INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 7, Issue 1)
Available online at: https://www.ijariit.com

Building a decentralized system to optimize the KYC process

Rohit Venkatraman
roven97@ymail.com
Amity School of Engineering and Technology, Noida, Uttar Pradesh

Piyush Aggarwal
piyush.aggarwal804@gmail.com
Amity School of Engineering and Technology, Noida, Uttar Prades

Ashish Jha
a.jha740417@gmail.com
Amity School of Engineering and Technology, Noida, Uttar Pradesh

Dheeraj Gupta
dheerajkumar92100@gmail.com
Amity School of Engineering and Technology, Noida, Uttar Pradesh

ABSTRACT

This paper aims to develop a reputation-based, decentralized system using Blockchain to optimize the KYC process. The system will aim to optimize the core know-your-customer (KYC) verification process for financial corporations and improve the customer experience by reducing costs and increasing transparency of the data shared. The current KYC process costs more than $500 million for banks, most of which is spent to secure the customer's personal data. In the proposed system, the KYC verification process will only be conducted once for each customer by a validator in the Blockchain, regardless of the number of financial corporations with which that customer intends to work. Owing to Blockchain, the core KYC verification result can be securely shared by the customers with all the financial corporations that they intend to work with. This process produces a gain in efficiency, reduces cost, improves customer experience, and increases transparency throughout onboarding a customer for the financial corporation.

Keywords: *Blockchain, Proof-of-Reputation, Decentralized Ledger, KYC*

1. INTRODUCTION

With the amount of data in our world increasing drastically, it is vital to have a system to tackle the increasing risks associated with it. According to a recent report [1], it is estimated that 20% of the world's data has been collected in the past couple of years. It is safe to assume that data is continuously being collected and analyzed in this Big Data era, leading to innovation and economic growth. Companies and organizations use the personal data they collect to personalize services, optimize the corporate decision-making process, predict future trends, and more. Therefore, data acts as a valuable asset in our economy [2]. While we all enjoy the benefits of a data-driven society, public concern about user privacy is growing. Centralized organizations – both public and private, gather large quantities of personal and sensitive information. An individual has little or no control over the data stored about them and how it is used. To address this issue, leading companies implement their proprietary authentication software based on the OAuth protocol [3], and researchers have been developing models that focus on privacy concerns regarding personal data. In recent years, a new section of accountable systems has emerged. The first such system was the Bitcoin blockchain, which allowed users to transfer currency, such as bitcoins, securely without a centralized regulator. Since then, many projects[4] have demonstrated how these blockchains can serve other functions requiring trusted computing.

The success of Bitcoin led more and more researchers to study the potential applications of blockchains in general. Blockchains are consensus-driven, i.e., many computers are connected to the network, and to reduce the potential of an attack, those adding transactions to the Blockchain must compete to solve a mathematical proof. The results received are shared with all the other nodes on the network. The computers, or nodes, connected to this network must agree on the solution, hence the term "consensus." Different blockchains make use of different consensus such as Proof-of-Work, Proof-of-Stake, Proof-of-Authority, and Proof-of-Reputation.

No single node can take control of the information present on the Blockchain. Therefore, we need not trust a single entity since we rely on agreement by many entities instead. The beauty of this is that the transactions being recorded in the chain can be publicly published and verified. Any node can view the Blockchain contents and verify that the events recorded into it took place. Unlike Bitcoin, Ethereum [5] is not purely digital money, but it does not mean that one cannot transact on that Blockchain. Ethereum's Blockchain can handle accounts and transactions like Bitcoin's Blockchain. However, it can also store and execute newly coded programming logic. These logical codes are written, executed, and stored historically on the Ethereum blockchain forever for future reference called smart contracts.

The surged regulatory cost incurred due to the know-your-customer (KYC) verification process in banking is one of the enormous challenges that the banking sector is currently facing.

The costs of performing KYC and complying with the regulations imposed by respective governmental authorities is estimated to be more than $500M[6]. The fines can further augment these costs levied on financial institutions due to their misconduct concerning anti-money-laundering (AML) and KYC regulations. The sources of extra costs do not stop here, as financial institutions cannot conduct business with corporate entities that have not completed the complete KYC process. Since the process is long and tends to increase with the corporate entity's size, the starting point of the business relationship between a customer and a financial corporation is delayed, which results in customers' displeasure with the process itself.

This paper aims to propose a new method for the KYC verification process. We introduce a system based on Blockchain (Ethereum) that provides a solution to the increased costs of the KYC procedure and the lack of customer satisfaction. The key reason for using Blockchain is that it allows us to create a chronological, decentralized, interbank ledger in which financial institutions process the KYC verification tasks for a customer. Upon verifying the result of the process that has already been conducted for that customer, the institution can avoid conducting duplicated KYC verification tasks. In particular, the system allows customers to carry out the full KYC process with only one financial corporation and later share their result of that KYC process with any other financial institutions that they expect to work with.

The proposed system's main improvement over the current KYC process is that it only needs to be carried out once by each customer, rather than once by each corporation working with that customer. It provides complete control to the customer over their data and reduces the cost of the KYC process without compromising the security of the system. It also respects the privacy of the participants and increases transparency in case of a conflict. Additionally, the Proof-of-Reputation consensus mechanism allows the immutable exchange of information across participating corporations with honest interbank collaboration.

Organization. Section II discusses the KYC problem we solve in this paper; section III provides an overview of the approach we use, whereas section IV describes in detail the implementation; section V discusses the results after implementing the system using the Ethereum blockchain; section VII discusses future extensions of the system, and concluding remarks are found in section VI.

2. THE KYC PROBLEM

Know-Your-Customer (KYC) refers to the steps taken by financial corporations to establish customer identity and gain insights into the customer's activities. The primary goal is to satisfy that the source of the customer's funds is legitimate. The corporations also assess money laundering risks associated with that customer to monitor the customer's activities. Data privacy has always been an issue for the customer in a data-driven environment like today. Respecting data and transactions' privacy is a core tenant of any project, and the companies understand that. In the current scenario of performing a KYC check, Customer-Identification-Program is used to mandate that any individual conducting financial transactions needs to have their identity verified. A critical element to a successful Customer-Identification-Program is a risk assessment, both on each account's institutional level and procedures. Additionally, a Customer-Due-Diligence makes sure any potential customer is worthy and is a critical element of effective management of the risks and protection against criminals, terrorists, and corrupt politically exposed persons. Generally, the most forward-thinking financial entities worldwide have started to look towards more sophisticated biometrics to simplify KYC verification and improve security. Biometrics rely on a person's biological characteristics to verify identity, which can come through fingerprints, vein patterns, iris scanning, and facial recognition. All this data acquired by the banks is stored in a centralized database used by these banks. The costs to secure these databases are north of $500M per year to protect their customers' data. Throughout this paper, we address the customers' privacy concerns- is the data shared safely? And the financial corporation's concern- how to secure personal data in a much more efficient manner?

3. PROPOSED MODEL

We propose a general blockchain-based system that aims to solve the significant challenge of storing and securing a customer's data for various financial corporations that the previous generations of a centralized system have failed to resolve. We will focus on the application of this system on a peer-to-peer network. Blockchain technology is a peer-to-peer approach for linking a sequence of transactions or events together such that it makes them immutable. This was initially described by Nakamoto and implemented for the virtual currency Bitcoin [7]. In Bitcoin, users exchange money using transactions. When a user problem. It suffers when it comes to providing high performance makes a transaction he broadcasts the transaction to all the nodes in the network. A particular group of nodes, called miners, collect broadcast transactions and attempt to incorporate them into a block that satisfies a cryptographic hash function. The process of producing a block is computationally intensive and probabilistic. Given a proposed block, each miner has a fixed and independent probability of successfully producing a block that satisfies the hash function for each unit of computation time. This is called the Proof-of-Work consensus mechanism. Blocks are also linked together by chaining the hash of the previous block with each subsequent one. Thus, an attacker must control a significant proportion of the computation power (typically 51%) to produce one false block, and faking transactions back into the past is exponentially hard. Ethereum, like Bitcoin, currently works on this Proof-of-Work mechanism, but their team is pursuing a switch from Proof-of-Work to Proof-of-Stake in which less energy is consumed. Currently, all the Bitcoin blockchain's mining activities consume much power, almost 510,000,000 GH/S [8] of computational power, which is unsustainable. While Proof-of-Work depends on computing power to mine blocks, Proof-of-Stake depends on a node's stake in the network. Typically it is the amount of currency the node holds. The more stake of the node, the more authority it has over validation. Therefore, a set of validators are chosen who take turns proposing and voting on the next block. While Proof-of-Stake manages to reduce energy consumption costs by choosing validators based on their stake in the network, it has its fair share of and acts by making assumptions about its peers' interests being in line with the network. Also, there is no additional accountability over Proof-of-Work for anonymous actors who can gather much wealth on the network.

Resolving these issues, our proposed system uses a consensus model, similar to Proof-of-Stake, that depends on the validators' reputation to keep the network secure called Proof-of-Reputation. Once a node passes verification and proves reputation, it is voted into the network as an authoritative node, and at this point, it can sign and validate blocks. This accounts

for the transparency and accountability problem of the Proof-of-Stake consensus mechanism. Each validated node goes through a verification process to ensure that its identity is correct. This is automated through the use of smart contracts on the Blockchain. Additionally, the smart contracts are used to implement an access-based method[9] for the banks to modify an existing customer's data. Our system focuses on ensuring that users own and control their data. The system accepts the users as the data owners and the validators(financial corporations) as guests with delegated permissions.

4. IMPLEMENTATION OF THE SYSTEM ON

4.1 Blockchain

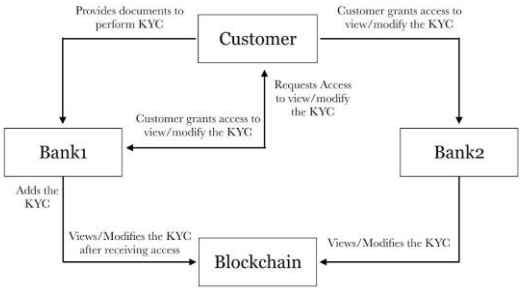

Fig. 1: Our proposed system

The overview of our system is illustrated in Fig 1. The entities comprising our system are customers interested in the corporations' financial services, and the financial corporations, the providers of such services, require processing personal data for operational and business-related reasons. These corporations also act as the entities entrusted with maintaining the Blockchain and the distributed data store and shall be called validators.

The proposed network has two goals – to provide a generalized reputation-based system that can be implemented into any network and provide an access-based method to access the personal data. According to Guy Zyskind[9], a simple function that assigns a trust score to the nodes in a blockchain gives more weight to *trusted* nodes and computes blocks more efficiently. Our system uses a method that assigns a validator's reputation as a function of the number of valid KYC's performed by them.

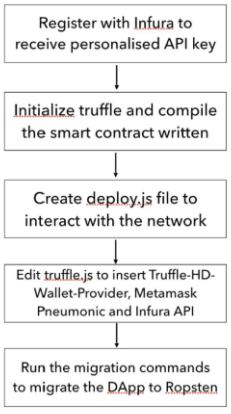

Fig. 2: Implementation of Ropsten

The system is implemented in the Ethereum's Ropsten network. To do so, we need to connect with Ethereum through a node. This can be done with Infura. Infura is a gateway in order to connect with Ethereum Blockchain. The user first signs up on Infura and can then use the API. The API provides a set of public nodes removing the need to have a local or maintained client fully synchronized with the leading Ethereum network. For Ethereum to grow, the Blockchain falls on its users to develop the infrastructure that keeps the network running. Across the Ethereum network, there is a need for utilities to lower entry barriers and simplify access to Ethereum data. Truffle/ web3 provides the simplest way to integrate Infura. Whenever the page is loaded, the initial javascript file that initializes web3.js and Infura runs on the Truffle framework, and thus, Truffle connects to the Ethereum network through Infura(see Figure 2).

The system is designed as follows. The Blockchain, broadly, accepts two types of transactions: ***T****access* used for access management; and ***T****data*, for data storage and modification. These operations can be integrated into different Software Development Kits (SDK) that applications/services can use in their development process. Our model is designed and implemented using the Ethereum Blockchain. Ethereum offers anonymous transactions and is not controlled or regulated by any centralized body. Its language- *Solidity* is used to develop *smart contracts.* Smart contracts are self-executing contracts that follow a protocol to allow the performance of credible transactions without third parties. This is often why Ethereum is termed to be development friendly. For our model, we developed a smart contract that interacts with our system to access and modify Blockchain data using transactions like ***T****access* and ***T*** *data*[9].

To understand their functions, imagine a scenario where the user signs up for the first time. A new, shared identity is generated and added to the Blockchain. After that, whenever any organization requests to view their data, an access request ***T*** *access* is sent to the customer for the first time. The user decides whether or not to provide access to the organization, i.e., the user's data cannot be accessed without the user's consent. At any point, the user can revoke the access they had granted to the organization earlier, thus ensuring complete transparency. The following flowchart shows how access is granted to the bank for viewing the customer's sensitive data.

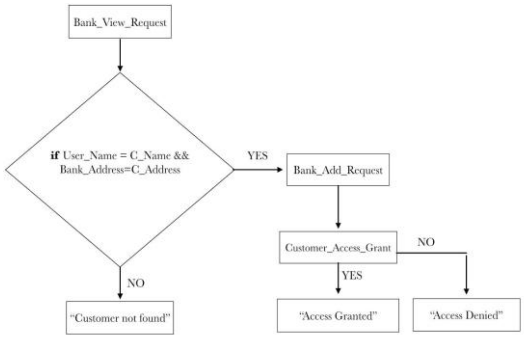

Fig. 3: Flowchart depicting access granting

Bank_View_Request is the procedure through which the bank adds a request to view the customer's data. Customer's data (*such as their username and the bank's address that added them to the network*) through which we uniquely identify the customer from the pool of customers is passed with the procedure. Once the customer is found, the request is sent to the customer, and the customer decides to grant access to the bank. Data collected from the customer (e.g., KYC documents) is encrypted using a shared encryption key and is then sent to the Blockchain. A ***T*** *data* transaction consists of the organization's request to access the data stored over the blockchain network to modify it. The data stored in the Blockchain is sufficiently randomized across the validators (nodes) and replicated to ensure high availability. It is instructive to note that alternative off-blockchain solutions, too, could be considered for storage.

Our system's final component, which makes it more secure and reliable, is the Proof-of-Reputation consensus mechanism. There are various methods to calculate the reputation of an entity. However, in our system, each validator is assigned a reputation, which we compute as a function of the number of valid KYC's performed by that validator. We propose to solve the issue of quantifying reputation by removing the human opinion from the transaction. Unlike most previous generation reputation systems where validators' reputation is community-controlled, our proposed reputation system is based on the integrity of the client of the validators. The function through which we compute it is given as:

$$f(KYC\ count) = f(KYC\ count-1) + 1/(KYC\ count)$$

Where KYC count represents the number of KYC's performed by the validator, the function does not exceed ten, and upon successful entry of multiple KYC's by the validator, falls in the range of [1,10]. The value of the function is zero at f(0). A similar function is used with a negative sign to compute the decrease in the reputation of a validator whenever they are caught falsifying customer data.

5. RESULTS

As discussed in section III and section IV, only the customer has control over their data in the given model. Upon developing the system using the Ethereum blockchain, we were able to verify the first part of our hypothesis, which is the Ethereum blockchain acted as an access-moderator between the customers and the banks. Only when the customer provided access to the banks to access their data could the banks do so. The decentralized essence of the Blockchain combined with digitally-signed transactions ensures that an adversary cannot pose as the bank or corrupt the network, as that would imply the adversary forged a digital-signature or gained control over the majority of the network's resources. Similarly, an adversary cannot learn anything from the public ledger, as only hashed pointers are stored in it. The other part of our hypothesis, which is primarily based on using a bank's reputation as the *stake*, was also visualized. By setting up our reputation-based system, we saw an increase in a validator rating every time they entered a valid customer into the network. Their reputation took a toll when a customer was found fraudulent and was forced to quit the network. This reputation-based system brings something new to the table, which may never have existed before; that is, it brings different banks(*validators*) together to maintain the integrity of the network. If anyone were found cheating, then their reputation would go down, and in the real world, this has proven, from time-to-time, to be a compelling downfall factor many for large corporations.

6. CONCLUSION

This paper has discussed a system based on the distributed ledger technology(*Blockchain*), working on the Proof-of-Reputation consensus to optimize the current KYC scenario for financial corporations. We have shown how a system that could be implemented into Ethereum's network or, in essence, various networks is possible. Our system also enables a blockchain as an access-control moderator. We discuss how the system would be implemented and demonstrate how our proposed system solves many of the issues faced by the current procedure to perform KYC and store personal(*sensitive*) data.

Overall, this paper aimed to propose a system that solves most issues faced in the current method. Companies can thus, focus on utilizing data without being overly concerned about adequately securing and compartmentalizing them. However, this is just the foundation of the idea. There is a lot more research to be conducted in the future in various areas to ensure that this reputation-based system can solve all the problems in the real world.

7. FUTURE WORK

The most crucial area of research to be performed soon is on expanding the scalability of this system. This paper has so far operated under the assumption that the given Blockchain has a fixed number of nodes(*participants*). We are yet to explore the possibilities of a massive, network-wide deployment of this system and its issues.

We cannot yet answer questions such as whether a validator who acts honestly on one network can be assumed to act honestly on all networks they interact with, or when does a past reputation for a validator become irrelevant. However, with more research, we are hopeful of resolving these questions and more besides.

8. REFERENCES

[1] ScienceDaily. Big data, for better or worse: 90% of the world's data

[2] K Schwab, A Marcus, JO Oyola, W Hoffman, and M Luzi. Personal data: The emergence of a new asset class. In An Initiative of the World. Economic Forum, 2011.

[3] Cynthia Dwork. Differential privacy. In Automata, languages and programming, pages 1–12. Springer, 2006.

[4] Jon Evans. Bitcoin 2.0: Sidechains and ethereum and zerocash, oh my!, 2014.

[5] Staples, Mark; Chen, Shiping; Falamaki, Sara; Ponomarev, Alex; Rimba, Paul; Tran, An Binh; Weber, Ingo; Xu, Sherry; Zhu, John; "Risks and opportunities for systems using blockchain and smart contracts"; Sydney, Australia: CSIRO; 2017

[6] Dr. Gavin Wood, "Ethereum: A Secure Decentralized Generalised Transaction Ledger", Byzantium Version e94ebda , 2018

[7] Arasa R, Ottichilo L (2015) Determinants of know your customer (KYC) compliance among commercial banks in kenya. J Econ Behav Stud 7(2):162–175

[8] J. Poon and T. Dryja. (2016) The bitcoin lightning network: Scalable off-chain instant payments. Accessed 26, January 2018.

[9] I. Eyal, A. E. Gencer, E. G. Sirer, and R. Van Renesse, "Bitcoin-ng: A scalable blockchain protocol," in Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation, ser. NSDI'16. USENIX Association, 2016, pp. 45–59. [Online].

[10] Available: http://dl.acm.org/citation.cfm?id=2930611.2930615

[11] Guy Zyskind, Oz Nathan, Alex' Sandy' Pentland. Decentralizing Privacy: Using Blockchain to protect Personal Data

[12] Dr. John Callahan CTO at Veridum . The Forbes Web.22October,2018.https://www.forbes.com/sites/forbestechco uncil/2018/07/10/know-your-customer-KYC-will-be-a-great-t hing-when-it-works/

[13] Wazen M. Shbair, Mathis Steichen, "Blockchain Orchestration and Experimentation Framework: A case study of KYC," University of Luxembourg, 2018

[14] Jose Parra Moyano, Omri Ross. KYC Optimization using Distributed Generalised Transaction Ledger", Byzantium Version e94ebda [Online]. 2018

