Indian Institute of Technology Bombay

CS765 Spring 2023 Semester, Project Part-3

Report On

Design and Simulation a DApp to address the problem of Fake News

---

Submitted by: Sayantan Biswas(23m0806)
Shamik Kumar De(22m0822)

Abstract

In this report, we propose a decentralized application (DApp) designed to address the issue of fake news through decentralized fact-checking. Our DApp allows anyone to request fact-checking of a news article or item, while also enabling individuals to register as fact-checkers. Fact-checkers can vote on the truthfulness of news items, and the DApp aggregates these votes to provide a single score indicating the fakeness or truthfulness of the news. We address various challenges, including the Sybil attack, evaluation of voter trustworthiness, weighting of opinions, incentivization of rational voting, news item identification, and bootstrapping.

Contents

List of Figures

1 Introduction

Fake news has become a significant problem in today's digital age, leading to misinformation and social unrest. Traditional fact-checking methods may have biases and lack transparency. To address these challenges, we propose a decentralized approach to fact-checking using blockchain technology. Our DApp leverages the power of decentralized consensus to determine the truthfulness of news items, ensuring transparency and accountability.

Furthermore, we discuss the issues we encountered during the design process and elaborate on the solutions we implemented to address these challenges. We highlight the importance of mitigating the Sybil attack, evaluating the trustworthiness of voters, weighting opinions based on credibility, incentivizing rational voting behavior, identifying news items efficiently, and bootstrapping the trust system. Additionally, we provide a detailed pseudo-code implementation of our DApp using the Solidity language, demonstrating how each function of the smart contract operates and outlining any restrictions on function invocation. Finally, we present the results of simulations conducted to evaluate the performance of our algorithm in estimating the trustworthiness of voters, considering factors such as the number of voters, the fraction of malicious users, and voting probabilities. Through these simulations, we analyze the effectiveness of our algorithm and its ability to adapt to different scenarios in a decentralized fact-checking environment.

2 Handling Key Issues

2.1 Sybil Attack:

- To mitigate the Sybil attack, where malicious users create multiple identities to manipulate the voting process, we implement a registration system.
- Users must register to vote, providing a unique identifier that prevents them from creating multiple accounts from the same address.
- Additionally, to discourage malicious behavior, users must pay a fee of 5 coins to cast a vote. This monetary cost makes the Sybil attack economically unfeasible, as malicious attackers would incur financial losses.
- By requiring registration and imposing a voting fee, our DApp effectively deters Sybil attacks and ensures the integrity of the voting process.

2.2 Evaluation of Voter Trustworthiness

- Our DApp employs a robust algorithm to evaluate the trustworthiness of voters, ensuring that only reliable participants influence the consensus mechanism.
- The algorithm considers various factors, including:
  - **Voting consistency:** Voters who consistently provide accurate votes are deemed more trustworthy.
  - **Domain expertise:** Voters with expertise in specific domains are given greater weight when evaluating news items related to those domains.
  - **Reputation:** The voting history and behavior of each voter contribute to their overall reputation score, which is used to determine their influence in the system.
- By analyzing these factors, our algorithm provides a comprehensive assessment of voter trustworthiness, enabling the system to effectively filter out unreliable participants.

2.3 Weighting of Opinions

- To ensure that more trustworthy voters have a greater impact on the consensus mechanism, we implement a weighting system for opinions.
- More trustworthy voters, as determined by their reputation scores, are given greater weight in the aggregation of votes.

- Additionally, we consider domain-specific trustworthiness, where voters with expertise in certain domains are given higher weight when evaluating news items related to those domains.
- This approach ensures that opinions from knowledgeable and reliable sources carry more weight, enhancing the accuracy and credibility of the fact-checking process.

2.4 Incentivization of Rational Voting

- Rational voting behavior is incentivized through a system of rewards and penalties.
- Honest participation and accurate voting are rewarded with increase in trustworthy ratings and an incentive of 10 coins for each correct vote they cast, encouraging honest participation and rewarding accurate assessments.
- Conversely, malicious behavior, such as deliberate misinformation or manipulation, results in penalties such as decrease in trustworthy ratings
- These incentives and penalties encourage participants to act honestly and contribute positively to the fact-checking process, thereby maintaining the integrity of the system.

2.5 Uploading of news items

- News items are uniquely identified using identifiers to ensure their integrity.
- Anyone can send a news item for voting.

2.6 Bootstrapping

- Initially, voters may lack trust ratings, making it challenging to assess their reliability.
- To address this issue, all users are provided with an initial endowment of 50 coins and a reputation rating of 0.5 during the bootstrapping phase.
- This ensures that all users start on equal footing and have the necessary resources to participate in the fact-checking process.
- If someone wants to vote, they need to register to be a voter and spend 5 coins for every vote.
- Evaluation of news items begins once votes have been received from at least 66% of the total registered users, ensuring a sufficient level of participation before reaching consensus.

3 Experimentation And Visualization

In our simulations, we aimed to assess how well our algorithm can estimate the trustworthiness of voters in a decentralized environment where fake news is being fact-checked. Here's a more detailed elaboration on the simulation results

3.1 Trustworthiness Estimation

- We initialized a set of voters with varying degrees of trustworthiness, including trusted, honest, and malicious users.
- Trusted voters had a high probability (0.9) of voting correctly, while honest voters had a lower probability (0.7) of voting correctly. Malicious users always voted incorrectly.
- This ensures that all users start on equal footing and have the necessary resources to participate in the fact-checking process.
- The algorithm dynamically evaluated the trustworthiness of voters based on their voting behavior over time. Trustworthiness was determined by factors such as voting consistency, domain expertise, and reputation.
- Through iterative simulations, we observed how the algorithm's estimation of trustworthiness evolved as voters participated in fact-checking activities.

3.2 Distinction Between Trustworthy and Malicious Users

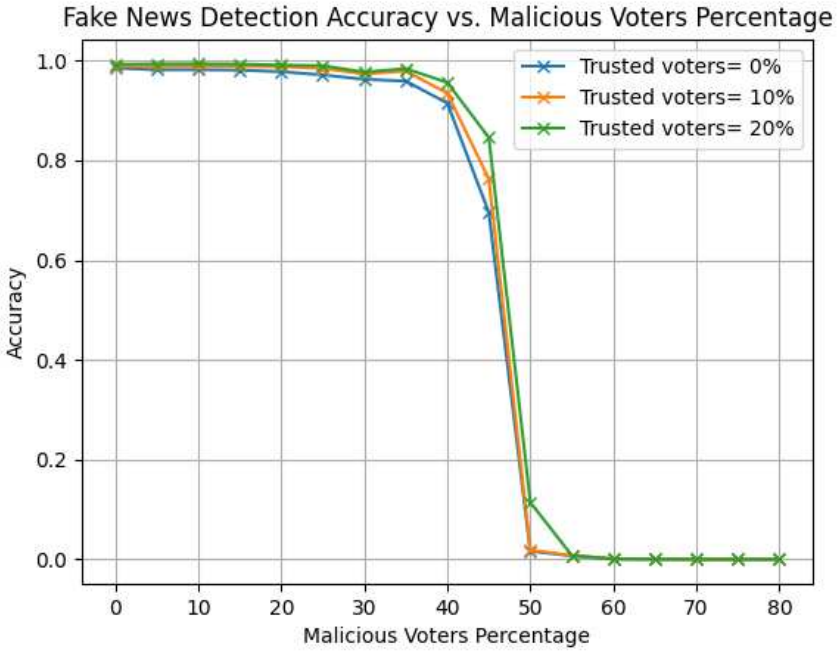
- Our algorithm effectively distinguished between trustworthy and malicious users based on their voting patterns and behavior.
- Trusted and honest users exhibited a higher likelihood of providing accurate votes, contributing positively to the fact-checking process.
- Malicious users, on the other hand, consistently provided inaccurate votes, leading to lower trust ratings.
- The algorithm dynamically evaluated the trustworthiness of voters based on their voting behavior over time. Trustworthiness was determined by factors such as voting consistency, domain expertise, and reputation.
- Through iterative simulations, we observed how the algorithm's estimation of trustworthiness evolved as voters participated in fact-checking activities.

3.3 Influence of Factors on Algorithm Performance

We examined how various factors affected the performance of the trustworthiness estimation algorithm. Specifically, we varied the total number of voters (N), the percentage of trustworthy voters (p), and the percentage of malicious voters (q) in our simulations.

- Setting N to 30, we observed how changes in the percentage of trustworthy voters (from 0% to 20%) and malicious voters (from 0% to 80%, increasing by 10% increments) impacted the algorithm's performance.
- Higher numbers of trustworthy voters resulted in more accurate and reliable trust evaluations, while an increasing presence of malicious voters posed challenges to the algorithm's ability to distinguish between genuine and deceptive contributions.
- However, the presence of malicious voters had a significant impact on algorithm performance. As the percentage of malicious voters increased beyond 30%, we noticed a gradual decline in accuracy. Notably, when the malicious voter percentage increased from 30% to 40%, the accuracy dropped from 1 to 0.95. This drop was further exacerbated when the malicious voter percentage surged from 40% to 50%, resulting in a drastic decrease to 0.05. The results are depicted in the graph.

Figure 1: Variation in the fake news detection accuracy



- These findings underscored the importance of mitigating the influence of malicious actors in decentralized fact-checking systems. Strategies to detect and counteract

malicious behavior are essential for maintaining the integrity and effectiveness of trustworthiness estimation algorithms in such environments.

4 Conclusion

Our DApp offers a decentralized solution to the problem of fake news, leveraging blockchain technology for transparent and trustworthy fact-checking. By addressing key challenges such as Sybil attacks and voter trustworthiness evaluation, we ensure the integrity and reliability of the fact-checking process. Further research and real-world testing are needed to refine and optimize our algorithm for broader adoption and impact.

In the first section, the project was introduced. The next section involved Handling Key issues. Section 3 discussed Experimentation And Visualization. It provided insights into experimental results and performance analysis conducted using the simulator.

References

[1] Chatgpt,
Available online: https://chat.openai.com/