DistProv-Data Provenance in Distributed Cloud for Secure Transfer of Digital Assets with Ethereum Blockchain using ZKP

Navya Gouru, GITAM (Deemed to be University), Visakhapatnam, India
https://orcid.org/0000-0001-9321-1461
Nagalakshmi Vadlamani, GITAM (Deemed to be University), Visakhapatnam, India

ABSTRACT

The importance and usage of the distributed cloud is increasing rapidly over a traditionally centralized cloud for the storing and exchanging of digital assets between untrusted parties in many business sectors. Storing the digital assets in the distributed cloud is considered superior to traditional cloud computing in terms of environmentally friendly, cost, security and other technical dimensions. In this article, a contemporary architecture DistProv is proposed where an open source distributed cloud IPFS is used to store and transfer the digital assets between the consignor and consignee. These two are untrusted parties exchanging sensitive documents secured by cryptographic algorithms with permission-based access verified by ethereum smart contracts using zero-knowledge proof (ZKP) and simultaneously publishing the provenance data about the digital asset as a transaction on the blockchain. This article also discusses on verifying the integrity of the digital assets and authentication of the consignor and thus preserving a strong CIA triad.

KEYWORDS
Blockchain, CIA Triad, Cryptography, Data Provenance, Digital Assets, Distributed Cloud, Ethereum Smart Contracts

INTRODUCTION

Trust is one of the persisting issues in centralized cloud computing where the most extensively publicized security breaches are related to involvement with the third parties. Sustaining centralized data centers farms involves various challenges like maintaining lots of servers, availability and uptime round the clock, maintenance cost and performance issue and better staffing productivity. In contrast, distributed cloud storage does not rely on the server farms but every node that is the part of the network, rent out their excess hard disk space to store the information and also each node that provides storage space is incentivized based on the size they rent for storage. Distributed cloud has advantages like eco-friendly, involves high upload and download speed, rewarding the host servers and distributed geographically. Data Provenance is a major influential component in cloud security as it detects data tampering and uncovers unauthorized access. Data provenance is the mechanism that derives information about the lineage of data from its original sources. Apart from tracking, ensuring the integrity and accuracy and securely maintaining the tamper-proof data provenance is a challenge as data provenance may also contain sensitive information.

DOI: 10.4018/IJOSSP.2019070101

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Blockchain provides secured data provenance as the information stored on the blockchain is immutable, reliable and secured. The blockchain is a distributed ledger that stores digital records as transactions in a block with chronological order. The transaction is verified by the nodes across the distributed blockchain network before publishing in a block. The blockchain is structured in a single linked list in a linear way where the first block is called a genesis block. The superior features of blockchain include immutability where once the data is published it cannot be edited or deleted, it has no central mechanism that leads to no central point of failure, each block in the blockchain is identified with a cryptographic signature and holds retention property of all the transactions. Bitcoin (Nakamoto, 2008). A distributed digital currency is the world well-known implementation of Blockchain that uses a consensus mechanism to verify, confirm and record a transaction to transfer value in bitcoin.

In this paper, the authors propose a control flow diagram built on blockchain that is used for tamper-proof data provenance collected from storing and accessing the digital assets on a distributed cloud and transferring these assets between unreliable parties. These digital assets are encrypted and digitally signed by the consignor, who is the sender or owner of the document and also set permission on whom to access these assets. These permissions are set on ethereum smart contracts and the provenance data is verified and validation before publishing it to the blockchain. The authorization between the blockchain nodes accessing the verification script which resides on DistProv server is done with zero-knowledge proof to protect against unauthorized users.

The other sections of this paper are formulated as follows. The Background section provides an overview and background concepts related to DistProv that includes data provenance, distributed cloud storage, public blockchain vs private blockchain, ethereum smart contracts and zero-knowledge proof. The DistProv Control Flow Diagram section describes the DistProv control flow diagram that explains how Digital Assets stored on the Distributed cloud can be transferred securely with cryptography and publishing the provenance data on the blockchain using ZKP. The DistProv implementation with LegalProv section explains the implementation of proposed DistProv with a use case called LegalProv. The related work section compares DistProv with other approaches related to data provenance on the cloud using blockchain. Finally, the conclusion is discussed.

BACKGROUND

Data Provenance

Data provenance is very crucial as it captures the execution of workflows and helps to debug any malicious activities to a system or a network or an application level. There are various provenance methods proposed during the usage of local servers that are prior to cloud invention and also during the era of cloud computing. PASS (Reddy and Holland et al., 2006) is the earliest proposed system that automates the collection and maintenance of data provenance in a relational data storage system. PASS gathers the provenance details at the operating system and file system level and provides a detailed log on how the data workflow changes but have limitations like interoperability challenges and does not explain about data provenance security. OPM (Moreau and Freire et al., 2008) The Open Provenance Model conceptually explains the exchange of provenance information between systems based on a shared provenance model but does not deal with storing of provenance data in repositories and protocols to query the provenance repository. PASSv2 (Reddy and Braun et al., 2009) is a new version of PASS that tracks data provenance in multiple levels of abstraction within a system and results in a unified provenance infrastructure that stores the application-specific provenance. SProv Hasan and Sion et al., 2009) is system architecture that ensures data integrity with digital signature and confidentiality with encryption for the provenance data captured at the kernel level, file system layer or application layer but does not discuss on querying the provenance data. In Provenance for the cloud (Reddy and Macko et al., 2010), three different protocols have been discussed for implementing provenance for the data stored in the cloud. Here in the first protocol, both the data and its provenance is stored in
Related Content

An Empirical Comparison of Machine Learning Techniques in Predicting the Bug Severity of Open and Closed Source Projects
www.igi-global.com/article/empirical-comparison-machine-learning-techniques/78560?camid=4v1a
The Social Order of Open Source Software Production
www.igi-global.com/chapter/social-order-open-source-software/21187?camid=4v1a

Open Source Software to Enhance the STEM Learning Environment
www.igi-global.com/chapter/open-source-software-to-enhance-the-stem-learning-environment/120983?camid=4v1a

Competition-Based Learning: A Model for the Integration of Competitions with Project-Based Learning using Open Source LMS
www.igi-global.com/chapter/competition-based-learning/120952?camid=4v1a