Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Information Processing and Management 58 (2021) 102570

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Enhancing Traceability of Infectious Diseases: A Blockchain-Based Approach

Peng Zhu [a], Jian Hu [a], Yue Zhang [b], Xiaotong Li [c,*]

[a] Nanjing University of Science and Technology, Nanjing 210094, China
[b] California State University, Northridge, Northridge, CA 91330-8372, USA
[c] University of Alabama in Huntsville, AL 35899, USA

ARTICLE INFO

Keywords:
Blockchain
Distributed systems
Infectious diseases
Timestamp
Traceability

ABSTRACT

The global pandemic of COVID-19 has brought significant attentions to three important features of disease direct reporting systems: traceability, reliability, and effectiveness. A traditional disease direct reporting system has a central node of control, with a hierarchical structure that goes up from locals (cities and counties) to regions and eventually reaches a central data repository. Such systems are often prone to easy data loss, arbitrary or unauthorized data changes, and unreliable traceability to individual nodes. Blockchain, as a new disruptive technology, provides a potential solution. Leveraging blockchain's features of decentralization, unforgeability, whole-process traceability, we develop a method for disease information tracing with key components including infectious disease information collection, information chain-style storage, and information query. Our blockchain-based infectious disease traceability method can promptly collect disease information and form the disease information time series blockchain. We demonstrate that the information chain constructed is authentic and transparent, and it can be queried and maintained at any node in the system. Consequently, the infectious disease information on the blockchain can be monitored and queried any time, thereby greatly facilitating the tracing of the propagation paths of infectious diseases.

1. Introduction

Infectious diseases such as the flu and SARS have features of fast outbreak, high infection rates, and potentially serious illness (Smith, 2006). Such infectious diseases pose increasingly severe challenges to public health. The COVID-19, a more recent example, spreads to the globe in merely a few months, posing a gigantic threat to people's lives and the normal activities of the human society. This global pandemic demonstrates once more that, given the large scale movement of personnel made easy by convenient transportations, controlling and coping with pandemics are increasingly more difficult (Walker et al., 2020). The key to the control and suppression of infectious diseases is to accurately infer and trace the root source of diseases at the onset (Baldominos et al., 2020). Traceability of an infectious disease means collecting and storing information about early patients of the disease, thereby identifying the root source of the infectious disease based on the propagation laws in epidemiology and the relationship network of patients' human contacts (Tsui et al., 2013).

To trace the source of an infectious disease, it is necessary to have information upload from multiple nodes, with authentic and

* Corresponding author: Xiaotong Li, University of Alabama in Huntsville, AL 35899, USA.
E-mail addresses: pzhu@njust.edu.cn (P. Zhu), hhujame@qq.com (J. Hu), jeff.zhang@csun.edu (Y. Zhang), lixi@uah.edu (X. Li).

https://doi.org/10.1016/j.ipm.2021.102570
Received 13 July 2020; Received in revised form 25 February 2021; Accepted 1 March 2021
Available online 6 March 2021

transparent data that is tamper-proof, and fast query and tracing capabilities at the node level. These features are hard to achieve in traditional disease reporting systems that emphasize hierarchical upward reporting and downward management by the superior level. The hierarchical structure of the traditional system cannot satisfy the great needs for effective and immediate source tracing of diseases (Inayatulloh & Theresia, 2016).Thus, it is urgently needed to design a highly efficient disease direct reporting system that is more capable of handling disease reporting from the onset of a pandemic. Blockchain, a new disruptive technology, can be adopted to implement a distributed bookkeeping system that cannot be tampered with(Yuan et al., 2018). A record, once logged to a blockchain, cannot be altered. As the time stamp function of the blockchain can accurately record the time of the information storage, CDCs, hospitals, and other institutions can upload infectious disease information in a faster and more secure manner. Blockchain also allows every node on it to easily trace back to the root, eliminating the need to pass the information in a level-by-level, hierarchical manner. Therefore, with regard to disease direct reporting systems, blockchain can be potentially a very useful tool for disease information storage and root source tracing.

Incorporating the blockchain technology into infectious diseases reporting systems, we propose an architecture of such a system and use simulation to demonstrate its effectiveness. The key features of our proposed method include: 1) Down-shift the functions of the central node to all nodes, to achieve the storage and root-source tracing of pandemic parameters such as confirmed cases, suspected cases, infected time of cases, infected location of cases, direction of propagation, etc. 2) Starting from root source tracing, leverage the blockchain technology's function of distributed node maintenance, hash tamper proofing, and time stamps, to implement a feasible solution for infectious disease source tracing. 3) Expound the process flow for disease information collection, storage, query, and tracing, using Python coding to build a blockchain environment to construct a simulated source tracing scenario to test the proposed method. 4) Articulate the process flow of infectious disease information collection, storage, query, and tracing. 5) Improve the orderly accumulation of the information of infectious diseases' confirmed and suspected cases, and infectious time and location, providing more accurate inputs for the disease control authorities in disease monitoring and pandemic response (Liu et al., 2020).

After a comparative analysis, we demonstrate the main novelty and contributions of our study as follows: 1).The proposed method can promptly collect disease information and construct a time series infectious disease blockchain. The chain of information is authentic and transparent, allowing easy maintenance and monitoring for nodes at all levels. Furthermore, the disease information on the blockchain can be monitored and queried by any node throughout the whole process, facilitating early warning and source tracing of infectious diseases. 2) Because a smart contract is used to manage the access permissions of all nodes on the infectious disease chain, nodes with different identities have different rights to manage information. The smart contract is also used to set the early warning threshold of infectious diseases, and a whole network broadcast mode is used. As a result, our approach enjoys several advantages over the traditional hierarchical reporting and early warning mechanisms. 3) Instead of using the traditional POX consensus mechanism algorithm, our method uses PBFT mechanism to verify infectious disease information during the process of date storage on the blockchain, reducing computing power consumption and the likelihood of 51% attacks. 4) Our simulation of the disease information storage and tracing processes clearly demonstrates traceability enhancement through data visualization.

2. Related works

In this section, we will review the literature to identify related research in the new advancement of blockchain technology and its applications in source tracing. Tracing and provenancing are important steps in fighting a pandemic after the outbreak. The word "provenance" originated from the French word "provenir" meaning "origin, source," bearing the meaning of "from the records at the end point tracing back to its earliest origin" (Olsen & Borit, 2013). Traceability technology was first used in meta data tracing in database and workflow areas. Through the recording of data generation and analysis, one can achieve the counter-discovery of data according to the path of data (Lee, 2019). This article's focus is information traceability enhancement in the context of infectious disease reporting. Although the core idea of information traceability inherits the concept of data traceability, it is actually more about tracking and querying specific information content. Such a traceability method realizes a two-way traceability which traces from beginning to the end and vice versa. Its most prominent features are security and traceability (Lu & Xu, 2017).

Epidemic information traceability is an application of information traceability in the understanding and control of infectious diseases' propagation. Through continuously collecting, storing, examining, and analyzing the dynamic data of infectious diseases' propagation and spatial distribution, it is possible to develop effective measures in time for disease control. Fast and accurate disease monitoring and control, with reliable information traceability, have become a research emphasis of many researchers and related organizations.

Lin & Heffernan (Lin & Heffernan, 2011) constructed a monitoring and tracing system for the Highly Pathogenic Asian Avian Influenza (H5N1, "Avian flu") based on mobile devices. Their system allows veterinary doctors, healthcare experts, and farmers (who are in frequent contacts with the carrier of the flu virus) from all over the world to upload (i.e., report) the health condition of animals. To avoid a flu pandemic, decision makers can follow and monitor the health conditions of animals, and adopt interference measures based on the information reported. Foley, Wilkerson, *et al.*(2010) studied mosquitos, the important propagation media of malaria, yellow fever, and other infectious diseases, and constructed mosquito's information collection and distribution model database based on a global network. Researchers conducted tracing on the species and original location of spotting of mosquitos, and assessed the risks of infectious diseases based on such mosquito-mediating diseases. In responses to the challenges of the tracing and information sharing for infectious diseases, Hu, Zeng, *et al.* (Hu et al., 2007) designed an infectious diseases information sharing and analysis system, based on infectious disease informatics. Their system focuses on the accuracy of analysis, task performance, user satisfaction, and system usability. Anderson et al., (2020) argued that the key to control of diseases likeCOVID-19 is to trace and control their dissemination and propagation. Keeping records of infectious individuals and tracing the person's previous contacts (thus quarantining any infected

individual) are the most effective coping method.

The above infectious diseases reporting systems or methods, though relatively mature, still have several problems, such as tracing is not sufficiently timely, data is stored in one database prone to data loss and hard to be maintained in a distributed manner. When an epidemic has an outbreak, it is under the monitor of (higher-level) decision makers as well as under the scrutiny of the public (in cases that decision makers fail to disseminate the information in a timely manner, the public might panic when learning about the outbreak). Therefore, the assurance of data authenticity/data integrity, and transparency, are also concerns for the general public. In summary, a good infectious disease direct reporting system should meet two key requirements: first, data storage must have integrity and tamper proofing capability that assure information credibility, and must be efficient for fast tracking and tracing; second, nodes of various healthcare institutions can participate real-time in the uploading, integration, tracing, and utilization of the disease data. Blockchain, with its resistance to tampering, can well serve as the technical foundation for the information storage and tracing of infectious disease direct reporting systems (Zhang et al., 2018).

Blockchain is a promising technology for the infectious disease traceability system because it has several desirable characteristics, such as the irreversible time vector, smart contract (agreements automatically executed without third party participation(Leeming et al., 2019), and consensus algorithm(George et al., 2019).The concept of blockchain was first brought forth by Satoshi Nakamoto in an article “Bitcoin: A peer-to-peer electronic cash system”(Nakamoto, n.d.). In this article, Nakamoto described a brand-new Bitcoin digital currency system that is decentralized, that it does not need to rely on any trusted authority. Blockchain is the foundation technology of bitcoin. Literally, the data container in blockchain is a Block; newly added data is packaged into a block in a batch manner at set intervals. Every piece of data and every block must be agreed and reached a consensus upon by participants who are eligible to perform data recording operations in the blockchain(Yang et al., 2019). In a blockchain, each block includes the cryptographic hash of the prior block in the blockchain, linking the two (Swan, 2015), eventually forming a Chain, thus the name "blockchain." In essence, as pointed out by Maxmen (Maxmen, 2018), blockchain is like a decentralized, distributed database with strong cryptographic algorithms that assure its data not be modified or forged. The time stamps in a blockchain allow the blockchain to record, in time sequence, the delivery and confirmation history of each transaction, and the other related information and data in the transaction, thereby realizing chain-wise arrangement of data (Zhu et al., 2020).

In recent years, blockchain technology has become increasingly prominent, many scholars believe that this technology has different potential applications in any industry, market, institution or government organization (Berdik et al., 2021). At present, the application domains of blockchain include intelligent social applications (Esposito et al., 2021), intelligent Internet of vehicles transportation (Oham et al., 2021; Khalid et al., 2021; Campanile et al., 2021), cloud computing and storage (Li et al., 2020; Baniata et al., 2021), and terminal storage technology and Internet of things system (Yu et al., 2021; Zhao et al., 2020). The application of blockchain traceability mainly focuses on supply chains (Leng et al., 2019; Ahmed & Broek, 2017; Queiroz & Fosso Wamba, 2019), fake-proof queries of healthcare product safety (Tanwar et al., 2020; Hardin & Kotz, 2021), and copyright tracing (Jing et al., 2021). There is little research in blockchain’s applications in infectious disease information tracing. Kuo, Kim, and Ohno-Machado (Kuo et al., 2017), leveraging the blockchain technology, designed a decentralized privacy-preserving cross-institution disease classification framework called Health-Chain. Their framework focuses on the integration and sharing of infectious disease data resources, and the issuance of early warning of infectious diseases through a smart contracts function. However, it does not sufficiently address information storage and tracing issues. In the wake of COVID-19 pandemic, Mashamba-Thompson & Crayton(2020) proposed a self-test and tracing system of COVID and other new infectious diseases, using low-cost blockchain coupled with artificial intelligence (AI). But their work is only in theoretical design phase and does not have concrete models or methods. Chamola, Hassija, et al.(Chamola et al., 2020)explored the application of blockchain in the epidemic process of infectious diseases. They believe that blockchain technology can effectively carry out digital monitoring and management of patients with COVID-19, thus reducing some of the burden of hospital staff. Their study, however, is more about the prospect of blockchain in the field of infectious diseases than empirical research.

In summary, as an emerging technology, blockchain has attracted broad research interests in tracing. Due to its features of decentralization, data reliability, and transaction anonymity, blockchain can support safe and transparent tracing of data. It can also assure the accuracy and integrity of the data stored in each participating node(Sultana et al., 2020). Therefore, blockchain technology can be applied in the context of infectious disease study and control, to enhance the information storage and tracing during the propagation of infectious diseases. Highlighting blockchain technology’s role in addressing the information traceability issue for disease direct reporting, we perform several empirical tests through simulation experiments to demonstrate the value of this emerging technology in the prevention and control of infectious diseases.

3. Proposed blockchain-based solution

In this section, we will describe the solution for tracing and query in the application of blockchain to infectious diseases direct reporting systems. Though methodological model construction, we will expound the application of blockchain in tracking and tracing along infectious disease chain in the propagation of a disease. The method proposed by our study mainly deals with the participating entities, the access privilege, agreement mechanism, storage, and traceability.

3.1. Model description

At present, when the disease reporting system makes a report, the lower level entities usually report the discovered disease cases to the immediate upper level. The upper level, after discussions among its own decision makers, decides whether to further escalate the report. Finally, the information reaches the central node (top-level decision makers/institutions), where the epidemic status is

recorded and managed. At the same time, when the infectious disease data value exceeds the threshold, the highest level department will also give a downward early warning in turn. Such a "hierarchical reporting" system has the problem of low efficiency and the possibility of infectious disease information being tampered by local nodes. When a blockchain is introduced to this process, people can utilize the mutuality of the network model to allow each individual node entity to participate in autonomous data management. In such a blockchain based mechanism, the distributed node entities can reach consensus and establish mutual trust, without a third-party trust authority. After incorporating a blockchain into a disease reporting system, any node in the blockchain enjoys the same privilege, and it can thus exchange infectious disease information with any other nodes in the chain, forming a direct reporting mode for the infectious disease. Similarly, when the infectious disease data value exceeds the threshold value, with the cooperation of the highest level department, the smart contract in the blockchain will automatically broadcast in the whole blockchain network. All nodes at any levels participating in the infectious disease blockchain will receive early warning information at the same time and the first time. Figure 1 demonstrates the disease information tracing solution model proposed in this study.

In this model, infectious disease information includes case information, number of cases, checkup reports and abnormal reports, which are collected by the nodes of medical institutions at all levels in different regions. After confirming the authenticity and effectiveness of infectious disease information through a consensus mechanism and data verification, the information is stored on the blockchain for the whole network nodes to query and trace. The consensus process is guaranteed by relevant laws and regulations to reduce the problem of untrue information sources of infectious diseases to a certain extent. Each entity in the model plays a different role, and different roles will be assigned different data operation permissions by the smart contract. Meanwhile, the smart contract is responsible for setting the warning threshold of infectious diseases and undertaking the operation of automatically broadcasting early warnings throughout the network. Finally, the whole process of infectious disease blockchain is supervised by decision-making departments and the highest CDC. The details of the proposed blockchain-based method for enhancing infectious disease traceability are provided as follows:

Participating Entities: The tracing of infectious diseases requires many participating entities and their concerted efforts; such entities include health institutions (such as hospitals designated for caring for the infected patients, the CDC of related cities/regions, etc.), decision makers (Administration and the national CDC), and ordinary participants (researchers, healthcare experts, and the public). Some entities are on both the inputting end as well as the receiving end (for example, hospitals).In the process, regional hospitals and CDCs mostly play the role of information generators; they are the originating points of infectious disease

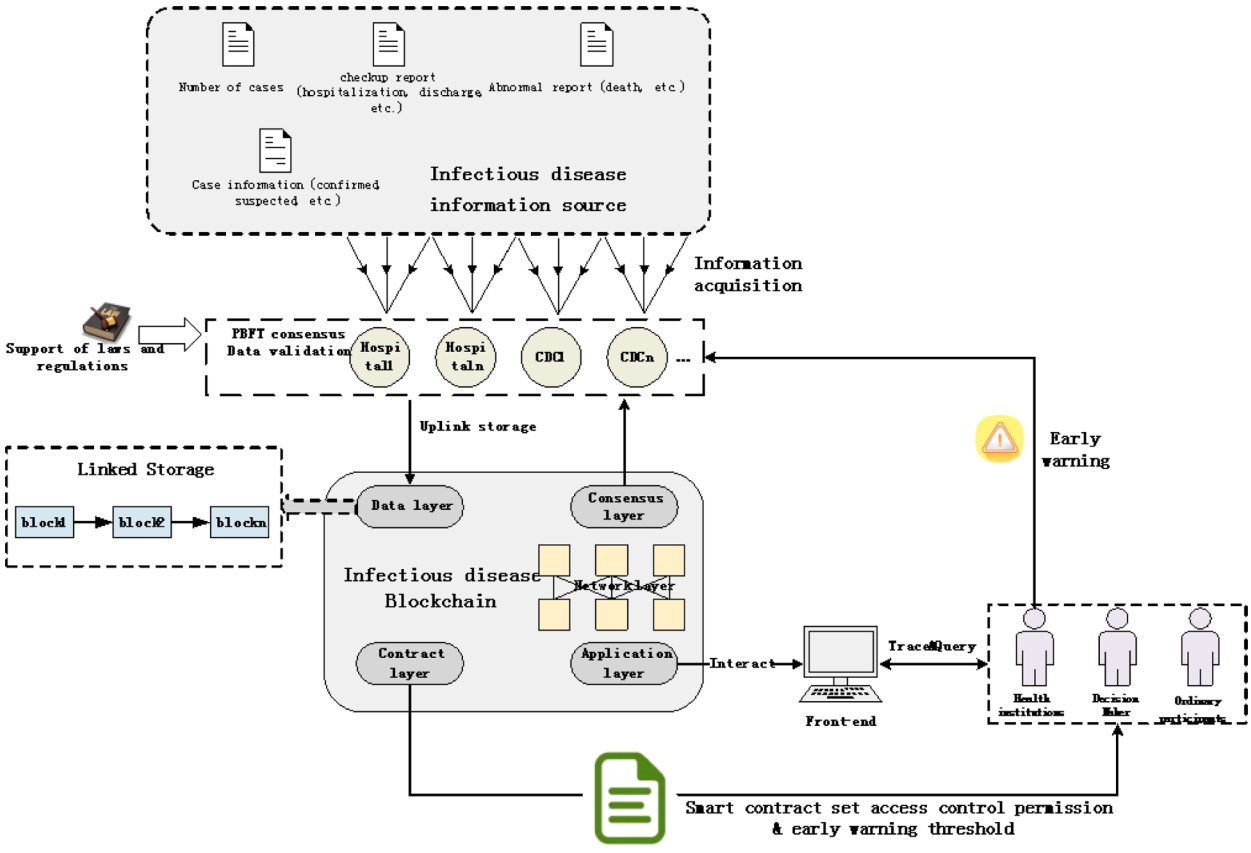

Fig. 1. Blockchain-Based Disease Information Tracing Model

information chain, and they are responsible for the collection of disease information and for the storage of the information on the chain. The data/information to be stored on the chain is published through the P2P network of the blockchain. After data/information is agreed upon by the nodes in the consensus layer, it will be recorded in the data layer in a chain-wise manner (França et al., 2020). The Ordinary participants (e.g., social groups, medical experts, academic researchers) and the decision makers are the destinations of the information flow, and they are responsible for receiving information. Through interacting with the application layer, they work on the front-end platform (Dapp/web mode) to obtain the propagation status and the initial scenario of the outbreak. Among the players, hospitals and CDC can also be in the role of end-points for receiving information. While all the node entities can play the functions of information manager, different node entities have different functions and privileges. Leveraging the smart contract to set the access control, the disease information stored on the chain can be verified and managed. This process can form an environment for authentic and transparent disease reporting. In addition, each participating entity must register a node on the blockchain, and obtain a unique address identity. This ID consists of pubic key and private key in asymmetric cryptography; each storage transaction or infectious disease information transmission is associated with a specific node account.

Objective: During the outbreak of infectious diseases, all the information related to infectious diseases will be uploaded to the infectious disease blockchain by the relevant nodes. The monitoring node and any node can obtain the current information by tracing the historical infectious disease data, so as to take targeted prevention and control measures. It can also form a monitoring mechanism: the immutable feature of the blockchain ensures the authenticity and reliability of all infectious disease information, and it can track and investigate the responsibility of relevant parties according to the sequence of time stamp records.

System workflow: As shown in the above figure, in the proposed model, when an infectious disease breaks out, different types of infectious disease information (case information, number of cases, abnormal reports, etc.) will be collected by the nodes of local medical institutions. After node consensus and data validation, it is stored in the infectious disease blockchain. The stored information can be used for traceability and other research analyses, or for generating early warning signals, or just for ordinary participants' queries and research. All information and data are entered from the front-end, stored in the blockchain, and can be called back through the front-end interface for viewing. When all information and data are arranged chronologically in the blockchain, they are considered "non-tamperable" and "traceable".

System architecture: The hierarchical structure of applying blockchain to infectious disease traceability system mainly includes data layer, network layer, contract layer, consensus layer and application layer. Data layer is the core layer, which is used to store

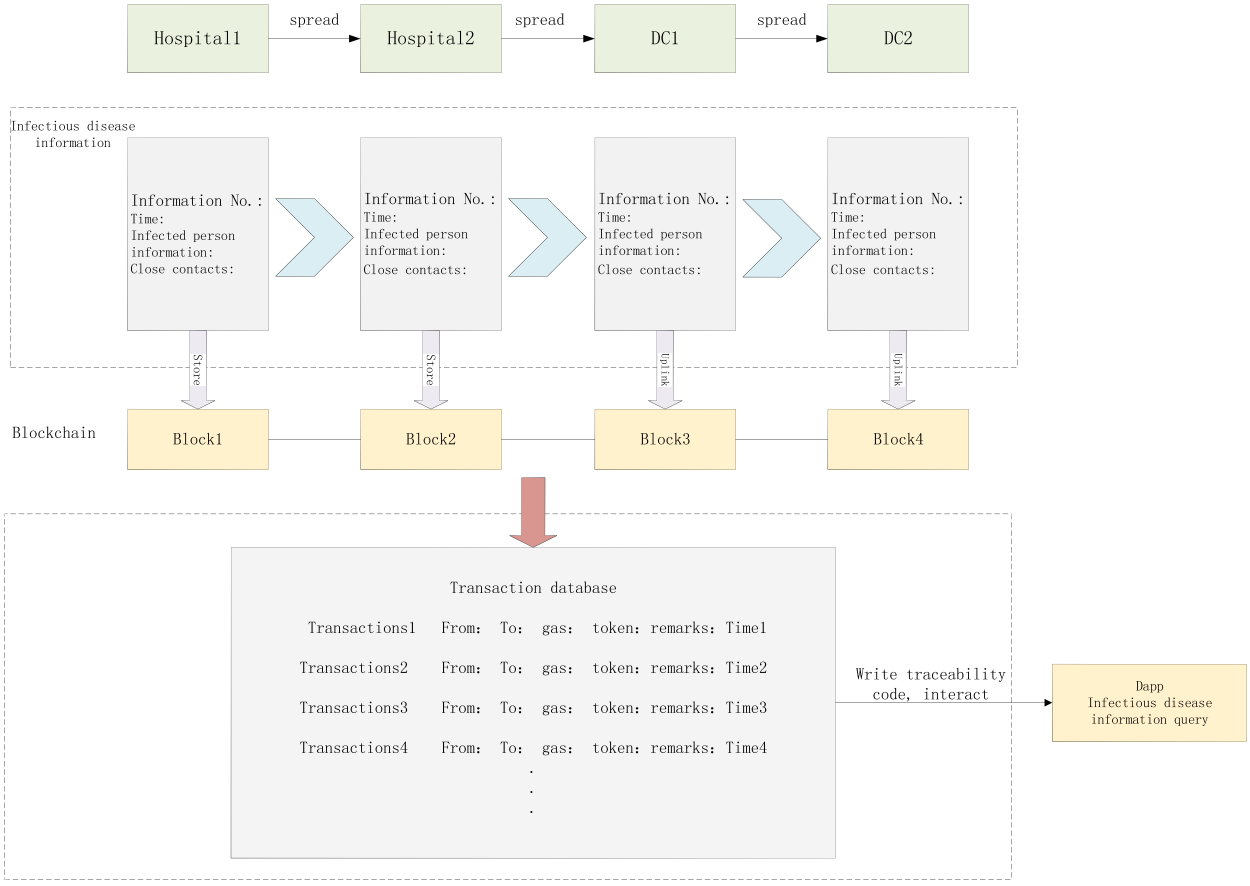

Fig. 2. Infectious disease information storage and tracing

the collected infectious disease information; P2P network layer is the basis of information transmission; consensus layer is used for node consensus and data verification of information to be stored. The contract layer is the core of the whole system method, and the smart contract is included in it. It encapsulates the code to transfer the system functions, and uses the smart contract to set the operation permissions and thresholds in advance to complete data management and early warning. The application layer enables users to interact with the underlying blockchain and provides an indirect interface for all participating nodes to query, trace and supervise. In this paper, we will explain the method at four levels. In the data layer, we store the infectious disease information; in the application layer, we trace the stored information; in the consensus layer, we carry out the consensus mechanism and data verification; and in the contract layer, we use the smart contract to set the early warning threshold and the permission set.

3.2. *Infectious disease information storage and traceability functions*

Blockchain-based infectious disease information storage is to leverage blockchain technology to achieve information recoding and storage during the time of an epidemic. In such a disease direct reporting system, each entity can conduct the autonomous node upload of disease-related data or information. This process is demonstrated in Figure 2. In Figure 2, Hospital 1 handles the infectious disease information – confirmed cases, suspected cases, infected patients' information, etc. - that the hospital captured. The time of the upload is recorded automatically with the time stamp in the blockchain. The blockchain would store the uploaded information as a fixed-length hash value, and mark it with a time stamp. Through the agreement on the blockchain, the other nodes on the blockchain would be able to see the disease information uploaded by Hospital 1. Hash function is a function that can map data of any length to a fixed-length function value. In the disease information upload on the blockchain, a hash function can perform hashing on the data and obtain a fixed-length hash value (Liu & Li, 2020). Every piece of data has a hash value that is unique and irreversible. When the input data has a slightest difference/change, the resulted hash value will change significantly. Therefore, when it is needed to verify whether the information on a blockchain was modified, it only takes a hash function computation on the original disease information, and compare the resulted hash with the hash stored on blockchain; if the two values are the same, the disease data can be believed to be intact (no tampering or forging). Therefore, when the next hospital (Hospital 2) needs to upload new disease case(s), the hospital can compare the previous piece of information with its hash, to assure the integrity of data. If Hospital 2 also has new disease information to upload to the blockchain, the new information unloaded would still be stored as a hash. The tamper-proof disease information storage provide assurance for the truthiness of tracing, from this tamper-proof storage method.

For the update of infectious disease information, due to the chain storage of blockchain, all the information needs to be recorded in chronological order and cannot be modified. Therefore, when any health institution node verifies that the stored infectious disease information needs to be updated again, it will send an information update request to the infectious disease blockchain system for consensus verification. The updated infectious disease information that passes the consensus will be recorded on the blockchain with the current upload time. Then, the system will broadcast to the national CDC and management department to verify and update the information at the terminal. Finally, the national CDC and relevant management departments broadcast it to all participating nodes.

On the function of tracing, the information collection and storage in the disease reporting system is not one that lower-level nodes collect data to submit to the upper-level nodes: it is not a hierarchical sequence. Instead, all nodes nationwide that can collect related data would take part in the uploading and storage operations, thus forming a horizontal transfer process of disease information. From the beginning of the disease information record to the final end, every time the previous node finishes its recording starts a data transfer process, which is a transaction on the blockchain. Every disease transfer transaction would have a new block being generated to record the transaction process, including the transaction addresses of the transfer-out party and the transfer-in party, the transaction amount, transaction time, and attached contents (this information is unmodifiable). The From-address represents the unique address label of the last node, and the To-address represents the unique address label of the current node. Therefore, based on a block of information of the transfer of disease information, we can clearly know the disease information storage process (e.g., party inputting, information input) that happened right before the information recording with the node of the current institution.

Since the blockchain is connected by the hash function value generated by the previous block that connects the blockchain and forms the sequential arrangement of past transactions, any node can trace back based on the immediate last output node's transaction address, and to query all the transaction information that contains this specific transaction address, until it reaches the destination transaction input node. As shown in Figure 2, once the disease information "transaction" was successful between previous node and next node, the information would be stored in the blockchain, which would form a "transaction" database. Every transaction is a block;

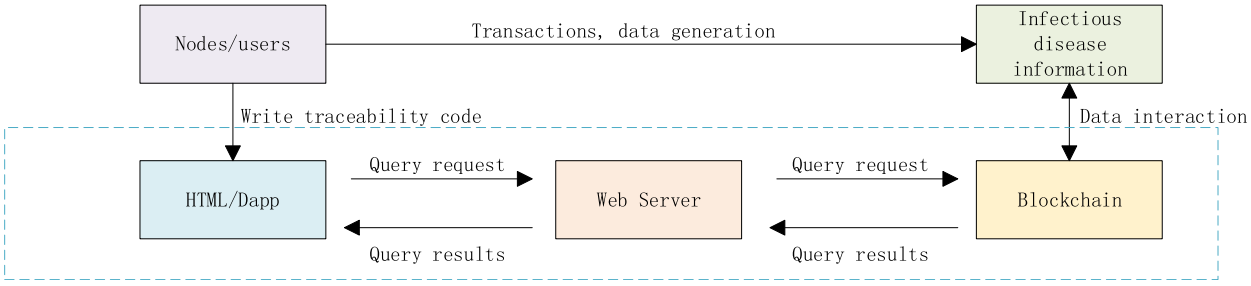

Fig. 3. Interaction process of tracing

each block is distinguished by a self-own hash value. In the specific block here, there are information contents of from sender, to receiver, time the transaction happened, and message of transaction, etc. The input entity in the previous block is the output entity of the next block, thus connecting the blockchain. By inputting the address of the current transaction exporter, the transaction information containing the address in the blockchain can be traced, so as to query all infectious disease infection information that meets the requirements of the conditions, and finally find the initial infection status when the infectious disease spreads.

The final outcome of tracing can be obtained by the relevant entity satisfying the authority through the front-end page or DAPP software (a distributed application). When related entities request queries from a web server, the web server issues a query request to blockchain tracing system by using SDK or API approaches. Finally, the blockchain traceability system judges the source according to the transaction confirmed by all parties, returns the query results, and displays the data information in the window mode on the front-end visualization platform, as shown in Figure 3.

3.3. Consensus mechanism

In a disease tracing direct reporting system that is decentralized and distributed, how to reach consensus is a critical issue. The key is in the consensus layer. In this study we adopt PBFT(Practical Byzantine Fault Tolerance) consensus algorithm to reach consensus for all nodes. The main principle is that one node in the system will be taken as the master node, while all other nodes are child nodes. All nodes in the system will communicate with each other, and the ultimate goal is that everyone can reach consensus with the principle of minority subordinate to majority (Wang et al., 2019). PBFT algorithm, as the consensus algorithm for disease tracing blockchain, has the following features:

- PBFT algorithm and POW and POS algorithms have different methods to reach consensus; it does not need large amount of computing power, and it does not rely on virtual currency as the criterion to measure voting rights
- PBFT algorithm only allows the maximum of (N-1/3) dysfunctional or malicious nodes for the normal execution of one consensus process. In other words, among all nodes in the system, for every consensus computation, there must be at least (2N+1)/3 normal nodes. These nodes must be in a relatively safe and stable operation environment (H. Wang & Song, 2018).
- The entities participating in the infectious disease tracing blockchain are mostly hospitals, CDC, or government executive departments. They possess certain credibility to the public, and they are usually tightly monitored by multiple regulating institutions/departments. The probability that these entities would perform malicious conducts is far lower than many other blockchain systems. At the same time, every provincial level CDC and hospitals all have better functioning networks, servers, and databases; these together provide a relatively safe and stable operating environment for the normal operation of the PBFT algorithm. Therefore, PBFT algorithm is very suitable for blockchain-based infectious disease tracing.

The working principle of the model proposed in this study is: a specific node makes a request to make a record, each node would judge whether it should take charge to generate the current block, based on the formula (1),

$$B = L\%(N-1) \tag{1}$$

where

B – the node to general the block,
L – length of the blockchain,

Because the main node needs the request from user end applications, the main node itself does not participate in the generation of the block, so as to pursue the balanced load among all nodes. If certain central node figures that it needs to generate a block at the present time, this node would then collect certain amount of disease data that has been checked and verified. The node will package the collected data into blocks within one minute and send them to the master node. Through the PBFT algorithm's three stages, this new block is appended to the end of the blockchain of this node. Through the above process, it can be assured that when malicious or problem nodes is less than (N-1)/3, the consensus process can be accomplished successfully.

3.4. Early warning of infectious diseases based on smart contract

The key objective of early warning of infectious diseases is to identify the abnormal situations where case numbers exceed the predetermined threshold level. Early warning in early stages of infectious diseases can alert all the organization nodes to reduce the risk of large-scale pandemic outbreaks. Early warning of infectious diseases is closely linked to the traceability of infectious diseases. The National CDC can monitor potential abnormal conditions based on the infectious disease information uploaded by the subordinate regional nodes. The traditional early warning system of infectious diseases has obvious deficiencies in terms of timeliness. Potential concealment and tampering of data at all levels of nodes make it very difficult to ensure the accuracy of early warnings. To improve the efficiency and accuracy of early warnings, and to facilitate joint maintenance and rapid responses of nodes at all levels, we propose to implement early warning of infectious diseases using a smart contract function on blockchain. Smart contract is a computing transaction protocol used to execute contract terms (Hu et al., 2021). The protocol can automatically complete the relevant operations to achieve the predetermined conditions without the intervention of a third party, achieving the effect of quick triggers and efficient

responses.

Specifically, when any participating node uploads new infectious disease case information to the infectious disease traceability blockchain system in real time, the system will judge whether the new information triggers early warning specified by the rules of the blockchain smart contract. If yes, the blockchain system will broadcast an early warning to all participating nodes. As each node continuously uploads the data of infectious disease cases, the early warning smart contract will compare the number of infectious disease cases in the current observation period in real time. Based on the existing infectious disease data, the smart contract can use a moving percentile method to dynamically calculate the historical baseline of infectious disease, that is, the threshold for early warning. The early warning smart contract will be released to the distributed infectious disease traceability blockchain according to the early warning logic and threshold of infectious diseases. The smart contract includes early warning logic, threshold, early warning broadcast, risk record and early warning account book. All judgments are automatically executed by the smart contract, and the national CDC is responsible for supervision and management.

As the outbreak of infectious diseases is usually very rapid, the early warning of linkage is also a very important component of the traceability of infectious diseases. After the early warning contract of the blockchain is triggered, the raw data and the synchronous discovery of epidemic outbreaks can be shared in real time through full network broadcasting, transforming the traditional hierarchical early warning arrangement into a more agile and flexible one.

3.5. *Participating entities and access privilege*

The control and management of access rights directly affect security and integrity of the data stored on the blockchain (Putz et al., 2021). The method proposed in this study aims at providing the control on authorization, and providing different access privileges to the three participating entities –Health institutions such as hospitals and CDC, decision making departments, and ordinary participants (researchers, and the public); the different access privileges are granted based on different roles of different entities. We adopt the role-based access control (RBAC) model to implement the access rights control and management, and use a smart contract to complete the functions of initial permission setting, node identity authentication and role matching. We define three different level of rights/privileges: monitor, storage and distribute, and query, as shown in Figure 4. RBAC maps users to their roles; it can define different roles, the relationship among roles, and the related privileges, with which the security of data can then be assured (Ghafoorian et al.,

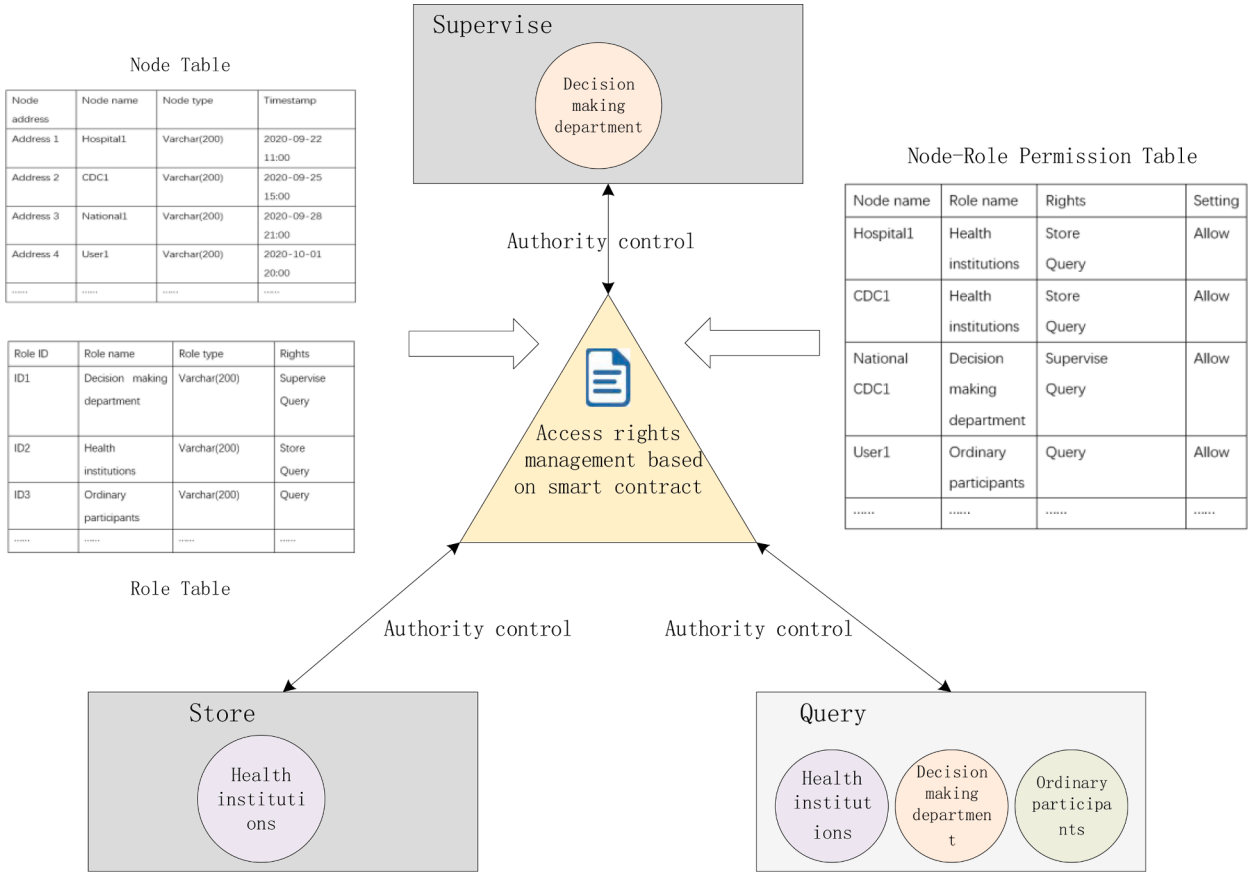

Fig. 4. Access control management

2019).

The roles and privileges of each entity can be summarized as follows:

Health institutions

Health institutions are the nodes that are the hospitals and CDC in local regions. These are the most important nodes in the process. All infectious disease information is collected and stored onto the blockchain by these nodes; these nodes can also function as the ultimate party in tracing to query the disease records. Having the rights of storing and querying, they can, within their authorization time frame, append, distribute, and view infectious disease information.

Decision making department

Decision making department nodes are usually run and controlled by the highest central government departments, including the state CDC, and related state ministries. They are charged to monitor and manage the infection chain; they can also obtain disease information from the infection chain as the information receiver, in order to formulate rational measures and disease control solutions. Within their authorization time frame, they do not have the privilege of uploading and modifying information.

Ordinary participants

Ordinary participants are researchers, healthcare experts, and the public. Within their authorization time frame, they can only perform queries of disease information. This type of nodes mostly utilizes the disease information queries to conduct scientific research or to be updated with the development of the epidemic; they are not responsible for the management of the nodes. They perform all the needed operations through user end applications.

4. Simulation details

In order to test the blockchain-based infectious disease tracing solution, we wrote a program using Python 3.7 to implement the algorithm to simulate the storage and tracing of infectious disease data. The coding was conducted in the Pycharm professional development environment.

From the procedural view, when a node needs to upload disease information, the blockchain will transform infectious disease information content in the form of hash function value as a unique identifier of a block. If the content of the information is altered, the later block would judge whether the information maintains integrity based on the hash value; if hash is no consistent, the block would be marked as invalid. In the process of tracing, those blocks containing Time and Message that satisfy the specified address conditions would be displayed. The tracing process would also use the final input party's address to trace back to the initial output party's address, with each address uniquely corresponding to one medical institution. The detailed implementations of Algorithms 1-7 are presented as follows.

4.1. Using smart contract to set access rights and early warning threshold

Algorithm 1.

Algorithm 1
Access rights and early warning threshold setting

```
Input: Roles, Rights, Threshold
1 Roles and rights are used to set access right sets, and threshold is used as trigger conditions for early warning
2Struct Roles(address){
3     address User address;
4unit Health institutions;
5unit Decision making department;
6unit Ordinary participants}
7mapping (address => rights);
7Struct rights{
8Health institution => function (store, query)
9Decision making department => function (supervise, query)
10Ordinary participants => function (query)
11function early warning {
12uint i;
13uint j;
14   int threshold = Prescribed threshold for infectious diseases
15int temp
16 temp = Number of infectious diseases (Sum of institution[i] data[j])
17if temp >= threshold {
18     Whole network broadcast warning}
19 End
```

Algorithm 2
Information upload

Input: Address, SP state, Timestamp, Information hash value
1*Address* is the identity mark of the node in Infectious disease blockchain; it is unique. *Timestamp* and *Information hash value* represent the uploading information.
2if *Address*= = Nodes on the chain then
3if SP state == Not Available information then
4 Emit an event to notify every node that an Infectious disease information is uploading with *Timestamp* and *Information hash value* is submitted.
Generate first block.
SP state = information submitted.
5End
6else
7 Preview an error after returning the contract to its previous state.
8End
9End
10else
11 Preview an error after returning the contract to its previous state.
12End

Algorithm 3
Node consensus

Input: Client, master node, other nodes
1 *Client* is responsible for sending information uplink request
2 *master node* is responsible for the preliminary review of messages
3 *other nodes* are the main nodes of message consensus
4*Client* sends infectious disease information uplink request to the *master node*, and the *master node* verifies
5If Information signature==true **then**
6 broadcast a pre-prepare message containing information to *other nodes*
7 the *other nodes* validate the pre-prepare message
8If Verification passed **then**
9 broadcast prepare message
10If*other nodes* received more than 2f + 1 prepare messages **then**
11 broadcast commit
12If *other nodes* received more than 2f+1 commit **then**
13 reach a consensus, store the message and return the result to the *Client*
14Else illegal request discard
15Else illegal request discard
16Else illegal request discard
17Else illegal request discard
18 End

Algorithm 4
Forming blockchain

Input: create_genesis_block, add_block, bc, b1, b2
1 *create_ genesis_ block* is the initial block; *add_ block* is a newly added block; *bc* represents the blockchain structure, containing the initial block; *b1* is new block 1; *b2* is the new block 2.
2class IDBlockchain:
3def _init_(self):
4 self.blocks=[*create_genesis_block*()]
5def *add_block*(self,data,timestamp):
6 prev_block=self.blocks[len(self.blocks)-1]
7 new_block=IDBlock(data,datetime.datetime.now(),prev_block.hash)
8 self.blocks.append(new_block)
9 return new_block
10 if _name_ == '_main_':
11*bc*=IDBlockchain()
12*b1*=bc.add_block("data1",datetime.datetime.now())
13*b2*=bc.add_block("data2",datetime.datetime.now())
14 for b in bc.blocks:
15 print("prev Hash:{}".format(b.previous_hash))
16 print("Time:{}".format(b.timestamp))
17 print("Data:{}".format(b.data))
18 print("Hash:{}".format(b.hash))
19End

Algorithm 5
Tampering with infectious disease information

```
Input: b, b1
    b indicates block modification data
    b1 is the block stored on the chain
      def change_hash(b):
    b.data=" Modified data1"
  for i,b in enumerate(bc.blocks):
        print("prev Hash:{}".format(b.previous_hash))
        print("Time:{}".format(b.timestamp))
        print("Data:{}".format(b.data))
        print("Hash:{}".format(b.hash))
  if b.previous_hash and b.previous_hash!=bc.blocks[i-1].hash then
          print("invalid block")
  else
          print("valid block")
  change_hash(b1)
  End
```

Algorithm 6
Tracing the infectious disease information of matching address

```
Input: Block, Node, Message, Result
    Block contains all the stored infectious disease information and the from / to Deliverer
    Node is the transaction address in the blockchain
    Result is the address to be traced
    Starting from block1
    If Node == Result then
    If Block==Block1
     Output the current block message (from)
    Else
     Output current block message and next block message (to, from)
    Else
     traverse block2 down
     until blockn
    End
```

Algorithm 7
Tracing the initial node from the final node

```
Input: Block, Blockn, Block1
    Block contains all the from / to Deliverer
    Blockn is the last block in traceability query
    Block1 is the first block
    Starting from Blockn
    Output the current block address (to)
until Block1
If Block==Block1
    Output the current block address (from)
End
```

4.2. Infectious disease information upload

Algorithm 2.

4.3. Consensus among nodes

The information uploaded to the blockchain must go through consensus and verification. Only information passed such consensus and verification can be uploaded to the blockchain. When a user app terminal requests a connection to the chain, it needs to go through the consensus process of pre-preparation, preparation, and commitment to return the results. (Algorithm 3.)

4.4. Formation of back end blockchain structure

All nodes upload the collected infectious disease information in the form of blocks, thus forming a blockchain structure. (Algorithm 4.)

4.5. Disease information tampering

The following showcases the attempt to tamper the disease data uploaded to the blockchain; such operation can be tested through the visualization results of the effectiveness of the subsequent blocks. (Algorithm 5.)

4.6. Infectious disease information tracing (I)

Based on the provided institution's address, perform a quick tracing of all transferable disease information that contains such a given address. (Algorithm 6.)

4.7. Infectious disease information tracing (II)

Starting from the last block before tracing, back track the addresses of all uploading nodes, eventually reach the initial node. (Algorithm 7.)

5. Testing and verification

Based on the details of the algorithm discussed in the previous section, we test the algorithm and provide the final visualization of the outcome about the information query and traceability of infectious diseases.

5.1. The Output of infectious disease storage and tampering

The infectious disease information that is provided by two input parties ("Institution 1: There are 50 confirmed cases and 100 suspected cases. Specific patient information: 1. Zhang San……"&"Institution 2: There are 70 confirmed cases and 160 suspected cases. Specific patient information: 1. Li Si……") has been stored on blockchain, with time stamps, and a hash value is generated from the submitted disease information. So a new block is connected to the initial block's value, forming the blockchain. The process is indicated in Figure 5.

We change the disease information of one of the blocks in the example (changed the information in the second block to "Institution 1: There are 10 confirmed cases and 50 suspected cases. Specific patient information: 1. Zhang San……", as compared to the original information "Institution 1: There are 50 confirmed cases and 100 suspected cases. Specific patient information: 1. Zhang San……"). Through the comparison of the current hash value of the block with the hash value of untampered block, tamper can be discovered, and the blocks would be marked as invalid block or valid block respectively. Valid block means that the pre hash value displayed in the current block corresponds to the previous block, that is, the data information of the previous block has not been tampered with. At this point, the previous block is real and valid. Of course, if the data of the previous block is modified, the hash value of the block will change. Thus, it will not match the prev hash value on the next block, and it will be displayed as an "invalid block". This is shown in

```
prev Hash:
Time:2020-06-18 10:32:25.213203
Data:Genesis Block
Hash:89eb0ac031a63d2421cd05a2fbe41f3ea35f5c3712ca839cbf6b85c4ee07b7a3
--------------------------------------------------------------------------------------------------------
prev Hash:89eb0ac031a63d2421cd05a2fbe41f3ea35f5c3712ca839cbf6b85c4ee07b7a3
Time:2020-06-18 10:32:35.213203
Data:Institution 1:There are 50 confirmed cases and 100 suspected cases. Specific patient information: 1. Zhang San……
Hash:309b3fc946108afcb7829a37d59118f46ba4d187becde24f57c0b6741a25973f
--------------------------------------------------------------------------------------------------------
prev Hash:309b3fc946108afcb7829a37d59118f46ba4d187becde24f57c0b6741a25973f
Time:2020-06-18 10:32:45.213203
Data:Institution 2:There are 70 confirmed cases and 160 suspected cases. Specific patient information: 1. Li Si……
Hash:1c6ae44839d0b8d17ddd56c86be86272d1c00d215ed75f209cb8df8020932c2b
--------------------------------------------------------------------------------------------------------

Process finished with exit code 0
```

Fig. 5. Infectious disease uploaded to blockchain

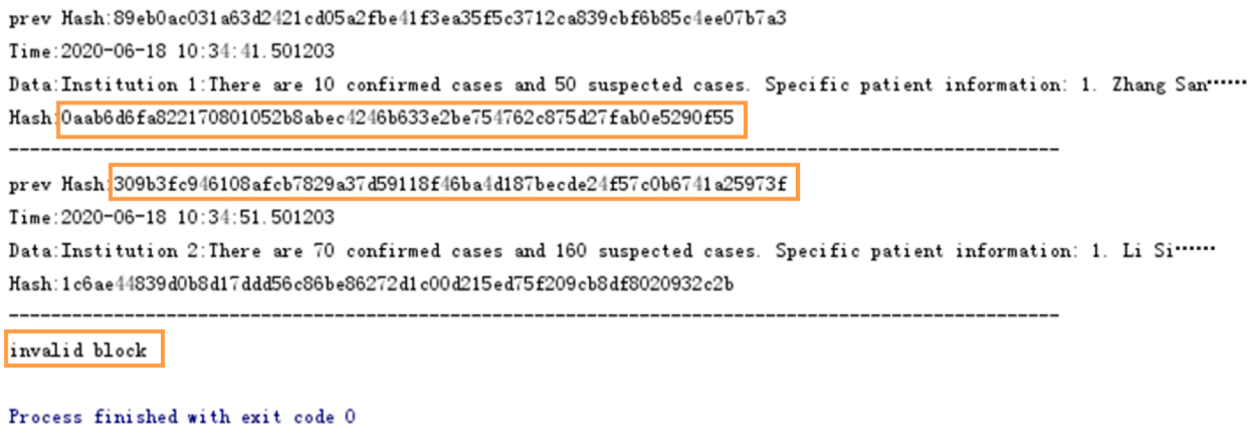

Fig. 6. The result after attempted tampering of a block

Figure 6.

From the result we can see that when the data on the block was tampered, the original block's corresponding hash value completely changed, resulting in the failure for the succeeding blocks' hash value to match with the preceding one. As a result, all succeeding blocks become invalid.

5.2. *The outcome of infectious disease traceability*

Explanation on the addresses below: The nodes participating in the uploading and transfer on the blockchain are healthcare institutions (hospitals and local CDC) of involved city or local regions. Each node possesses one unique blockchain address. In the simulation in our study, the following node addresses are involved:

- Hospital1: 96be0ac031a54d2421cd05a2fef23f3ea35f5c3124ca839cbf6b85c4ee07b7c5
- Hospital2: bc8ce83fb7306a4aec4074dd676de84fe7e9614159db325ff02f474f55d01659
- DC1: d0c99c4809defac64084dea4b620e3cfd7ca0271bf5ba6c6e235ef8e427522bc
- DC2: 92507c5730268562541975030303051250d73ebcbec6ec8b8e5842af1fdff4ffc9
- DC3: dd60f05174f2641f054bb4a7cabc510ac46c065f50a3e39abdcf9268c84f02e9

Through the traverse of all information on the blockchain, we queried all transaction information containing the designated address:92507c5730268562541975030303051250d73ebcbec6ec8b8e5842af1fdff4ffc9.

The result is shown in Figure 7:

At the same time, take the last To-address as the query point, we can perform a forward tracing over all outputting parties' transaction addresses, that is, from the final receiving party address to the initial inputting party address. This can achieve the related information tracing over the complete propagation process of the infectious disease. The result is shown in Figure 8, which is for Hospital 1, the earliest institutional node that uploaded the information of the infectious disease.

In the storage and tamper-proof phases of infectious disease, every block has one and only one unique identification value. No node can tamper the data that has been stored, all information is authentic and cannot be forged. Blockchain-based disease direct reporting system can transparently maintain and present the most accurate data to the society. On the front of traceability, blockchain technology can anchor on a designated hash transaction address, and trace all propagation processes of disease information involving this address as the inputting and outputting parties. This can achieve the periodic management of disease information on its chain of propagation. Such a system can provide complete data/information to decision makers and participating entities, allowing them to keep up with the current epidemic status, conduct further and deeper analyses of the disease-related data.

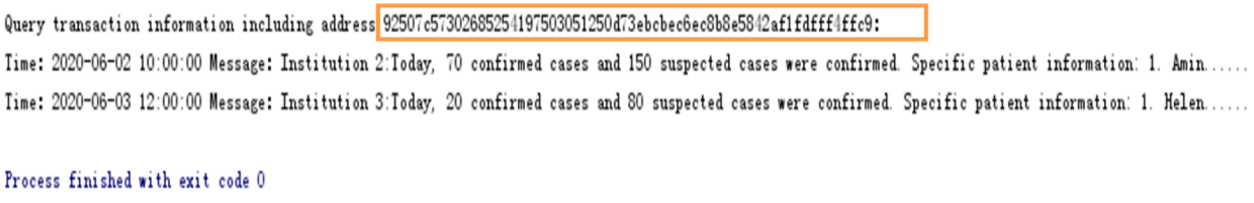

Fig. 7. The result of specified node

```
Trace initial mechanism node:
    DC3: dd60f05174f2641f054bb4a7cabc510ac46c065f50a3e39abdcf9268c84f02e9
    DC2: 92507c57302685625419750303051250d73ebcbec6ec8b8e5842af1fdff4ffc9
    DC1: d0c99c4809defac64084dea4b620e3cfd7ca0271bf5ba6c6e235ef8e427522bc
    Hospital2: bc8ce83fb7306a4aec4074dd676de84fe7e96144159db325ff02f474f55d01659
    Hospital1: 89eb0ac031a63d2421cd05a2fbe41f3ea35f5c3712ca839cbf6b85c4ee07b7a3


Process finished with exit code 0
```

Fig. 8. Initial institutional node traceability

6. Analyses and Discussions

In this section, we conduct analyses on the security, efficiency, and functions of the proposed method, to further assess the blockchain-based solution, and to verify its reliability and feasibility. We also discuss the potential challenges to the implementation of our proposed method.

6.1. Security analysis

A blockchain may be subject to a type of attacks called "51% attacks". A blockchain's integrity will be compromised when an attacker controls more than 50% of resources on a blockchain (Ye et al., 2018). "51% attacks" mostly occur in situations where the PoX series consensus mechanism is adopted. Using the consensus algorithm of PBFT, our proposed method can greatly reduce the likelihood of this type of attacks (PBFT can provide (n-1)/3 fault tolerance while ensuring availability and safety). Moreover, most of the physical nodes participating in the infectious disease blockchain are hospitals or CDCs, which have a certain degree of credibility and are strictly supervised by the national CDC, the Health Bureau or other government agencies. Thus, malicious behaviors and dishonest nodes occur with much smaller probabilities. In addition, our proposed method adopts P2P architecture, which can effectively prevent single-point attacks. With the combined maintenance by all nodes, the system's robustness can be assured. At the same time, through determining the identity of each user and setting fine-grained role-based access, we can assure that different stakeholders play different roles on the tracing chain, assuring data privacy on the chain.

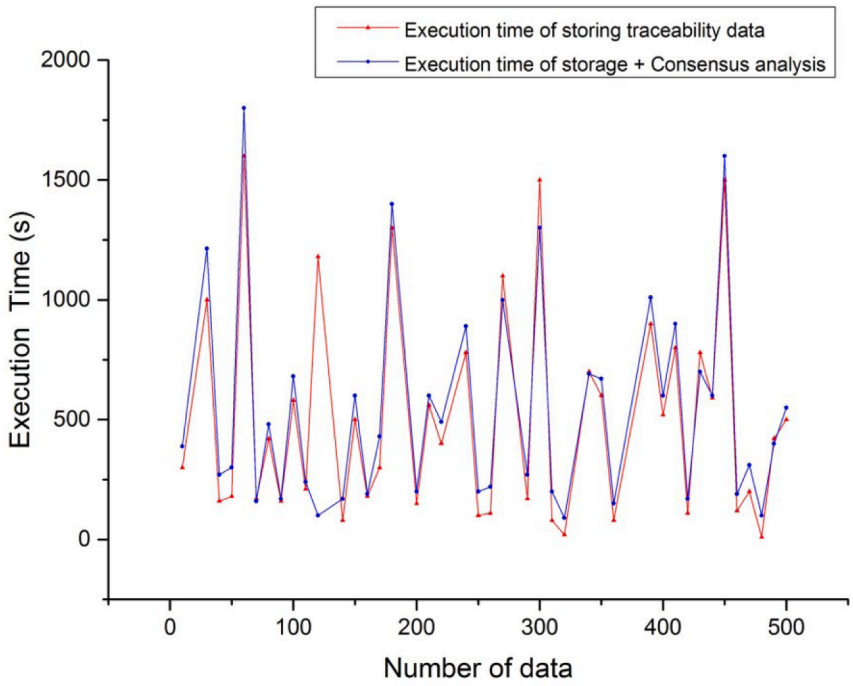

Fig. 9. Consensus mechanism storage

6.2. Efficiency analysis

The simulation platform of this paper is: Inter (R), Core i5 CPU, Memory 16GB, and the Operating system is Windows10 64 bit. We used 150 nodes in the simulation to analyze the efficiency and performance of this scheme. Performance is crucial for evaluating the effectiveness of a structured blockchain (Xu et al., 2021). In data storage, the generation of each block needs a consensus operation in a given time interval (Figure 9).As the number of dishonest nodes increases, the storage efficiency will decrease because of the frequent needs of the consensus operation. In future research, therefore, we can try to add a credit integration mechanism based on PBFT consensus algorithm. The nodes participating in block consensus generation will be weighted according to credit scores, and rewards and punishments will be imposed for block successful verification or malicious behavior. Dishonest nodes with low credit can join and leave the consensus process in a benign and dynamic manner, so as to shorten the confirmation time and improve the storage efficiency. With the long-term operation process, the credit rating of nodes with high error rate is reduced, and the low error rate of primary nodes is combined with the simplified consistency protocol and incentives. Within the upper limit of error nodes, block generation can be completed more efficiently than simply using PBFT.

In addition, the storage efficiency of a blockchain depends on the characteristics of data to be stored and the execution time used for consensus and verification. Comparing the storage efficiency of our traceable data type with those in JSON format (such as "ID": {}, "Infection information":{"confirmed number", suspected number"…}, "Time"…), we can see (as shown in Figure 10) that using JSON data format can reduce the demand for computing resources, thereby reducing the execution time of block generation and improving blockchain storage efficiency.

On tracing, our experiment tracked and traced 500 pieces of disease data; the resulted run time and data query efficiency are presented in Figure 11. From the figure we can see that, conducting tracing of disease information stored on blockchain can achieve high efficiency reaching the time magnitude of millisecond.

6.3. Effectiveness analysis

(1) Assessment of method

The informationization of the control of infectious disease has been fairly mature. With regard to the confidentiality and shareability, however, informationization still has many tasks yet to be completed. Table 1 lists some challenges facing infectious disease data storage, suggesting several blockchain-based solutions addressing the challenges.

(2) Comparison analysis

We compare the blockchain-based solution in this paper with hierarchical reporting systems (traditional method), and we further compare the existing medical information blockchain with our proposed infectious disease information blockchain. The major medical information blockchains are Factom (*Business Processes Secured by Immutable Audit Trails on the Blockchain » Brave New Coin*, n.d.), MedRec (Azaria et al., 2016) and Model Chain (Kuo & Ohno-Machado, 2018). A comparison of these products and our proposed method is presented in Table 2.

As shown in Table 2, in terms of the storage and query mode of infectious disease information, the existing hierarchical reporting system stores infectious disease information based on the traditional central database. While the traditional system doesn't require consensus mechanisms and significant computing, it is vulnerable to database attacks, information tampering and leakage. In addition, the hierarchical reporting system can only submit and report infectious disease information by regions in order of the hierarchy, and the organizations at all levels can only obtain the operation authority of the information at certain levels. The early warning process is also a top-down process. As shown in Figure 12, while the traditional system can ensure the screening and verification of infectious disease information to a certain extent, it usually suffers from low efficiencies of storage and querying during the outbreak of infectious diseases. Our paper shows how to leverage blockchain technology to improve infectious disease systems (as shown in Figure 13), changing the system from the single hierarchical structure to the whole network sharing, circular gear linkage structure. In such an environment, nodes at all levels can directly communicate with the highest central level, upload and query infectious disease information. At the same time, when the infectious disease case number exceeds the predetermined threshold, an early warning will be broadcast to the whole blockchain network, and nodes at all levels can respond expeditiously.

Because of the characteristics of blockchain-based network architecture, the integrity of infectious disease information can be more conveniently ensured. A hash function algorithm can be used to verify the chain hash code of infectious disease information stored on the chain, which ensures that the information is difficult to be maliciously tampered by some unauthorized nodes (perhaps for the sake of political gains or information concealment). At the same time, the smart contract of the blockchain can be set by the highest regulatory agency (the highest government / National Center for Disease Control and Prevention) in the initial stage, which can set the operation permissions of infectious disease information in accordance with different roles. Different user nodes can automatically identify and complete the relevant operations only with the specified permissions. The use of the consensus mechanism also ensures the authenticity and correctness of infectious disease information before storage on the chain. Although a consensus mechanism consumes some computing power and incurs some extra costs, it can lead a significant improvement in security. The consensus mechanism of POX series is used in the schemes of Medrec and Model Chain, which requires significant miner calculation, and its verification efficiency is relatively low. In this paper, PBFT consensus mechanism is applied to the infectious disease information blockchain, which improves efficiency (also mentioned in Section 6.2). Last but not least, in terms of application scope, most solutions

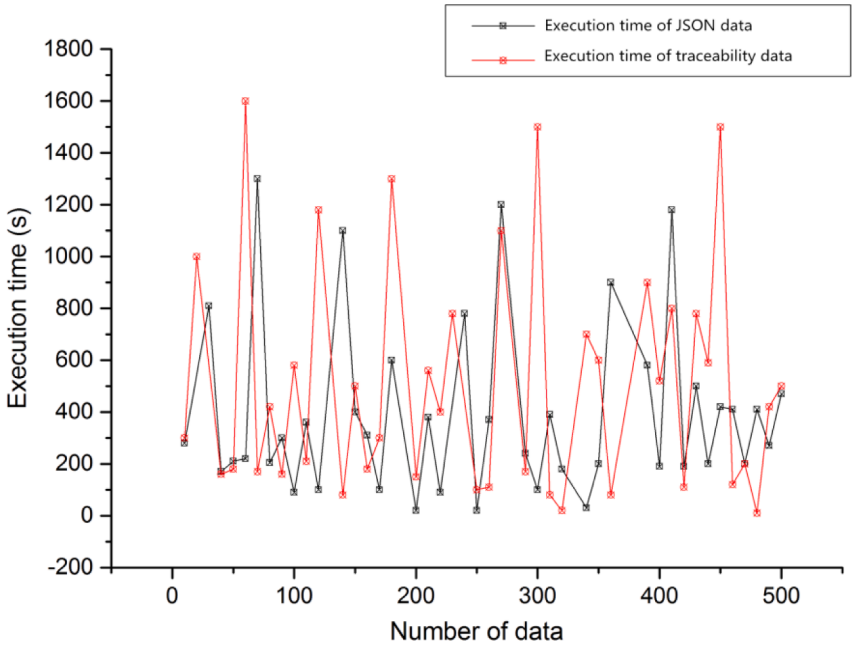

Fig. 10. Comparison of storage efficiency

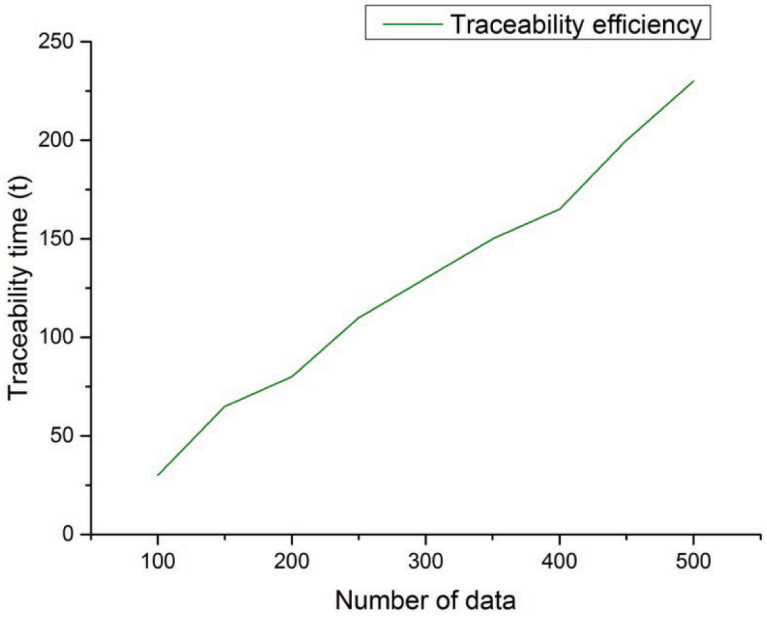

Fig. 11. Infectious disease tracing efficiency

of blockchain + traceability focus on medical information and supply chain information (very few focuses on infectious disease information). Thus, this paper can provide insights into new ways of infectious disease information storage and traceability.

To sum up, compared with methods used in similar research, the method proposed in this paper can significantly improve the security of infectious disease information, enhance storage and query efficiencies, identify legitimate authorized nodes for different data operation permissions, mitigate many problems caused by using a central database, and enable nodes at all levels to manage infectious disease information on a more coordinated basis.

Table 1
Problems and Solutions

Type	Existing problems	Model response plan
Data management and Authority	Monopoly Authority of Management Center	Refine the permission classification of each node, and associate the permission with the operation of the node
	Data centralized management storage	Decentralize the data generated by each operation of the user node and store it on the blockchain of all consensus nodes
	Vulnerable to malicious attacks	The data on blockchain is encrypted to ensure that the privacy of users will not be threatened. The data off the chain is backed up by multiple nodes to effectively prevent data loss from malicious attacks
Sharing and Security	Information island data protection	Blockchain PBFT consensus algorithm not only ensures data security, but also enables data to be shared among more than ten nodes, breaking the information island

Table 2
Comparison

Model/ system	Blockchain	Consensus mechanism	Whether to pay	Computational power	Security	Efficiency	Decentralization	Applicability
Hierarchical system	No	Null	No	Null	Lower	Common	No	Infectious Diseases
Factom	Yes	Null	No	Null	Low	Common	Yes	Bitcoin
MedRec	Yes	POW	Yes	Large	Common	Low	Yes	Medicine
Model Chain	Yes	POI	Yes	Large	High	Low	Yes	Healthcare
This paper	Yes	PBFT	No	Little	Higher	High	Yes	Infectious Diseases

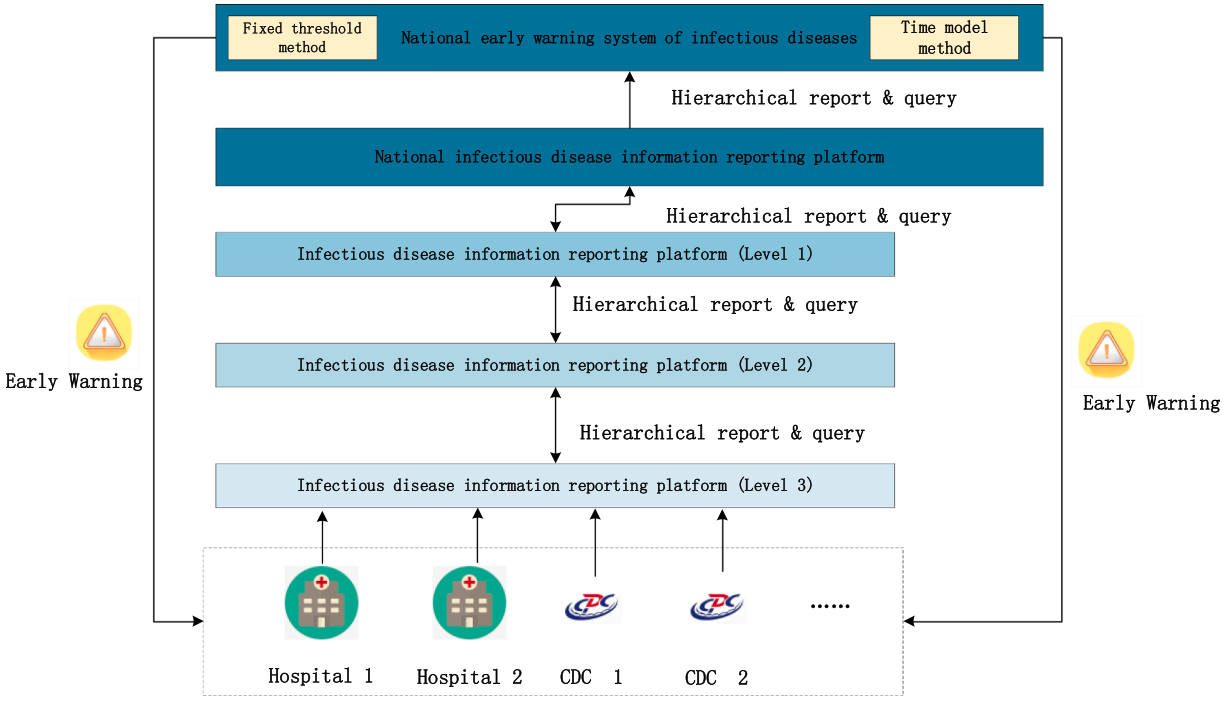

Fig. 12. Traditional infectious disease system

6.4. Potential challenges

The infectious disease information tracing method proposed in this study can effectively solve the common problems of traditional infectious disease reporting systems. As an emerging technology, blockchain still faces many challenges when applied to various arenas. First, while the disease information stored on blockchain possesses such desirable characteristics as reliability and unmodifiability, the credibility of data before uploading still relies on the network sources at the time of uploading. The authenticity of data still needs to be examined. This requires the use of some node consensus mechanism and the support of relevant laws and regulations.

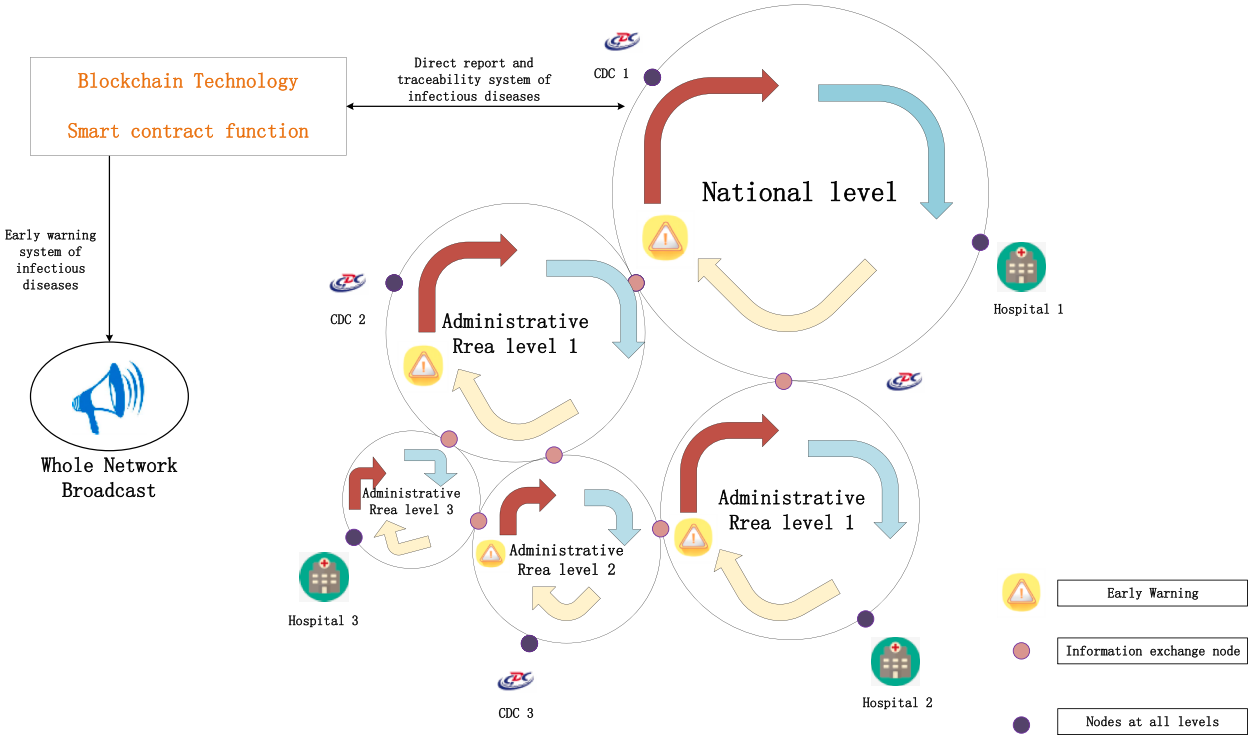

Fig. 13. Infectious disease system based on Blockchain in this paper

Second, for updating information on the infectious diseases blockchain, the current blockchain-based scheme can only send the update requests on the basis of the original storage records, and the highest level CDC or relevant departments need to verify and update the information on the back-end before broadcasting it to the whole network. In future research, we plan to adopt a parallel dual blockchain structure, so that two infectious disease information structures are reserved between adjacent blocks, and only one chain is destroyed by data modification. Third, all data and information involved are stored on a blockchain that requires high-capacity storage resources. Future research should target this problem and introduce distributed databases such as IPFS, to divide disease information into digest and record and store the two separately, which can make the method more efficient. At last, we employ Python to write the code to simulate a blockchain and to conduct method verification. The simulations are conducted on a single machine system (not tested on a cluster deployment). Thus, it is an important research direction to explore and identify better ways to verify the effectiveness of our blockchain-based method in a distributed computing environment.

7. Conclusion

The outbreaks of infectious diseases, as emergent public health incidents, are eminent threats to the human society. The goal of infectious disease tracing is to make sure that, in the process of disease propagation, all disease information can be truthfully and reliably recorded and stored, and be transferred to decision makers and other stakeholders in an open and transparent manner. The emergence of blockchain technology not only makes data storage more transparent and non-tamperable, but also makes it easier for the information on a blockchain to be queried and traced. Our study proposes a blockchain-based method for infectious disease information tracing. To demonstrate our method's feasibility, we expound our method's benefits in disease information storage, query, and tracing. Our simulations results suggest that the proposed method can meaningfully contribute to the improvement of infectious disease reporting systems. As the blockchain technology advances, we will undoubtedly witness its broader applications in many fields. As far as the application of blockchain in infectious disease information tracing is concerned, many exciting research issues remain to be investigated. Our study can also be extended in several directions. For example, an extension of our study is to use Hyperledger Fabric (an application of blockchain 3.0) to implement our method. We encourage other researchers to explore blockchain technology's capabilities to further enhance the traceability of infectious diseases.

CRediT authorship contribution statement

Peng Zhu: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Jian Hu:** Writing – review & editing, Methodology. **Yue Zhang:** Writing – review & editing. **Xiaotong Li:** Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

The authors have no competing interests to declare.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 71874082, Humanities and Social Sciences Foundation of Ministry of Education of China under Grant 18YJA870021, and Philosophy and Social Sciences Foundation in Universities of Jiangsu Province of China under Grant 2016SJD870005.

References

Ahmed, S., & Broek, N. (2017). Blockchain could boost food security. *Nature, 550*(7674), 43. https://doi.org/10.1038/550043e.

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. In *, 395*. *The Lancet* (pp. 931–934). Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(20)30567-5.

Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In *Proceedings - 2016 2nd International Conference on Open and Big Data, OBD 2016* (pp. 25–30). https://doi.org/10.1109/OBD.2016.11.

Baldominos, A., Ogul, H., Colomo-Palacios, R., Sanz-Moreno, J., & Gómez-Pulido, J. M. (2020). Infection prediction using physiological and social data in social environments. *Information Processing and Management, 57*(3). https://doi.org/10.1016/j.ipm.2020.102213.

Baniata, H., Anaqreh, A., & Kertesz, A. (2021). PF-BTS: A Privacy-Aware Fog-enhanced Blockchain-assisted task scheduling. *Information Processing and Management, 58*(1). https://doi.org/10.1016/j.ipm.2020.102393.

Berdik, D., Otoum, S., Schmidt, N., Porter, D., & Jararweh, Y. (2021). A Survey on Blockchain for Information Systems Management and Security. *Information Processing and Management, 58*(1). https://doi.org/10.1016/j.ipm.2020.102397.

Business Processes Secured by Immutable Audit Trails on the Blockchain » Brave New Coin. (n.d.). Retrieved February 23, 2021, from https://bravenewcoin.com/insights/business-processes-secured-by-immutable-audit-trails-on-the-blockchain.

Campanile, L., Iacono, M., Marulli, F., & Mastroianni, M. (2021). Designing a GDPR compliant blockchain-based IoV distributed information tracking system. *Information Processing and Management, 58*(3). https://doi.org/10.1016/j.ipm.2021.102511.

Chamola, V., Hassija, V., Gupta, V., & Guizani, M. (2020). A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact. *IEEE Access, 8*, 90225–90265. https://doi.org/10.1109/ACCESS.2020.2992341.

Esposito, C., Ficco, M., & Gupta, B. B. (2021). Blockchain-based authentication and authorization for smart city applications. *Information Processing and Management, 58*(2). https://doi.org/10.1016/j.ipm.2020.102468.

Foley, D. H., Wilkerson, R. C., Birney, I., Harrison, S., Christensen, J., & Rueda, L. M. (2010). MosquitoMap and the Mal-area calculator: New web tools to relate mosquito species distribution with vector borne disease. *International Journal of Health Geographics, 9*. https://doi.org/10.1186/1476-072X-9-11.

França, A. S. L., Amato Neto, J., Gonçalves, R. F., & Almeida, C. M. V. B (2020). Proposing the use of blockchain to improve the solid waste management in small municipalities. *Journal of Cleaner Production, 244*, Article 118529. https://doi.org/10.1016/j.jclepro.2019.118529.

George, R. V., Harsh, H. O., Ray, P., & Babu, A. K. (2019). Food quality traceability prototype for restaurants using blockchain and food quality data index. *Journal of Cleaner Production, 240*, Article 118021. https://doi.org/10.1016/j.jclepro.2019.118021.

Ghafoorian, M., Abbasinezhad-Mood, D., & Shakeri, H. (2019). A Thorough Trust and Reputation Based RBAC Model for Secure Data Storage in the Cloud. *IEEE Transactions on Parallel and Distributed Systems, 30*(4), 778–788. https://doi.org/10.1109/TPDS.2018.2870652.

Hardin, T., & Kotz, D. (2021). Amanuensis: Information provenance for health-data systems. *Information Processing and Management, 58*(2). https://doi.org/10.1016/j.ipm.2020.102460.

Hu, P. J. H., Zeng, D., Chen, H., Larson, C., Chang, W., Tseng, C., & Ma, J (2007). System for infectious disease information sharing and analysis: Design and evaluation. *IEEE Transactions on Information Technology in Biomedicine, 11*(4), 483–492. https://doi.org/10.1109/TITB.2007.893286.

Hu, T., Liu, X., Chen, T., Zhang, X., Huang, X., Niu, W., Lu, J., Zhou, K., & Liu, Y. (2021). Transaction-based classification and detection approach for Ethereum smart contract. *Information Processing and Management, 58*(2). https://doi.org/10.1016/j.ipm.2020.102462.

Inayatulloh, & Theresia, S (2016). Early Warning System for infectious diseases. In *Proceeding of the 2015 9th International Conference on Telecommunication Systems Services and Applications, TSSA 2015*. https://doi.org/10.1109/TSSA.2015.7440435.

Jing, N., Liu, Q., & Sugumaran, V. (2021). A blockchain-based code copyright management system. *Information Processing and Management, 58*(3). https://doi.org/10.1016/j.ipm.2021.102518.

Khalid, A., Iftikhar, M. S., Almogren, A., Khalid, R., Afzal, M. K., & Javaid, N. (2021). A blockchain based incentive provisioning scheme for traffic event validation and information storage in VANETs. *Information Processing and Management, 58*(2). https://doi.org/10.1016/j.ipm.2020.102464.

Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association, 24*(6), 1211–1220. https://doi.org/10.1093/jamia/ocx068.

Kuo, T. T., & Ohno-Machado, L. (2018). Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks.. In *arXiv*. https://xueshu.baidu.com/usercenter/paper/show?paperid=5605ba36eff3d7bff22a863653cebd59&site=xueshu_se&hitarticle=1.

Lee, D. (2019). Big Data Quality Assurance Through Data Traceability: A Case Study of the National Standard Reference Data Program of Korea. *IEEE Access, 7*, 36294–36299. https://doi.org/10.1109/ACCESS.2019.2904286.

Leeming, G., Ainsworth, J., & Clifton, D. A. (2019). Blockchain in health care: hype, trust, and digital health. In *The Lancet, 393* pp. 2476–2477). Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(19)30948-1.

Leng, J., Jiang, P., Xu, K., Liu, Q., Zhao, J. L., Bian, Y., & Shi, R. (2019). Makerchain: A blockchain with chemical signature for self-organizing process in social manufacturing. *Journal of Cleaner Production, 234*, 767–778. https://doi.org/10.1016/j.jclepro.2019.06.265.

Li, J., Wu, J., Jiang, G., & Srikanthan, T. (2020). Blockchain-based public auditing for big data in cloud storage. *Information Processing and Management*, (6), 57. https://doi.org/10.1016/j.ipm.2020.102382.

Lin, Y., & Heffernan, C. (2011). Accessible and inexpensive tools for global HPAI surveillance: A mobile-phone based system. *Preventive Veterinary Medicine, 98*(2–3), 209–214. https://doi.org/10.1016/j.prevetmed.2010.10.003.

Liu, X., zhou, yanju, & Zongrun, W. (2020). Can the development of a patient's condition be predicted through intelligent inquiry under the e-health business mode? Sequential feature map-based disease risk prediction upon features selected from cognitive diagnosis big data. *International Journal of Information Management, 50*, 463–486. https://doi.org/10.1016/j.ijinfomgt.2019.05.006.

Liu, Z., & Li, Z. (2020). A blockchain-based framework of cross-border e-commerce supply chain. *International Journal of Information Management, 52*, Article 102059. https://doi.org/10.1016/j.ijinfomgt.2019.102059.

Lu, Q., & Xu, X. (2017). Adaptable Blockchain-Based Systems: A Case Study for Product Traceability. *IEEE Software, 34*(6), 21–27. https://doi.org/10.1109/MS.2017.4121227.

Mashamba-Thompson, T. P., & Crayton, E. D. (2020). Blockchain and Artificial Intelligence Technology for Novel Coronavirus Disease-19 Self-Testing. *Diagnostics, 10*(4), 198. https://doi.org/10.3390/diagnostics10040198.

Maxmen, A. (2018). AI researchers embrace Bitcoin technology to share medical data. *Nature, 555*(7696), 293–294. https://doi.org/10.1038/d41586-018-02641-7.

Nakamoto, S. (n.d.). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Retrieved January 30, 2021, from www.bitcoin.org.

Oham, C., Michelin, R. A., Jurdak, R., Kanhere, S. S., & Jha, S. (2021). B-FERL: Blockchain based framework for securing smart vehicles. *Information Processing and Management, 58*(1). https://doi.org/10.1016/j.ipm.2020.102426.

Olsen, P., & Borit, M. (2013). How to define traceability. In *Trends in Food Science and Technology, 29* pp. 142–150). Elsevier. https://doi.org/10.1016/j.tifs.2012.10.003.

Putz, B., Dietz, M., Empl, P., & Pernul, G. (2021). EtherTwin: Blockchain-based Secure Digital Twin Information Management. *Information Processing and Management, 58*(1). https://doi.org/10.1016/j.ipm.2020.102425.

Queiroz, M. M., & Fosso Wamba, S. (2019). Blockchain adoption challenges in supply chain: An empirical investigation of the main drivers in India and the USA. *International Journal of Information Management, 46*, 70–82. https://doi.org/10.1016/j.ijinfomgt.2018.11.021.

Smith, R. D. (2006). Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Social Science & Medicine, 63*(12), 3113–3123. https://doi.org/10.1016/j.socscimed.2006.08.004.

Sultana, T., Almogren, A., Akbar, M., Zuair, M., Ullah, I., & Javaid, N. (2020). Data Sharing System Integrating Access Control Mechanism using Blockchain-Based Smart Contracts for IoT Devices. *Applied Sciences, 10*(2), 488. https://doi.org/10.3390/app10020488.

Swan, M. (2015). Blockchain Thinking : the Brain as a Decentralized Autonomous Corporation [Commentary]. In *IEEE Technology and Society Magazine, 34* pp. 41–52). Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/MTS.2015.2494358.

Tanwar, S., Parekh, K., & Evans, R. (2020). Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications, 50*, Article 102407. https://doi.org/10.1016/j.jisa.2019.102407.

Tsui, K. L., Wong, Z. S. Y., Goldsman, D., & Edesess, M (2013). Tracking infectious disease spread for global pandemic containment. *IEEE Intelligent Systems, 28*(6), 60–64. https://doi.org/10.1109/MIS.2013.149.

Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B. A., Cucunubá, Z., Olivera Mesa, D., Green, W., Thompson, H., Nayagam, S., Ainslie, K. E. C., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N. F., Cattarino, L., …, & Ghani, A. C. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science, 0035*(June), eabc0035. https://doi.org/10.1126/science.abc0035.

Wang, H., & Song, Y. (2018). Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain. *Journal of Medical Systems, 42*(8), 1–9. https://doi.org/10.1007/s10916-018-0994-6.

Wang, Y., Cai, S., Lin, C., Chen, Z., Wang, T., Gao, Z., & Zhou, C. (2019). Study of Blockchains's Consensus Mechanism Based on Credit. *IEEE Access, 7*, 10224–10231. https://doi.org/10.1109/ACCESS.2019.2891065.

Xu, X., Sun, G., Luo, L., Cao, H., Yu, H., & Vasilakos, A. V. (2021). Latency performance modeling and analysis for hyperledger fabric blockchain network. *Information Processing and Management, 58*(1). https://doi.org/10.1016/j.ipm.2020.102436.

Yang, Q., Lu, R., Rong, C., Challal, Y., Laurent, M., & Wang, S. (2019). Guest editorial the convergence of blockchain and IoT: Opportunities, challenges and solutions. In *, 6. IEEE Internet of Things Journal* (pp. 4556–4560). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JIOT.2019.2921235.

Ye, C., Li, G., Cai, H., Gu, Y., & Fukuda, A. (2018). Analysis of security in blockchain: Case study in 51%-attack detecting. In *Proceedings - 2018 5th International Conference on Dependable Systems and Their Applications, DSA 2018* (pp. 15–24). https://doi.org/10.1109/DSA.2018.00015.

Yu, G., Zhang, L., Wang, X., Yu, K., Ni, W., Zhang, J. A., & Liu, R. P. (2021). A novel Dual-Blockchained structure for contract-theoretic LoRa-based information systems. *Information Processing and Management, 58*(3). https://doi.org/10.1016/j.ipm.2021.102492.

Yuan, R., Xia, Y. Bin, Chen, H. B., Zang, B. Y., & Xie, J (2018). ShadowEth: Private Smart Contract on Public Blockchain. *Journal of Computer Science and Technology, 33*(3), 542–556. https://doi.org/10.1007/s11390-018-1839-y.

Zhang, G., Li, T., Li, Y., Hui, P., & Jin, D. (2018). Blockchain-Based Data Sharing System for AI-Powered Network Operations. *Journal of Communications and Information Networks, 3*(3), 1–8. https://doi.org/10.1007/s41650-018-0024-3.

Zhao, Q., Chen, S., Liu, Z., Baker, T., & Zhang, Y. (2020). Blockchain-based privacy-preserving remote data integrity checking scheme for IoT information systems. *Information Processing and Management*, (6), 57. https://doi.org/10.1016/j.ipm.2020.102355.

Zhu, P., Hu, J., Zhang, Y., & Li, X. (2020). A Blockchain Based Solution for Medication Anti-Counterfeiting and Traceability. *IEEE Access, 8*, 184256–184272. https://doi.org/10.1109/access.2020.3029196.