

Monetization of Private Data Using Blockchain Ensuring Control and Ownership of the User Data

Prabhu D¹, Ragavendiran B¹, Karthik Balaji B¹

¹Student, Department of Computer Science and Engineering, Velammal Engineering College, Chennai, Tamilnadu, India.

Corresponding Author: bgragavendiran@gmail.com

Abstract: - In the digital era, data is one of the most valuable assets and is worth more than any other commodity that exists, Data is an essential resource that powers the information economy, and the data with such value is being siphoned from the users without their knowledge and is being sold in data markets. We are providing a way to give power and control to users by which they can sell and have control over their data. We are developing A web extension that collects users DNS propagation data for each site they visit and streams it to the data unions in real-time in exchange for cryptocurrencies like data coins, Ether, bitcoin, etc., without compromising the user's real identity leveraging the block chain, Ethereum framework, and Streamr data marketplace. The DNS propagation data is different for each ISP as they hold different SRV records, A data map of interlinked DNS propagation data could be used by data analytics companies for a lot of reasons.

Key Words— *Blockchain, DApps, ERC20, Streamr, Crypto, Data markets.*

I. INTRODUCTION

Data is power and it is one of highest valued assets. Data from a lot of reputable and reliable sources are bought and sold in data marketplaces Online, but general population is unable to monetize their private data without compromising their original identities in the web. Analytics companies collect and analyze user's data in their own servers which is also prone to data breaches. Currently data analytics companies collect and sell users data through their own applications and hoard it on their servers which are prone to data breaches despite the high level of encryption security and authentication parameters.

Data breaches affecting millions of users have become far too common. There has been several incidents due to which private information of millions of users consisting of details from their credit card numbers and passwords to their social security/Identity numbers have been leaked.

Blockchain is a system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system. A

blockchain is essentially a digital ledger of transactions that is duplicated and distributed across the entire network of computer systems on the blockchain. Each block in the chain contains several transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. The decentralized database managed by multiple participants is known as Distributed Ledger Technology (DLT). Blockchain is a type of DLT in which transactions are recorded with an immutable cryptographic signature called a hash.

The new Decentralized Applications differ from their traditional counterparts in the sense that instead of a client-server architecture, a client-blockchain model is adopted. Clients issue requests on the app and blockchain nodes perform the computations, returning the result back to the client. Leveraging the blockchain and Ethereum network the data is no longer saved in a single server to be breached the data we collect is decentralized through a network of Hyper ledger's (Block systems) and DLT's with a regenerative 128bit long hash code generated for each record of the user thereby eliminating the human as well as the external factors which could decode or trace back the origin or misuse the data. So, we ensure the highest level of authentication while maintaining user's privacy and control over the data that they wish to sell.

Manuscript revised July 31, 2021; accepted August 01, 2021.

Date of publication August 03, 2021.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898

II. SYSTEM ARCHITECTURE

The System Architecture for the proposed system is divided into three segments, Frontend, Backend and Marketplace.

Frontend and Backend are the two most popular terms used in web development. These terms are very crucial for web development but are quite different from each other. Each side needs to communicate and operate effectively with the other as a single unit to improve the website's functionality.

A. Frontend

The part of a website that the user interacts with directly is termed the front end. It is also referred to as the 'client side' of the application. It includes everything that users experience directly: text colors and styles, images, graphs and tables, buttons, colors, and navigation menu. HTML, CSS, and JavaScript are the languages used for Front End development. The structure, design, behavior, and content of everything seen on browser screens when websites, web applications, or mobile apps are opened up, is implemented by front End developers.

The web extension installed in the user's browser is designed with HTML, CSS and JS along with CORS permission from an authenticated browser extension license acts as the frontend. The Frontend displays the amount of data streamed by the user to the data market and the equivalent crypto returned.

Metamask Wallet: Metamask wallet is the cryptocurrency wallet of the user through which the user could be uniquely identified, the funds would be directly deposited into the crypto wallet of the User.

B. Backend

Backend is the server-side of the website. It stores and arranges data, and also makes sure everything on the client-side of the website works fine. It is the part of the website that you cannot see and interact with. It is the portion of software that does not come in direct contact with the users. The parts and characteristics developed by backend designers are indirectly accessed by users through a front-end application.

Node.Js and Express JS are hosted in a live remote VPS in sync with the transactions made on the data market. The Backend part ensures authorization token for each data that is to be broadcasted into the data markets infrastructure. The authorization token generated for a data transaction remains universally unique through which the user has access to the record. In addition to generating the authorization token the

backend also tracks and synchronizes the equivalent crypto transferred to the cryptocurrency wallet of the User.

C. Data Marketplace

Streamr is an open-source, crowdfunded decentralized platform for real time. It's a PUBSUB model in which it transports streams of messages from data publishers to subscribers, appearing to the user as a global publish/subscribe messaging service. It is similar in functionality to message queues, enterprise messaging systems, MQTT brokers, Apache Kafka. The main feature is publishing messages to make them available to subscribers. Subscribing to messages to process them as they occur. Storing messages for later use. So now when the data is passed to the streamr community they will analyze the data received from us and after that, in return, the user will the appropriate Ethereum from them. Those data will be stored. Popular data will be subscribed by other users and the rate will be increased. Instead of a centralized party, the Network is run by its community of users, and it heavily relies on cryptography to remove the need for trust. This creates a permission less and neutral network for real-time data, which is highly suitable for use cases such as data sharing, data monetization, and the decentralized web, extending the capabilities of the underlying internet without compromising its open and decentralized nature. Also by adopting this extension the data which is supplied to the streamr are genuine since the user is logged in to the Ethereum account as well as the URL that is being fetched is from the actual browser but not the hardcoded generated URL's. So by this, the value for the data is increased on both sides. The 35 Currency is added to the account where the user is logged in, at the beginning of the module.



Fig.1. PUBSUB model

III. DATA MONETIZATION

We collect DNS propagation data of each site that user visits via their ISP and we create a location hash map of all the DNS name server availability record and publish it to the streamr marketplace, upon proper analytics and recording. The data we collect can lead to the following benefits:

- Reducing the webpage loading times by a significant factor of time by cross-linking different ISP communication channels with better access to the same DNS records.
- Can be to connect to a website even if the nearest DNS record match fails to get updated as multiple ISPs will have redundant access to a different active DNS record for the same website.

This data is decentralized and is held by each ISP which is updated and shared with ICANN on demand. With the advent of our web extension, we can collect the data in a decentralized fashion through which researchers and analytics companies can find a platform to create new interlinking pathways to reduce the existing latency margins of the internet to a significant extent.

IV. PROCESS FLOW

On installation, the web extension requests the user to connect his ethereum wallet to our web extension upon establishing a successful authentication and connectivity with the Metamask or the crypto wallet of the user, the Extension starts listening to the browsers web search index as and when a website is redirected. In the search bar of the web browser the website is indexed on a local cache and the website address is passed to the Node JS backend for parsing upon successful web scraping of the required data from the DNS propagation verification site the IP of the ISP used by the user and the DNS records that are available for that ISP is then formatted into a JSON format and then the data is published to the marketplace via the streamr API.

V. CONCLUSION

The development of a Blockchain-based real-time data monetization platform is explained in this report. This project demonstrates a way for users to control and safely monetize their data without compromising their private information. This project eliminates the role of a third-party data collector who monetizes and benefits by selling user's data. The web extension can be installed on any web browser that supports a cryptocurrency wallet, the project can be further developed into a standalone big data analytics platform powered by the blockchain infrastructure which ensures credibility and anonymity of the data source.

REFERENCES

- [1]. Hocht and T. J. Lampoltshammer, "Social implications of a data " market," in Proceedings of the Conference for E-Democracy and Open Government (CeDEM17). (P. Parycek and N. Edelmann, eds.), pp. 171– 175, Edition Donau-Universitt Krems, 2017.
- [2]. Hassani, Hossein & Huang, Xu & Silva, Emmanuel. (2018). BigCrypto: Big Data, Blockchain, and Cryptocurrency. 2.
- [3]. B.-P. Ivanschitz, T. J. Lampoltshammer, V. Mireles, A. Revenko, S. Schlarb, and L. Thurnay, "A Semantic Catalogue for the Data Market Austria," in Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMANTiCS 2018), September 10-13, 2018, Vienna, Austria, pp. 1–4, 2018.
- [4]. M. Chowdhury, A. Colman, A. Kabir, J. Han, and P. Sarda, "Blockchain as a notarization service for data sharing with the personal data store," in Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/ 12th IEEE International Conference on Big Data Science And Engineering Trust Com/Big Data SE) (n/a, ed.), pp. 1330–1335, 08 2018.
- [5]. The Framework of Consensus Equilibria for Mining-Pool Games in Blockchain Ecosystems, 30 Jun 2020.

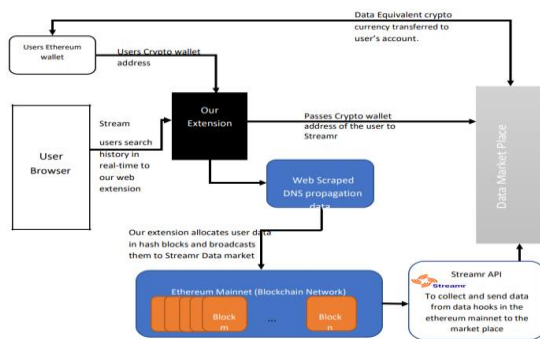


Fig.2. Process Flow Network