Scientific Journal of Samarkand University

Volume 2020 Article 5

5-22-2020

Algorithms for data reliability based on a cryptographic distributed database (blockchain)

Fayzullo M. Nazarov
Samarkand state university, fayzullo-samsu@mail.ru

Akmal R. Akhatov
Jizzakh branch of National University of Uzbekistan, akmalar@rambler.ru

Farkhod F. Meliyev
Samarkand state university, a-istam@mail.ru

Algorithms for data reliability based on a cryptographic distributed database (blockchain)

Cover Page Footnote

UDC: 681.306

ALGORITHMS FOR DATA RELIABILITY BASED ON A CRYPTOGRAPHIC DISTRIBUTED DATABASE (BLOCKCHAIN)

A.R.Akhatov[1], F.M.Nazarov[2], F.F.Meliyev[2]
[1]Jizzakh branch of National University of Uzbekistan,
[2]Samarkand State University
Email: akmalar@rambler.ru, fayzullo-samsu@mail.ru

Abstract. This research paper presents algorithms for solving the problem of increasing data flow by forming a distributed database. Based on this, methods for optimizing the storage and processing of data in computer networks are presented. A distributed registry and cryptographically sensitive data processed by blockchain technology are considered. In the general case, blockchain is a multifunctional and multi-level information technology designed to reliably record various procedures and operations carried out on the network, a technology for reliable distributed storage of records of all ever completed transactions. The blockchain is a chain of data blocks, the volume of which is constantly growing as new blocks are added with records of the most recent transactions.

Keywords. Distributed Database, Cryptographic Encryption, Network Technologies, Blockchain Technology, Data Security, Algorithm RSA.

Алгоритмы для обеспечения достоверности данных на основе криптографической распределенной базы данных (блокчейн)

Аннотация. В данной исследовательской работе представлены алгоритмы решения проблемы увеличения потока данных путем формирования распределенной базы данных. На основании этого представлены методы оптимизации хранения и обработки данных в компьютерных сетях. Рассматриваются распределенный реестр и криптографически важные данные, обработанные по технологии блокчейн. В общем случае блокчейн - это многофункциональная и многоуровневая информационная технология, предназначенная для надежной записи различных процедур и операций, выполняемых в сети, технология надежного распределенного хранения записей всех когда-либо выполненных транзакций. Блокчейн представляет собой цепочку блоков данных, объем которых постоянно растет по мере добавления новых блоков с записями самых последних транзакций.

Ключевые слова. Распределенная база данных, криптографическое шифрование, сетевые технологии, блокчейн технология, защита информации, алгоритм RSA.

Kriptografik taqsimlangan ma’lumotlar bazasi (blokcheyn) asosida ma’lumotlar ishonchliligini oshirish algoritmlari

Annotatsiya. Ushbu ilmiy ishda taqsimlangan ma'lumotlar bazasini shakllantirish orqali ma'lumotlar oqimini oshish muammosini hal qilish algoritmlari keltirilgan. Shundan kelib chiqqan holda, kompyuter tarmoqlarida ma'lumotlarni saqlash va qayta ishlashni optimallashtirish usullari keltirilgan. Taqsimlangan reestr va blockchain texnologiyasi tomonidan ishlov berilgan kriptografik ma'lumotlar ko'rib chiqiladi. Umumiy holda, blockchain - bu tarmoqda amalga oshiriladigan turli xil proseduralar va operatsiyalarni ishonchli qayd etish uchun ishlab chiqilgan ko’p funksiyali va ko'p darajali axborot texnologiyasi, hamda tarmoq ichida amalga oshiriladigan barcha bitimlarning yozuvlarini ishonchli taqsimlangan saqlash texnologiyasi.. Blokchain - bu ma'lumotlar bloklari zanjiri bo'lib, ularning hajmi doimiy ravishda o'sib boradi, chunki yangi bloklar eng so'nggi tranzaksiyalar yozuvlari bilan tarmoqga qo’shiladi.

Kalit so'zlar. Tarqatilgan ma'lumotlar bazasi, kriptografik shifrlash, tarmoq texnologiyalari, Blockchain texnologiyasi, ma'lumotlar xavfsizligi, RSA algoritmi.

1.Introduction

Based on the above considerations, addressing security issues in network systems through blockchain technology is a priority. The task of algorithmic protection of distributed databases by dynamic creation of cryptographic algorithms was defined. Analysis of the blockchain technology (blockchain or a chain of blocks) shows that important advantages of the potential use of transaction blocks built in accordance with certain rules in electronic document management systems are security by encrypting transactions for subsequent confirmation, the inability to create unauthorized changes due to the

dependence of the current blockchain state from previous transactions, transparency and reliability of procedures through public and distributed looking, as well as the interaction of many users among themselves without the use of "trusted intermediaries"[1,2].

On the other hand, in the conducted studies, it is noted that in blockchain technologies it is rather difficult to achieve system performance. For example, in a Bitcoin system, the time required to add one block is about 10 minutes, this is due to the decentralization of the system: it is necessary that information about a new transaction be distributed to about 80% of the network [2,3]. Along with this, there is also an acute question of the amount of memory for storing information. Using the same bitcoin as an example, it was found that each user stores 80 GB, and about 16.1 million users (according to 2016), i.e. according to these data, the amount of required memory is 1.3 exabytes in total.

Important advantages of the potential use of transaction blocks built according to certain rules in systems by limiting and delaying electronic documents are ensuring security by encrypting transactions for subsequent confirmation, the inability to make unauthorized changes due to the dependence of the current blockchain state on previous transactions, transparency and reliability of procedures due to public and distributed storage, as well as the interaction of a large number of users between without the use of "trusted intermediaries"[4].

Researches show that when using existing algorithms for adding blocks in any system, it is possible to achieve the requirements of decentralization, openness of the entered data, the inability to change the data once entered into the system. However, mathematical-cryptographic information protection must be developed for each designed system separately.

2. Encryption schema

This is the process of encoding a piece of information that only authorized parties can access. This can be used to ensure the confidentiality of blockchain data by encrypting it. There are many encryption schemes that can be used on the blockchain. Symmetric key encryption is used in the Hyperledger matrix for the confidentiality of smart contract and Blockchain for the Smart Home [4, 7]. Although finding and calculating encrypted data is a big problem, there are many existing methods that can be utilized for this purpose. Some of these methods, such as searchable encryption for searching the encrypted data in the cloud, are already used in the allowed blockchain [4], and fully homomorphic encryption and functional encryption can also be used to calculate over encrypted data in the blockchain. Cryptocurrency Monero [5, 8] uses (half) additive homomorphic encryption along with range validation methods, but only supports value transactions.

To ensure the confidentiality and authenticity of data at the same time, authenticated encryption can be used on the blockchain. In authenticated encryption, two peers establish a connection, they both share their public keys and calculate the shared secret, which is used as a symmetric key for the authenticated encryption algorithm. The recently completed CAESAR [1, 4] cryptographic contest has identified a portfolio of six ciphers for authenticated encryption.

When this article was written in June 2019, none of these ciphers has been deployed in any blockchain system. Broadcast encryption can be used on the blockchain to ensure the anonymity of the recipient nodes of the blockchain gives an offer to use blockchain for availability and accountability for IoT. This happens because each user in the group receives an encrypted message, although only users with the correct authority or key can decrypt it.

3.Cryptographic model for data reliability

Studies have shown that using the existing algorithms for adding blocks in any system, it is possible to fulfill the requirements of decentralization, openness of the entered data, and the inability to change the data entered into the system once. However, mathematical and cryptographic information protection should be developed for each designed system separately.

The idea of a peer-to-peer system provides a centralized solution using cryptography, mathematical rules, and general rules for conducting transactions between in the system. According to some experts, this problem can be partially solved by using a digital signature, but this is possible only if there is a trustee who controls double spending, which deprives the advantages of this approach [5]. In the architecture of the system proposed in this study, the electronic document data is represented by a sequence of records that can be supplemented. Records along with supporting information are stored in blocks. The blocks are stored as linked chains. Each user is represented by a node that stores all available data streams and communicates with other nodes.

One of the main problems in such an architecture will obviously be ensuring the reliability of electronic documents, which determines the need for effective encryption algorithms [5,6]. They must guarantee sufficient cryptographic strength for information on the network, as well as enable the implementation of a digital signature.

Encryption is a reversible data transformation that forms ciphertext from plain text. Decryption is the opposite of encryption. And together it is a cipher a cryptographic method used to ensure the confidentiality of data, including an encryption algorithm and a decryption algorithm.

A cipher is a set of reversible mapping functions E_{K_1} of a set of plaintexts M onto a set of ciphertexts C depending on the selected encryption key K_1 from the set K_ε, as well as the corresponding inverse decryption functions D_{K_2}, K_D that map the set of ciphertexts to the set of plaintexts:

$$E_{K_1},\ k_1 \in K_\varepsilon : M \to C,\ D_{K_2}, k_2 \in K : C \to M,$$
$$\forall k_1 \in K_\varepsilon \exists k_2 \in K : \forall m \in M : E_{k_1}(m) = c, c \in C, \quad (1)$$
$$D_{K_2}(c) = m.$$

We can say that encryption is a reversible function of two arguments: message and key. Reversibility is the main condition for the correctness of encryption, according to which each encrypted message Y and key K corresponds to one original message X. The legal user B (on the receiving side of the communication system) receives the message Y and performs the decryption procedure.

The controller is a one-way function that confirms membership without revealing an individual identity in the base set. It can be used on the blockchain to create other cryptographic primitives, such as commitment, ring signatures, and zero-knowledge evidence. The Merkle tree, used in many cryptocurrencies, is suitable for a more complete class of cryptographic provision, which is a space-time efficient data structure for checking membership in a set. Figure 2 shows how blockchain transactions are represented in the Merkle tree, and the Merkle root is stored in the blockchain block structure. Non-Merkle provisions are classified as RSA provisions and elliptical curve controllers [7, 3].

In system, a controller A is computed by the network overall information commitments $(c_1, c_2, \ldots c_n)$ along with the appropriate membership witnesses for each item in the set. The witness w is computed by the accumulation information with the exception of one. In this way, during Zeroinformation spend transaction, a user proves the knowledge of one information by using that witness. This witness w and controller A are publicly verifiable without any trusted third party. Controller A in Zeroinformation is defined as:

$$A = u^{c_1 c_2 c_3 \ldots c \ldots c_n} \mod N \quad (2)$$

where the integers A, u and N are known to everyone. The information c is a Pedersen commitment of an information serial number s and the random number z. Zeroinformation uses Random Number Generator

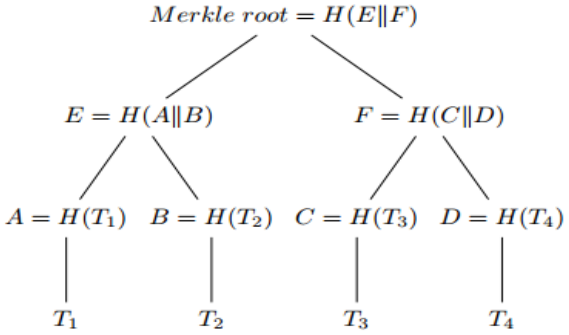

Figure 1: Merkle tree of blockchain transactions.

(RNG) to generate different s and z to find c using Pedersen commitment until c is prime. The witness w of an information c is defined as the accumulation of all information with the exception of c:

$$A = u^{c_1 c_2 c_3 \ldots c \ldots c_n} \mod N$$

Controllers can also be employed for range proofs in blockchain. Controllers are used in [8] to design a stateless blockchain where to participate in consensus, the node only needs a constant amount of storage.

Oblivious Transfer is a two-party protocol between a sender S and a receiver R. The general type of oblivious transfer is k-out-of-n oblivious transfer $\binom{n}{k} - OT$, , where $k < n$, in which S holds n messages and R retrieves simultaneously k of them without letting S know about which k out of n messages R

received. Oblivious transfer is introduced by Rabin [7] in which a sender sends a message to a receiver with probability $\frac{1}{2}$. The protocol is called $\frac{1}{2}-OT$, and it is as follows:

- Sender S chooses two large primes *p, q* and computes $N = pq$ and then the sender generates RSA public *key(e,N)* such that e is relatively prime to $(p-1)(q-1)$;
- S computes cipher text *c* over message *M* as $c = E_{(e,N)}(M) = M^e \bmod N$ and sends *e,N,c* to receiver *R*;
- R chooses a random $x \in Z_N$ and sends $a = x^2 \bmod N$ to *S*;
- S computes four square roots of *a* mod *N* and chooses one of the roots *y* at random and sends it to R;
- R checks whether $y^2 \equiv a \bmod N$ and if $y \equiv x \bmod N$, then *R* will be able to factor *N* and, hence, be able to decrypt *c* to recover *M*.

$\frac{1}{2}-OT$ is complete for secure multi-party computation. Oblivious transfer has been realized in secure multiparty computation to create private and verifiable smart contracts on blockchain [8]. Oblivious transfer can also be utilized for the exchange of secrets, private information retrieval, and building protocols for signing contracts. There has been loads of work done in oblivious transfer, and some of these works have been applied in blockchains such as Searchain and APDB (for automated penalization of data breaches using cryptoaugmented smart contracts). We will develop distributed database modeling using this method.

4. Algorithms creating a database architecture based on blockchain technology

We offer a database architecture where the data of electronic documents is represented by a sequence of records that can be supplemented. Records along with supporting information are stored in blocks. Blocks are stored as a singly linked list. Each participant is represented by a node, which stores the entire current data array and contacts other nodes. Nodes can add new entries to the end of the list, and also inform each other about list changes [9,6].

When storing data, especially when solving problems of recording and displaying information, it is advisable to use distributed databases. The distributed database algorithm is as follows:

- Each user's computer performs the function of a server;
- Data on user servers are tied to the main database;
- Data storage is distributed.

The introduction of a system for assessing the rating and creative activity of students in higher education through distributed registry technologies contributes to solving the problems of a large data stream and ensuring information security.

The platform of the system of ratings and assessments of students' creative activity will be posted on the servers of higher educational institutions. Copies of data on all university students will be generated in the main database [10].

The logical scheme of the software package will be executed according to the following schedule.

The advantage of a distributed database is that data is stored in a blockchain, and each copy of the data is stored in the main database. If a certain part of the data chain is damaged, this part of the chain restores its activity using the main database. If the main database is damaged, the main database restores its activity using the data contained in the data chain.

Thanks to this process, it is possible to ensure the security of the database. Secreting data in a database with blockchain technology using a cryptographic method using a dynamic key ensures complete data reliability.

Using cryptographic method with a dynamic key, based on blockchain technology, secretfying data and placing them in a database is implemented by the following algorithm:

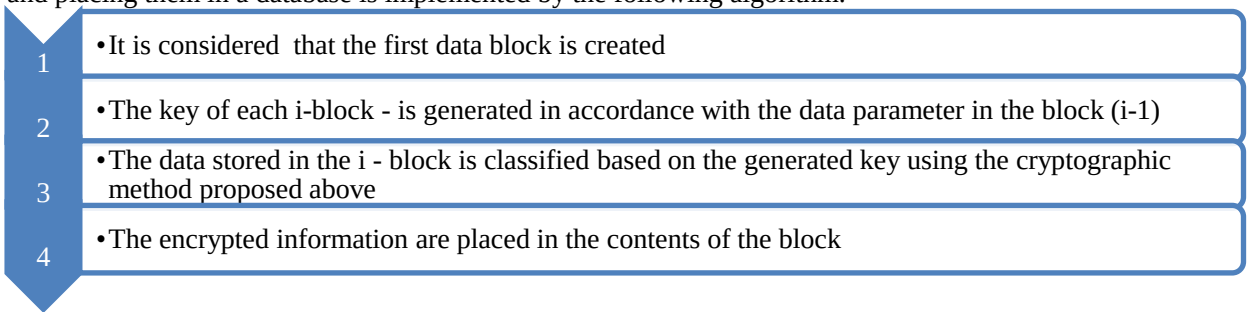

Figure 2: Blockchain-based dynamic key encryption algorithm.

Data security in a database created by the developed algorithm is considered high. In the process of organizing the database, you can choose the cryptographic method, based on the appeal to the knowledge base. When accessing the database from an unauthorized user or system, access to this information will be impossible, since in order to determine the data for the i-block, it will be necessary to open the encrypted text of the i-1 block. The data of each block will depend on the data of the previous block.

Search information becomes unnecessary at the time, for the complete opening of a data block an average of 1000 MB will have to spend several years.Ensuring the reliability of the data of information systems operating on the basis of network technologies is an urgent issue. To solve this problem, it is necessary to introduce blockchain technology in the form of the scheme described above. We use a stochastic model to create blocks.

5.Experimental calculation

The experience of creating database blocks, based on the above models, is obtained. It is clear that, for our cryptographic block model the strong stability (i.e.the fork possibility equals to 0) is guaranteed only if the meannetwork transit time t_n approaches 0. On the other hand, the perpetual eventual growth of the greatest common prefix is still possible while $0 \le t_n \le t_b$, although more or less frequent depending on dimensions. Clearly, the lower the number n of processes and the ratio $r = \frac{t_n}{t_b}$ are, the better the consistency is. To validate this intuition while refining the characterization of stability, we introduce three indicators, which can be seen as three complementary metrics of stability (next we call the absolute blockchain the most advanced among all locally viewed blockchains according to the order:

- consensus possibility (Security possibility): the possibility that all processes agreed on the absolute blockchain (higher is better);
- stability rate: the mean proportion of processes agreed on the absolute blockchain (higher is better);
- worst process delay: the mean length difference between absolute blockchain and the greatest common prefix (lower is better).

(In each case, upper, middle and bottom values are accordance the consensus possibility (security possibility), the consistency rate and the worst process delay.)

Table 1.

Values of stability indicators for several dimensions assignments.

r/n	2	3	4	6	10	20	40	60	100
0.1	0.913 0.955 0.094	0.868 0.938 0.143	0.839 0.930 0.180	0.803 0.922 0.220	0.761 0.914 0.271	0.702 0.909 0.347	0.657 0.908 0.415	0.635 0.908 0.442	0.598 0.907 0.505
0.2	0.837 0.912 0.189	0.762 0.884 0.279	0.718 0.870 0.338	0.660 0.858 0.418	0.587 0.844 0.533	0.505 0.832 0.671	0.455 0.831 0.771	0.412 0.830 0.853	0.368 0.832 0.945
0.5	0.686 0.823 0.424	0.559 0.766 0.614	0.479 0.735 0.754	0.391 0.705 0.918	0.304 0.683 1.080	0.231 0.663 1.264	0.180 0.656 1.373	0.168 0.655 1.406	0.088 0.646 2.111
0.7	0.602 0.769 0.607	0.453 0.698 0.895	0.369 0.665 1.070	0.280 0.627 1.311	0.192 0.598 1.617	0.118 0.579 1.973	0.073 0.572 2.316	0.054 0.568 2.500	0.037 0.559 2.761
0.99	0.515 0.715 0.844	0.347 0.625 1.238	0.264 0.586 1.476	0.173 0.538 1.810	0.109 0.513 2.169	0.054 0.484 2.615	0.026 0.475 3.021	0.017 0.467 3.279	0.009 0.463 3.554

To ease the results understanding, we assume a perfect symmetry of the network and a perfect fairness between processes, formally:

$$\forall (\mathrm{i},\ \mathrm{j}) \in \rho^2, \begin{cases} t_{n,i} = t_{n,j} \\ m_i = m_j \end{cases}$$

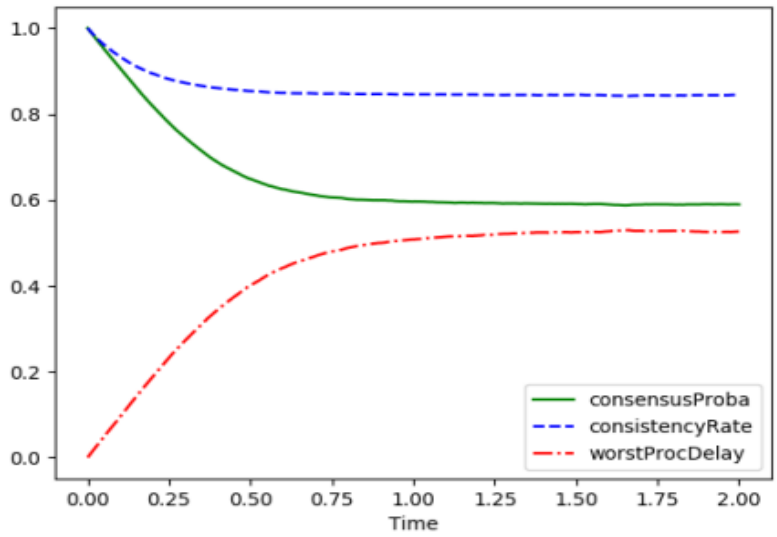

Figure 3. Time develop of the consistency indicators for *n = 10* and *r = 0.2*

Figure 4 shows the time development of the three indicators for *n = 10* and *r= 0,2* estimated running a cryptographic block model (100000 histories performed in 9.5 minutes on a single core of an i7-6700HQ CPU).

6.Conclusion

In conclusion, we note that ensuring the reliability of information in turn leads to the solution of other related problems, namely, optimization of process management by reducing costs, achieving data exchange efficiency and minimizing errors, which in many ways, allows blockchain technology to be made. Algorithms were developed to generate a database in a distributed database by dynamically creating cryptographic methods. The formation of the database information system in a distributed database and the possibility of transferring this data are checked. In practice, blockchain technology requires high-performance servers and supercomputers for cryptographic encryption and storage of this data.

References

1. A. Dorri, M. Steger, S. S. Kanhere and R. Jurdak, Blockchain: A distributed solution to automotive security and privacy, IEEE Communications Magazine , 55 (2017), 119–125.

2. J. A. Garay, A. Kiayias and N. Leonardos, The bitcoin backbone protocol: Analysis and applications, in EUROCRYPT (2), 9057 (2015), 281–310.

3. Mayank R., Danilo G., Katina K. SoK of Used Cryptography in Blockchain. Department of Information Security and Communication Technologies, Norwegian University of Science and Technology. P 54. 2019.

4. Wang L., Shen X., Li J., Shao J., Yang Y. Cryptographic primitives in blokcheyns. Journal of Network and Computer Applications, vol. 127, P. 43 – 58, 2019.

5. A. S. Elmaghraby and M. M. Losavio, Cyber security challenges in smart cities: Safety, security and privacy, Journal of Advanced Research , 5 (2014), 491–497.

6. Akhatov A.R., Nazarov F. M. Methods of Implementation of Blockchain Technologies on the basis of Cryptographic protection for the Data processing System with Constraint and Lagging into Electronic Document Management. Herald of Computer and Information technologies. vol.10. P. 3–13. Moscow.2019.

7. Pierre-Yves Piriou., Jean-Francois Dumas. Simulation of stochastic blockchain models. Chatou, France. P.[1-8]. 2018.

8. Duffield E., Schinzel H., Gutierrez F., Transaction locking and masternode consensus: A mechanism for mitigating double spending attacks. CryptoPapers.info, 2014, [Online; accessed 3-Jun-2019].

9. Pedro Franco. The Blokcheyn. Understanding Bitinformation: Cryptography, Engineering and Economics. John Wiley & Sons, 2014. 288 p.

10. Коблиц. Н. Курс теории чисел и криптографии - М., Научное издателство ТВП, 2001 г., 260 стр.