Data Cooperatives for Trusted News Sharing in Social Media

Abiola Salau(✉), Ram Dantu, Kritagya Upadhyay, and Syed Badruddoja

Department of Computer Science and Engineering, University of North Texas, Denton, TX 76207, USA
{abiolasalau,kritagyaupadhyay,syedbadruddoja}@my.unt.edu, ram.dantu@unt.edu

Abstract. The voluntary gathering and pooling of personal data by individuals via legal fiduciaries called data cooperatives is gaining a lot of attention as an approach to secure data management. Data cooperatives and blockchain are an excellent combination since they share fundamental features like decentralization and democratic design. In this paper, we leverage the power of blockchain to design a trusted news-sharing system for social media. We prove our concept by implementing a consumer news coop network on the Ethereum blockchain where members can voluntarily pool news information about their neighborhood for their benefit and also receive incentives in the form of an improved reputation for sharing credible news stories. We enforce honest behavior among the participants by implementing a trust and reputation scheme based on EigenTrust. Our results show that the blockchain approach to implementing a data cooperative is efficient with respect to memory consumption, scalability, and cost while also providing improved trust among participants. Furthermore, the reputation mechanism is effective in ensuring that malicious participants are severely penalized and removed from the system, while honest participants are rewarded. This approach can be used in a much bigger setup like Twitter so that the credibility of a shared post can be verified by a consensus before being shared on the network, thereby mitigating the spread of misinformation.

Keywords: Blockchain · Distributed Ledger Technology · Data cooperative · Reputation System · Social media · News sharing

1 Introduction and Problem Motivation

1.1 Data Cooperatives

Data cooperatives "refer to the voluntary collaborative pooling by individuals of their personal data for the benefit of the membership of a group or community" [1]. The data cooperative serves as a fiduciary for the data subjects, mediating between them and the data companies to help them negotiate the

W. Meng and W. Li (Eds.): BlockTEA 2022, LNICST 498, pp. 106–127, 2023.
https://doi.org/10.1007/978-3-031-31420-9_7

control and use of their personal data. The success of a data cooperative relies so much on the credibility of the kind of data that is pooled and how trustworthy the participants are in sharing credible data or information [1]. The kinds of resources that are pooled in a data co-op may be of varying types. For example, a community could come together to start a co-op where they pool personal health data, community news information, etc. [2]. Human beings are naturally inclined to attach some level of credibility to an individual based on how they perceive the reputation of that person, either through direct interactions or based on others' recommendations. In an online community such as a data co-op where members may not know each other enough to have some form of a trust relationship, a mechanism must be adopted to enable users to confidently interact or transact on the platform without fear of distrust [3]. *In a data cooperative platform, trust among the participants is necessary for the overall functioning of the co-op* [4,5]. *However, establishing and maintaining effective trust without a central control inherent in traditional systems where members may choose to be untrustworthy or malicious for their benefit is of concern—this is our focal point in this paper.* Specifically, but without loss of generality, our focus will be on news-sharing data co-ops where the members of the network can pool news about happenings in their community. This kind of system can help to create awareness about the events and security situations that exist in a community.

1.2 Fake News

The dissemination of news information used to be the sole responsibility of traditional media houses like TV, radio, and newspapers since they have the resources and are trained in fact-finding on a subject before broadcasting a news story [6]. However, since the advent of social media networks, every user now has the platform to share news posts and other information so that all their followers can read them in real-time. This real-time sharing and access to news information can be important for a thriving community. For instance, if people in a community come together to voluntarily share information about happenings in their respective neighborhoods. This gives them more knowledge about their community in a way that can help improve growth and innovation in the community, as shared data is the backbone of the knowledge economy [1]. The shared information or data can cover areas like health, weather forecasts, security, services, resources, and many more. Some benefits of this form of sharing may include: increased usage for an underutilized asset since a member can pool an asset currently underutilized and other members can benefit from it; a source of revenue for the data providers as the pooled data can be bought by service providers or researchers, and knowledge sharing among participating members.

Building on the concept of news information sharing on the social media networks and data cooperatives introduced in [1], we introduce our notion of a consumer news cooperative or co-op. A consumer news cooperative is a consumer-owned news-sharing platform where members of a community pool news information in their locality for the benefit of the members and the community at

large. Members of this consumer news co-op can be professional news organizations, fact-checking organizations, or even individual citizens of a community. Media professionals coming together to share news information and resources have been existing for a long time with the Associated Press[1] as an example of media cooperatives formed by five newspapers. The Banyan project[2] is another such co-op that aims at providing daily coverage of events with trust, relevance, and respect in members' communities. To succeed in this form of cooperative, [7] pointed out some questions that need to be answered: what returns are the members going to get for their participation in the co-op, and how do we ensure trust in the system?

In this paper, we use blockchain technology as a bedrock for building a data co-op and leverage the inherent properties of the technology to establish trust among the participating members and also incentivize them through improved reputation, ensuring everyone in the co-op benefits either directly from the news information or resources shared or in the form of incentives from transactions on the blockchain.

1.3 Why Blockchain for Data Cooperatives?

A major challenge in any online community such as a data co-op is establishing and maintaining trust among participating members and thus, indirectly encouraging cooperation. The data co-op is modeled such that risk and reward are shared among its members [1]. They are also decentralized in the sense that there is no single member in a position of control or above the others. This is a key attribute of a democratic setting, and it is fundamental to the modeling of a co-op [8]. Coincidentally, this decentralized notion of power and democratic control is also a core attribute of the blockchain. The benefits of applying blockchain technology to this are multifold.

- It can provide a trusted mechanism for operational activities such as decision-making and record-keeping without the need for physical proximity in a secure and immutable manner, which is important in a news-sharing platform.
- The immutable property of the blockchain will provide a means to prove the provenance of a pooled data source and other transactions on the system if the need arises.
- Being inherently distributed, it aligns with the concept of a cooperative since the members may not necessarily be in proximity and each member will still have the same view of the distributed ledger.
- Blockchain, using smart contracts will help to enforce the laws and regulations binding the data cooperative and will also help to ensure that punishments and rewards mechanisms are automatically handled [9,10].

To implement a blockchain-based data cooperative, the design has to correlate to blockchain systems such that it will have a ledger that will hold the detail

[1] Associated Press, ap.org/en-us/.
[2] banyanproject.coop/.

of transactions, a consensus mechanism to agree on the pooled data (depending on the goal of the co-op), how new members can join and the requirements and a mechanism to build and maintain trust in the system and also a way of incentivizing participants to encourage them to keep being on the network [11]. We have addressed these fundamental elements in this work.

1.4 Our Contributions

The main contributions of this paper are summarized as follows:

- We prove our concept by implementing a consumer news sharing prototype using blockchain as the underlying technology for a news pooling data co-op and our results show that the system is effective, scalable, and secure.
- Then, we integrate a reputation scheme to track malicious members and dishonest consensus nodes as well as incentivize honest behavior on the system which is a gap in related papers that we studied.
- We propose the use of a modified Eigentrust model which takes the transaction history of a user into account in the reputation computation for the reputation computation.
- Lastly, we evaluate our system experimentally, discuss its performance in terms of efficiency, scalability, and security and present defense mechanisms against potential attack.

1.5 Outline

The rest of the paper is organized as follows: Sect. 2 describes the existing literature and works relating to ours while also mentioning the differences between our work and these existing works. Section 3 describes the proposed system architecture and framework in detail. Section 4 talks about the experimental setup for the project and the implementation. In Sect. 5, results and discussion are provided, as also the security analysis. We conclude the paper in Sect. 6 by discussing future works and summarizing the key concepts and ideas of this paper.

2 Background and Related Work

The idea of data providers getting better control over their data is gradually developing due to the declining level of trust between users and service providers [12]. With the volume of data generated daily by users of the internet and especially the social media networks, individuals are increasingly concerned about the privacy of their data and how the social network platforms handle their data [12,13]. [7] describes a data cooperative as a community where individuals come together to pool their personal data for the benefit of its members. The authors further discussed the advantages communities can get from such a collaboration, among which is the economic growth a community can derive from having access to its data and being able to analyze it.

There are existing forms of data cooperatives [8] with the goal of individuals collaborating for the common good of the group while ensuring the members have control of their personal data. One such platform is MIDATA[3] where account holders can actively contribute to medical research by granting controlled access to their personal data. HAT[4] is a micro-server platform that gives the right of personal data to individuals through ownership of their personal data. Enigma [14] which is a peer-to-peer platform where different parties can jointly run computations on their data while keeping the data private. A common concept behind these platforms is to ensure that the data owners have absolute control over their data and can grant control-based access to whomever they want to.

In the world of news information broadcast, the Associated Press and the Banyan Project are examples of media cooperatives where news media organizations can come together to pool their resources for better news coverage or facility sharing. The media cooperatives report news based on the interests of their members and their geographical location[5] The application of blockchain as a technology for "sharing" has gained much attention with copious research work in literature. For a comprehensive review on the blockchain and trust in a sharing economy, we refer readers to paper [15]. However, we highlight some of these existing works that are similar to ours. Worthy of mention are the works of [6,16–19] where the authors propose the use of blockchain in sharing and analyzing news, while some others propose a hybrid of blockchain and ML or DL to tackle fake news [20–23] but our paper has a fundamental difference from these papers in the implementation of a reputation scheme based on Eigentrust [24] and a penalty system, which are missing from these previous works. Although our work leverages the concept introduced in [16], we have addressed some of the gaps identified in the paper.

In this paper, we present a data cooperative blockchain framework that can help to address the challenge of declining trust among members of an online community. It proposes a novel implementation of a consumer news data co-op using blockchain as the underlying technology, with comparable results (see Sect. 5) It also develops a novel reputation tracking system for participants to encourage and ensure honesty in the community and a punishment system to penalize dishonest players, effectively removing malicious members from the community after a limited number of rounds.

3 System Architecture

A high-level overview of the framework architecture is depicted in Fig. 1. Community members can interact with the blockchain using a news-sharing distributed application (DApp) on their client device. The news shared by a user is sent as a transaction to the blockchain through a smart contract implemented to manipulate the blockchain. This interaction is achieved through the Python web3 API.

[3] midata.coop/en/home/.
[4] Hubofallthings.com/main/what-is-the-hat.
[5] Cooperatives of the Americas - http://www.aciamericas.coop/Who-we-are.

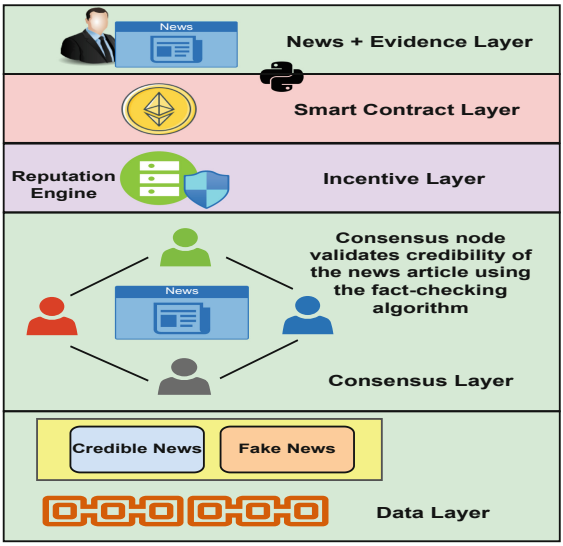

Fig. 1. Overview of the Blockchain-based Consumer News Co-op Framework. We have a client-accessible news-sharing DApp at the topmost application layer, which interacts with the blockchain through a smart contract on the second layer. The incentive layer includes a reputation engine that computes the reputation values of each validator that participates in the consensus process. The transaction data is stored on the blockchain ledger in the data layer.

In the smart contracts layer, the smart contracts have different functionalities based on the option selected by the user. An Eigentrust-based reputation engine in the incentive layer computes the trust level and reputation of the participants taking into account their historical behavior on the network.

The consensus algorithm layer can be any of several consensus algorithms like Proof of Stake [25], Proof of Reputation [26,27], Proof of Review (PoR) [28], etc., depending on implementation requirements. Note that if the consensus algorithm is reputation-based as in [26], the reputation engine may be omitted or embedded into the consensus layer if required.

3.1 Participants and Their Roles on the Blockchain

There are three types of nodes: news sharing nodes, verification nodes, and management nodes.

– ***News Sharing Nodes*** are registered members of the cooperative that post news content. Examples may include TV channels, radio, newspapers, or any citizen of the community, etc., as shown in Fig. 2. A news-sharing node shares

"newsworthy" information with the network. Each node will be able to communicate with the blockchain through an HTTP web client using a REST API to post a newsworthy article on the blockchain.

- ***Verification Nodes.*** In simple terms, the verification nodes, via an external oracle service, verify the transactions on the network and validate the credibility of the transactions by reviewing the underlying news information and its supporting evidence. If the news information is validated as credible, it is broadcast to all users of the co-op; otherwise, it is discarded.
- ***Management Nodes*** are responsible for the disposal of misinformation detected by the verification nodes. If this architecture is used in a social media network design on a consortium blockchain, social network providers like Facebook, and Twitter can be elected as management nodes since their only interest is in the design and successful running of a social media platform.

3.2 Smart Contract

To design the news co-op on a blockchain, we implemented two smart contracts using the Solidity language. A registration smart contract is used for user registration when joining the co-op, creating a digital identity profile with a UserID and digital signature for the new user, while another smart contract, called the Function smart contract, implements the logic for news sharing and management. The reputation value update is also done in this smart contract. In the Function smart contract, three key functions are implemented: a news sharing function, which creates the logic for the user to post news; a verification and management function, which is run by the verification nodes and management nodes; and a reputation function, which is the implementation of our modified Eigentrust algorithm [24,31] that computes the reputation of each of the nodes based on their transaction history and the credibility of the news posted by a news sharing node.

As depicted in Fig. 1, a sharing node can post news information along with IPFS[6] links to supporting evidence (such as verifiable media files) by running the news sharing function from its DApp. On receipt of this transaction, nodes interested and available online to verify the transaction, and run the verification and management algorithm on their DApp. This algorithm collates a set of userIDs comprising all the interested nodes, $u_i \in U$ and orders them according to their reputation, t_i on the system. It then randomly selects the userIDs with the top 20% reputation into a set of verification and management committees. From this set, one is randomly chosen as the management node while the others are verification nodes. The other users that were not selected for the committee will play no further part at this time until another news item has been posted.

The verification nodes review the transaction and, based on the evidence given, assign a flag of 0 for False or 1 for True to the transaction, sign it with their respective signature, and send it to the management node. The management node gathers all the signed transactions and checks for the majority flag. If a

[6] ipfs.io/.

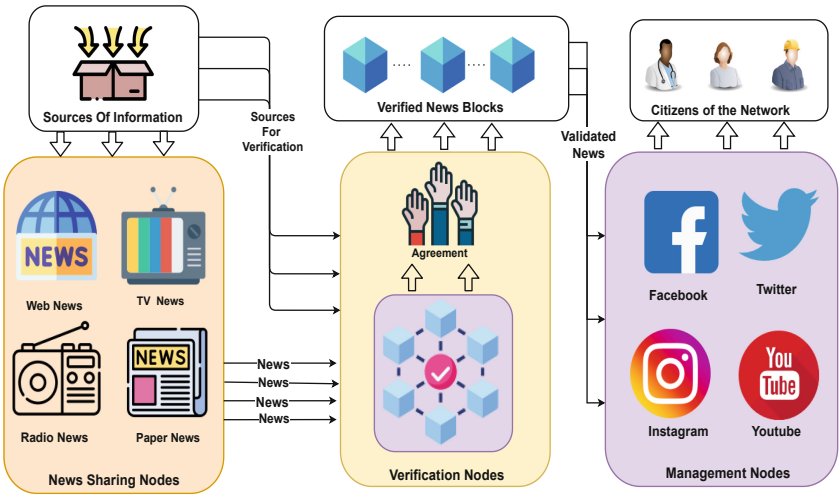

Fig. 2. Participants in the Blockchain. This figure shows a high-level overview of the news sharing ecosystem, with the news sharing nodes sharing a piece of newsworthy information along with supporting evidence, which is verified by the verification nodes via oracle services and broadcast to the citizens of the network or disposed of by the management nodes based on its credibility.

majority of the verification nodes assign a flag of 1, then the news is broadcast to the community and the transaction hash is added to the blockchain, but if the majority assigns a flag of 0, then the news information is tagged as false and not broadcast to the community. Lastly, if there is no majority, then the transaction is tagged as "undecided" and not added to the chain until the sharing node can re-share with more supporting evidence.

After the management node has completed its task, the reputation algorithm collects the userIDs of all the users that played a role in the last transaction, which includes the sharing node and the members of the verification and management committee, into a reputation set R, then computes and updates their respective reputations, as described in the reputation system subsection of this paper.

3.3 Event Flows

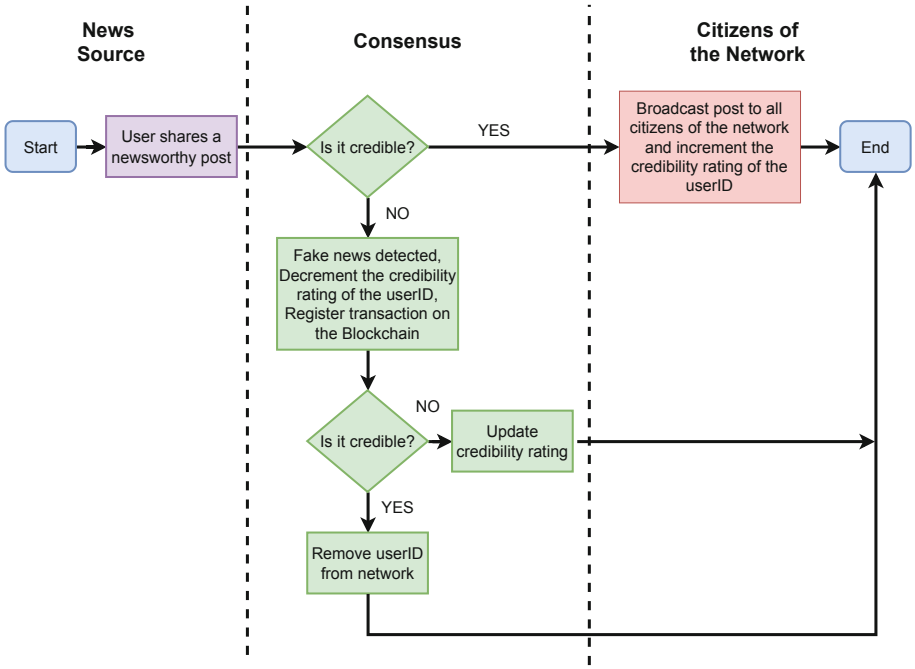

Fig. 3. Flow chart of Events. A user, through its client DApp shares a news transaction on the blockchain. The blockchain environment activates validators who validate the news credibility through an oracle that feeds the implemented smart contract. Depending on the credibility of the transaction based on the validator's outcome, the news is either broadcast and meta-data added to the ledger or discarded by the management nodes. The reputation values of validators are updated accordingly using Eq. 4.

Figure 3 shows the flow of events in the consumer news co-op. For example, if Bob shares a news post, it is sent as a transaction to the blockchain where the verification nodes will review the news information based on the supporting evidence and, if adjudged credible, the news information is broadcast to all users of the co-op and Bob's credit rating is incremented. However, if the news is not credible, the information is discarded and not broadcast to all users, and their reputation is updated with a decreased. We defer further discussion on the reputation computation to Sect. 3.4. Transactions meta-data will be stored on the blockchain ledger with details including the userID, timestamp, news source, etc. while the news article itself is stored off-chain for memory management. This would consume less space on the blockchain and also give the opportunity to delete news content that is illegal (e.g., a death threat). Note that in such a case, the meta-data remains on the blockchain as it is immutable. This feature

also helps to track any misinformation that may have been wrongly reported as credible. Real-time APIs will be able to query the reputation of a userID which can be used to notify users whenever a news post is shared from that userID. The reputation feature incentivizes the users to avoid posting false information.

3.4 Reputation System

To begin, we assume that the reader is familiar with the standard equations for trust computation in EigenTrust (see [24]). We modify the equations to address its limitations and the inherent vulnerabilities identified by Fan et al. in their work, Eigentrust++ [29] to make it more attack resilient. The Eigentrust approach has been well established in the literature [24,30–32], as an effective trust model in P2P systems due to its algorithmic approach to enhancing the overall system security by detecting malicious peers, making it easy to punish such behavior, and encouraging honest peers, thereby encouraging cooperative behavior and enforcing trust in the system.

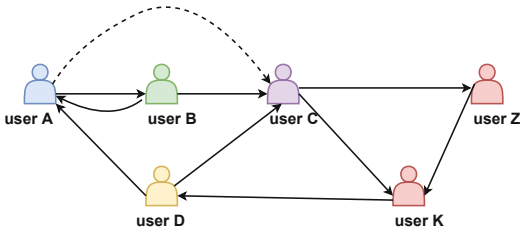

Fig. 4. Trust transitivity principle showing the relationship between users

Equation 4 shows the formulation of the global trust value (or reputation) by combining a direct trust value between a user and other users it had direct transactions with and a recommended trust value, which is a type of trust that is based on transitive trust [33]. As described [3,33] and illustrated in Fig. 4, a direct trust is what exists between user A and user B and it is derived using Eq. 1. If user A trusts user B, user B trusts user C, and user B recommends user C to user A, then user A can derive a measure of trust in user C based on the recommendation, and due to its direct trust in user B, this type of trust is derived trust. The derived trust is computed as shown in Eq. 3.

$$c_{i,j} = \begin{cases} \frac{max(s_{i,j},0)}{\sum_k max(s_{i,k},0)} & \text{if } \sum_k max(s_{i,k},0) \neq 0; \\ p_i, & \text{otherwise} \end{cases} \tag{1}$$

where p_i represents a case where peer i may be new and does not trust any other peer, it will have to choose the pre-trusted peers and its value is given as:

$$p_i = 1/|P|, \tag{2}$$

where P is the set of pre-trusted peers.

$$c_{i,k} = \sum_{j} c_{i,j} c_{j,k} \tag{3}$$

To address the possibility of malicious users colluding to assign arbitrary high trust values to each other and arbitrary low trust values to good peers in a distributed setting, a proliferation parameter a is used, thereby recalculating the current reputation of each peer as in Eq. 4, where a represents the probability of a peer having an interaction with any other, therefore relying on the pre-trusted peers. We also incorporate an Additive Increase and Multiplicative Decrease penalty component to it such that the reputation value of a malicious node is drastically reduced by half, resulting in its quick removal from the system when its reputation is below an acceptable threshold.

$$t_i = (1 - a)(c_{1,i} t_1 + c_{2,i} t_2 + ... + c_{n,i} t_n) + a p_i \tag{4}$$

where a is a constant ≤ 1.

In Eq. 1, $c_{i,j}$ refers to the direct trust between user i and j, $s_{i,j}$ is the ratio of the satisfactory transactions to the total transactions between users i and j, computed as

$$s_{i,j} = \begin{cases} \frac{sat_{i,j}}{sat_{i,j} + unsat_{i,j}} & \text{if } sat_{i,j} + unsat_{i,j} > 0; \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

In Eq. 3, $c_{i,k}$ is the trust derived between users i and k because of the transitive trust between user i and other users that have direct trust with user k.

In Eq. 4, t_i is the global trust, also known as reputation of user i as perceived by the network, and $c_{n,i}$ is the local trust between users n and i computed from Eqs. 1 and 2.

At the completion of a news sharing, verification, and management round, the reputation engine recomputes and updates the global trust of all the participants in that round. If the news information is flagged as false, then the news-sharing user is penalized with a reduction in its reputation. But if found credible by a minimum 51% majority, then its reputation rises. Likewise, for the verification nodes, the erring minority will have their reputation reduced to serve as a punishment for not adequately verifying the news for its authenticity before making a decision, while the majority will have their reputation raised for the honest work they performed. This will help to ensure that the nodes carefully verify the news and the supporting evidence before assigning a tag and thus discourage malicious behaviors from the nodes while also encouraging cooperation and trust-building for the success of the data co-op.

3.5 Incentive Structure

This work relies on a simple incentive structure in the form of improved reputation scores for well-behaved nodes in the system. This is based on the assumption

that humans do not require any monetary incentive to report the news happening in their locality. However, some approaches may be explored to enforce the incentives, such as the game-theoretic approaches studied in papers like [34–36] where the nodes can self-police themselves with a reward and penalty system such that if any participating node can verifiably report another malicious node, it gets a reward greater than the normal increment in its reputation for a round, while the dishonest node has its reputation either decreased by half or set to 0, depending on the game strategies. Another good approach is to have the reputation of the nodes tied to a monetary stake invested in the system so that they get to lose their money when they behave dishonestly, while honest nodes get to earn monetary rewards as well.

4 Experimental Setup and Implementation

Technologies software/tools that were used for this project include: i) Remix Web IDE ii) Truffle.js iii) Web3.js and Web3.py iv) Node.js v) Ganache vi) Ropsten test net vii) Solidity

The application will be developed on an Ethereum-based blockchain network; a smart contract will be written using Solidity, and a graphical user interface will be developed for interaction with users. The API services will be designed using a web3.js client in Node.js. Through this, the user can connect to the blockchain to post news transactions or request the reputations of users. Ganache was used as a local blockchain during the development stage of the smart contract. Performance results were measured on the Ropsten test net. The Metamask plug-in was used to interact with the blockchain for user accounts.

Data-set for the blockchain implementation was collected from Twitter through its developer API[7]. The dataset is described as a trending topic, e.g. #Covid19, #USElections2020, #RusyaUkrayna, etc., which involves several newsworthy tweets from various information sources. Possible news sources include newspapers, magazines, TV channels, radio stations, blogs, etc. A total of 5000 tweets were collected on the trending topic "#US Elections" between October 28, 2020, and November 4, 2020. Tweets that were not in English were removed, and duplicates were also removed. This reduced the total number of tweets in our test data set to 4153. For the initial set of users, we created multiple accounts on the Ropsten test net, starting with 50 accounts, and assigned random initial reputation scores with a minimum of the average reputation value to the nodes. This is based on the assumption that, as obtainable in reality, the developers of the system are less likely to want to destroy the system. We also started with 10 news-sharing users, each sharing a news message from the available dataset to the network, and in further iterations, we allowed new users to join and continue to increase and observe the system as we add new users and reputations being computed. In the next section, we present and discuss the results of our experiments.

[7] Twitter API - developer.twitter.com/en/docs/twitter-api/enterprise/search-api/overview.

5 Results and Discussion

The system testing was done on the Ropsten test net and performance data was generated on metrics that measure the relevance of blockchain as a platform for the pooling of news information on consumer news co-ops. Results from [37], show that evaluating the system on the Ropsten test net gives similar performance results when compared to the Ethereum main net only with increased transaction receipt times on the main net, which may be attributed to the volume of transactions in the pool at the time of making the transaction.

The transaction times on the Ropsten Test Net show that, on average, the time taken for a user to share a news post on the blockchain network is 14.24 s, with times ranging between 3.61 s and 43.25 s. This variation may be attributed to the size of the news information being shared and the time it takes to mine each transaction on the blockchain. This mining time is also dependent on the amount of gas offered by the sender of the transaction. This variation can also be attributed to traffic on the blockchain network [2]. However, the read time is close to real-time, which is important in a quick assessment of how credible the news source is.

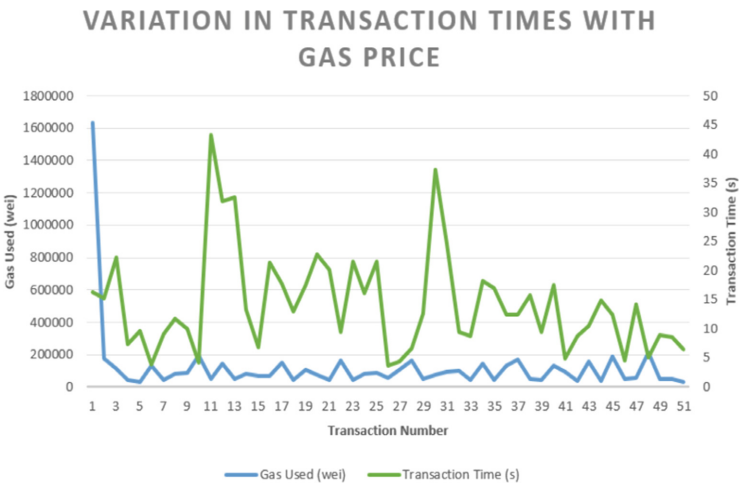

Fig. 5. Variation Between the Gas Used and Transaction Receipt Times for News Sharing Transactions. Transaction 1 with a gas used of 1630773 Wei is the gas consumed for the deployment of the smart contract. Observe that the gas used for the transactions with the exception of the deployment transaction is almost constant. This is because of the limitation on the number of characters acceptable per tweet on Twitter.

Figure 5 shows the variation between the gas used and transaction receipt times for each news-sharing transaction on the blockchain. A total of 50 tweets were used for the experiment and each of the tweets representing a piece of

news information to be shared is depicted on the horizontal axis as the transaction number while the left vertical axis represents the gas used in Wei and the right vertical axis shows the latency in sharing the news post on the blockchain. We notice that the time taken per transaction differs. The variation may be attributed to the time taken to validate and add the transaction to the blockchain.

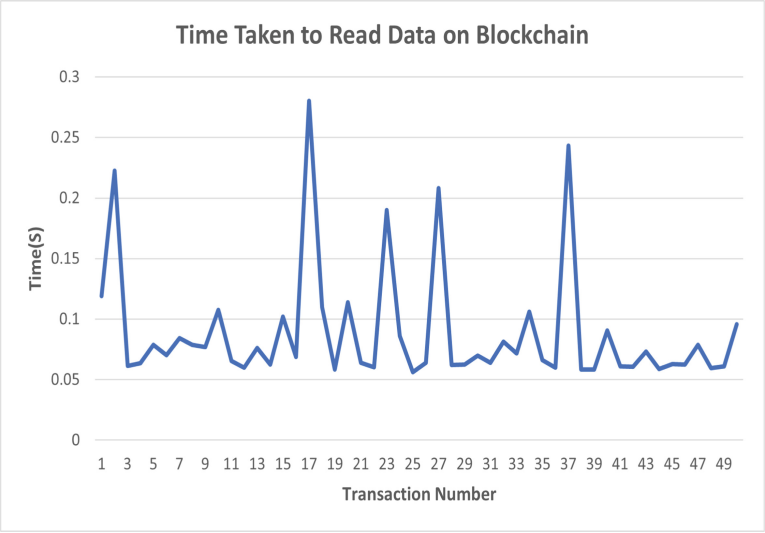

Fig. 6. Time Taken To Retrieve Reputations of Users On The Blockchain. An average of 0.09 s was used for reading the reputation of a user, with a minimum of 0.06 s and a maximum of 0.28 s.

A total of 1630773 Wei of gas was consumed for smart contract deployment. The news sharing function, which is one of the two write functions in the smart contract, consumed an average of 90707 Wei of gas, with consumption volume varying between a minimum of 33271 Wei and a maximum of 212781 Wei. The other write function, which is the user registration function, consumed an average of 20724 Wei of gas, with consumption volume varying between a minimum of 10844 Wei and a maximum of 47125 Wei. The amount of gas used for the deployment of the smart contract on the blockchain is dependent on the logic in the functions of the smart contract, i.e., the more complex the function logic, the higher the gas consumed. Another function in the smart contract is the one that reads the reputation of a user. This function only reads from the blockchain, and reading from the blockchain does not consume gas.

Figure 6 shows a plot of the time taken to read the reputations of users on the blockchain. For every piece of news shared by a user, its reputation is computed based on the outcome of the validation of the news information by the verification nodes. A user can query the system for its reputation score. We simulate this by

querying the reputation scores of different users and the time taken to get the ratings was plotted as depicted in Fig. 6. We may observe from the points on the plot that the credit ratings were retrieved in close to real-time, unlike the time taken in the news sharing plot of Fig. 5. This difference is because the function that generated the plot of Fig. 6 is a read function, and it does not require any transaction cost, so it can read from the blockchain in about real-time.

The graph of Fig. 7 shows that the reputation engine severely penalizes any user who shares malicious information on the network and encourages the sharing of credible information by a steady rise in the user's reputation. The approach is a modified Eigentrust algorithm that has an *Additive Increase and Multiplicative Decrease* (AIMD) reward strategy, making the reputations of malicious users decrease swiftly to 0 and getting them removed from the data co-op. The figure shows the comparison between the reputation calculations and penalties between a completely credible community member and a partially credible member.

This discussion is based on the memory consumption of the ledger as it is an integral part of a blockchain-based application. The ledger size increases with an increasing amount of newsworthy information shared on the blockchain. The data set used for this study is tweets from Twitter users, and this has a maximum permissible number of characters of 280. With this, the size of the news information shared is only a maximum of 350 bytes since only the user ID of an integer data type and a username of no more than 64 characters are

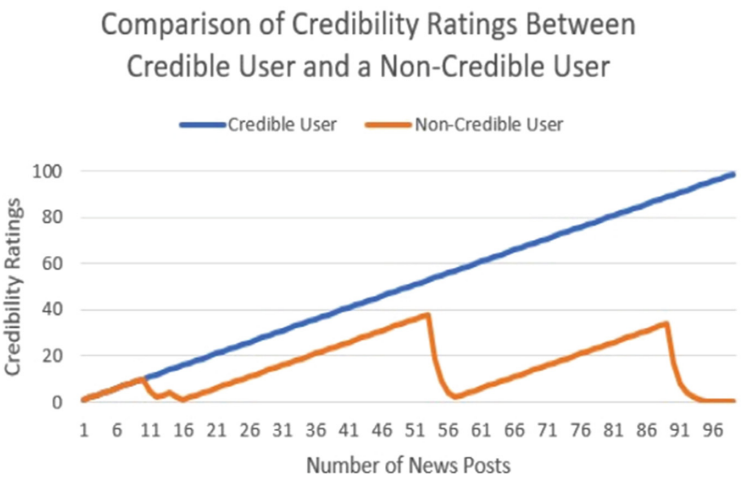

Fig. 7. Comparison Between The Reputation Calculations And Penalties Between A Fully Honest Member And A Partially Malicious Member. We can observe that the reputation of the honest user steadily increases while that of the malicious user drastically reduces after being severely punished by the 11th, 55th, and 91st rounds. Notice also that the reputation value of malicious users decreased to zero after the 93rd round and would be effectively removed from the system.

stored in addition to the news text. The blockchain also maintains data about the reputation of users.

Storing a user profile and its credit rating on the blockchain only consumes about 97 bytes per user considering the data types of the variables used. After the addition of 10,000 used, the ledger size will be 0.97 Mb, which is very small. [38] reports that, as of May 2020, there is an average of 6000 tweets per second on the Twitter social media platform. It is also reported that this has been the average number of tweets since the year 2014. According to Ethereum Transaction growth chart[8] the highest number of the 1,406,016 transactions occurred on Thursday, September 17, 2020, achieving a transaction rate of 16.27 Tx/s. This is very small compared to the number of average tweets per second on the Twitter social media platform.

5.1 Security Analysis

In this section, we will describe the defense of our model against potential threats and attacks relating to similar systems. We argue that, provided the reputation computation engine is accurate, it reflects the probabilities that the selected consensus nodes (the verification and management nodes) are not corrupted [27]. Thus, the verification and management node selection processes will always select an honest majority committee, and the outcome of the consensus will be added to the blockchain, proving the blockchain's safety.

To show the guarantee of an honest majority-of-reputation in the committee, we follow the approach presented in [39,40]. In similarity to [39] we employ Hoeffding inequality [41] rather than Chernoff bound in this context since the set of reputation values for the nodes is not equally distributed. In essence, given a set of reputation vectors randomly selected, it allows us to provide an upper bound on the probability that the sum of reputation vectors deviates from its expected value by more than a certain constant ϵ. Concretely, given a family $Rep = \{r^{n(\kappa)}\}_{\kappa \in N}$ and a set of stakeholders $n = n(\kappa)$ participating in a certain round of consensus, for all sufficiently large k's, using Hoeffding inequality, the average of the reputations is expected to be greater than $\frac{1}{2} + \omega(\sqrt{\frac{\log n}{n}})$, or that the expected number of honest parties should be greater than $\frac{n}{2} + \omega(\sqrt{n \log n})$. κ is a system security parameter that determines the difficulty level of selection.

Lemma 1 ***(The Hoeffding Inequality)*** [41]**.** *Let $X_1, \cdots, X_m$ be m independent random variables, each ranging over the (real) interval [0, 1], and let $\mu = \frac{1}{Q} \cdot E[\sum_{i=1}^{Q} X_i]$ denote the expected value of the mean of these variables. Then, for every $\epsilon > 0$, $Pr\left[\left|\frac{\sum_{i=1}^{Q} X_i}{Q} - \mu\right| \geq \epsilon\right] \leq 2e^{-2\epsilon^2 m}$.*

Theorem 1 ***(Honest Majority)*** [39]**.** *With overwhelming probability in the security parameter κ for some constant $\Delta > 0$, adversary $\mathcal{A}$ controls at most an $1/2 - \Delta$ fraction of the reputation of stakeholders in P_{sel}. Where P_{sel} is the set of nodes selected into the formed committee.*

[8] Etherscan - https://etherscan.io/chart/tx.

Assumptions:

1. *Rep* $= \{\mathbf{r}^{n(\kappa)}\}_{\kappa \in N}$ and a polynomial $n = n(\kappa)$, for all sufficiently large κ, the average of the reputations is greater than: $\frac{1}{2} + \omega(\sqrt{\frac{\log n}{n}})$, or equivalently, that the expected number of honest parties is greater than: $\frac{n}{2} + \omega(\sqrt{n \log n})$
2. Reputation vector $\mathbf{r} = (r_1, \cdots, r_n)$. Let $\boldsymbol{I} \leftarrow r$ be subset of $\boldsymbol{I} \subseteq [n]$ with $i \in \boldsymbol{I}$ chosen with probability $1 - r_i$ and the probabilistic choice of $\boldsymbol{I}$ is given to a distinguisher.

Claim:
With the assumptions stated above, if it holds that $\sum_{i=1}^{n} r_i > \lfloor \frac{n}{2} \rfloor + \omega(\sqrt{n \log n})$ then there exists a negligible function $\mu(\kappa)$ such that for every κ, $Pr\left[|\boldsymbol{I}| \geq \lfloor \frac{n}{2} \rfloor\right] \leq \mu(\kappa)$.

Proof:
Fix n and let $n = n(\kappa)$
For every $i \in [n]$, let X_i be a random variable that is 1 if P_i is honest and 0 otherwise.

$$Pr[X_i = 1] = r_i$$

Let $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$,
by linearity of expectations,

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^{n} r_i$$

Intuitively, we will have an honest majority when $|\boldsymbol{I} < n/2|$ (or when $\sum_{i=1}^{n} X_i \geq \lfloor \frac{n}{2} \rfloor + 1$).
Let $\Delta = (\sum_{i=1}^{n} r_i) - \lfloor \frac{n}{2} \rfloor = nE[\bar{X}] - \lfloor \frac{n}{2} \rfloor$.

Following Lemma 1,

$$Pr\left[\sum_{i=1}^{n} X_i \leq \lfloor \frac{n}{2} \rfloor\right] = Pr\left[\sum_{i=1}^{n} X_i - nE[\bar{X}] \leq \lfloor \frac{n}{2} \rfloor - nE[\bar{X}]\right]$$

$$= Pr\left[\sum_{i=1}^{n} X_i - nE[\bar{X}] \leq -\Delta\right] = Pr\left[\sum_{i=1}^{n} X_i - nE[\bar{X}] \leq -n\frac{\Delta}{n}\right]$$

$$= Pr\left[\frac{\sum_{i=1}^{n} X_i}{n} - E[\bar{X}] \leq -\frac{\Delta}{n}\right] \leq 2e^{-\frac{2\Delta^2}{n}}$$

By the assumption in the claim, $\Delta = \omega(\sqrt{n \log n})$, we have $\frac{\Delta^2}{n} = \omega(\log n)$ Hence,

$$Pr\left[|\boldsymbol{I}| \geq \lfloor \frac{n}{2} \rfloor\right] = Pr\left[\sum_{i=1}^{n} X_i \leq \lfloor \frac{n}{2} \rfloor\right] \leq 2e^{-\frac{2\Delta^2}{n}} < 2e^{-\omega(\log n)} \tag{6}$$

which is negligible in n. Thus, it holds that $e^{-\omega(\log n(k))}$ is a function that is negligible in κ as required. □

Figure 8 visualizes Eq. 6. It shows the probability of an attacker controlling the majority of reputation with changing committee size. It is worthy of note that as the committee size increases, the attacker's chances drastically reduce.

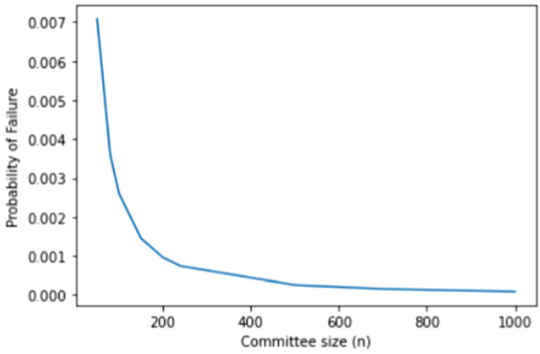

Fig. 8. Probability of an adversary having a majority in a sampling committee of size n

Next, we address the possibility of a Sybil attack on our system. In our proposed blockchain-based solution, the reputation of an individual plays a vital role in the successful operation of the data cooperative since it is a major criterion in the selection of nodes for system-specific tasks such as validating transactions, adding a new block to the chain, and so on. Thus, the reputation computation system is a major target for an attack.

Assumptions on the Reputation System

1. A reputation must be earned. It cannot be purchased, traded, or spent.
2. The reputation of a node is the aggregation of all reputation adjustments for the current round for that node.
3. Peers should not be responsible for directly computing their respective reputation values, and neither should they be able to modify these values.
4. Reputation values must be stored on the blockchain and be verifiable by all peers.

Sybil Attack: Saturation. In a Sybil attack with a saturation strategy, the attacker registers a large number of new users on the system with the aim of using its control over them to subvert the system [42]. The goal of the attacker is to either bring down the reputation of honest nodes or increase the reputation of a certain node that it controls, which has to improve its chances of getting selected for system tasks and then subvert the system.

Mitigation: This type of attack is designed to fail on our proposed system since the newly-joined nodes will have a reputation of zero and will require time and good behavior on the system to build their reputation before being chosen for system tasks. In addition, if a lot of nodes are joining the network at the same time, this would make it easily detectable and it could be a trigger for a transition to more stringent selection conditions.

Sybil Attack: Wait then Attack Strategy [42]**.** This is a more dangerous strategy than the previously discussed saturation strategy. Here, the attacker's nodes act honestly until they have a high reputation and then switch to a bad action when they have a majority in a selected consensus committee.

Mitigation: Due to the randomness in our committee selection protocol, the chances of the attacker nodes having a majority in the formed committee is negligible (see Theorem 1). In addition, with our multi-tier reputation protocol, even nodes with a high reputation might have to wait a long time before getting a chance to get selected since nodes with lower reputations are also given a fair chance of selection by the committee.

6 Conclusion and Future Work

In this work, we showed the development of a novel blockchain-based approach to the data cooperative concept using a consumer-owned news cooperative use case. Applying blockchain to data co-ops solves the challenge of incentivizing the news providers, as this was a major setback in the traditional system. Also, the addition of a reputation system ensures that the participants are honest since they are penalized when they post any misinformation. This will provide a platform for a transparent, multi-party system where all participating nodes can post newsworthy information and get rewarded with an improved credit rating.

For the blockchain to serve as the preferred platform for peer-to-peer news sharing platforms, it will have to support the number of news posts reported by [38] per second. Theoretically, with a block gas limit of 12,472,493[9] with the gas cost of around 21,000 for each transaction, we achieve approximately 594 transactions per block. With the current block time of 13.08 s[10], Ethereum can theoretically support 45.69 TXN/s. This number is obviously below the requirement for real-time support for news sharing. We will explore other blockchain platforms and consensus algorithms like proof-of-stake, where validators can be

[9] etherscan.io/chart/gaslimit.
[10] ethstats.net/.

selected based on stakes or permission from blockchain platforms depending on production requirements. This will enable us to examine how well the blockchain platform will scale with an increasing number of participants.

In future work, we aim to explore the application of game theory in the design of the consensus mechanism as well as the reward and penalty system. Some of the existing work in this area include [34–36]. We also plan to implement access controls in the system since it is an integral aspect of any data cooperative. A user should be able to control who gets access to their data, for what duration, and also be able to revoke this granted access. We will also further evaluate the system against threats and attacks that could result in the possibility of forking the blockchain. The work of Kleinrock et al. [27], showed how a consensus committee can be selected using a fair lottery tier-based approach such that new nodes whose reputation is naturally low would have a fair chance of being selected into the consensus committee while simultaneously ensuring that older yet highly reputable nodes have a higher probability of being chosen than the new nodes. This approach can be useful in our work since we only consider the top 20% reputable nodes and have no chance of newly joined nodes.

References

1. Pentland, A., Hardjono, T.: 2. Data Cooperatives in Building the New Economy (2020)
2. Salau, A., Dantu, R., Upadhyay, A.: Data cooperatives for neighborhood watch. In: 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 1–9 (2021). https://doi.org/10.1109/ICBC51069.2021.9461056
3. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. **43**(2), 618–644 (2007)
4. Data Co-Ops Workshop: Executive Summary of a December 22, 2019 Workshop Hosted at the Hebrew University of Jerusalem, The Federmann Cyber Security Research Center (2019). https://csrcl.huji.ac.il/sites/default/files/csrcl/files/data_co-ops_summary.pdf
5. Salau, A., Dantu, R., Morozov, K., Upadhyay, K., Badruddoja, S.: Multi-tier reputation for data cooperatives. In: Pardalos, P., Kotsireas, I., Guo, Y., Knottenbelt, W. (eds.) MARBLE 2022. LNOR, pp. 253–273. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-18679-0_14
6. Kim, B., Yoon, Y.: Journalism model based on blockchain with sharing space. Symmetry **11**(1), 19 (2019). https://doi.org/10.3390/sym11010019
7. Kaiser, J.E.G.: Media Cooperatives: Challenges and Opportunities (2019). https://medium.com/@jgksfconsulting/media-cooperatives-challenges-and-opportunities-e6803c0716ae. Accessed 30 Jan 2021
8. Pentland, A., Hardjono, T., Penn, J., Colclough, C., Ducharmee, B., Mandel, L.: Data cooperatives: digital empowerment of citizens and workers (2019)
9. Upadhyay, K., Dantu, R., He, Y., Salau, A., Badruddoja, S.: Paradigm shift from paper contracts to smart contracts. In: 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 261–268. IEEE (2021)

10. Upadhyay, K., Dantu, R., He, Y., Badruddoja, S., Salau, A.: Can't understand SLAs? Use the smart contract. In: 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 129–136. IEEE (2021)
11. Salau, A., Dantu, R., Morozov, K., Badruddoja, S., Upadhyay, K.: Making blockchain validators honest. In: The Fourth International Conference on Blockchain Computing and Applications (BCCA). IEEE (2022)
12. Madden, M.: Public perceptions of privacy and security in the post-snowden era (2014)
13. World Economic Forum: Rethinking Personal Data: A New Lens for Strengthening Trust (2014). http://reports.weforum.org/rethinkingpersonal-data
14. Zyskind, G., Nathan, O., Pentland, A.: Enigma: decentralized computation platform with guaranteed privacy. ArXiv abs/1506.03471 (2015)
15. Hawlitschek, F., Notheisen, B., Teubner, T.: The limits of trust-free systems: a literature review on blockchain technology and trust in the sharing economy. Electron. Commer. Res. Appl. **29**, 50–63 (2018)
16. Balouchestani, A. Mahdavi, M., Hallaj, Y., Javdani, D.: SANUB: a new method for sharing and analyzing news using blockchain. In: 2019 16th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC), pp. 139–143 (2019). https://doi.org/10.1109/ISCISC48546.2019.8985152
17. Saad, M., Ahmad, A., Mohaisen, A.: Fighting fake news propagation with blockchains. In: 2019 IEEE Conference on Communications and Network Security (CNS), pp. 1–4 (2019). https://doi.org/10.1109/CNS.2019.8802670
18. Islam, A., Kader, M.F., Islam, M.M., Shin, S.Y.: NEWSTRADCOIN: a blockchain based privacy preserving secure NEWS trading network. In: Patel, D., et al. (eds.) IC-BCT 2019. BT, pp. 21–32. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-4542-9_3
19. Paul, S., Joy, J.I., Sarker, S., Ahmed, S., Das, A.K.: Fake news detection in social media using blockchain. In: 2019 7th International Conference on Smart Computing & Communications (ICSCC) pp. 1–5. IEEE (2019)
20. Katal, A., Singh, J., Kundnani, Y.: Mitigating the effects of fake news using blockchain and machine learning. In: 2021 2nd International Conference for Emerging Technology (INCET), pp. 1–7 (2021)
21. Jaroucheh, Z., Alissa, M., Buchanan, W.: TRUSTD: combat fake content using blockchain and collective signature technologies. ArXiv, abs/2008.13632 (2020)
22. Agrawal, P., Anjana, P.S., Peri, S.: DeHiDe: deep learning-based hybrid model to detect fake news using blockchain. In: International Conference on Distributed Computing and Networking 2021 (ICDCN 2021), pp. 245–246. Association for Computing Machinery, New York (2021)
23. Torky, M., Nabil, E., Said, W.: Proof of credibility: a blockchain approach for detecting and blocking fake news in social networks. Int. J. Adv. Comput. Sci. Appl. **10**(12), 321–327 (2019)
24. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in P2P networks. In: Proceedings of the 12th International Conference on World Wide Web (2003)
25. Nguyen, C.T., Hoang, D.T., Nguyen, D.N., Niyato, D., Nguyen, H.T., Dutkiewicz, E.: Proof-of-stake consensus mechanisms for future blockchain networks: fundamentals, applications and opportunities. IEEE Access **7**, 85727–85745 (2019)

26. Gai, F., Wang, B., Deng, W., Peng, W.: Proof of reputation: a reputation-based consensus protocol for peer-to-peer network. In: Pei, J., Manolopoulos, Y., Sadiq, S., Li, J. (eds.) DASFAA 2018. LNCS, vol. 10828, pp. 666–681. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91458-9_41
27. Kleinrock, L., Ostrovsky, R., Zikas, V.: Proof-of-reputation blockchain with Nakamoto fallback. In: Bhargavan, K., Oswald, E., Prabhakaran, M. (eds.) INDOCRYPT 2020. LNCS, vol. 12578, pp. 16–38. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-65277-7_2
28. Zaccagni, Z., Dantu, R.: Proof of review (PoR): a new consensus protocol for deriving trustworthiness of reputation through reviews. Cryptology ePrint Archive, Report 2020/475 (2020)
29. Fan, X., Liu, L., Li, M., Su, Z.: EigenTrustp++: attack resilient trust management. In: 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), pp. 416–425 (2012)
30. Gao, S., Yu, T., Zhu, J., Cai, W.: T-PBFT: an EigenTrust-based practical Byzantine fault tolerance consensus algorithm. China Commun. **16**, 111–123 (2019)
31. Abrams, Z., McGrew, R., Plotkin, S.: A non-manipulable trust system based on EigenTrust. SIGecom Exch. **5**(4), 21–30 (2005)
32. Kurdi, H.A.: HonestPeer. J. King Saud Univ. Comput. Inf. Sci. **27**(3), 315–322 (2015)
33. Jøsang, A., Gray, E., Kinateder, M.: Simplification and analysis of transitive trust networks. Web Intelli. and Agent Sys. **4**, 139–161 (2006)
34. Nojoumian, M., Golchubian, A., Njilla, L., Kwiat, K., Kamhoua, C.: Incentivizing blockchain miners to avoid dishonest mining strategies by a reputation-based paradigm. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) SAI 2018. AISC, vol. 857, pp. 1118–1134. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01177-2_81
35. Samanta, A.K., Sarkar, B.B., Chaki, N.: Quantified analysis of security issues and its mitigation in blockchain using game theory. In: Dutta, P., Mandal, J.K., Mukhopadhyay, S. (eds.) CICBA 2021. CCIS, vol. 1406, pp. 3–19. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75529-4_1
36. Liu, Z., et al.: A survey on blockchain: a game theoretical perspective. IEEE Access **7**, 47615–47643 (2019). https://doi.org/10.1109/ACCESS.2019.2909924
37. Muttavarapu, A.S., Dantu, R., Thompson, M.: Distributed ledger for spammers' resume. In: 2019 IEEE Conference on Communications and Network (2019)
38. Sayce, D.: The number of tweets per day in 2020. Accessed 26 Nov 2020
39. Asharov, G., Lindell, Y., Zarosim, H.: Fair and efficient secure multiparty computation with reputation systems. In: Sako, K., Sarkar, P. (eds.) ASIACRYPT 2013. LNCS, vol. 8270, pp. 201–220. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-42045-0_11
40. Larangeira, M.: Reputation at stake! A trust layer over decentralized ledger for multiparty computation and reputation-fair lottery. Cryptology ePrint Archive (2021)
41. Hoeffding, W.: Probability inequalities for sums of bounded random variables. In: Fisher, N.I., Sen, P.K. (eds.) The collected works of Wassily Hoeffding. Springer Series in Statistics, pp. 409–426. Springer, New York (1994). https://doi.org/10.1007/978-1-4612-0865-5_26
42. Biryukov, A., Feher, D., Khovratovich, D.: Guru: universal reputation module for distributed consensus protocols. Cryptology ePrint Archive (2017)