INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Distributed Ledger Technology based data minimization of digital identities

Shashwat Dhananjay
shashwat.dhananjay14@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Vinay Kumar
gupta.vinay1601@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Harshit Agrawal
harshit.nic@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Prajakta Patankar
prajaktaa.patankar@gmail.com
MIT Academy of Engineering, Pune, Maharashtra

Dr. Shitalkumar A. Jain
dean.cr@mitaoe.ac.in
MIT Academy of Engineering,Pune, Maharashtra

ABSTRACT

The avatars can be created and managed by a prime identity and can be used by an individual in a self-sovereign way to provide access to their information through the method of minimization or to cast the usage of the data and the statistical or marketing use of the same. User has the choice of sharing minimum attributes required for KYC purpose based on different avatars of the Blockchain, also track the flow of data through various concerned agencies.

Keywords— *Security, Data breach, Personally identifiable information, Right to Privacy, Blockchain DLT, Digital KYC, Data controller, Data processor*

1. INTRODUCTION

In this time of technology, Digital identities are one of the most complex and crucial issues to solve. Arising a major question: How do we guarantee the gentility of someone over the claims as identified by him without undergoing verification? This comes in line with the multiple daily verifications customers undergo who want to set up a bank account. They have to go through an incommodious procedure of Know-Your-Customer (KYC). Hence, there is a crucial need to design a solution for the immutable storage of such data to strengthen trust and security.

Traditionally the proofs involved in authentication have often been provided by government identification certificates. This information when shared is prone to falsification or even alteration.

(a) Digital assets are represented as a token which typically does not bear inherently any value but linked to an existing asset that could be of value to the owner.
(b) Distributed Ledger Technology refers to a technology which allows transactions to be enrolled, shared, traced or co-occur across a distributed network. It employs an algorithm method to synchronize data across the network in an immutable manner, which is a fast-evolving approach to share data across multiple ledgers.
(c) Blockchain, used in some distributed ledger is a particular type of data structure which store and transmit data as blocks, which are connected to one another as a digital chain.
(d) Depending on whether nodes need permissions to modify any ledger from any entity, DLs are permission less or permission. Based on the ledger access by only participating nodes or anyone, they're partitioned as public or private Distributed Ledgers.
(e) Digital currencies distinct from e-money is a digital representation of value, denominated in their units of account. Whereas Crypto currencies which relies on cryptographic technique is a subspace of digital currencies to achieve consensus, for example, ether and Bitcoin.

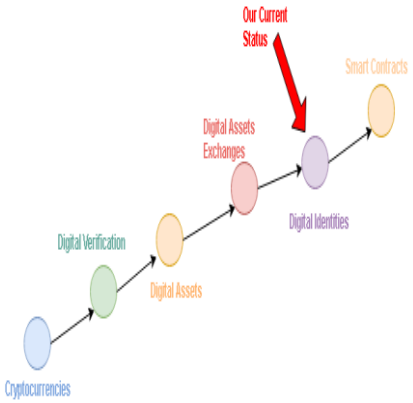

Fig. 1: Growth

Bitcoin wields in the Cryptocurrencies and Digital Verifications stages, Bitshares in the Decentralized Assets Exchanges the sector and Ethereum in the Smart Contracts stage. For our proposed system, it's the link between blockchain and real-world applications, aims to build an environment that remains around these elements.

2. RELATED WORK

Review on paper Blockchain + RFID = Total product lifecycle management, presented by Rain RFID Alliance meeting. This paper tells how blockchain can be benefited at all stages of product's lifecycle i.e. Supply chain, manufacturing, product life and end of product life. This product lifecycle looks into how and what goes into products where it comes from, how it is performing during use and how it is managed at the end of the lifecycle. The principal for which the cycle is traced is provenance track, trace/performance, recalls and recycling.

Review on paper Decentralizing Privacy: Using Blockchain to Protect Personal Data published by Guy Zyskind, Oz Nathan, Alex 'Sandy' Pent land. Personal and sensitive data is at risk in the hands of third-parties, where it is vulnerable to attacks and misuse. Users should have control of their data without compromising security or limiting companies' and authorities' ability to provide personalized services. The proposed system in this Paper enables this by combining the power of blockchain, deployed as an access-control moderator with an off-blockchain storage solution. The need for middlemen and third-party is completely eliminated and user trust and transparency are increased. The proposed system benefits Companies to take fewer efforts and save resources required for properly securing and compartmentalizing the data of the user and thus they can utilize the resources to focus on utilizing the data.

Review on paper Bitcoin-NG: A Scalable Blockchain Protocol published by Ittay Eyal, Adem Efe Gencer, Emin Gün Sirer, and Robbert van Renesse, Cornell University. This paper introduces Bitcoin-NG (Next Generation), It gives a new set of rules for blockchain to scale and enhance. Bitcoin-NG is an excessively complicated fault tolerant blockchain system that is efficient to extreme use and provides the same level of features as in Blockchain. In addition to these, this system introduces new parameters in increasing security and working of bitcoin like blockchain systems. Bitcoin-NG was deployed on a size which is 15% of the present Bitcoin without changing the clients at both ends. The results came out to be significantly good which depended on the bandwidth and performance of each independent nodes.

Review on the paper design of the Hyperledger Blockchain material conferred by Christian Cachin (IBM analysis – Zurich). The Hyperledger material may be a permission blockchain, ASCII text file, supports sturdy security and identity options. The corroborative peers run a BFT agreement protocol for corporal punishment a replicated state machine that accepts 3 forms of dealings as operations: Deploy transaction, Invoke dealings, question dealings. The blockchain's hash chain is computed over the dead transactions and also the ensuing persistent state. Validation of transactions happens with the replicated running of the chain code and provided fault assumption with underlying BFT agreement, that among the n conformational peers at the most $f < n/3$ might "lie" and behave willy-nilly, however, others can execute the chaincode properly. A standard resolution to strain non-deterministic transactions that square measure provably oblique is obtainable and has been enforced within the SIEVE protocol. Membership among the corroborative nodes running BFT agreement is presently static and also the setup needs manual intervention. Support for dynamically dynamical the set of nodes running agreement is planned for a future version

Review on paper Comparison of Ethereum, Hyperledger Composer Fabric material and Corda published by Martin Valenta, Philipp Sandner. The Paper compares and analyses there, Hyperledger Composer Fabric and Corda aspect by aspect. The examined DLTs span a time. Fabric and Ethereum square measure each extremely versatile in several aspects. Ethereum's sensible contracts engine is extremely powerful, it is a generic platform appropriate for any reasonable application. However, Ethereum's permissionless mode of operation and total transparency prices performance, quantifiability and privacy. Material solves performance quantifiability and privacy problems by permission mode of operation and specifically by employing a BFT rule and fine-grained access management. Further, the standard design permits material to be custom. Corda is intended as DLT for the money services trade. It takes the extremely regulated surroundings into consideration by augmenting sensible contracts with legal prose. Corda exclusively focuses on money services transactions and thence is best compared to Fabric. However, it'd be attainable that Fabric, thanks to its modularity, are often tailored to match Corda's feature set. Efforts square measure being created to integrate Corda into the Hyperledger project. Corda, therefore, can't be seen as a contender to Fabric however a lot of as a compliment.

Review on paper Performance defining and reducing the cost for Hyperledger Composer Fabric Blockchain Platform published by Parth Thakkar, Senthil Nathan, Balaji Viswanathan. In this paper, the parameters on which the performance of blockchain are discussed. There are a number of factors which decide how Hyperledger composer Fabric Blockchain Platform is going to perform like block size, endorsement policy, channels, resource allocation, and state info selection on the dealing outturn. The paper proposes two solutions, in the first part various metrics of Hyperledger composer Fabric Blockchain Platform are analysed and in the second part based on this analysis and evaluation optimization of Hyperledger Blockchain Platform is done. For performance benchmarking in-depth study of configuration parameters by varying latency and throughput is done. Identifying bottlenecks and optimizing those is the second task.

Review on paper Blockchain technology and Trusted Computing: Problems and Solutions for Hyperledger Composer Fabric given by Marcus Brandenburger, Christian Cachin, Rüdiger Kapitza, Alessandro Sorniotti. A transaction cannot be kept secret because its data is replicated in a network among all the nodes. To overcome this, blockchain has been combined with Trusted Execution Environments (TEEs) for running applications demanding privacy. This paper explores pitfalls arising from the combinations of TEEs and blockchains. As TEEs are stateless they are vulnerable to many attacks. However, the non-final consensus protocols in blockchains, like the proof of works in Ethereum and other, the contract running must be handled rollbacks by default. Hence, TEEs are not directly used for securing blockchain runnings. This solution works only when the consensus decisions are finalized. Our system provides a solution for problems carried by the execute order validate architecture of Hyperledger Composer Fabric and prevents rollback attacks on TEE based execution as much as possible.

Review on paper Supporting Data privacy on Hyperledger Composer Fabric with Secure Multiparty Computation published by Fabrice Benhamouda, Shai Halevi, Tzipora Halevi. Hyperledger Composer Fabric needs to be extended to support private data. This Paper support adding private-data to Hyperledger Composer Fabric using secure Multiparty Computation (MPC). In this, the peers securely store on the chain code encryption of their personal data, uses secure multiparty computation whenever personal data is needed in a transaction process. The demo of the proposed solution is deployed over

Hyperledger Composer Fabric v1.0, implementing a bidding system. The Paper finds two basic services which should be joined to Hyperledger Composer Fabric to support the given solution.

Review on Paper collaborating mobile healthcare application and integrating blockchain for sharing of data published by Xueping Liang, Juan Zhao, Sachin Shetty, Danyi Li, Jihong Liu. This Paper proposes a unique centralized sharing of health data solution by using a Permission based and decentralized blockchain to secure privacy using channel formation scheme and upgrade the identity management using the service supported by the blockchain. Deployment of a mobile application to collect data related to health from personal gadgets, manual/automatic input, and reliable medical devices, and duplicate data to the cloud for data sharing with healthcare providers and health insurance companies. Moreover, for taking into account scalability and performance, data processing based on tree and batching methods are adopted to manage a large number of data sets of health data collected and uploaded by the IoT devices.

This paper was published by Fabrice Benhamouda, Shai Halevi, Tzipora Halevi, describing the problem in the architecture of Hyperledger Composer Fabric is that every peer has the complete view of the shared ledger which makes it difficult for transactions in which any peer wants to keeps its data private. Insurance companies want to detect whether one single person is ensured by many companies in a short duration of time and claiming for the same accident to all the companies. This would be only possible if all the companies in a Blockchain share the relevant data for fraud detection and keep other sensitive data private. For implementing the above Hyperledger structure for detecting insurance frauds we need to put encrypted data on the ledger so that only the concerned peers get to access the data. So in the first place, we need to put encrypted data on the ledger by adding the encrypted data in the endorsement policy, and clients who are giving the proposals to multiple peers need to have access of encryption keys of the organization for encrypting their data.

3. EXISTING SYSTEM

Currently, to prove our identity in the physical world, documents like voter ID is used, asserting facts such as our age, name, or eye colour. This personal information doesn't already exist on the Internet. To prove our identity/KYC were forced to share extra information. This happens through our documents, or else use a patch of username-password systems. The personal information of an individual ends up being duplicated or shared across the Internet. Daily we deal with so many siloed systems, users need to compromise with their privacy and underline security.

A customer, who intends to have a bank account, needs to carry the documents needed for KYC procedure to the branch. The documents are verified and scanned to fulfil the purpose. If there's no record of the customer in the KYC database, the data and images are uploaded, and the record is created for which later the consumer is given a key to his record. There are several stages of doing KYC:

- Gathering of personal information
- Acquiring proofs for personal information
- Storing personal information
- Sometimes Background checks as CDD or Client Due
- Diligence is carried out to find little more information about the individual.

During the KYC process citizens end up giving unnecessary additional information than required to agencies. For example, where only mobile number verification is required, Users by providing Aadhaar/PAN/Passport details to agencies allow all personal information, family details, medical details, and picture. This Personally Identifiable Information is very sensitive. It often gets much complicated to claim to revoke of such personal information by a citizen under Fundamental Right to Privacy.

4. PROPOSED METHOD

The solution proposed uses Blockchain based KYC technology. This solution effectively addresses Distributed Ledger Technology (DLT) to enable the exchange of cryptographically signed credentials. Digital signatures are widely being accepted legally in jurisdictions around the world. Emsigner is an example being accepted by Govt. of India. However, it needs two keys. The private key kept secret by the issuer which is used to sign the document. The second key, called the public key, verifies whether the signature and documents are genuine.

There is a need for a standard way to verify the public key of the users for universal acceptance of digital credentials, which would then prove the genuineness of the KYC documents. The mining technology proves the solution of a mathematical puzzle, based on possibility, which requires high computational power.

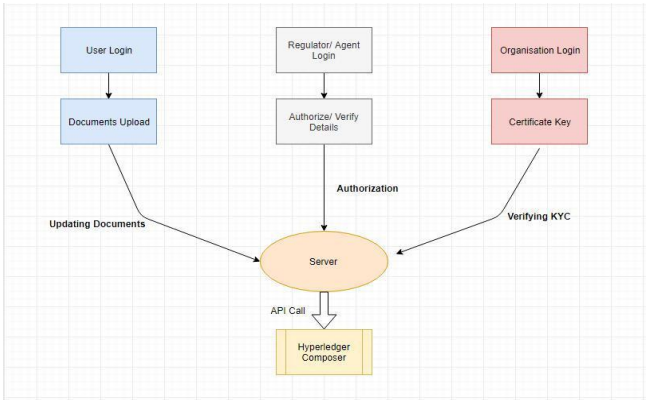

Fig. 2: System architecture

There is a need for a standard way to verify the public key of the users for universal acceptance of digital credentials, which would then prove the genuineness of the KYC documents. The mining technology proves the solution of a mathematical puzzle, based on possibility, which requires high computational power.

- The hash function helps to find input knowing the output.
- Integer factorization representing a number as a product of two numbers.
- Every transaction is chained to the prior via a digital hash, whether singly or in blocks.
- The consensus algorithm is used to replicate Validated Transactions across all machines.

All transaction is digitally signed in the Blockchain. Cryptographic ledger of this immutable records results in making it very difficult, almost impossible to modify previous transactions or malicious future ones. The implementation of DLT part of the solution is directly on Hyperledger composer rather than using a private Blockchain.

The solution to the mathematical equation or PoW problem is called hash. With the expansion in the network, it faces large difficulties as the complexity of the job is a sensitive issue, the

algorithms need high hash power to solve it. All Blockchains in the network represents a cryptographic triple play:

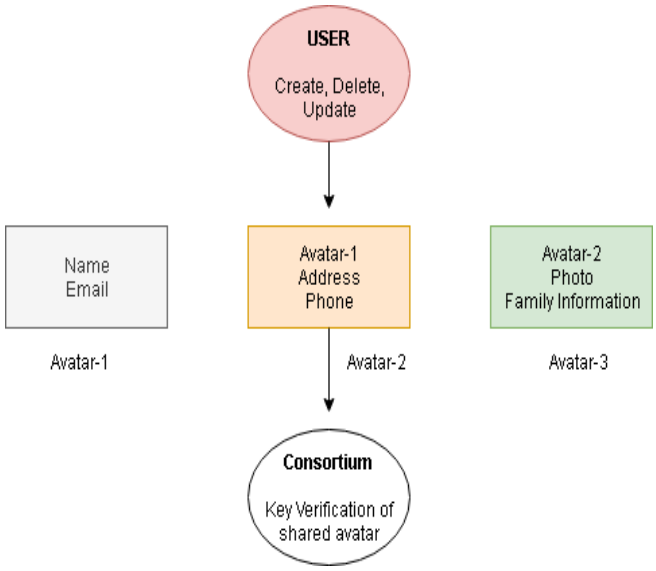

Fig. 3: Avatars

Due to the existence of a large mining community, it would be harder to corrupt and much more difficult for any attempt to change the information on the Blockchain. Second, preventing the risk of any other party to view the system, the regulator is removed from the system. Lastly, to introduce more efficiency, data can be stored only at the organization. Having a digital KYC allows the users to check data in a privacy safe way.

Changing the current system will save private information, lower transaction costs limit the opportunity for crime. As well to overcome the aggregated cost of the KYC process to increase transparency, prioritizing the privacy of the participants. Rather than process carried by each working organization for each user it only needs once by each user, improving over today's system.

How does Digital Identity as Avatar KYC works in real-time?

As Digital Identities does not rely on any third party or central entity, user can autonomously control it for identity verification. The proposed system can be integrated with various organizations to let users complete the KYC procedures as given below:

(a) A person seeking to open account need to verify his identity
(b) For verification of the person's identity, the information used need to be recorded and maintained.
(c) To determine if a person is known to suspect for any government-provided lists.

For the completion of this procedure, there are two stages: Stage 1, Zero-Knowledge Proof, a concept for checking whether something is true or false without the aim to look at the data, where user will be provided choice to enable Digital ID and a token for the same will be issued which can be attached to Digital ID. The output of the verification could be directly presented to the intended organization instead of the original data itself.

In stage 2, user can use the private key for authorization of KYC seeking organization to access certificate if there's a need, and complete information of KYC will be stored in Blockchain. Instead of showing their identity to the relevant authority for verification, Digital identity will let people know that the ID has been verified.

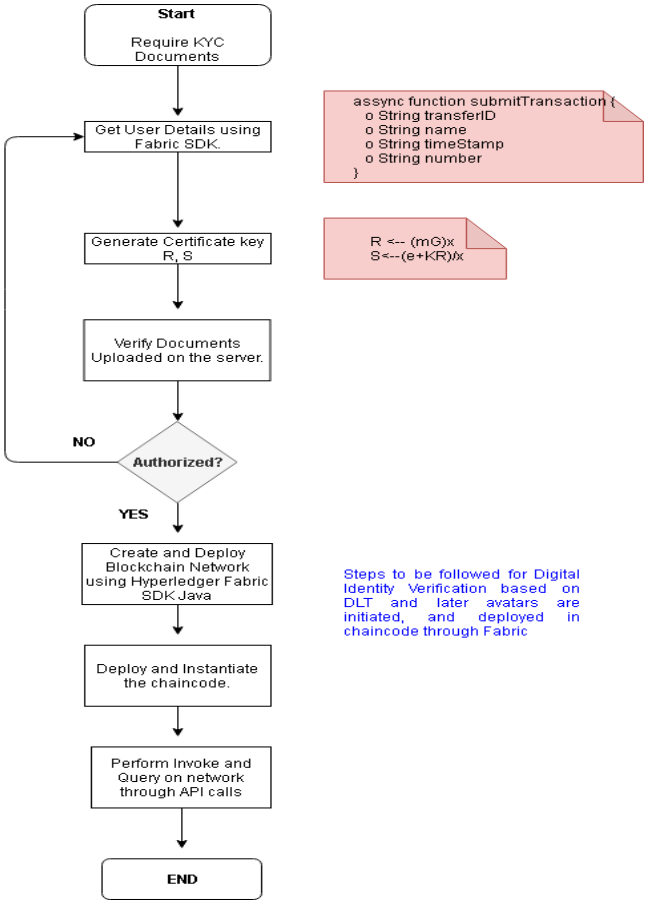

Fig. 5: Flow graph

Blockchain-based technology can solve many of the problems of digital identification by providing transparent, efficient and reliable solutions for government's agencies, banks and financial stakeholders to share trusted certificates. A new transparent frontier is near, where users can finally protect their own data, while authorities can effectively identify users safely and secures them.

5. CONCLUSION

Blockchain has created an infrastructure which is something completely new which ensures trust. Identity is the addition of attributes of data that describe being in a simple and unique way. Neither the data nor the unique attributes of it must be controlled by someone other than the users. Initial approaches to solving digital identities, such as electronic ID cards or other solutions, have failed. Unless we use blockchain technology.

6. REFERENCES

[1] "Bitcoin Mining and its Energy Footprint", Karl Dwyer and David Malone - https://karlodwyer.github.io/publications/pdf/bitcoin_KJOD_2014.pdf
[2] Egelund-Müller B, Elsman M, Henglein F, Ross O (2017) Automated execution of financial contracts on blockchains. Business & Information Systems Engineering.
[3] https://doi.org/10.2139/ssrn.2898670
[4] European Central Bank (2012) Virtual currency schemes. https://www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf.
[5] Autonomous Research LLP, "Block Chain: Backoffice Block Buster". https://www.autonomous.com/fintech/d9335db1-bf1a-4ab2-8d1d-a36cb747a6ae
[6] Nick Szabo, "The Idea of Smart Contracts" (1997). http://szabo.best.vwh.net/smart_contracts_idea.html

[7] CoinDesk, "Understanding The DAO Attack", by David Siegel, 25 June 2016. https://www.coindesk.com/understanding-dao-hackjournalists/

[8] CoinDesk, "CoinDesk Research: Ethereum Hard Fork Had Little Impact on Sentiment". By Bradley Miles. 17 November 2016. https://www.coindesk.com/coindesk-research-spotlight-studyq3-ethereum-hard-fork/

[9] U.K. Government Office for Science. "Distributed ledger technology: beyond blockchain". A report by the UK Government Chief Scientific Adviser. 19 January 2016. https://www.gov.uk/government/publications/distributed-ledgertechnology-blackett-review

[10] https://hyperledger.github.io/composer/latest/

[11] https://searchcio.techtarget.com/definition/distributed-ledger

[12] https://blockgeeks.com/guides/what-is-blockchain-technology/

[13] https://github.com/IBM/blockchain-application-using-fabric-java-sdk

[14] https://hyperledger-fabric.readthedocs.io/en/release-1.4/fabric-sdks.html

[15] https://www.youtube.com/watch?v=vCTabgkvfS0

[16] https://medium.com/coinmonks/getting-started-with-hyperledger-composer-34cb7228d44c

[17] https://www.pwc.in/consulting/financial-services/fintech/fintech-insights/digital-identity-changing-the-way-financial-institutions-connect-with-consumers.html

[18] https://searchcio.techtarget.com/definition/distributed-ledger

