Springer manuscript No.
(will be inserted by the editor)

Review of Blockchain Technology in Cloud Computing

Vikash Tripathi
Department of Information Technology
Indian Institute of Information Technology, Lucknow.
lit2016004@iiitl.ac.in

Abstract Cloud computing gives application developers the ability to marshal virtually infinite resources with an option to pay-per-use and as needed and does not require upfront investments in resources that may never be optimally used. Thus cloud computing has been dramatically adopted in all IT environments for its efficiency and availability.While cloud computing optimises the use of resources, it does not provide an effective solution for the secure hosting of large data applications. However Blockchains public and distributed peer-to peer ledger capability can benefit cloud computing services which require functions such as, assured data provenance, auditing, management of digital assets, and distributed consensus. In this paper, we discuss the concept of blockchain technology and its hot research trends. In addition, we also briefly discuss the concept of integration of blockchain with cloud platforms in order to improve the security of data storage as well as resource, data and user management in both environments.

Keywords Blockchain · Cloud Computing · Distributed Database · Information Security

1 Introduction

The blockchain model has been gaining popularity since 2008, when the first electronic money protected through the cryptographic mechanisms (cryptocurrencies) was introduced. The first cryptocurrency to use a blockchain based approach was Bitcoin. Blockchain can be successfully utilised in diverse areas, including the financial sector and the ICT computational environment, such as computational clouds.
In this paper, we define the generic model and the main characteristics of the blockchain network. We present it as a reference infrastructure, which can be easily combined with other large-scale distributed computational environments.

Address(es) of author(s) should be given

2 Blockchain

Blockchains can be defined as distributed digital ledgers of cryptographically signed transactions that are grouped into blocks. Each block is cryptographically linked to the previous one after validation and undergoing a consensus decision. Thanks to encryption technologies, the single point of failure caused by an authorised third party has been overcome when it comes to verifying the authenticity of transactions.

The blockchain model leverages many features of the Peer2Peer (P2P) model. This broker-free approach enables users to not incur avoidable costs related to third-party centralised authorisation operations. In this model, security standards are higher and transactions are committed faster as they are automatically accepted and saved by multiple agents. It makes it harder for hackers to exploit vulnerabilities of the system, thus reducing costs of security-related tasks. Furthermore, transactions can be easily made public and open access. Figure 1 shows the basic components of the blockchain P2P architecture.

2.1 Ledgers

Ledgers in blockchain are simply a set of transactions. Each node has a local copy of this set of transactions, i.e., the ledger. By the same token, a blockchain is usually composed of a set of nodes. The exchange of goods and services has been stored historically in analogue

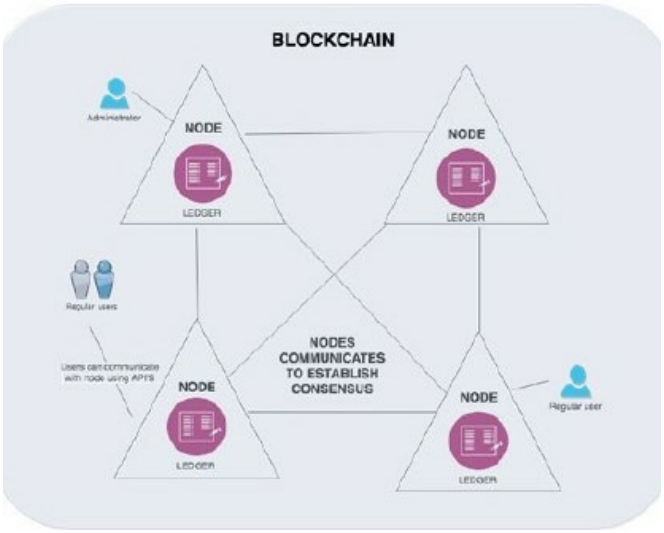

Fig. 1 Blockchain Distributed Architecture

(pen and paper) ledgers. In blockchain, distributed ledger is a type of decentralized database that stores records one after another. Each record in the ledger includes a time stamp and a unique cryptographic signature. The complete transaction history of the ledger is verifiable and auditable by any legitimate user.

There are two different types of distributed ledger in practice:
1) Permissionless Ledger
2) Permissioned Ledger
The key benefits of permissionless ledger are that it is censorship-resistant and transparent. However, the permissionless ledger also known as public ledger, has to maintain complex shared records and it consumes more time to reach the consensus compared to the permissioned ledger also known as private ledger.

2.2 Blocks

Each of the nodes in the blockchain may receive candidate transactions submitted by end-users. These transactions are then propagated to other nodes in the working group network. This operation, however, does not actually save the transaction in the blockchain. Subsequent to this process, mining nodes need to add the aforementioned transactions to the blockchain. Until then the committed transactions wait in the transaction pool (a queue).

As mentioned before, the mining nodes are responsible for keeping the blockchain up-to-date by publishing freshly committed blocks. This process performs the actual operation of adding transactions to the blockchain. Thus, a block is composed of validated transactions. To this end, the providers of transactions, who are shown in the input values of each transaction, must cryptographically sign the transaction to ensure its legitimacy, meaning that each of them had access to the appropriate private key. No blocks containing invalid transactions will be accepted in the blockchain. To this aim, the rest of the mining nodes in the network check the validity of each and every transaction in the published block. Once a block is created, it must be hashed. To this purpose, a 518 digest, which represents the block, will be created. The immutability of data is ensured by this method since even a change in a single bit of the block would drastically change the generated hash. In addition, a copy of the hash of every block is shared among all the nodes in order to improve security. This system prevents any change since every node can check if the hash matches. Each block typically consists of the following components

- The block number, also known as block height
- The current block hash value
- The previous block hash value
- The Merkle tree root hash
- A timestamp
- The size of the block
- A list of transactions within the block

The generated hash is stored in a data structure called Merkle tree instead of the header of the block. The hash values of the gathered data are combined by the Merkle tree until there is a singular root, called Merkle tree root hash.

3 Cloud Computing

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. In simple terms, cloud computing means storing and sharing of resources, software, and information over the internet. It is easier to divide it into two sections: the front end and the back end.
The front end consists of the clients computer or computer network and the application essential to access the cloud computing system. It is not necessary that all cloud computing systems have the same user interface. On the back end of a cloud system, there are various computers, servers and data storage systems that make up the cloud.

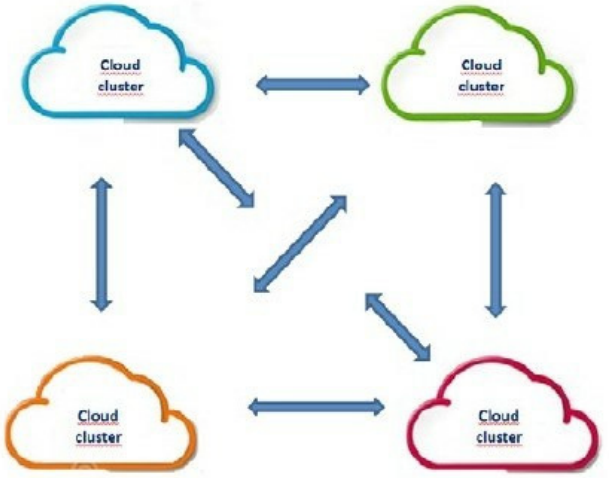

Fig. 2 P2P-based 'Many clouds' Architecture

4 Integration of blockchain with cloud environments

Cloud computing assembles large networks of virtualized services: hardware resources (CPU, storage, and network) and software resources (databases, message queuing systems, monitoring systems, load-balancers). In the industry, these services are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Cloud computing services are hosted in large data centres, often referred to as data farms.

Based on resource and data management and the related security and privacy issue, we can distinguish three main types of cloud platforms: (i) public cloud, (ii) private cloud, and (ii) hybrid cloud. Public clouds offer unlimited access to shared data and resources for a wide group of users, but there is no guarantee that users data will be protected. Access to resources and data in private clouds is restricted and each user must be validated through strong authorisation and authentication procedures. Private cloud clusters are usually owned by enterprises and work under specific cloud standards. Hybrid clouds seem to be an ideal model of integration of the many private clouds into a joint global infrastructure. Such integration is done through the upper level public layer.The main problem with that model is to reach an agreement among private cloud providers to work under a unified public cloud standard. Therefore, the many cloud model, where the distributed private cloud clusters are connected by using the standard P2P network (see Figure 2), is a much more realistic scenario.

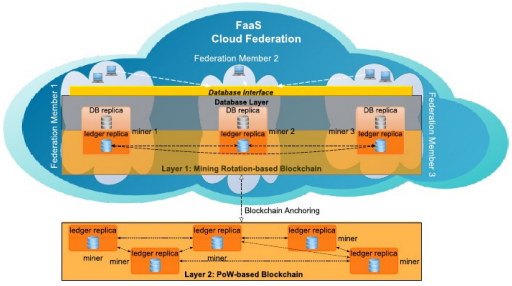

Fig. 3 A blockchain-based database proposal for a Cloud Federation

There are two main methods of integration of the cloud with blockchain platforms:

1) Using cloud for the development of blockchain applications and supporting the integration with enterprise networks (private clouds) to facilitate storage, replication and access to transactional data.

2) Using blockchain methods to improve the security of task, user and data management in the clouds.

5 Recent developments

5.1 Threats to Data Integrity

The threats to the data integrity in the context of cloud federation can be multiple and variegated. Our focus is on the database storing the governance data of a federation, hence on data whose corruption critically affects the whole federation and its security. The threats we consider span from malicious alterations of data, to data updates without all the involved members informed. More specifically, we can enumerate the following threats:

T1: An attacker violates the integrity of the data by directly altering (part of) the database.

T2: A federation member updates the database without informing the other members.

T3: Multiple federation members collude to maliciously altering (part of) the database.

Work Done -

1) Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments: In the context of cloud computing environments, the blockchain could be exploited to realise a database ensuring strong integrity

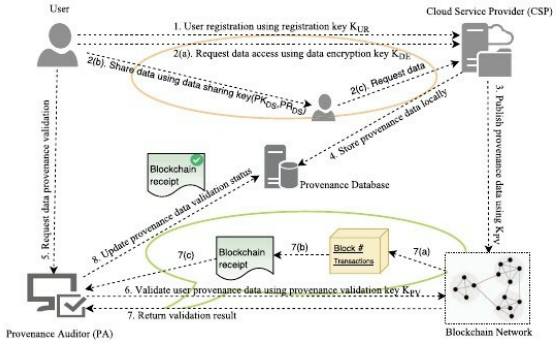

Fig. 4 ProvChainSystem Interaction

guarantees. In particular, the blockchain could be used to store the logs of database operations, thus to avoid the data threats presented. The proposed blockchain-based database distributed on three clouds member of a federation, with 2-layered structure as shown in Figure 3. The member clouds operating on the database issue operations through the Database Interface. The operations are first logged via appropriate evidences by the the first-layer blockchain, then they are executed on the distributed DB replicas. The first-layer blockchain is permissioned, the miners thus achieve consensus by means of the so-called mining rotation consensus mechanism. Once all miners have signed the operations, they can become part of the blockchain: all the miners add these operations to their local ledger, and apply them to their local replica. In particular, at certain intervals of time, a witness transaction containing the hash of the first-layer blockchain up to the current operation is sent to the second-layer blockchain, and these hashes act as forensics evidence for proving and validating the integrity of the data stored in the first-layer blockchain.

2) ProvChain: Cloud data provenance is metadata that records the history of the creation and operations performed on cloud data object. Secure data provenance is crucial for data accountability, forensics and privacy. Blockchain-based data provenance can provide tamper-proof records, enable the transparency of data accountability in the cloud, and help to enhance the privacy and availability of the provenance data.

ProvChain operates mainly in three phases:
(1) provenance data collection,
(2) provenance data storage, and
(3) provenance data validation.
The implementation of ProvChain (see Figure 4) is conducted using a three layer architecture, comprising of data storage layer, blockchain layer, and provenance database layer. The functions for each layer are described as follows.

-Data Storage Layer: ProvChain is implemented to support cloud storage applications. A cloud service provider is used, but more than one can be easily used.

-Blockchain Network Layer: Blockchain network to record each provenance data entry is used. Each block can record multiple data operations. Here we use file as a data unit, so we record each file operation with username and file name. File access operations include Create, Share, Change and Delete.

-Provenance Database Layer: We build an extended database locally for recording the file operation as well as querying. In ProvChain, the service provider can assign a provenance auditor to verify the data from the blockchain network. The response is a blockchain receipt that gets validated and appended in the database.

ProvChain provides a real-time auditing for all data access in the cloud storage application. File as a data unit and all the operations are used on the cloud data objects that are audited as well as recorded using blockchain. In this way, evidence for all cloud data access events can be collected and monitored.

5.2 Secure Blockchain Solutions in cloud computing

1) Blockchain Technology in Cloud Computing : A Systematic Review. If the user data is disclosed in the cloud computing environment, monetary and psychological damages can occur due to the leak of users sensitive information. The security of the saving and transmitting data, such as confidentiality and integrity, in the cloud computing environment is mainly studied. Note, however, that studies on privacy protection and anonymity are not sufficient. Blockchain is a representative technology for ensuring secrecy. If combined with the cloud computing environment, blockchain can be upgraded to a convenient service that provides stronger security. User anonymity can be ensured if the blockchain method is used when saving the user information in the cloud computing environment. An electronic wallet is installed when using the blockchain technology. If the electronic wallet is not properly deleted, the user information can be left behind. The remaining user information can be used to guess the user information. To solve this problem, we propose a solution that installs and deletes the electronic wallet securely.

2) Security Implications of Blockchain Cloud with Analysis of Block Withholding Attack: The blockchain tech-

nology will be beneficial to cloud services which have a strong desire for assured data provenance and support cloud auditing. To enable data integrity over the public ledger in a blockchain cloud, cryptographically enforced blocks join in the blockchain after a consensus is reached in the decentralized network, where transactions in the blocks are authenticated by peers of the network.
The combination of cryptographic mechanism and decentralized public ledger allows to build any kind of application on top of the blockchain without worrying about trust components of users and maliciousness in the blockchain enabled cloud system.

Cloud computing allows users to remotely store their data into the cloud and provides on-demand applications and services from a shared pool of configurable computing resources. The security of the outsourced data in the cloud is dependent on the security of the cloud computing system and network. The protection of data exchanged within the cloud infrastructure currently relies on PKI based signatures.

3) Data Provenance in the Cloud: Assurance of data transfer within intracloud and inter-cloud environments is very crucial. Typical assurance of data focuses on ensuring the confidentiality, integrity and availability of the data contents. However, assurance of the ancestry of the data (where the data came from) is a challenge in cloud environments. Data provenance addresses such issue based on the detailed derivation of data objects. If true data provenance existed in the cloud for all data stored on cloud storage, distributed data computations, data exchanges and transactions, detecting insider attacks, reproducing research results, and identifying the exact source of system/network intrusions would be achievable. Unfortunately, the state-of-the art in data provenance in cloud does not provide such assurances. Data provenance will be very critical for cloud computing system administrators to debug break-ins to the system or network. Cloud computing environments are typically characterized by data transfers between diverse system and network components. These data exchanges could take place within a data center or across federated data centers. The data does not usually follow the same path due to multiples copies of the data and diversity of paths taken to ensure resilience.

5.3 Blockchain based distributed control system for Edge Computing

Edge computing proposes a novel model for poviding computational resources close to end devices that are

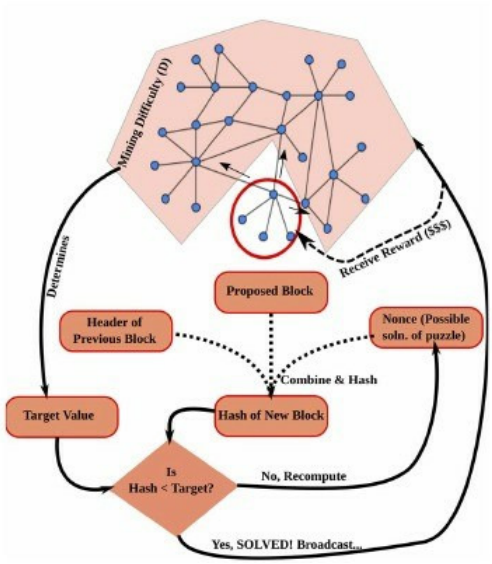

Fig. 5 Block Mining Process

connected to the network. It has numerous applications in Internet of Things, as well as smart grids, healthcare, smart home, etc. The emergence of Internet connected smart devices which can be accessed and controlled remotely via computer networks, has raised the expectations for new and enhanced intelligent computing services. However, security and privacy of users data, and specific requirements regarding personal data protection, have challenged the effectiveness of the centralized computing model available on the public cloud infrastructures. A new computing model, with the emphasis on decentralization, and the collaboration of individual work units to achieve a common goal, has been devised for Internet of Things (IoT) applications, and for other domains such as the smart grid, healthcare, connected vehicles, etc. For many IoT applications which require mobility support, location awareness and low latency, there is a need of a new platform, one which can provide computational resources to both large-scale sensor networks which monitor the environment, as well as intelligent services based on data processing and cloud resources integration.

Edge computing nodes can be seen as members of a decentralized network which provides compute, storage and networking services to end devices. Since smart devices are usually inadequate in computation power, battery, storage and bandwidth, IoT applications and services are usually backed up by strong server backends, which are mostly deployed in the cloud, since cloud computing is considered as a promising solution to deliver services to end users and provide applications

with elastic resources at low cost.

Within the blockchain context, smart contracts are scripts stored on the blockchain. (They can be thought of as roughly analogous to stored procedures in relational database management systems). Since they reside on the chain, they have a unique address. A smart contract can be triggered by addressing a transaction to it. It then executes independently and automatically in a prescribed manner on every node in the network, according to the data that was included in the triggering transaction. (This implies that every node in a smart contract enabled blockchain is running a virtual machine (VM), and that the blockchain network acts as a distributed VM).

For many IoT applications, a distributed automation system can be implemented as a hierarchical structure with two tiers, with the higher level performing supervision and strategic decisions, and the lower level having direct control of devices and processes.

5.4 Securing Smart Cities Using Blockchain Technology and Cloud Computing

A smart city uses information technology to integrate and manage physical, social, and business infrastructures in order to provide better services to its dwellers while ensuring efficient and optimal utilization of available resources. With the proliferation of technologies such as Internet of Things (IoT), cloud computing, and interconnected networks, smart cities can deliver innovative solutions and more direct interaction and collaboration between citizens and the local government. Despite a number of potential benefits, digital disruption poses many challenges related to information security and privacy. This paper proposes a security framework that integrates the blockchain technology with smart devices to provide a secure communication platform in a smart city.

Originally developed to support crypto-currency, the blockchain can be utilized for any form of transactions without an intermediary. The benefit of blockchain is that an attacker has to compromise 51 percent of the systems to surpass the hashing power of the target network. Thus, it is computationally impractical to launch an attack against the blockchain network. The following example demonstrates working procedures of the blockchain technology. Let A and B be two entities in a blockchain based parking system and A is paying parking fee to B, the parking authority. This transaction is represented online as a block including information such as block number, proof of work, previous block, and transaction records and this block is broadcast to every entity in the network. The other entities verify the block and if more than 50 percent of the entities approve the block then the transaction is confirmed and added to the chain. After that, the fee is transferred from entity A to authority Bs account.

Security Framework:

- Physical Layer: A smart city devices are equipped with sensors and actuators which collect and forward data to the upper layer protocols.
- Communication Layer: In this layer, smart city networks use different communication mechanisms such as Bluetooth, 6LoWPAN, WiFi, Ethernet, 3G, and 4G to exchange information among different systems. The blockchain protocols need to be integrated with this layer to provide security and privacy of transmitted data.
- Database Layer: Using private ledgers to ensure scalability, performance, and security for realtime applications like traffic systems in a smart city.
- Interface Layer: This layer contains numerous smart applications which collaborate with each other to make effective decisions. For example, a smart phone application can provide location information to the smart home system so that it turns on the air conditioner 5 minutes prior to reach at home.

The main advantage of using blockchain is that it is resilient against many threats. Further, it provides a number of unique features such as improved reliability, better fault tolerance capability, faster and efficient operation, and scalability. Thus, integration of blockchain technology with devices in a smart city will create a common platform where all devices would be able to communicate securely in a distributed environment.

6 Conclusion

In this paper, we have identified the requirements and research questions to be addressed to realise a blockchain-based database for cloud computing environment. We also collected the recent developments in the field, and tried to summarise the efforts so far made in this field. We also tried to identify the Integration of blockchain with cloud environment in bi-directional ways.

References

1. Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments, Edoardo Gaetani,

Leonardo Aniello, Roberto Baldoni, Federico Lombardi, Andrea Margheri, and Vladimiro Sassone, Supported by European Commissions H2020 Programme under the SUNFISH project, grant N.644666, 2017.

2. Blockchain based distributed control system for Edge Computing Alexandru Stanciu, 2017 21st International Conference on Control Systems and Computer Science.

3. ProvChain: A Blockchain-based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability Xueping Liang , Sachin Shetty, Deepak Tosh, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla, 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing.

4. Securing Smart Cities Using Blockchain Technology Kamanashis Biswas Vallipuram Muthukkumarasamy, 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems.

5. Security Implications of Blockchain Cloud with Analysis of Block Withholding Attack Deepak K. Tosh, Sachin Shetty, Xueping Liang, Charles A. Kamhoua , Kevin A. Kwiat, Laurent Njilla, 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing.

6. Blockchain and Its Coming Impact on Financial Services Kurt Fanning and David P. Centers, The Journal of Corporate Accounting Finance ,July August 2016

7. Blockchain Technology in Cloud Computing : A Systematic Review Ketki R. Ingole, Sheetal Yamde , International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 Apr-2018

8. Il-Kwon, L. Young-Hyuk, K. Jae-Gwang, L. and Jae-Pil, L.

9. The Analysis and Countermeasures on Security Breach of Bitcoin. Proceedings of the International Conference on Computational Science and Its Applications. Guimares, Portugal. Springer International Publishing: Cham, Switzerland(2014, June 30-July 3).

10. Blockchains and Smart Contracts for the Internet of Things Christidis, K. and Michael, D. (2016) IEEE Access, 4. pp. 22922303.

11. BlockChain Technology: Beyond Bitcoin Michael Crosby, Nachiappan , Pradan Pattanayak, Sanjeev Verma, Vignesh Kalyanaraman, Applied Innovation Review Issue No. 2 June 2016.

12. 13.Blockchain Technology: Transforming Libertarian Cryptocurrency Dreams to Finance and Banking Realities ,Ittay Eyal, Cornell University

13. Managing IoT Devices using Blockchain Platform Seyoung Huh, Sangrae Cho, Soohyung Kim, ICACT2017 February 19 22, 2017

14. "17 Blockchain Disruptive Use Cases." Everis NEXT. Everis NEXT, 02 June 2016. Web. 03 Jan. 2017.

15. 2012, pp. 1623.5] D. K. Tosh, S. Shetty, X. Liang, C. Kamhoua, K. Kwiat, and L. Njilla, Security implications of blockchain cloud with analysis of block withholding attack, in Intl. Symposium on Cluster, Cloud and Grid Computing. IEEE/ACM, 2017.