1 Introduction

This section lays down the foundation for the building a system for proper KYC and using the data to provide authorization to different web services. Section 1.1 presents an insight into what KYC, BLockchain, smart contracts are and their implications in the current socio-economic.

1.1 Context

More and more are services are being shifted online to enhance their availability and customer reach. After the inauguration of Digital India Policy by the government of India in 2014, the rate of digitalization in the country has been increasing at an brisk pace. Also the availability of cheap internet means more and more people on internet, storing more data online and trusting the internet even more.The government regulations of KYC and PAN validations has increased the trust of financial institutions on this data and also to sharing of data between them.
The rate and effectiveness of cyber attacks have also increased in these past five years calling for an efficient way to not only storing but also sharing and using this data. A key challenge here is proving authorization with this data and sharing it among different institutions .

1.1.1 Blockchain

A blockchain is a secure, distributed transaction ledger which stores all the transactions occurring in the network Crosby et al. (2016). By this blockchain maintains an auditable and immutable history of all the transaction which have happened on the network. The blockchain is not limited to storing transaction only. It can also be used to store any code (as in Ethereum blockchain) or any data. As the name suggests, blockchain is nothing but a chain of blocks which storescertain records. It similar to linked lists in data structures. The new blocks are continuously created and added to the chain by special nodes (known as miners or validators). Since, blockchain is a distributed database, multiple copies of data exist in the network which forms a Peer to Peer network . This distributed nature of blockchain creates an in-built redundancy. The loss of one node or a couple of nodes will not have any impact on the whole network.

1.1.2 Know Your Customers

KYC is the process of business identifying and verifying their customers Limehouse (1999). The entire objective of KYC for any organization like the bank is to prevent them from being used by criminal elements for money laundering activities and terrorism financing. Financial Industries like Banks complete their KYC while opening new bank accounts and are required to periodically update their services.

1.2 Problem/Motivation

1. Any group of hackers can lay claim to data through attacking less secured sites and then use that data to take advantage of government policies through there official websites.

2. The authorization process can be tricked to compromise on identity. Thus with the major policies going online, it can be used mischievously to take undue advantage.

3. Almost same data is saved on multiple locations, as each website maintains its own copy of data, thus compromising on network storage.

2 Review of key related research

2.1 Background

A considerable amount of research work is currently in progress on the use of Block chains to address the various issues pertaining to data. Block chains are though to be the appropriate answer to the questions posed on Confidentiality, Consistency and availability of user data. We explore the possibilities of using block chains as a data storage and single point of authentication for different web services.

2.1.1 Types of authentication mechanisms

The authors of Bonneau et al. (2012) have presented various types of authentication mechanisms. These categorizations of authentication are all widely used in the their respective networking technologies and have been proved to be vulnerable to cyber attacks. Major types of mechansisms are given below:

- **Passwords**

- **Hard Tokens**
- **Soft Tokens**
- **Contextual Authentication**
- **Bio metric Authentication**

2.1.2 Uses of Block chains

Blockchain since its inception have been used to solve various real world problems, and till date research is going on to explore the utility of blockchain. Some of it's use case are:

- **Blockchain and IOT**
- **KYC with Blockchain**
- **Blockchain for biomedical applications**
- **DNS management with Blokchain**

3 Objectives

In order to make an authentication mechanism that addresses such issue we need to store data at a single location. But if all the data is kept at a central location, it gives birth to central point of failure which may disrupt not only isolated websites, but the complete system as a whole. Thus we need decentralization as well to assist the mechanism.

The data has to be shared amongst the various organizations/web services requiring it, but not without the consent of the user. There has to be a proper mechanism where in the organization can request for the data and the user approves the request before any actual transfer of information takes place.

The aim of the research is to develop an authentication mechanism that is decentralized, easily usable by multiple clients, promotes faster verification, reduces the redundancy of data. The decentralised authentication system should have the following features:

1. To develop a system that allows the user to store all his KYC(or any other sensitive information) on blockchain.
2. To use the blockchain as a sign on option for different web services.
3. To allow the different institutions(web services) to request for the data.
4. Allow the user to approve every request that is made by participating institutions.

4 Methodology

We plan to adopt the following work strategy in order to address the identified shortcomings in the recent related research along with the action plan to meet the set research objectives.

4.1 System Design

Ethereum was used to develop decentralized application and the smart contract was deployed on Rinkbey Test Network for testing which has been described below.

4.1.1 Entities within the system

The proposed system has two types of entities :

1. Users : User is any person who wants to upload his data on the blockchains and thus form a relationship with the clients.

2. Third Party apps (Clients) : The client on the other hand is an entity with which the user shares his/her data in order to enjoy the services provided by it.

4.1.2 Application Model and Security

A user posses private data like Aadhaar card, passport, address proof, his credit card details and contact information which needs to be shared with some client within the network. A different client might need different documents for authorization but mostly, they belong to a similar category which means that the same documents can be shared among clients.

In the model, the relationship between a customer and a client can have the following states:

1. The client wants to access the users information but does not have permission to do so from the user.

2. The client wants to access the users information but is not part of the network or is still not verified by the verifier to be a part of the network.

3. The client has permission from the user to access his information

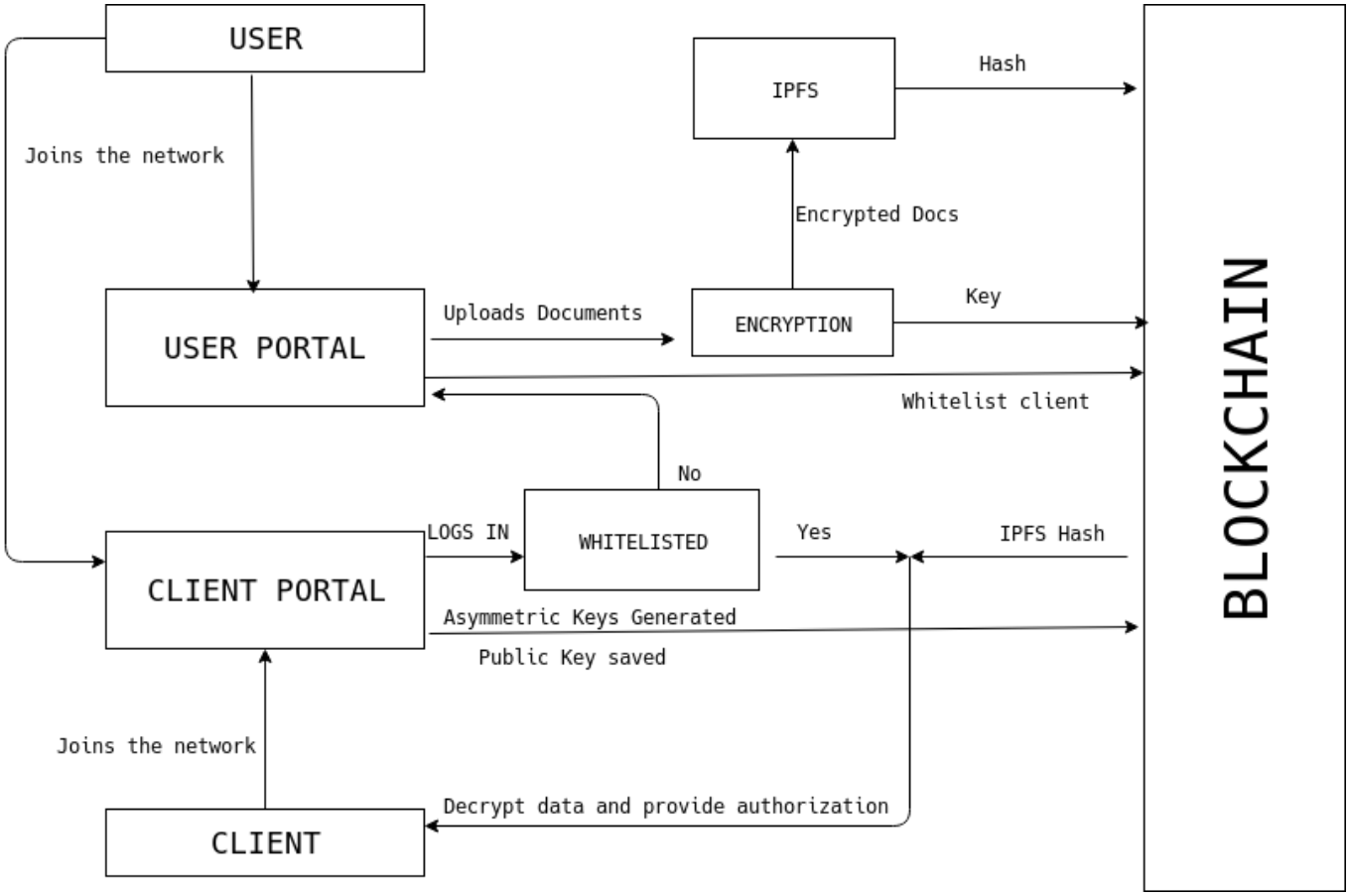

Figure 1: System Design and Model Overview.

4.1.3 Anonymity of relationship and privacy

Its important to ensure that users have total control over their information. Thats why the users have control over with whom they want to share their information and if necessary can revoke that consent.
Only after successful consent from the user, any client can read users information. The user can also revoke consent anytime but it will only apply to the future updates to the information as access to the previous version of information cannot be undone.

5 Expected research outcome

We expect to build a trusted decentralised authentication mechanism that is always available, without the use of central servers. It is required to maintain availibility and consistency of the data stored on it. We also need to enforce a system that allows sharing

of data between multiple clients.

Some of the expected outcomes of the research will be:

- Easy uploading of KYC details.
- Privacy, Consistency and Availibility of KYC detailed.
- Sharing of KYC data to the clients only after the due approval.
- Minimise the latency in communication between user, client and blockchain.
- More Secure than current password based authentication
- No single point of failure.

6 Security Analysis

The system proposed is heavily based on Blockchains, little insights into the internal working of blockchain is required to analyse and measure the security of the system. The core of any blockchain is it's consensus algorithm.

6.1 Consensus Algorithm

A consensus protocol has three key properties based upon which its applicability and efficacy can be determined.

1. Safety A consensus protocol is determined to be safe if all nodes produce the same output and theoutputs produced by the nodes are valid according to the rules of the protocol. This is also referred to as consistency of the shared state.
2. Liveness - A consensus protocol guarantees liveness if all non-faulty nodes participating in consensus eventually produce a value.
3. Fault Tolerance A consensus protocol provides fault tolerance if it can recover from failure of a node participating in consensus.

6.1.1 Approaches to consensus

- PoW(Proof of work) is a consensus strategy used in the Bitcoin network. In a decentralized network, someone has to be selected to record the transactions. In PoW, each node of the network is calculating a hash value of the block header. The block header contains a nonce and miners would change the nonce frequently to get different hash values. The consensus requires that the calculated value must be equal to or smaller than a certain given value. Nodes that calculate the hash values are called miners and the PoW procedure is called mining in Bitcoin.In the decentralized network, valid blocks might be generated simultaneously when multiple nodes find the suitable nonce nearly at the same time. As a result, branches may be generated . However, it is unlikely that two competing forks will generate next block simultaneously.In PoW protocol, a chain that becomes longer thereafter is judged as the authentic one.

- PoS(Proof of stake) is an energy-saving alternative to PoW.Miners in PoS have to prove the ownership of the amount of currency. It is believed that people with more currencies would be less likely to attack the network. The selection based on account balance is quite unfair because the single richest person is bound to be dominant in the network.

7 Results

The results achieved so far are in consistency with the objectives of the proposed system. The research is still in progress hence not all the objectives are achieved, the partial results have been enumerated below :

- **Easy upload of KYC Data** The primary motive of our system was to upload documents to blockchain, where they can be shared with other organisations which the user permits. The only limitation to this feature was the amount of data to be written on the Ethereum blockchain. Ethereum has the concept of gas, which is a measure of cost of writing data to blockchain.

  As can be inferred from Figure 4, the amount of data to be written on the blockchain has to be minimised and thus storing all KYC documents on the blockchain will not be a scalable implementation.

The system uses IPFS to counter this problem. The files are uploaded to IPFS, and the address of the file is hashed and stores into the blockchain , reducing the amount of data written onto the blockchain and thus saving the cost.

Table 1: Comparison of Gas Cost.

Sr. No.		**IPFS**	**Direct Upload**
1	**Size**	32B	50KB
2	**Gas Cost**	55600 units	88960000 units
2	**Ethereum Used**	0.001112 ether	1.7792 ether

- **Privacy of Uploaded Data** The primary objective of the system is to provide privacy to any data that user uploads on the blockchain. To fine tune the proper strategy we used and analysed two very popular encrytion algorithms, **RSA** and **ECC**

  As per the analysis, it was found that ECC provides better security than RSA for any given key size. RSA is based on the difficulty of factoring large integers , whereas ECC is based on detecting the separate logarithm of a random elliptic curve.
  Simple numerical tests reveal that the level of security provided by a 3072 bit key in RSA can be achieved using 256 bit key in ECC. To measure the security the notion of bit security been devised where

  n-bit security means $2^{(n)}$ operations are needed to break it.

- **No single point of failure** The proposed system has to be robust against any kind of hardware flaws and failures. As in case of centralised authorisation schemes, if the central DB goes off the network due to some reasons, the whole system becomes unusable, thus allowing a single point of failure for the whole system.

  In the system proposed , a consensus algorithm was chosen such that any node can participate in the mining process, and thus hold a replica of data, without any identity management to make sure that there always is a large pool of miners. To

Table 2: Key Size ECC VS RSA

Sr. No	**Bit Security**	**ECC**	**RSA**
1	90 bits	160 bits	1024 bits
2	112 bits	224 bits	2048 bits
3	128 bits	256 bits	3072 bits
4	192 bits	384 bits	7680 bits
5	256 bits	521 bits	15360 bits

cause any failure in such a system at least 51% of the nodes must fail, making our system free from any sort of Single point of failure.

References

[1] Bonneau, J., Herley, C., Van Oorschot, P. C. and Stajano, F.: 2012, The quest to replace passwords: A framework for comparative evaluation of web authentication schemes, *2012 IEEE Symposium on Security and Privacy*, IEEE, pp. 553–567.

[2] Crosby, M., Pattanayak, P., Verma, S. and Kalyanaraman, V.: 2016, Blockchain technology: Beyond bitcoin, *Applied Innovation* **2**, 6–10.

[3] Limehouse, D.: 1999, Know your customer, *Work study* **48**(3), 100–102.