Enhancing Web Spam Identification through Blockchain-Powered Crowdsourcing Incentives

1st Noah Kader
Rensselaer Polytechnic Institute)
Troy, New York
noahkader@gmail.com

2nd Inwon Kang
Rensselaer Polytechnic Institute)
Troy, New York
kangi@rpi.edu

3rd Oshani Seneviratne
Rensselaer Polytechnic Institute)
Troy, New York
senevo@rpi.rdu

***Abstract*—The proliferation of spam websites on the internet has necessitated the development of machine learning models to automate their detection. However, the dynamic nature of spam websites and the implementation of sophisticated evasion techniques by spammers often lead to low accuracy in these models. Crowdsourcing has emerged as a popular method for collecting and labeling large datasets to address this challenge. In this paper, we delve into the utilization of crowdsourcing to enhance model accuracy in identifying spam websites. We propose a novel incentive mechanism based on blockchain technology that rewards users for providing accurate labels while penalizing them for inaccurate ones. Our incentive mechanism is meticulously designed to stimulate high-quality contributions from users, motivating them to submit reliable data and preventing unscrupulous actors from benefiting from the system. Our research demonstrates that incentivizing crowdsourcing can effectively improve the data quality used to train machine-learning models while providing monetary incentives to contributors. Our findings indicate that the proposed incentive mechanism enhances the ratio of trustworthy participants while reducing the influence of malicious actors. Moreover, this approach showcases the potential for extension to other domains where machine learning models play a crucial role in decision-making processes.**

***Index Terms*—Information systems, Spam detection, Incentive schemes**

I. Introduction

Web spam broadly refers to deceptive or malicious techniques individuals or entities employ to manipulate search engine rankings, deceive users, or engage in illicit online activities. Web spammers aim to deceive search engines and users into perceiving their content as valuable or legitimate, even though it may be of low quality or carry malicious intent [1]. Web spammers often excessively use targeted keywords or phrases within the content or metadata of web pages to manipulate search engine rankings. This practice artificially inflates a web page's relevance for specific search queries, even if the content is not genuinely valuable [2]. Spammers may also hide keywords or links by using techniques such as setting the text color to match the background, positioning them offscreen, or using tiny font sizes. These hidden elements are intended to deceive search engines by including additional keywords [3]. There are also various tactics to manipulate the number and quality of inbound links to spam websites. These tactics include buying or exchanging links, participating in link farms or link exchange networks, and using automated programs to generate large volumes of low-quality or irrelevant backlinks [4], [5]. Spammers often copy or "scrape" content from legitimate websites and republish it on their sites. This technique aims to deceive search engines into considering the copied content as original, leading to undeserved search engine rankings and potentially stealing traffic from the original source [6]. Doorway pages, also known as bridge or gateway pages, are low-quality web pages optimized to rank well for specific search queries [7]. They often provide users with little or no valuable content but act as entry points to redirect visitors to other websites, including spam or malicious sites. Some web spammers employ techniques to distribute malware or engage in phishing activities. They may create web pages or deceptive advertisements that trick users into downloading malicious software, disclosing personal information, or engaging in fraudulent transactions. In the context of online advertising, web spammers engage in click fraud to generate illegitimate clicks on pay-per-click ads, which can be done using automated bots or by incentivizing individuals to click on ads, aiming to drain advertisers' budgets or artificially increase ad impressions and click-through rates [8].

Web spam techniques continue to evolve as search engines and security systems enhance their detection methods. Search engine algorithms and spam detection systems employ sophisticated techniques to identify and penalize web spam, striving to provide users with high-quality and trustworthy search results [1], [9]. The increasing prevalence of spam websites has created a pressing need for accurate web spam detection models. However, these models often encounter difficulties due to the dynamic nature of spam websites and the utilization of sophisticated evasion techniques by spammers. Consequently, there is a critical demand for innovative approaches that can enhance the effectiveness of web spam detection.

A. Contributions

We propose a mechanism that incentivizes users to contribute high-quality data, thereby improving the overall accuracy of our web spam detection models. Leveraging insights from previous research, we develop an improved incentive mechanism that aligns with users' needs for a robust and reliable dataset collaboratively constructed by the community. Our objective is to create an effective incentive mechanism that not only encourages users to provide accurate and valuable data but also discourages the submission of inaccurate or

malicious information by untrustworthy actors. By undertaking this research, we demonstrate the potential of incentivized crowdsourcing as a viable blockchain-based strategy for enhancing the accuracy of web spam detection models. The proposed incentive mechanism addresses the inherent limitations of existing models and offers a promising solution for leveraging crowdsourced data to combat the proliferation of spam websites. By integrating blockchain technology, we aim to foster transparency, trust, and accountability within the incentive mechanism, further strengthening the integrity of the collected data.

The contributions of this paper extend beyond the development of a novel incentive mechanism. We aim to showcase the efficacy of incentivized crowdsourcing in improving the data quality used to train web spam detection models while providing monetary incentives to the contributors. Additionally, our research explores the impact of the proposed mechanism on balancing the participation of trustworthy actors and mitigating the influence of malicious actors within the crowdsourcing ecosystem. Furthermore, we discuss the potential applicability of this approach to other domains where machine learning models play a pivotal role in critical decision-making processes. Through a rigorous examination and analysis of our proposed incentive mechanism, we seek to establish its efficacy and demonstrate its potential to revolutionize the field of web spam detection. By enhancing the accuracy and reliability of the spam detection models, we aim to contribute to the ongoing efforts to create a safer and more trustworthy online environment for users across various digital platforms.

II. Previous Works

Previous research has extensively explored the integration of blockchain technology within machine learning, specifically focusing on generic spam URL detection and blockchain-based crowdsourcing methodologies. Farooq et al. [10] presents a comprehensive survey of techniques commonly used by spam websites, serving as a valuable reference for our work. Additionally, several studies have examined the potential of blockchain in augmenting machine learning models for spam detection purposes. Our proposed approach represents an extension of these existing methodologies, harnessing the unique advantages of blockchain technology and crowdsourcing techniques to create a novel framework that maximizes the potential benefits of both domains, specifically focusing on incentivizing participants and enhancing data quality.

A. Spam URL detection

Spam detection on the web is a difficult problem, especially in a setting with limited resources. Past works have found efficient features that can be extracted from the source HTML of the domain URL of a website and have shown good performance. Jelodar et al. [11] propose a framework to learn and improve features using regex pattern matching systematically. The proposed approach is iterative, where the features are first encoded and used to train the model, which then finds better features while improving accuracy. While this approach demonstrates the effectiveness of keyword matching, its complexity may hinder its practicality. Introducing a blockchain-based incentive mechanism can incentivize participants to contribute their expertise in feature selection and improvement, resulting in a more streamlined and effective feature set.

Ntoulas et al. [6] discuss and experiment with various features extracted from web pages. These include repeated mentions of keywords and visible elements by styling them appropriately using Cascading Styling Sheets (CSS) or the number of anchor texts (text links to other pages). The authors showcase a high performance of over 90% precision accuracy in detecting spam with proposed features and a decision tree classifier. However, the performance of these features may vary in different contexts due to evolving spamming techniques.

However, by incorporating a blockchain-based incentive mechanism, we can tap into the collective intelligence of a community, allowing for continuous adaptation and improvement of feature sets as new spamming techniques emerge.

Markines et al. [12] define a specific spam problem called 'social spam' and define several features associated with it, such as styling features or the number of links on a page, and show their performance of around 97% in experiments. The authors introduce a feature called 'plagiarism' to mark how close website contents are, and they mention features like 'blurring' elements on the website (similar to the hiding from before) that can be a good feature. However, their focus on a specific type of spam limits the generalizability of their approach. By leveraging a blockchain-based incentive mechanism, we can engage a diverse group of contributors, encouraging the exploration of various spam types and improving the overall robustness of the model.

Mamun et al. [13] contribute a new dataset (ISCX-URL2017)[1] on spam data on the web. The authors also make use of network features to analyze this dataset. While the network features are irrelevant to our work, this dataset is much newer than the others that appear frequently, so we used the URLs available in this dataset to create our feature set. We build a low-computational spam filtering application using features from these past works. Furthermore, by integrating a blockchain-based incentive mechanism, we can incentivize participants to contribute additional datasets and feature insights, enabling a more comprehensive analysis encompassing network-based and content-based features.

Our mechanism stimulates collaboration, promotes continuous feature-set improvement, and encourages exploring various spam characteristics. By leveraging the collective intelligence and motivation of participants, our proposed approach holds the potential to overcome the challenges associated with limited resources and evolving spamming techniques, leading to more robust and effective spam detection models.

B. Blockchain Based Crowdsourcing for Spam Detecting

Harvey et al. [14] discuss the advantages of using blockchains, where the authors argue that blockchain tech-

[1] https://www.unb.ca/cic/datasets/url-2016.html

nology offers near-zero transaction cost, transparency, and permissioned identity sharing. These benefits can be applied to personalized spam detection, as users can share their unique identity with smart contracts and engage in data exchange at a minimal cost. The technology has several benefits, such as automation and connection to sophisticated machine-learning techniques through oracles [15]. However, improvements can be made to enhance scalability and efficiency, ensuring the system can handle many user interactions without compromising performance. We also noted that almost all the current work in this space is related to email spam detection rather than specifically to web spam detection. This section outlines existing work, its limitations, and how our methodology helps address them.

Sheikh et al. [16] propose a blockchain-based email system to discourage spammers. By adding a small fee to every email, spammers are disincentivized from sending large batches of emails due to the high total cost. While this approach provides a disincentive for mass spamming, scalability remains a concern, as each email triggers a transaction on the blockchain. Optimizations should be explored to minimize transactional overhead and ensure the system can handle a high volume of email transactions effectively to improve this method.

Choudhari et al. [17] propose a blockchain-based email system where transactions are attached to each email, and the recipient decides whether the email is spam or not, determining the refund of the transaction. If a wallet has been tagged as spam, the system will show the following emails from the address associated with the wallet as spam. While this approach can help identify spammers, it relies on individual recipient judgments, which may introduce biases or subjective evaluations. Enhancements can be made to incorporate more sophisticated spam detection techniques or community-based voting systems to ensure more reliable spam identification, which is the focus of our work in this paper.

Nguyen et al. [18] propose a blockchain-based geo-marketplace where users can sell arbitrary data tagged with their geo-location. The authors propose a system in which the users can list their data in a marketplace curated by a trusted authority, which allows buyers to search for and buy the data. Spammers are filtered by posting the hashed location values on the blockchain before their data is listed. Malicious actors cannot modify the location data associated with the sale because the data cannot be modified once posted on the blockchain. The use of blockchain ensures the integrity and immutability of location data. However, the scalability and efficiency of the marketplace can be improved by exploring mechanisms to handle a large number of data listings and transactions while maintaining low costs and fast processing times.

Xu et al. [19] proposed a blockchain-based crowdsourcing system that could help alleviate the problem of data shortage in machine learning by incentivizing data providers to contribute data to the system. The authors argued that utilizing blockchain's inherent security and transparency features could create a secure and transparent platform where data contributors can be compensated for their contributions. They also presented a detailed architecture of the proposed system and evaluated its performance through simulations. While the security and transparency features of blockchain enhance the system's trustworthiness, further research can focus on optimizing incentive mechanisms to ensure fair compensation for data contributors and incentivize high-quality data submissions. Additionally, scalability challenges should be addressed to accommodate a growing user base and increasing data demands.

Harris et al. [20] focused on the decentralization and collaboration aspects of blockchain technology and their potential to enhance machine learning. The authors discuss the challenges of centralized machine learning systems and how blockchain technology could overcome these challenges by providing a decentralized, secure, and transparent platform for machine learning. They proposed a blockchain-based collaborative machine learning framework that enables multiple parties to contribute their data and models, collaborate on model training, and share the results. While their proposed collaborative machine learning framework demonstrates the potential of blockchain, further research is needed to optimize the system's performance, ensure efficient model training, and address privacy, data integrity, and scalability challenges.

Kadadha et al. [21] proposed a machine-learning model for behavior prediction in blockchain-based crowdsourcing systems. They argued that utilizing blockchain transaction data could create a more accurate and trustworthy model for predicting behavior in such systems. The authors presented a detailed architecture of their proposed model, which utilized a combination of supervised and unsupervised learning techniques and evaluated its performance through simulations. While their approach shows promise, future work can focus on refining the model's accuracy and exploring advanced techniques to leverage blockchain data effectively for behavior prediction. Additionally, consideration should be given to the scalability and real-time processing requirements to handle large-scale crowdsourcing systems.

Overall, the past works highlighted the potential of blockchain technology in addressing various challenges in machine learning and crowdsourcing. Blockchain's inherent security, transparency, and decentralization features could help create more secure, trustworthy, and efficient systems for machine learning and crowdsourcing. However, further improvements are needed to optimize scalability, efficiency, privacy, and incentive mechanisms. These areas of improvement will contribute to developing more secure, efficient, and trustworthy systems for spam detection and other machine learning tasks, providing a solid foundation for this paper's proposed blockchain-based incentive mechanism.

III. System Overview

Our primary objective in this study is to enhance the accuracy of a web spam detection model by leveraging high-quality crowdsourced data. The proposed mechanism involves contributors who submit data to the model, with the potential

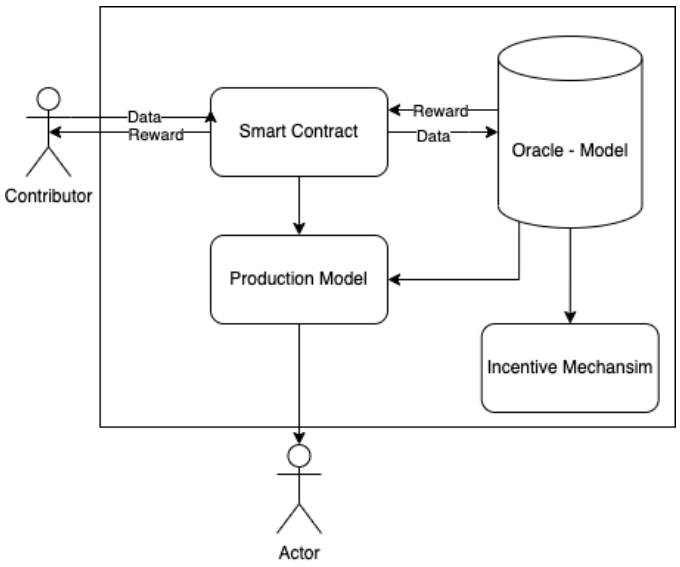

Fig. 1: System Overview

to earn profits based on the quality of their submissions. To ensure the system's integrity, we classify contributors into ***Good Actors*** and ***Bad Actors***. Additionally, users who wish to utilize the production model must submit a small fee in exchange for the right to access it.

Good Actors are contributors who operate in good faith, and while they are not restricted to submitting only non-spam websites, they must ensure that their submissions have accurate labels. On the other hand, Bad Actors aim to disrupt the model's integrity or exploit the mechanism for undeserved rewards by intentionally providing incorrect labels or repeatedly submitting the same data point that has already triggered rewards.

To encourage contributors to operate in good faith, we introduce a staking requirement whereby contributors must deposit a certain amount of currency before submitting data. The staking mechanism serves multiple purposes, including ensuring the smart contract balance and incentivizing contributors to provide high-quality submissions. The difference between the new model's accuracy and the base model's accuracy determines the weight assigned to each contributor. Based on this weight, the smart contract calculates and distributes rewards to the contributors. If no rewards are to be paid out, the contract retains the stake.

The system overview, depicted in fig. 1, illustrates the flow of operations. Contributors interact with a smart contract deployed on the blockchain, submitting their data and staking their currency. The smart contract then forwards the data to the web spam model, which processes the data and returns the weight to the contract. The model itself is not stored on the blockchain. The immutability of smart contracts ensures the safety of all monetary transactions, and contributors are the only entities authorized to withdraw their funds. Furthermore, users can download the enhanced model for a nominal fee, enabling them to leverage the improved accuracy in their web-based applications.

Moving forward, we will delve deeper into the details of the incentive mechanism, exploring the various components and algorithms that govern reward distribution and stake management.

IV. Incentive Mechanism

Our incentive mechanism, depicted in fig. 2, utilizes smart contracts to create a transparent and secure process for contributors to submit data and receive rewards based on the quality of their submissions. This mechanism is designed to encourage contributors to participate by offering monetary benefits. To evaluate the quality of the contributors' submissions, we use a *base dataset* that is a faithful benchmark for some spam-filtering models.

In our incentive mechanism, contributors must submit their data and a ***stake*** to the smart contract. The submitted data is then sent to an offline model trained on a base dataset. The objective is to evaluate whether the contributor's new data improves the model's accuracy on the base dataset. This evaluation is performed by calculating the accuracy of the current model on the base dataset and the accuracy of the new model trained on a combined dataset of existing data and new data. The ***weight*** is then calculated as the difference between the accuracy of the models trained on the old and new dataset on the base dataset.

The ***reward*** comprises the initial stake amount plus an additional reward proportional to the weight of the contributed data, incentivizing contributors to submit high-quality data that improve the model's performance. This ***reward*** is calculated as follows:

$$reward = stake + (weight * stake) \tag{1}$$

If the weight is positive, the new data will be added to the production model dataset, and the contributor will receive a reward. Conversely, if the weight is negative, the new data will be considered bad, and the contributor will lose their stake.

The smart contract handles all payouts, and users can access the improved model for a small fee for model inferencing purposes. By offering monetary rewards, contributors have a tangible incentive to submit high-quality data, thereby improving the overall accuracy of the web spam detection model. Additionally, using smart contracts ensures a transparent and immutable process, assuring contributors that their rewards will be distributed fairly and securely.

Special considerations have been taken to prevent malicious behavior and ensure the system's integrity. The stake requirement is a deterrent against potential bad actors attempting to exploit the mechanism for undeserved rewards. Furthermore, the model's offline training and the data evaluation contribute to the security and privacy of the contributed datasets, as they are not stored on the blockchain, reducing the risk of exposing sensitive information while benefiting from the advantages of blockchain technology.

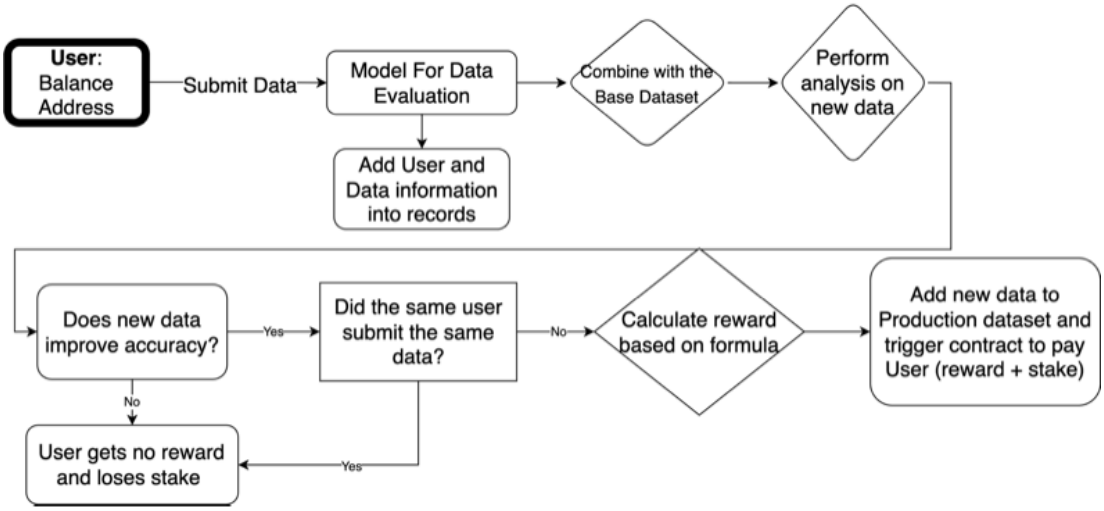

Fig. 2: Incentive Mechanism Codified in the Smart Contract

V. Evaluation

A. Dataset Description

To thoroughly evaluate the performance of our incentive mechanism, we conducted a simulation that emulated how it would function offline. We used the phishing websites dataset[2] to create three separate samples. The first sample, the base set, contained a reference of the entire dataset and performed at an accuracy of approximately 83%. The second sample was for good actors, which had to perform at a higher accuracy than the base set, and the third sample was for bad actors, in which the labels were flipped to ensure poor-quality data submissions.

B. Exploratory Data Analysis

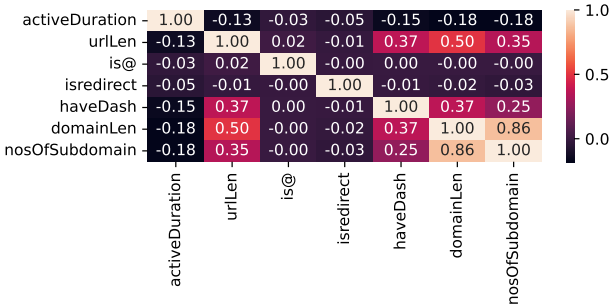

Fig. 3: Feature Correlation

The heatmap of the feature correlation can be found in fig. 3. The feature importance scores in fig. 4 show that features such as `domainlength`, `hasdash` (in the domain), and `isredirect` (if it has a double dash, there is a chance it is a redirect) are some of the most important features to name a few. The `active duration` (obtained from the whois API) and `nosOfSubdomains` (in URL) has the most significant impact on predicting spam websites. Since most of these features (expect active duration) are based on the URL patterns, it will be simple to parse new data inputs and does not require any API to get the website's metadata which can have external costs.

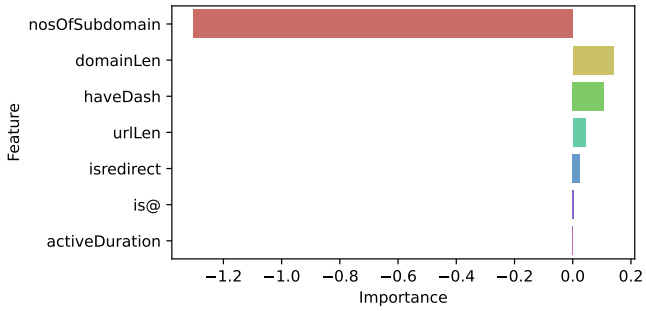

Fig. 4: Feature Importance

C. Simulation Setting

To evaluate how the dataset will make predictions, we ran accuracy tests on increasing samples of 10 up to 1000 samples. From fig. 5, the model will reach its max accuracy with tests of more than 200 samples. The results from this test provide evidence that these features identify spam vs. non-spam websites.

We then ran 50 iterations of users submitting data points to see how the balances and accuracy would perform. As shown in fig. 8, the incentive mechanism functioned as expected, with the balances of good actors increasing over time. Even if some of the submitted data were poor quality, good actors would still benefit from their contributions. As seen in fig. 7 and fig. 6, the accuracy of the production model will increase with the addition of the new "good" data.

[2]https://www.kaggle.com/datasets/aman9d/phishing-data

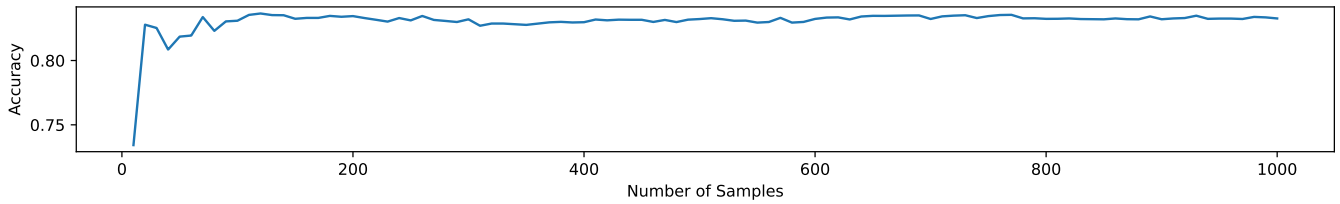

Fig. 5: Model Evaluation With Increasing Dataset Sizes

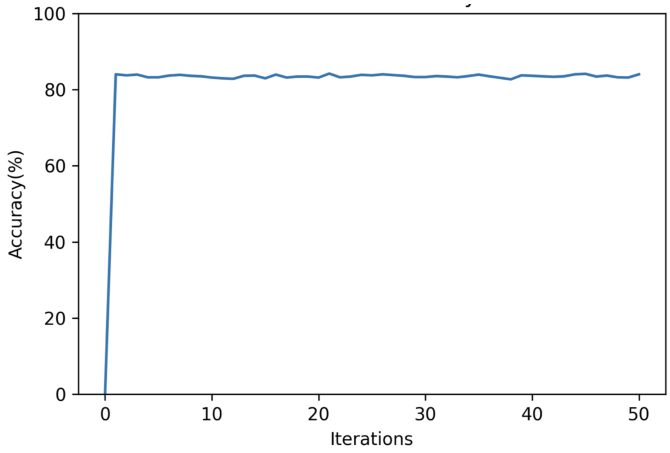

Fig. 6: Simulated Balances

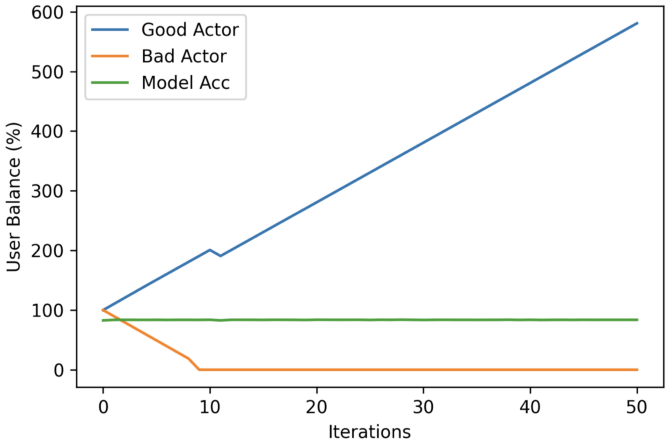

Fig. 8: User Balance

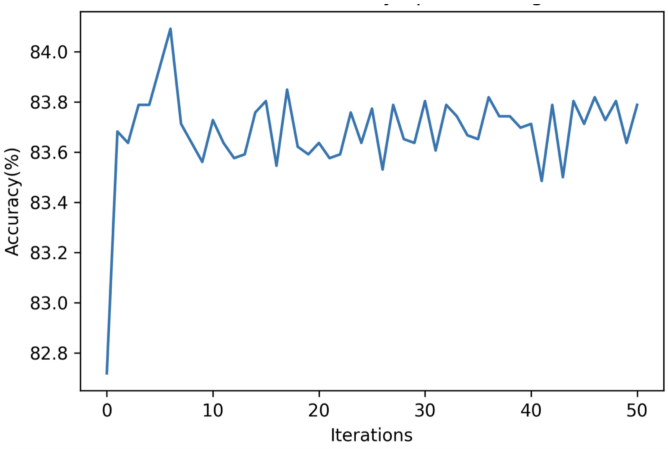

Fig. 7: Accuracy

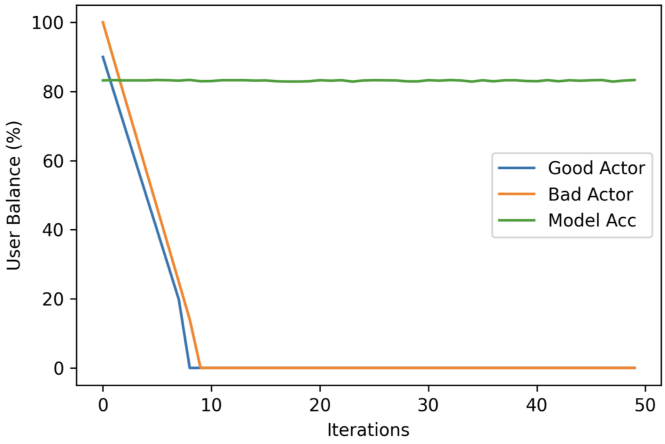

Fig. 9: User Balance of Actor Submitting Same Data

D. Detering Bad Actors

To deter bad actors from taking advantage of the system, we implemented measures to keep records of the data and which users submitted which data points. In fig. 9, we demonstrated that if a good actor attempted to repeatedly submit the same ”good” data point to manipulate the system, they would lose their stake, which was done by maintaining a record of who submitted which data point in the form `[Contributor:[domain1,domain2,...]`. Additionally, to discourage users from teaming up to scam the system, we included a mechanism to reduce rewards if the domain had already been seen multiple times in the system. For example, if the same domain were submitted three times, the weight of the reward would be adjusted by 1/3, i.e., `reward = stake + ((weight/3)*stake)`. These measures ensured that the incentive mechanism operated transparently and with integrity.

VI. Future Considerations

The proposed incentive mechanism for collaborative machine learning has great potential to motivate contributors to submit high-quality data. However, several areas for future research need to be addressed to enhance the mechanism’s

effectiveness and reliability further. It is recommended to use data from multiple sources to test the mechanism's performance across different datasets to ensure the robustness of the solution. Moreover, exploring different stake amounts and testing the mechanism on various models would provide a more comprehensive understanding of its behavior. It is essential to consider ways to protect the contributors' privacy while maintaining the mechanism's integrity to prevent privacy concerns. In addition, conducting a comprehensive evaluation of the proposed mechanism's effectiveness over an extended period is crucial to assess its long-term performance. Ensuring that the incentive mechanism functions as expected and remains sustainable even with increasing contributions over a long period is essential. Some concerns include the smart contract running out of funds or the potential negative impact on the accuracy of our model. Lastly, preventing collusion among contributors is critical to ensure the mechanism's fairness and reliability, and ways to detect and prevent it should be carefully considered.

VII. Conclusion

In conclusion, our proposed incentivized crowdsourcing mechanism for web spam detection presents a compelling and innovative solution to enhance the accuracy and reliability of existing models. By building upon the foundation laid by Harris and Waggoner [20], we have developed a streamlined and tailored approach that addresses the specific challenges of web spam detection.

Through the effective use of incentives, our mechanism motivates contributors to provide high-quality data while discouraging malicious actors from attempting to corrupt the system. By leveraging the collective intelligence of a diverse group of contributors, we can effectively crowdsource data and adapt to the ever-evolving landscape of spam websites and evasion techniques.

To further strengthen our findings, future research endeavors should focus on testing the robustness of the solution by incorporating data from multiple sources, exploring varying stake amounts, and conducting longitudinal evaluations to assess the mechanism's long-term effectiveness. Additionally, addressing privacy concerns and devising measures to prevent collusion among contributors are crucial aspects that warrant careful consideration to ensure the mechanism's fairness, reliability, and sustainability. Overall, our proposed incentivized crowdsourcing mechanism represents a promising avenue for improving web spam detection models and combating the escalating prevalence of spam websites on the internet. By fostering a safer and more trustworthy online environment, our solution has the potential to benefit both users and businesses, reinforcing the integrity and reliability of web-based interactions.

Resources: Our source code is available at: https://github.com/rpi-scales/crowd-spam.

References

[1] N. Spirin and J. Han, "Survey on web spam detection: principles and algorithms," *ACM SIGKDD explorations newsletter*, vol. 13, no. 2, pp. 50–64, 2012.

[2] Z. Gyongyi and H. Garcia-Molina, "Web spam taxonomy," in *First international workshop on adversarial information retrieval on the web (AIRWeb 2005)*, 2005.

[3] T. Urvoy, E. Chauveau, P. Filoche, and T. Lavergne, "Tracking web spam with html style similarities," *ACM Transactions on the Web (TWEB)*, vol. 2, no. 1, pp. 1–28, 2008.

[4] B. Wu and B. D. Davison, "Identifying link farm spam pages," in *Special interest tracks and posters of the 14th International Conference on World Wide Web*, 2005, pp. 820–829.

[5] Z. Gyöngyi and H. Garcia-Molina, "Link spam alliances," in *VLDB*, vol. 5, 2005, pp. 517–528.

[6] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in *Proceedings of the 15th international conference on World Wide Web*, 2006, pp. 83–92.

[7] K. Chellapilla and A. Maykov, "A taxonomy of javascript redirection spam," in *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, 2007, pp. 81–88.

[8] V. Dave, S. Guha, and Y. Zhang, "Measuring and fingerprinting click-spam in ad networks," in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, 2012, pp. 175–186.

[9] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *Journal of Big Data*, vol. 2, no. 1, pp. 1–24, 2015.

[10] S. Farooq, "A survey on adversarial information retrieval on the web," *arXiv preprint arXiv:1911.11060*, 2019.

[11] H. Jelodar, Y. Wang, C. Yuan, and X. Jiang, "A systematic framework to discover pattern for web spam classification," in *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2017, pp. 32–39.

[12] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in *Proceedings of the 5th international workshop on adversarial information retrieval on the web*, 2009, pp. 41–48.

[13] M. S. I. Mamun, M. A. Rathore, A. H. Lashkari, N. Stakhanova, and A. A. Ghorbani, "Detecting malicious urls using lexical analysis," in *International Conference on Network and System Security*. Springer, 2016, pp. 467–482.

[14] C. R. Harvey, C. Moorman, and M. Toledo, "How blockchain can help marketers build better relationships with their customers," *Harvard Business Review*, vol. 9, pp. 6–13, 2018.

[15] C. P. Matthew Pisano and O. Seneviratne, "Predictchain: Empowering collaboration and data accessibility for ai in a decentralized blockchain-based marketplace," in *ChainScience 2023*. Ledger Journal, 2023.

[16] S. A. Sheikh and M. T. Banday, "A cryptocurrency-based e-mail system for spam control," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[17] S. Choudhari and S. Das, "Spam e-mail identification using blockchain technology," *IEEE*, vol. 1, pp. 1–5, 2021.

[18] K. Nguyen, G. Ghinita, M. Naveed, and C. Shahabi, "A privacy-preserving, accountable and spam-resilient geo-marketplace," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019, pp. 299–308.

[19] H. Xu, W. Wei, Y. Qi, and S. Qi, "Blockchain-based crowdsourcing makes training dataset of machine learning no longer be in short supply," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.

[20] J. D. Harris and B. Waggoner, "Decentralized and collaborative AI on blockchain," in *2019 IEEE International Conference on Blockchain (Blockchain)*. IEEE, jul 2019. [Online]. Available: https://doi.org/10.1109%2Fblockchain.2019.00057

[21] M. Kadadha, H. Otrok, R. Mizouni, S. Singh, and A. Ouali, "On-chain behavior prediction machine learning model for blockchain-based crowdsourcing," *Future Generation Computer Systems*, vol. 136, pp. 170–181, 2022.