the table, combined with specially developed SDK tools, three core functions are jointly realized, constructing a complete medical information interaction system: patients’ self-query of medical information, hospitals’ uploading of medical information, and insurance companies’ information verification. In the patient self-query process, security is paramount.

The patient must provide a unique digital signature—acting as a digital ID card containing the patient’s identity information and authorization instructions. The system strictly verifies the signature; only when verification succeeds and the request is confirmed to originate from the legitimate patient is permission granted to access all medical information. Simultaneously, the patient can use their private key to decrypt the unique verification parameter’s ciphertext key_h , obtaining the unique verification parameter k . This parameter k serves as the “key” to the entire information interaction process, playing an irreplaceable role as the core verification basis in the subsequent functions of hospital information upload and insurance company information verification. The process of hospitals uploading medical information is rigorous and standardized, fully ensuring information authenticity and authority. First, the patient must provide the unique verification parameter k to the hospital, confirming their informed consent and authorization for the information upload. The system then combines k to convert specific medical information entries—including diagnosis results, treatment plans, and examination data—into a unique three-dimensional coordinate set string. This conversion not only provides encryption protection for original information but also standardizes data format for subsequent comparison.

After conversion, the patient and hospital provide their respective digital signatures: the patient’s signature confirms recognition of the medical information, while the hospital’s signature attests to the information’s source and authenticity. Only when both signatures pass the system’s signature-verification review is the three-dimensional coordinate set string and hospital public key address updated to the on-chain information table, enabling secure on-chain upload and permanent storage. The insurance companies’ information verification process is characterized by high efficiency and accuracy, aiming to quickly confirm patients’ relevant medical conditions. Similarly, the insurance company requires the patient to provide the unique verification parameter k , ensuring the patient consents to verification of specific medical information. The system combines k to generate a three-dimensional coordinate set string containing the medical entries to be verified—typically closely related to insurance coverage and claim conditions. By calling the smart contract, the system retrieves the corresponding improved Bloom filter string bf from the on-chain information table.

Finally, by comparing and verifying whether the newly generated three-dimensional coordinate set string matches any coordinates in bf , the system quickly determines if the patient has a relevant condition, providing an accurate basis for insurance underwriting, claims settlement, and other business decisions.

7. Simulation Results and Discussion

This section evaluates the Bloom filter and smart contracts developed for this solution. The testing environment is configured as follows:

- Processor: Intel Core i7-14700HX 2.10 GHz;
- Memory: 32 GB RAM at 5600 MHz;
- Operating Systems: Windows 11 and Ubuntu 20.04 (dual-boot configuration);
- Bloom Filter Testing: Implemented in Java using IntelliJ IDEA 2021.1 and JDK 1.8.0;
- Smart Contracts: Authored in Solidity.

7.1. Performance Testing of the Enhanced Bloom Filter

The testing phase primarily focuses on two evaluation metrics relevant to practical application scenarios:

1. **Collision Rate:** Calculated as $1 - (\text{number of unique coordinates} / \text{number of inputs})$. This metric represents the proportion of strings experiencing collisions within the total input (note: if multiple strings map to the same coordinate, all except the first are considered collided). Thus, this rate reflects the percentage of strings that collide with at least one other string (i.e., multiple strings mapping to a single coordinate). It differs from the conventional hash collision probability (the likelihood of two random strings colliding), as it specifically addresses multi-input collisions occurring in a single hash operation;
2. **Average Processing Time:** Derived as total processing time / total number of inputs.

According to World Health Organization (WHO) statistics, over 10,000 distinct disease names are currently documented worldwide. Consequently, the filter's performance was evaluated under input sizes of 5000, 10,000, and 50,000 randomly generated elements. Additionally, to assess the impact of varying prime number bit lengths, tests were conducted using prime factors of 4, 8, 16, 32, 64, and 128 bits under each input size condition. The specific results are illustrated in Figure 6.

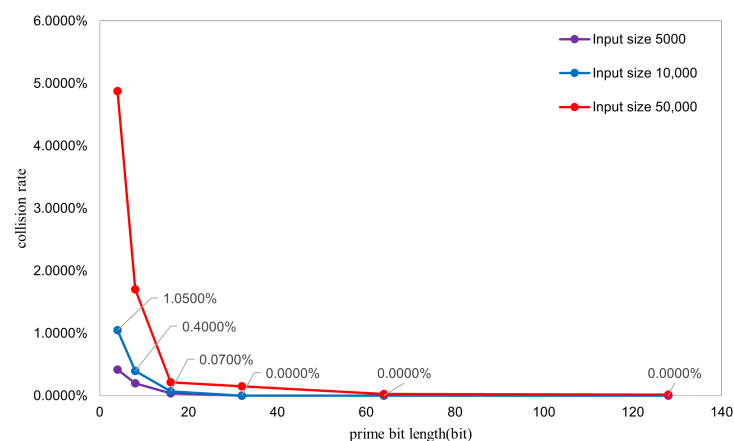


Figure 6. The collision rate of the three-dimensional hash function under prime numbers of different bit lengths.

Based on the test results, two prominent trends are readily observable:

1. **Increased Collision Probability with Larger Inputs:** As the input size increases, collision probability rises significantly. This aligns with the theoretical expectation for Bloom filters, where a fixed element space inevitably leads to higher collision probabilities as more elements are inserted.
2. **Decreased Collision Rate with Larger Prime Bit Lengths:** Higher prime bit lengths correlate with reduced collision rates. This occurs because larger primes yield outputs closer to an ideal uniform distribution. The underlying mechanism involves polynomial calculations within an expansive integer space (determined by prime p) prior to the modulo M operation. When larger primes are employed the following occurs:
 - Polynomial values occupy a broader numerical range before modulo reduction;
 - Increased base magnitude disperses hash values more widely across different strings;
 - Lower collision probability results from reduced hash value clustering.

Specifically, the three distinct hash functions utilize three separate large primes. Small primes increase collision susceptibility due to the following:

- Repetitive cyclic patterns inherent in polynomial computations with limited primes;
- Weakened independence between hash functions.

Conversely, larger prime bit lengths enhance inter-function independence because of the following:

- The primes themselves exhibit greater mutual independence;
- They lack small common factors.

Consequently, the probability of simultaneous collisions across all three hash coordinates decreases substantially.

In addition to influencing collision rates, prime bit lengths also exert a significant impact on processing time. To identify optimal prime bit lengths, we measured processing times for an input size of 10,000 across various prime bit lengths, with the results presented in Figure 7.

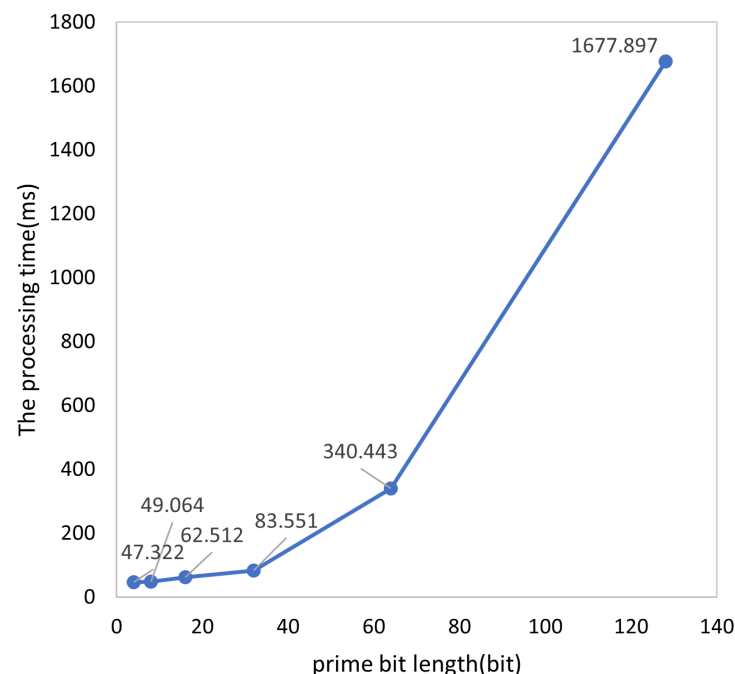


Figure 7. The processing time of the three-dimensional hash function under prime numbers of different bit lengths when the input quantity is 10,000.

Figure 7 clearly illustrates that processing time increases exponentially with larger prime bit lengths. Concurrently, we observe that 32-bit primes exhibit significantly lower processing times (below 100 ms), whereas 64-bit primes require over 300 ms. Considering the system's security requirements (higher prime bit lengths enhance resistance to cryptographic attacks), 32-bit primes strike an optimal balance between security and efficiency, making them the most suitable choice for this system.

For the Bloom filter in this solution, the selection of M significantly impacts collision rates. Typically, a larger M expands the sample space, thereby reducing collision probability. To determine an optimal M value, we performed comparative tests across distinct M configurations. Based on prior experimental results, the prime bit length was fixed at 32 bits, and the input size was set to 10,000. The test results are shown in Figure 8:

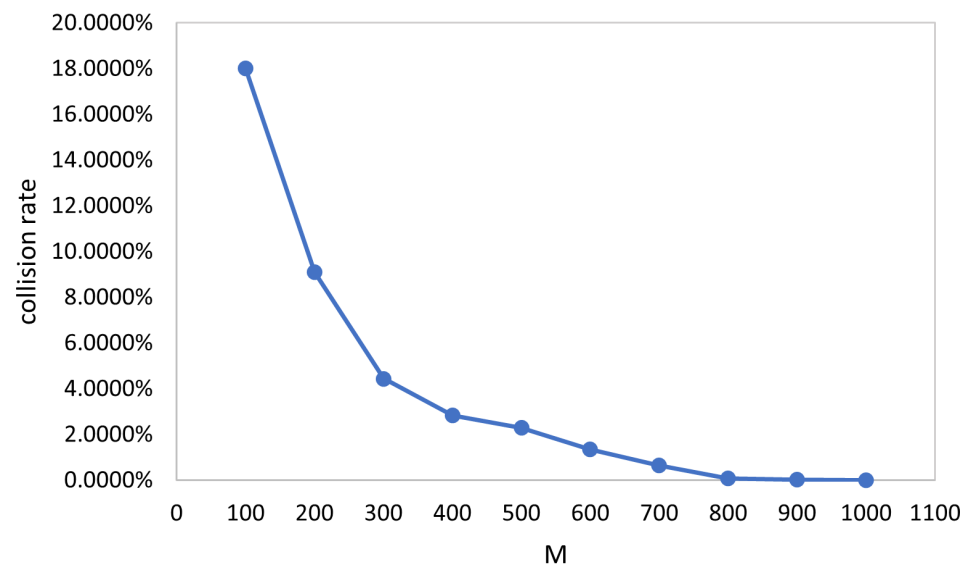


Figure 8. Impact of distinct M values on collision rate with 10,000 inputs under 32-bit prime configuration.

Based on these findings, it is evident that collision probabilities tend towards zero when M values fall below 800. Given practical storage requirements where coordinate storage for both the 800 range (0–799) and 1000 range (0–999) occupies three digits, selecting $M = 1000$ constitutes a rational engineering selection.

7.2. Comparative Analysis with Conventional Bloom Filters

To evaluate performance differences between the enhanced and conventional Bloom filters, we constructed a standard Bloom filter using the same triple polynomial hash functions. Both implementations were systematically compared across multiple dimensions, with key findings summarized in Table 2.

Table 2. Comparison table: traditional vs. enhanced bloom filters.

Metric	Enhanced Filter	Conventional Filter
Storage Consumption ($n = 100$)	1.27 KB	1.723 KB
False Positive Rate ($n = 10,000$)	0%	1.77%
Insertion Complexity	$O(1)$	$O(1)$
Query Complexity	$O(1)$	$O(1)$

1. Storage Efficiency:

Following optimization, storage of the entire data structure is unnecessary. Instead, a series of 3D spatial coordinates must be stored. Storing a single 3D coordinate requires three integers (12 bytes). Thus, storing coordinates for n elements demands $12n$ bytes of space. Given that the system stores medical insurance related entries (e.g., genetic diseases; infectious diseases) for individual patients, real-world data typically contain 100 entries. Therefore, we calculate the storage for 100 medical records, with collision probability measured at 0% in prior experiments. Since coordinates are stored as 13-character ASCII strings (1 byte per character), 100 records require 1300 bytes (1.27 KB). In practice, most patients exhibit single-digit condition counts, with few barely reaching double digits, reducing typical storage to 100 bytes.

For the conventional Bloom filter, its size can be calculated through the false positive rate calculation formula of the Bloom filter, which is expressed as follows:

$$p = \left(1 - e^{-\frac{kn}{m}}\right)^k \text{ where } p \text{ represents the false positive rate, } k \text{ denotes the number}$$

of hash functions employed, n denotes the quantity of elements, and m denotes the bit array size. To enable a direct comparison with the enhanced Bloom filter, we construct a conventional Bloom filter using three hash functions ($k = 3$) with an ultra-low false positive rate ($p = 0.001\%$), which aligns with the performance characteristics of the enhanced system. Solving this formula yields a required bit array size of $m = 13,781$ bits, corresponding to a storage requirement of 1723 bytes. Since this filter maintains a complete dataset representation, its storage footprint remains fixed at 1723 bytes irrespective of actual operational conditions.

2. Error Rates:

Given the divergent operational principles, direct comparative analysis of the enhanced filter as a conventional filter is infeasible. To enable valid comparison, both filters were constrained to identical storage allocation. When processing 10,000 inputs, the enhanced Bloom filter requires 12 KB for coordinate storage. Consequently, the conventional filter's bit array size was configured to the same size of 12 KB (96,000 bits) for performance simulation. Experimental results revealed the following:

- Conventional filter's false positive rate: 1.77%;
- Enhanced filter's collision rate: 0%.

3. Element Insertion:

For the enhanced Bloom filter, which stores a set of 3D coordinates, new elements are added by appending fresh 3D coordinates to the existing set, achieving a time complexity of $O(1)$. For the conventional Bloom filter, which maintains a complete bit array, insertion involves setting the corresponding k bit positions to 1, which also operates in $O(1)$ time.

4. Query Performance:

For the enhanced Bloom filter, querying involves searching the stored point set of 3D coordinates. Although this point set is stored in string format, assuming the exclusion of data structure conversion time (since both implementations require string-to-queryable conversion), lookups via a hash table implementation achieve a time complexity of $O(1)$. For the conventional Bloom filter, querying entails checking the corresponding bits in the bit array. Similarly, disregarding data conversion overhead, direct bit array querying also exhibits a time complexity of $O(1)$.

7.3. Smart Contract Testing

By carefully selecting on-chain data, the system significantly reduces the burden of unnecessary transactions on the blockchain network. The core functionality of our smart contract design focuses on data table maintenance, with the highest overhead occurring during the data table creation phase. Each creation requires initializing state variables and writing them on-chain, involving multiple SSTORE operations. Additionally, the contract initialization logic—including permission verification and data table structure generation—further increases computational overhead depending on the complexity of the logic.

To validate performance bottlenecks, we conducted a stress test on the registration function of the registry smart contract using a 4-node FISCO BCOS consortium chain. The test simulated a real-world scenario by inputting 10,000 registration records, each containing a user address, timestamp, and metadata hash (fixed 32 bytes), while maintaining a constant QPS (Queries Per Second) of 1000. Under these conditions, the system processed all 10,000 registrations in 13,582 ms, achieving a TPS (Transactions Per Second) of 736, which meets the system's design requirements.

Currently, the TPS is constrained by the single-threaded execution model of blockchain virtual machines (e.g., EVM). In practical deployments, sharding techniques can be employed to partition data tables by user addresses across different sub-chains, enabling

parallel processing of registration requests and further alleviating performance bottlenecks under high transaction volumes. According to FISCO BCOS's official data, when consensus nodes are deployed on 24-core, 32 GB RAM servers with 10Gbps network and optimized chain parameters, the network can achieve tens of thousands of TPS—sufficient for large-scale applications.

7.4. Main Computational Overhead of the BMIT System

The primary computational overhead of this system stems from SM2 key generation, SM2 data encryption, SM2 data decryption, SM2 signing, SM2 signature verification, and optimized Bloom filter calculations. The following is based on the system's actual workflow:

- SM2 encryption/decryption targets the unique verification parameter k , so the simulated data size is set to 1 KB.
- SM2 signing/verification primarily handles request information, with a simulated data size of 10 KB.
- The optimized Bloom filter's time consumption adopts the conclusion from the previous section, with a single-operation overhead of 83.6 ms under extreme conditions.

The test results are summarized in Table 3:

Table 3. BMIT system computational overhead table.

Component	Operation	Time Consumption
SM2 Cryptographic Module	Key Generation	0.8 ms
	Encryption	1.1 ms
	Decryption	1.5 ms
	Signing	1.0 ms
	Verification	1.1 ms
Optimized Bloom Filter	Single Calculation (extreme case)	83.6 ms

According to the system design, the entire process from registration requires the following:

- Three registration operations;
- One SM2 encryption;
- Two SM2 decryptions;
- Four SM2 signatures;
- Four SM2 signature verifications;
- Two optimized Bloom filter calculations.

Therefore, the total computational overhead is calculated to be approximately 181.7 milliseconds (ms).

8. Challenges in Practical Deployment

The Blockchain-based Medical Insurance Transaction (BMIT) system proposed in this paper adopts the architecture of “FISCO BCOS consortium blockchain + improved 3D Bloom filter + smart contract”, which has theoretically achieved trusted data sharing, privacy protection, and process automation. However, when deploying the system in real-world medical insurance scenarios, it still needs to confront three core challenges: interoperability, regulatory compliance, and implementation costs. These challenges not only determine whether the BMIT system can be transformed from a theoretical technical solution into a practical industry tool but also represent the key bottlenecks that need to be addressed in the process of promoting the practical application of this research.

8.1. Interoperability Challenges

The core value of the BMIT system lies in breaking down data silos between medical institutions and insurance companies and building an efficient cross-entity collaboration platform. Nevertheless, the “fragmented” nature of multiple institutions and systems in the current medical insurance ecosystem poses severe obstacles to the system’s interoperability.

From a practical perspective, the core business systems of medical institutions exhibit significant heterogeneity: Hospital Information Systems (HISs) widely used in tertiary hospitals, electronic health record systems in community hospitals, and underwriting and claims systems of insurance companies are all custom-developed based on the respective business needs of each institution. These systems differ fundamentally in critical aspects such as data formats (e.g., diagnostic coding; cost statistical dimensions), communication protocols (e.g., HTTP/HTTPS; WebSocket), and interface standards (e.g., HL7 FHIR; custom APIs). They cannot achieve direct interconnection and data exchange, forming “information islands” that hinder data sharing.

To address this issue, the BMIT system incorporates a “standardized adaptation” strategy during the design phase: it uses string format as the unified carrier for data interaction. Medical information such as diagnosis results and examination reports (e.g., “Type 2 diabetes, glycated hemoglobin 7.2%”) from hospitals, as well as underwriting rules from insurance companies (e.g., “Rejection of insurance for patients with systolic blood pressure 160 mmHg”), is required to be organized according to the string specifications preset by the BMIT system. As long as medical and insurance institutions comply with the same specifications, they can bypass the underlying differences of their systems and achieve cross-entity, cross-standard business collaboration. This means that each institution only needs to develop a set of middleware for converting “heterogeneous data to standard strings” to access the BMIT system, significantly lowering the technical threshold for interoperability.

In practical implementation, however, interoperability challenges have not been fully resolved. On one hand, some outdated business systems in medical institutions have closed-code architectures and rigid data formats, making it necessary to overcome compatibility issues during middleware development. On the other hand, medical and insurance institutions have clear business boundaries and complex divisions of data rights and responsibilities; some institutions have concerns about data sharing, requiring multi-party negotiations to clarify the scope of data usage and security responsibilities. These factors collectively make interoperability the primary obstacle to the deployment of the BMIT system.

8.2. Regulatory Compliance Challenges

As a core field safeguarding people’s livelihoods, medical insurance is subject to strict constraints from multiple regulations, including the *Data Security Law*, *Personal Information Protection Law*, and *Regulations on the Supervision and Administration of the Use of Medical Insurance Funds*. Among these, the *Regulations on the Supervision and Administration of the Use of Medical Insurance Funds* explicitly require medical insurance supervision departments to “monitor the medical insurance fund usage process in real time and accurately trace abnormal transactions”. This means that the BMIT system must provide supervision departments with audit access to on-chain data to ensure the compliance and security of fund usage.

The BMIT system selects the FISCO BCOS consortium blockchain as its underlying platform, which provides technical support for addressing regulatory compliance challenges. The “controllable access” feature of the consortium blockchain allows regulatory authorities and audit departments to join the blockchain network as observer nodes. Observer

nodes do not participate in consensus but can obtain all on-chain transaction data (e.g., medical insurance reimbursement applications; fund disbursement records) in real time, enabling full-cycle regulatory auditing of the fund usage process and meeting the stringent requirements of financial-grade supervision for “timeliness, accuracy, and traceability”.

Meanwhile, the distributed architecture, immutable data, and traceable transaction features of blockchain endow supervision with inherent advantages. Based on these features, the BMIT system can build a “penetrating supervision solution”: supervision departments can not only view the final transaction results but also trace the initiating entity, data source, and processing node of each medical insurance transaction, and even identify the trigger link of abnormal transactions (e.g., the upload node of false diagnosis certificates; the approval process of irregular reimbursements). This ensures that all data is in a state of being “supervisable, auditable, and traceable”.

In practical operation, however, regulatory compliance still faces a balance dilemma: on one hand, supervision requires sufficient access to on-chain data to achieve effective monitoring; on the other hand, medical insurance data contains a large amount of user privacy information (e.g., medical history; diagnosis and treatment records), which needs to be desensitized using technologies such as the improved 3D Bloom filter. How to find the optimal balance between “meeting regulatory needs” and “protecting user privacy” while avoiding excessive compliance reviews that affect business processing efficiency remains a core issue that the BMIT system needs to continuously optimize.

8.3. Implementation Cost Challenges

The implementation cost of a blockchain system is not a one-dimensional expenditure but a complex cost system centered on “deployment-operation-expansion”. Although the FISCO BCOS consortium blockchain platform adopted by the BMIT system has significant advantages in cost control, it still cannot fully avoid cost pressures under large-scale deployment.

8.3.1. Cost Advantages of the FISCO BCOS Consortium Blockchain

- **Open-source Features Reduce Upfront Investment:** As an open-source consortium blockchain platform, FISCO BCOS allows free access to its core code, eliminating the need to pay third parties for software licensing fees and directly reducing the upfront technical introduction costs of the project. At the same time, the open-source community has an active group of developers. When the BMIT system encounters technical problems (e.g., smart contract vulnerabilities; node communication failures), it can quickly obtain solutions with the support of community resources, reducing the trial-and-error costs and time costs of technical research and development.
- **High-performance Design Reduces Resource Consumption:** FISCO BCOS significantly improves the system’s transaction throughput (TPS) and processing efficiency by optimizing consensus mechanisms (e.g., parallel processing of PBFT consensus) and transaction processes (e.g., pre-execution optimization), which can meet the demand for “centralized processing of peak reimbursement applications” in medical insurance scenarios. The high-performance feature can reduce over-reliance on hardware resources; for example, in scenarios with 100,000 daily medical insurance transactions, there is no need for frequent server configuration upgrades, indirectly reducing hardware procurement and operation and maintenance costs.

8.3.2. Cost Pressures in Practical Deployment

Taking the “Pearl River Delta Credit Information Blockchain” (a project based on the FISCO BCOS consortium blockchain) as a reference: this project includes 15 nodes covering local credit information platforms, corporate credit reporting institutions, data source units,

and supervision departments. If domestic mainstream cloud server rental plans are adopted (4-core 8 G configuration; annual cost of approximately CNY 1500 per unit), the server rental cost alone reaches CNY 22,500. If a self-built node model is adopted, the implementing entity must bear additional costs for hardware procurement (servers; storage devices), software deployment (operating systems; security components), and dedicated operation and maintenance personnel, with the total cost approaching CNY 1,000,000.

For the BMIT system, the cost pressure of nationwide deployment will be further amplified. On one hand, medical insurance covers 31 provinces (autonomous regions and municipalities directly under the Central Government) in China, requiring far more nodes than the “Pearl River Delta Credit Information Blockchain” to connect medical institutions (hospitals; community health service centers), insurance companies, and supervision departments. This will lead to a multiplicative increase in server rental or procurement costs alone. On the other hand, the daily volume of national medical insurance transactions reaches tens of millions, requiring improved node performance (e.g., 8-core 16 G configuration or higher) and increased storage capacity (to meet the needs of long-term data retention), which further drives up hardware and operation and maintenance costs.

8.3.3. Balance Between Costs and Benefits

Despite the high implementation costs, the investment in deploying the BMIT system is justified from the perspective of the long-term benefits of the medical insurance industry. According to data from the National Healthcare Security Administration, a total of CNY 27.5 billion in medical insurance funds was recovered nationwide in 2024. This figure only reflects the verified losses from insurance fraud, while the scale of undiscovered losses may be even larger. By leveraging the immutability and traceability of blockchain, the BMIT system can fundamentally reduce insurance fraud behaviors such as false reimbursements and excessive medical treatment. In the long run, the losses saved for medical insurance funds will far exceed the system implementation costs.

In summary, the three challenges of interoperability, regulatory compliance, and implementation costs are the core obstacles to transforming the BMIT system from a theoretical concept to a practical solution. Future research needs to further optimize technical solutions for these challenges (e.g., developing universal middleware; designing dynamic privacy protection mechanisms) and establish a collaboration mechanism with industry entities to promote the gradual large-scale implementation of the system.

9. Conclusions

In the current development of the medical insurance industry, the traditional claim settlement model faces a series of critical challenges requiring immediate resolution. Its procedures are cumbersome and complex. For instance, in the case of insurance applications for patients with chronic diseases, insured individuals must travel back and forth between hospitals, medical examination institutions, and pharmacies to collect past medical records, medication lists, and physical examination reports. Sorting out these materials alone may take 1–2 weeks, and the underwriting cycle can be as long as 5–7 working days. Moreover, the underwriting error rate due to incomplete information exceeds 12%. Meanwhile, the issue of weak trust mechanisms is notable: the traditional insurance system relies heavily on the credit endorsement of centralized institutions, but security vulnerabilities in centralized databases pose significant risks to the industry. For example, in 2024, the internal system of a major insurance company was breached, leading to the leakage of sensitive information of millions of insured persons and directly triggering a class-action lawsuit. In the same year, there were incidents such as hospitals forging CT reports and large-scale insurance fraud. Additionally, inadequate protection of user privacy is a major drawback, with the risk

of privacy leakage persisting throughout data circulation. These problems have severely hindered the efficient operation and sustainable development of the industry.

Against this background, the BMIT (Blockchain-Based Medical Insurance Transaction) scheme proposed in this paper, as a novel medical insurance claim settlement solution based on blockchain technology, holds significant innovative significance and practical value.

The BMIT scheme leverages core features of the FISCO BCOS consortium chain network—such as distributed ledgers and immutability—and has carefully constructed a comprehensive framework integrating medical insurance purchase and verification. This framework clearly defines the functional boundaries of three key roles (patients, hospitals, and insurance companies) and establishes a secure and trustworthy ecosystem for insurance application and claim settlement.

Benefiting from blockchain technology, patients can exercise full control over their medical data through the BMIT system. Any party must obtain the patient-controlled private key and the unique verification parameter k to access the user's unique Bloom filter and thus check the user's medical information. As providers of medical services, hospitals are responsible for accurately inputting and processing patients' medical data within this framework. They protect user privacy by obfuscating medical information into three-dimensional coordinates via the improved Bloom filter and ensure the immutability of medical information through blockchain storage. Using the BMIT system, insurance companies can verify medical information requiring confirmation in the insurance policy with the user's authorization, accurately determining whether the user meets the policy requirements to decide whether to issue the medical insurance.

Furthermore, the immutability of blockchain technology fosters a unique "trust" environment. The unique technical architecture of blockchain enables each data record to be packaged into a "block," which is closely linked to the previous block via an encryption algorithm, forming a chain structure. To alter information in a block, one must not only tamper with the block itself but also synchronously modify the encryption verification information of all subsequent blocks—an operation nearly impossible in a blockchain network with distributed storage. Since each node in the network holds a complete copy of the ledger, any single-point malicious operation will be identified and rejected by other nodes, ultimately rendering the modification unsuccessful. This "collective supervision" mechanism makes it difficult for data to be tampered with or privately deleted once written, thus establishing a "code as trust" mechanism through the technology itself. It enables unfamiliar entities to collaborate directly based on immutable data without relying on intermediaries to guarantee data authenticity, significantly reducing process costs and mitigating the risk of intermediaries' errors or breaches of trust.

To address the practical requirement of user privacy protection in medical insurance systems, this paper proposes an innovative enhancement to the traditional Bloom filter. This enhancement not only enables the reasonable obfuscation of real medical data but also effectively eliminates the false positive issue, thereby significantly improving the accuracy of data verification.

Furthermore, the system assigns a dedicated hash function to each user, who retains the corresponding key. Only by holding this key can the Bloom filter of a specific user be generated and accessed. This design transfers data control to the user: through asymmetric encryption technology, users can manage the core parameters of their own Bloom filters, truly realizing the control over the sovereignty of their personal medical data and further enhancing privacy security.

Practical tests and verification showed that the improved Bloom filter eliminates false positives. When a 32-bit prime number, a modulus $M = 1000$, and 10,000 input entries were used, the error rate was 1.77% lower than that of the traditional scheme, with the

actual error rate approaching 0%—sufficient to meet the medical insurance industry’s requirements for data accuracy. Meanwhile, tests on the main computational overhead of the smart contracts and the BMIT system show that the smart contract can achieve 736 TPS, and the main computational overhead in the entire BMIT system process is only 181.7 ms (in the extreme case of 10,000 inputs to the improved Bloom filter; in actual deployment, this value should be much lower).

The current research findings also have limitations. First, due to cost constraints, only single-machine multi-node blockchain deployment was feasible, resulting in a lack of simulation verification for actual large-scale deployment. Second, in future practical deployment, leveraging the high scalability of FISCO BCOS, targeted contracts for actual insurance policies could be developed based on the current system, with insurance policies stored on-chain. It might even be possible to enable users to purchase insurance services directly on-chain without contacting insurance companies. However, this requires a practical deployment environment for in-depth research, making it one of the potential future research directions.

In summary, through the organic integration of blockchain technology and the improved Bloom filter, the BMIT scheme not only enables the secure, trustworthy, and efficient operation of medical insurance application and claim settlement processes but also achieves breakthroughs in user privacy protection. It provides a highly valuable solution for the digital transformation and innovative development of the medical insurance industry.

Author Contributions: Methodology, J.F. and L.L.; software, J.F.; writing—original draft preparation, J.F.; writing—review and editing, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon reasonable request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

BMIT	Blockchain-Based Medical Insurance Transaction
NS3	Network simulator 3
IoT	Internet of Things
pb	Public key address
pr	Private key
bf	Bloom filter string
pb_p	The patient’s public key address
pr_p	The patient’s private key
$sign_p$	The patient’s digital signature
pb_h	The hospital’s public key address
pr_h	The hospital’s private key
$sign_h$	The hospital’s digital signature
pb_i	The insurance company’s public key address
pr_i	The insurance company’s private key
$sign_i$	The insurance company’s digital signature
t	timestamp
key_h	The ciphertext of a user’s unique verification parameters

References

1. Elangovan, D.; Long, C.S.; Bakrin, F.S.; Tan, C.S.; Goh, K.W.; Yeoh, S.F.; Loy, M.J.; Hussain, Z.; Lee, K.S.; Idris, A.C.; et al. The use of blockchain technology in the health care sector: Systematic review. *JMIR Med. Inform.* **2022**, *10*, e17278. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Tomar, K.; Sharma, S. A proposed artificial intelligence and blockchain technique for solving health insurance challenges. In *Data-Driven Technologies and Artificial Intelligence in Supply Chain*; CRC Press: Boca Raton, FL, USA, 2023; pp. 31–57.
3. Najar, A.V.; Alizamani, L.; Zarqi, M.; Hooshmand, E. A global scoping review on the patterns of medical fraud and abuse: Integrating data-driven detection, prevention, and legal responses. *Arch. Public Health* **2025**, *83*, 43. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Tiwari, A. Blockchain technology for health insurance. In *Sensor Networks for Smart Hospitals*; Elsevier: Amsterdam, The Netherlands, 2025; pp. 389–410.
5. Radanović, I.; Likić, R. Opportunities for use of blockchain technology in medicine. *Appl. Health Econ. Health Policy* **2018**, *16*, 583–590. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Kar, A.K.; Navin, L. Diffusion of blockchain in insurance industry: An analysis through the review of academic and trade literature. *Telemat. Inform.* **2021**, *58*, 101532. [\[CrossRef\]](#)
7. Rahimi, N.; Gudapati, S.S.V. Emergence of blockchain technology in the healthcare and insurance industries. In *Blockchain Technology Solutions for the Security of IoT-Based Healthcare Systems*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 167–182.
8. Antwi, M.; Adnane, A.; Ahmad, F.; Hussain, R.; ur Rehman, M.H.; Kerrache, C.A. The case of HyperLedger Fabric as a blockchain solution for healthcare applications. *Blockchain Res. Appl.* **2021**, *2*, 100012. [\[CrossRef\]](#)
9. Ismail, L.; Zeadally, S. Healthcare insurance frauds: Taxonomy and blockchain-based detection framework (Block-HI). *IT Prof.* **2021**, *23*, 36–43. [\[CrossRef\]](#)
10. Liu, W.; Yu, Q.; Li, Z.; Li, Z.; Su, Y.; Zhou, J. A blockchain-based system for anti-fraud of healthcare insurance. In Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 6–9 December 2019; pp. 1264–1268.
11. Saldamli, G.; Reddy, V.; Bojja, K.S.; Gururaja, M.K.; Doddaveerappa, Y.; Tawalbeh, L. Health care insurance fraud detection using blockchain. In Proceedings of the 2020 Seventh International Conference on Software Defined Systems (SDS), Paris, France, 30 June–3 July 2020; pp. 145–152.
12. Mishra, A.S. Study on blockchain-based healthcare insurance claim system. In Proceedings of the 2021 Asian Conference on Innovation in Technology (ASIANCON), Pune, India, 27–29 August 2021; pp. 1–4.
13. Zheng, H.; You, L.; Hu, G. A novel insurance claim blockchain scheme based on zero-knowledge proof technology. *Comput. Commun.* **2022**, *195*, 207–216. [\[CrossRef\]](#)
14. Zhou, L.; Wang, L.; Sun, Y. MIStore: A blockchain-based medical insurance storage system. *J. Med. Syst.* **2018**, *42*, 149. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Al Amin, M.; Shah, R.; Tummala, H.; Ray, I. Utilizing Blockchain and Smart Contracts for Enhanced Fraud Prevention and Minimization in Health Insurance through Multi-Signature Claim Processing. In Proceedings of the 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Windhoek, Namibia, 23–25 July 2024; pp. 1–9.
16. Purswani, P. Blockchain-based parametric health insurance. In Proceedings of the 2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA), Langkawi Island, Malaysia, 10–11 July 2021; pp. 1–5.
17. Alnuaimi, A.; Alshehhi, A.; Salah, K.; Jayaraman, R.; Omar, I.A.; Battah, A. Blockchain-based processing of health insurance claims for prescription drugs. *IEEE Access* **2022**, *10*, 118093–118107. [\[CrossRef\]](#)
18. Namitha, T.S.; Rai, B.K.; Pallavi, P.; Pavana, R.; Pooja, M. Blockchain Based Health Insurance System. In Proceedings of the 2025 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 7–8 March 2025; pp. 1216–1221. [\[CrossRef\]](#)
19. Liang, X.; Zhao, J.; Shetty, S.; Liu, J.; Li, D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; pp. 1–5.
20. Raza, A.A.; Arora, B.; Irfan, M.T. Securing Health Insurance Claims with Decentralization and Transparency: A blockchain-based Approach. *Procedia Comput. Sci.* **2025**, *259*, 1918–1926. [\[CrossRef\]](#)
21. Kapadiya, K.; Ramoliya, F.; Gohil, K.; Patel, U.; Gupta, R.; Tanwar, S.; Rodrigues, J.J.; Alqahtani, F.; Tolba, A. Blockchain-assisted healthcare insurance fraud detection framework using ensemble learning. *Comput. Electr. Eng.* **2025**, *122*, 109898. [\[CrossRef\]](#)
22. Joginipalli, S.K. Leveraging blockchain for secure and transparent insurance claim processing. *J. Comput. Anal. Appl.* **2022**, *30*, 12.
23. Alnavar, K.; Babu, C.N. Blockchain-based smart contract with machine learning for insurance claim verification. In Proceedings of the 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECOT), Mysuru, India, 10–11 December 2021; pp. 247–252.
24. El-Samad, W.; Atieh, M.; Adda, M. Transforming health insurance claims adjudication with blockchain-based solutions. *Procedia Comput. Sci.* **2023**, *224*, 147–154. [\[CrossRef\]](#)

25. Kapadiya, K.; Patel, U.; Gupta, R.; Alshehri, M.D.; Tanwar, S.; Sharma, G.; Bokoro, P.N. Blockchain and AI-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects. *IEEE Access* **2022**, *10*, 79606–79627. [[CrossRef](#)]
26. Elhence, A.; Goyal, A.; Chamola, V.; Sikdar, B. A blockchain and ML-based framework for fast and cost-effective health insurance industry operations. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 1642–1653. [[CrossRef](#)]
27. Alhasan, B.; Qatawneh, M.; Almobaideen, W. Blockchain technology for preventing counterfeit in health insurance. In Proceedings of the 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021; pp. 935–941.
28. Guerar, M.; Migliardi, M.; Russo, E.; Khadraoui, D.; Merlo, A. SSI-MedRx: A fraud-resilient healthcare system based on blockchain and SSI. *Blockchain Res. Appl.* **2025**, *6*, 100242. [[CrossRef](#)]
29. Chouhan, G.K.; Singh, S.; Jaduvansy, A.; Rai, P.; Rath, A. A Blockchain Based Decentralized Identifiers for Entity Authentication in Electronic Health Records. In Proceedings of the 2025 International Conference on Intelligent and Cloud Computing (ICoICC), Bhubaneswar, India, 2–3 May 2025; pp. 1–6. Available online: <https://ieeexplore.ieee.org/abstract/document/11052007> (accessed on 14 October 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.