UIDAI Aadhar Enrollment with P2P Blockchain Technologies

Pragati Mynampati, Medha Gourayya, Shashidhara HR

Abstract*: Blockchain technologies are becoming more popular in securing the sensitive data such as government holding citizens' s wealth, health and personal information. A blockchain is a shared encrypted data of records, consisting of a ledger of transactions. As the data stored in blockchain is tamper proof, it is proposed to implement new Aadhar enrolments with P2P Blockchains and migrate the existing centralized Aadhar personnel's personal data from the conventional RDBMS / Big data system repositories to distributed ledger technologies by creating private blockchains. In this paper, we will discuss how to provide security for Aadhar card enrolment data using blockchain architectures. A blockchain-based Aadhaar would help UIDAI in truly complying with the data protection and privacy stipulations outlined in the Right to Privacy Act judgment

Keywords: Aadhar, Distributed Ledger Technologies, P2P Blockchains, UIDAI.

I. INTRODUCTION

Unique Identification Authority of India (UIDAI) body was created with an objective to create Unique Identification (UID) numbers, here after named as Aadhar, for all the citizens of India to eliminate duplicate and fake identities, verify and authenticate in an easy and cost-effective way. The system enrolls the individuals based on demographic and biometric details submitted at the enrollment center (EC) and issue UID numbers (Aadhar cards) for the purpose of unique identification. With the help of this UID number, the citizens who are eligible and below poverty line, avail the Government of India announced socio economic schemes time to time. Aadhar document is made compulsory for all the citizens of India for either opening a new bank account, in getting new telecom service provider connection or link the same to the existing bank accounts / telecom numbers or conducting periodic Know Your Customer (KYC) checks. UID number is also made mandatory for the application process of passport, driving license, travel reservations and property registrations services. Aadhar has become an essential document in all transactions of human life. The importance of this Unique Identification Number, which is generated based on personal information submitted during enrollment time, is growing day by day.

Revised Manuscript Received on December 12, 2019.

***** Correspondence Author

Pragati Mynampati, CSE, RNS Institute of Technology, Bangalore. pragati.1rn16cs044@gmail.com

Medha Gourayya, Asst Professor, CSE, RNS Institute of Technology, Bangalore. Email: medha1103@gmail.com

Shashidhara HR, Asst Professor, CSE, RNS Institute of Technology, Bangalore.

Keeping such large volumes of personal information of every citizen at the national level in a secured way is challenging.All these personal details of the citizens are stored in a centralized database, nationwide and are maintained by the IT service providers / contractors under the supervision of UIDAI body. The primary concern regarding Aadhaar is that a centralized identity database containing demographic and biometric data of over billions of people is an obvious honeypot for hackers. We had observed in the past and present that, the personal information of the citizens is compromised and is utilized slanderously for creating fake Aadhar documents, un-authorized authentications for identity verifications in various public and administration activities. Some incidents of malicious cyber-attacks were also observed in tampering this personal information. Now it is high time to prevent such data tampering, accessing the personal data of citizens by the unauthorized personnel, even UIDAI data security personnel and UIDAI data administrators under different levels of governance control cannot be trusted (prone to venality).

II. CURRENT UID ENROLLMENT SYSTEM

We would like to brief the following enrollments steps take place in present Aadhar system.

The enrollment data captured is encrypted and stored in enrolment center (EC) desktop disc storage system. This encrypted data is serialized, and data packets are FTPed to the enrolment processing center, CIDR centralized FTP server file system through secured virtual private network. This data is further decrypted for various manual and automatic validation checks at enrolment processing application as shown in below figure.

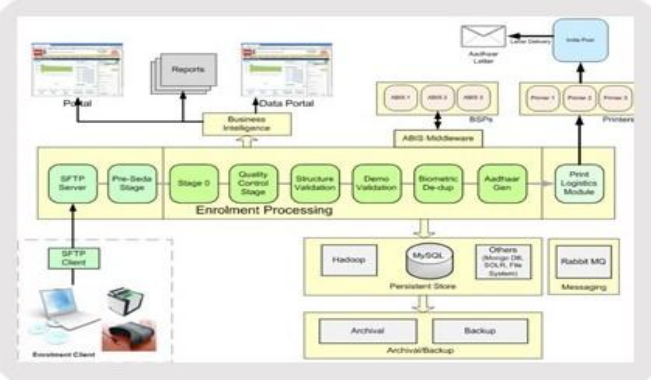

Fig1 : Present Aadhar enrolment processing Architecture

The backend UIDAI staff are involved in all these verification processes. The correctness of enrollment data mainly depends upon the efficiency and capabilities of EC authorities and UIDAI staff, who are responsible for UID generation.

The present enrolment application service is built with staged event driven architecture as shown in the above figure, and each stage is responsible for different kinds of business logic validations that include authentic EC enrolment authority checks, demographic and biometric data deduplication checks etc. All the rejected records are moved into error DBs whose records are manually verified and appropriate actions will be taken by the UIDAI staff.

After successful completion of all validation checks in the enrolment application, the unique identification number is generated.

III. CLIENT SERVER SYSTEMS VERSUS P2P BLOCKCHAINS

We propose to use P2P blockchains over conventional client server database systems in place of enrolment client at EC and enrolment server at CIDR because of the following advantages.

A. Decentralization

Since blockchain follows P2P architecture, there will not be single centralized application and database server. Therefore, even if one node isn't working, the rest of the nodes run the blockchains. In decentralization, the information saved in one peer system node is saved in all the nodes, hence promoting more reliability. If a hacker, tampers with the information in one of the nodes, then the information is recovered from the other nodes containing same replica of data. Whereas, in traditional RDBMS, the database is centralized i.e. under a single authority and the data can be easily tampered. And it can be a single point of failure.

B. Immutable

The data stored in ledgers can only be read or added new transactions to the ledger. The existing record stored in the ledger cannot be altered or deleted. This prevents any unauthorized data changes. Whereas, in RDBMS the data in the table can be altered making it prone to tampering.

C. Transparency

Any creation, updating of data in a distributed ledger is visible to all the participants involved in P2P architecture of blockchains. Any new data creation or updation to the distributed ledger, happens with the consensus of all the stakeholders involved in blockchain ecosystem.

D. Security

The data in the distributed ledger is encrypted and hence cannot be read by unauthorized persons. Whereas, in the traditional RDBMS, the data is not encrypted. Therefore, blockchains offers more protection than the traditional database systems.

IV. PROPOSED ENROLMENT SYSTEM WITH P2P BLOCKCHAINS

In Peer-to-Peer(P2P) based blockchain systems, the data is validated by many stake holders defined by the P2P processes and consensus algorithms runs, as per the majority of consensus, the data is recorded into shared distributed ledger with data encryption. The detailed Peer-2-Peer model of enrolment with block chains is shown in the below figure2. The online citizen enrolls for the Aadhar program by submitting the demographic details and uploading supporting documents related to POI, POA, POR at online portal as per the guidelines of UIDAI. The citizen gets date and time, nearest EC to appear for the enrolment process. The entered demographic details by the citizen may not be recorded into shared ledger directly as these details are not verified and validated by the enrolment staff. After citizen appearing in person at the EC, along with the captured biometric data, both demographic and biometric data after though rough verification is moved into shared ledger of blockchain. Since UID enrolment involves with large size of population, we may not suggest moving record by record to the next peer for his action in the blockchain. A group of enrolment records with data encryption appended by hash keys are moved. The most important and critical document for generating UID is Proof-Of-Identity, Proof-Of-Residence, the Municipality authority is added as another important stake holder in the P2P process. As per the validation logic written in smart contracts module of blockchain, the demographic details are verified against the Municipality records IT system. If the check is successful either by automatic or manual validation, the shared ledger record is updated with demographic details validation completed and moved to the next Peer. As per the POA document documents submitted, the shared ledger record is routed to the bank or telecom authority for his scrutiny with his IT system. The shared ledger record is updated with POA check completed. Finally, the enrolment records are routed to the CIDR backend staff for complete automatic biometric identity system (ABIS) deduplication validation checks. If the duplication of any biometric image found in any one of the geographically located distributed ledger sharded databases, it rejects for enrolment. The detailed architecture of distributed ledger technologies (DLT) is explained in section V. Finally, CIDAI authority waves flag to generate UID number.

The system will not allow any changes to be done to the attribute of the record, if any one of the stakeholders wants to modify, as the records are immutable by head and tail hash codes of the record. In the current system, the UIDAI staff have got full privileges to modify the content of record without any users, EC authorities consent.

In the current system, data packets are moved with batch of enrollment records from EC to CIDR, their order of data packets during reconstruction may get changed at CIDR and due to this EC may have to resend the packets once again to the CIDR data center. Sometimes, if the enrolment data is not available in EC desk top disc storages, the UIDAI may ask citizens to re-enroll for the registration. This can be eliminated by the P2P process.

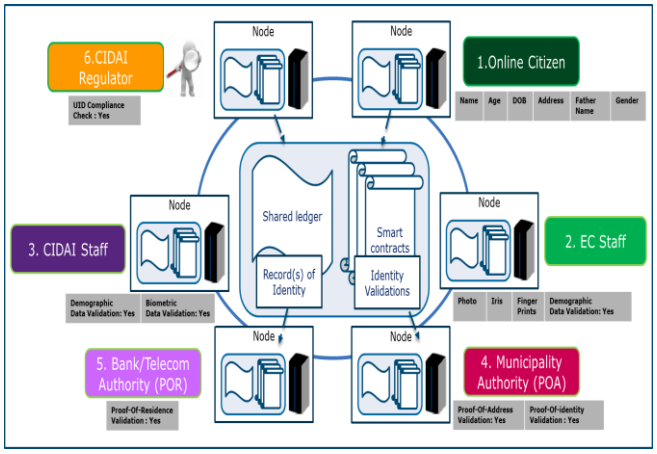

Fig2: Proposed P2P Blockchain model enrolment

As the enrollment master data is encrypted and hashes are added to a record or group of records. the IT service provider who is maintaining the system has no scope for tampering the records.

In a P2P blockchain models each enrollment record or group of few records are processed through smart contracts layer, where complete business logic built, in a workflow process model. If any error observed, this will be notified immediately, and action will be taken place and the same will be communicated to the user immediately. In the current Aadhar system, the enrolment processing takes place in batch mode at CIDR, Due to this enormous delay taking place in communicating the status of registrations to the citizens and PIN generations.

V. RESUTLS AND DISCUSSIONS

The migration of the data from the big data (Hadoop) which contains all the records of the individuals who enrolled for the Aadhaar, to blockchain can be implemented as follows.

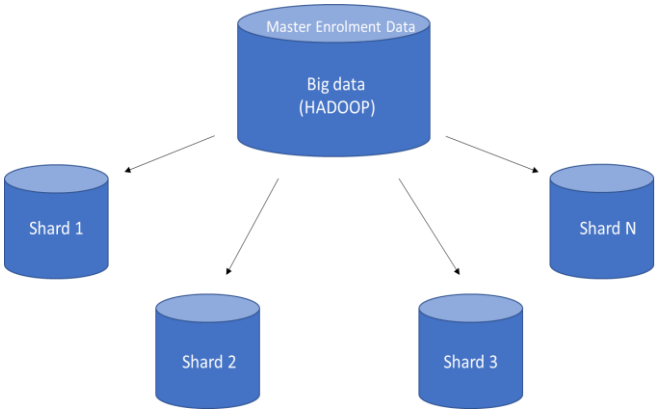

Fig 2: Sharding – Dividing the big database into small chunks

The traditional data is sharded, a process that divides the data into small chunks. From the figure illustrated above, the Hadoop data is divided into small shards based of different sharding strategies. Few strategies for sharding the Big Data platform are by based on number of enrolment records, by province or region wise of enrolment or by CIDR Authorities ownership, or by 12 digit UID Aadhar number wise etc.

A metadata table about shards is created for reference. The metadata table can contain information about shards such as, Shard ID, the size of shard, the starting Enrolment ID (EID) of the record, ending Enrolment ID of the record, location of the shard, owner of the shard etc. One sample metadata table is shown below.

Each shard is encrypted on the local system, so that the information isn't displayed. The peers have the decryption key to view and validate the data in the block. Only the content owner has the full authority over this process.

A hash (encrypted output string of fixed length) is generated for each shard. The hash is generated from shard's data or the encryption key. This hash is added to the ledger and shard metadata table to link the transactions to the stored shards.

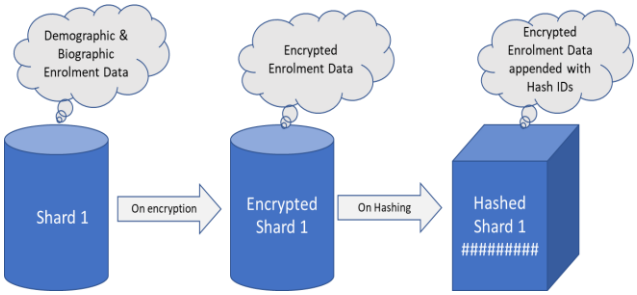

Fig3: Performing encryption and hashing on shard to transform to Distributed Ledger Block

The hash generated for each shard is appended on to the metadata table. The hash ID assigned to each shard helps in identifying the shard uniquely. The metadata table is shown below.

Shard ID	Starting EID	Ending EID	Hash IDs	Shard Location	Shard Authority
S0001	EID-0000001	EID-1000000	EF64976875JFHJ	ND 001	Louis Philips
S0002	EID-1000001	EID-2000000	HT98568648KITL	ND 002	Mary Rose
S0003	EID-2000001	EID-3000000	MD1348913DTUI	ND 003	Jean Marc
S0004	EID-3000001	EID-4000000	KL679006875JFHJ	ND 004	Ram Kumar
S0005	EID-4000001	EID-5000000	RT398048648KITL	ND 005	Haritha Rao

Fig4: Shard's Metadata Table

Now, the big data is divided into hashed shards. These hashed shards are replicated and are sent to the nodes in the P2P architecture

The shards are assigned to the nodes in such way so that the workload is reduced and the time consumption is minimized. For example, if a blockchain network contains 1000 nodes, and the data is partitioned into 10 shards, each shard is assigned to 100 nodes. Therefore, each node stores only 1/10th of data, but data is still verified across 100 nodes.

For simplicity, assume that there are 3 shards created and 6 nodes in the P2P architecture (shown in the figure). The nodes doesn't contain copies of all the 3 shards. Instead, the nodes are divided into a network of 2, which can handle a single shard. The other two nodes contain another shard and so on. This adds to multiple transactions taking place instead of one.

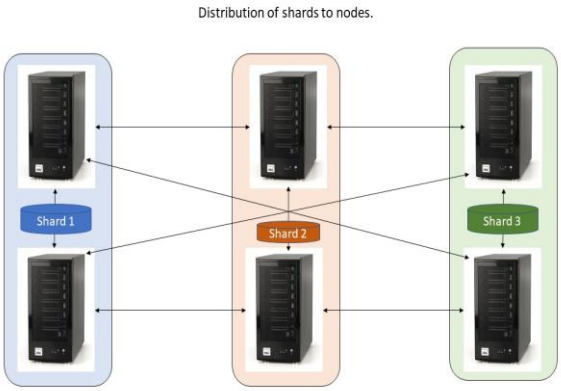

Fig 5: Distribution of shards to nodes

The nodes are in P2P architecture, meaning that the nodes can be anywhere globally or regionally. The P2P nodes with respect to Aadhaar can contain the following stakeholders:
UIDAI
Government Officials
Third party vendors (service providers)
Security Organizations
Technology vendor partners
Citizens
If a record in the shard is to be modified by the user, The user makes a new record which contains updated information. This updated record is then passed on to all the nodes to verify.
All the peers follow the consensus algorithm to validate the new record. The new record is then appended to the shard, which contained the old record.

VI. SUMMARY AND CONCLUSIONS

In this paper we have discussed about the present UIDAI Aadhar enrolment system and its functioning. By introducing P2P block chains technologies, how the enrolment processes and system can be improved. We also discussed current pitfalls of the present system and with the block chains based Aadhar system, how these pitfalls can be overcome. We also discussed about the idea of how the present big data-based enrolment platform can be migrated into distributed ledger-based technologies. A Proof-Of-Concept model must be built further based on this distributed ledger technologies, the security and robustness of the model must be evaluated.

REFERENCES

1. Zibin Zheng,Shaoan Xie,Hongning Dai, Xiangping Chen, and Huaimin Wang "An Overview of blockchain technology: Architecture, Consensus and future trends."
2. "Aadhaar technology and architecture: *Principles, Design, practices and key lessons" – March 2014*, UIDAI planning commission, Government of India.
3. Figure 1. "Aadhar technology and architecture", March 2014, page 43, Figure 3: enrolment module overview.
4. "Introduction To Aadhaar, Digital Identifiers" by Ashok Kumar.
5. Architecting World's Largest Biometric Identity System - Aadhaar Experience by Dr. Pramod Verma.

AUTHORS PROFILE

Pragati Mynampati has published paper on AI & ML techniques for fraud detection in Aadhar system. Her areas of interest include adoption of AI & ML techniques in life sciences. She has carried out couple of projects in this area like detection of Diabetic retinopathy by retina images and classification of white blood cells using deep learning techniques.

Ms.Medha Gourayya is currently an Assistant Professor in the department of computer science at RNS Institute of Technology, Bangalore. Her research areas include Machine Learning, Deep Learning and Natural Language Processing.

Dr. Shashidhara HR is currently an Associate Professor in the department of computer science at RNS Institute of Technology, Bangalore. His areas of interest and research include Data mining, Web mining and Information Retrieval.