Impacts of Trust Measurements on the Reputation-Based Mining Paradigm

Pouya Pourtahmasbi and Mehrdad Nojoumian
Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, Florida, USA
{ppourtah,mnojoumian}@fau.edu

***Abstract*—Trust and reputation play an essential role in the success of the reputation-based mining paradigm, and therefore, it is important to have a mechanism for monitoring, analyzing, and labeling participants' behavior as a measurement of trust. To improve the existing reputation-based mining paradigm, we designed and implemented a trust model that takes into account real-world trust forming habits and incentivizes participants to commit to honest mining strategies in the cryptocurrency system. While detection of dishonest mining strategies can be challenging, this trust model considers past behavior and it is defection sensitive, i.e., making it more difficult to attain a high reputation value, the more one commits to dishonest mining. We also observed that the success of this trust model relies on the performance of attacks' detection, which is a reasonable observation. Our trust model can be used in cryptocurrency simulations to promote cooperation among miners and create a trustworthy environment for all participants.**

***Index Terms*—Blockchain; mining attacks; dishonest mining.**

I. Introduction

The reputation-based mining paradigm is a cryptocurrency mining paradigm that regulates the mining process to ensure miners are held accountable for any dishonest behavior [9]. The measurement of trust is the critical factor behind the reputation-based mining paradigm and it incentivizes miners to conduct honest mining strategies. More specifically, the trust function captures the number of times a player cooperates and defects over time and makes this value public to the system.

In order to understand the trust function and its importance in the reputation-based mining paradigm, we must first define trust and reputation as these are two distinct and integral concepts [6]. In the academic community, trust has been defined as the expectation that a particular player holds about the future behavior of the other players [4]. In order to be a trustworthy player, one must be committed to fulfilling the legitimate expectations of others [1]. With this definition of trust in place, we can now use Abdul-Rahman et al.'s definition of reputation: "A reputation is an expectation about an agent's behavior based on information about or observations of its past behavior." In a paper by Mui et al. [5], reputation is defined as a "perception that an agent has of another's intentions and norms." Combining these two definitions of reputation, we now have a well-rounded understanding of reputation and how it relates to trust. It is interesting to note that because trust is a social phenomenon and reputation plays an integral part in determining the trustworthiness of a player, reputation can therefore be used as a form of social control. This social control is considered soft security according to [1] and it is expected that there may be malicious participants. With social control, the behavior of an agent is influenced by the other parties' ability to cooperate. If a participant does not cooperate, their reputation value will decrease and this low reputation value will be publicly available in the system. Other players will be less likely to trust a participant with a low reputation value as this indicates that the participant has defected in the past. This social reputation penalty, applied to the participant that defected, will significantly decrease the participants incentive to defect and make it more difficult to do so in the future. Without the ability to measure trust and make it publicly available as a reputation value, miners may not hold the adequate incentives to act trustworthy and therefore lack trust in the cryptocurrency system as a whole [3].

The main purpose of the reputation value or measurement of trust is to help players within the system decide who to trust and to detect the presence of a dishonest player. The measurement of trust must be able to collect, calculate, update, and distribute feedback about participants' behavior over time [8]. This ensures that the reputation value incentivizes participants to act honestly and cooperate over time. In addition, it is also important to integrate the two stages of trust into the trust function. Initial trust is the first stage of trust and it is usually formed in the first few interactions without having prior knowledge of past behavior. The second stage is continuous trust and it is related to the situation where initial trust has been formed and now the trust is being maintained over a longer period. The trust function must reflect real-world trust to be effective in the reputation-based mining paradigm.

A. Our Motivation

In our earlier research [7], [10], we proposed a trust modeling mechanism that is based on human reasoning factors. The specification for our trust model was inspired by preliminary/pilot data collection. We transformed the data into a mathematical model that could be used in different platforms and software systems.

978-1-6654-3924-4/21/$31.00 ©2021 IEEE

In this article, we propose a similar trust model that is based on the same premise, but is specifically designed for the evaluation of trust inside the reputation-based mining scheme. The detection of dishonest mining activities could be a challenging effort [2], [11]. Therefore, it is likely that a significant portion of dishonest mining activities remain undetected. To address this issue, we propose a trust modeling procedure that is defection sensitive. This means that the negative impact of the defections outweighs the positive impact of cooperation. This setting will incentivize players to avoid defections if they are willing to stay in the system for a long period of time.

B. Our Approach

Our trust model is designed to maintain the reputation history of player p by updating and saving only a few parameters. In our method, a defection not only decreases the reputation value, but it also decreases the growth rate of the reputation if player p cooperates in the future. In other words, when player p increases the number of times he has defected, he will have to spend exponentially more time cooperating in order to compensate for the reputation loss. After a few defections, as player p cooperates repeatedly and consecutively, the growth rate of his reputation will increase until it is restored to the original value. Even when the growth rate is restored to its original value, if player p defects again, the reputation value as well as the growth rate will drop dramatically and further defections will exponentially cause more negative impact on the reputation of player p.

For the reputation-based cryptocurrency mining paradigm, If miner m has a low hash power in conjunction with a negative reputation, miner m will have a much lower chance of making a profit in the cryptocurrency system.

II. Trust Model Protocols

Our trust model includes a set of step-wise procedures that are used for calculating the level of reputability for the player p. First the trust variable is calculated for player p, and then the reputation value is derived through a Sigmoid function that is bounded between -1 and 1, where -1 is the lowest and 1 is the highest possible reputation. The reputation value is denoted as r and it is calculated for the player p after each round of the game. When player p enters the system, $r_0 = 0$. This means that in the beginning the reputability of player p is not known since player p has not played yet and initial trust has not been formed. At each round i, the player p can either cooperate (denoted as $\mathcal{L}_i = 1$) or defect (denoted as $\mathcal{L}_i = 0$). The cooperation will be rewarded by an increase in the reputation. The amount of this increase depends on the trustworthiness of player p until round i. Likewise, the defections will be punished by a decrease in reputation in a similar way.

A. Parameters of the Model

In order to maintain the reputation for player p, our reputation function is required to update and maintain three parameters for player p as follows:

1) The Total Number of Defections: Denoted as α and is initially set to zero. It increments every time p defects.
2) The Trust Deficit Value: Denoted as λ_i, where i represents the round and $\lambda_i \geq 0$. Initially $\lambda_0 = 0$ and then the value of λ_i is calculated after each round. If player p always cooperates, the value of λ remains zero for all rounds. But if player p defects at round i, the value of λ_i increases. Unlike α, the value of λ will decrease as player p cooperates in the future rounds; however, the decrease is at a lower rate than the increase. The rate of decrease has an inverse relationship with α.
3) The Trust Variable: Denoted as x and x_i is the trust variable at the end of round i. Initially $x_o = 0$ and the value of x_i is calculated at the end of round i. The reputation of player p has a direct relationship with x_i.

B. Calculation Procedure

The complete procedure for calculating the reputation for player p at the end of round i is given bellow:

$$1)\ \alpha_i = \begin{cases} \alpha_{i-1} + 1 & ,\ \mathcal{L}_i = 0 \\ \alpha_{i-1} & ,\ \mathcal{L}_i = 1 \end{cases}$$

$$2)\ \lambda_i = \begin{cases} \lambda_{i-1} + ln(e + \lambda_{i-1}) & ,\ \mathcal{L}_i = 0 \\ \lambda_{i-1} - log_{e^{\alpha_i}+1}(1 + \lambda_{i-1}) & ,\ \mathcal{L}_i = 1 \end{cases}$$

$$3)\ x_i = \begin{cases} x_{i-1} - \big(ln(e + \lambda_i)\big)^e & ,\ \mathcal{L}_i = 0 \\ x_{i-1} + (e-1)^{-\lambda_i} & ,\ \mathcal{L}_i = 1 \end{cases}$$

$$4)\ r_i = \begin{cases} -1 & ,\ r_{i-1} \leq \epsilon - 1 \ \textit{and}\ \mathcal{L}_i = 0 \\ 1 & ,\ r_{i-1} \geq 1 - \epsilon \ \textit{and}\ \mathcal{L}_i = 1 \\ \frac{2}{1+e^{-\gamma x_i}} - 1 & ,\ \textit{otherwise} \end{cases}$$

In the above procedure, γ is a constant and positive steepness parameter. Larger values for γ results in a higher steepness for the Sigmoid function, and consequently, the rate of change becomes faster when x is in the proximity of zero. Likewise, smaller values for γ results in a lower rate of change for the reputation. ϵ is a small positive and constant parameter. $1 - \epsilon$ and $\epsilon - 1$ set an upper and a lower bound for the calculation of the reputation. Therefore, when $1-\epsilon \leq r_{i-1} < 1$ and the player p has cooperated in round i, the function returns 1 and when $-1 < r_{i-1} \leq \epsilon - 1$ and the player has defected in round i, the function returns -1. This procedure prevents the divergence of x_i, and consequently, x_i will always remain within the proximity of the usable range of Sigmoid function.

C. Further Adjustments

In our trust calculation procedure, we used e as a constant as well as the base for the logarithm function for the calculation of λ and x, however, for further adjustments and tweaks, these constants may be changed. It is also worth mentioning that the base of the logarithms in the second and third steps of the procedure, when $\mathcal{L}_i = 0$, are not required to be the same value. By modifying these values, the intensity of rewards and punishments can be adjusted.

III. Effectiveness of Our Trust Model

We utilize the following strategies that are used by malicious players to disrupt a reputation system or decode its behavior.

1) Cooperate for some time to build a high reputation value and then defect on costly transactions.
2) Cooperate for some time and then take a mixed strategy of cooperation-&-defection to decode the behavior.
3) Take a mix strategy of cooperation-&-defection and then cooperate for some time to build a high reputation value.

The reputation value starts at zero then, the value changes after each round of game based on the behavior of the player. An increase in reputation value between round i and $i+1$ is the result of a cooperation and a decrease in reputation value between round i and $i+1$ is the result of a defection.

Figure 1 represents a player who cooperated for 15 rounds then defected three times in row. The decline of the reputation started slowly since the player was considered trustworthy. When the player defected for three consecutive iterations, a dramatic decline in the trust value occurred. After the decline of the reputation, the player continued by cooperating two times in row, however the reputation value did not significantly change. To compensate for the loss of trust, the player is required to cooperate for a significantly greater number of times than the number of times he cooperated in the beginning of the game. This is due to the fact that the player has shown to be untrustworthy after round 15.

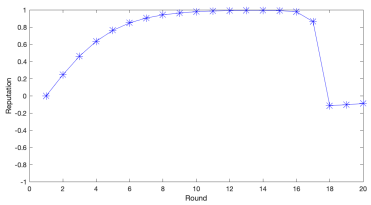

Fig. 1: Effectiveness of our model against the 1st strategy.

Another player, whose reputation progression is shown in Figure 2, also cooperated for a number of times in the beginning of the game but then eventually non-consecutively defected. The graph shows that after the third non-consecutive defection, the reputation value declines at a significantly faster pace. The drop is still not as dramatic as the blue graph example since the defections were not consecutive, but if the trend continues the reputation loss will keep accelerating.

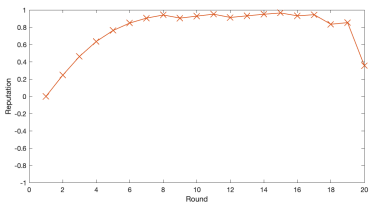

Fig. 2: Effectiveness of our model against the 2nd strategy.

Finally, Figure 3 shows the change of reputation for a player that defected in the beginning of the game. It is evident that the trust model is less tolerant to the non-consecutive defections taking place in the beginning of the game. With two cooperations and three defections, the reputation value of the player dropped down to -0.75. Then the player had to cooperate for thirteen times in order to establish a high reputation value. Defections committed upon entering the system are harshly penalized due to the fact that past behavior is not known and the reputation has not yet been established. In other words, the initial trust has not been established so that player is essentially making a bad first impression in the system. Once that initial trust is established and is reflected in a low reputation value, the player will have to cooperate many times in order to restore their reputation as a trustworthy player. This reflects the stage of continuous trust, which can only be realized over a longer time period. In contrast, the first two players gained the same positive reputation for only four cooperations at the beginning of the game, proving themselves to be trustworthy players. Even when they defected, their reputation did not drop as quickly as the player represented in Figure 3 because they had built up a reputation and essentially made a positive first impression for initial trust. This comparison highlights the long term effect of defections on the growth of the reputation value and also how the trust function captures the nuances of trust over time.

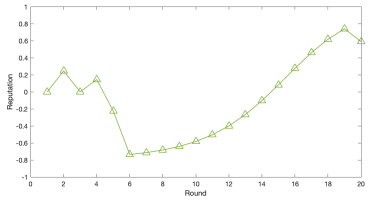

Fig. 3: Effectiveness of our model against the 3rd strategy.

IV. Empirical Assessment

To evaluate the performance of our trust function in action, we implemented a cryptocurrency mining simulation in which a percentage of miners who hold high hash powers conduct block withholding attacks against their pools. In this model, every miner has a public reputation value that is updated by the pool manager using our trust model. The pool managers run a statistical test in cycles to detect whether block withholding attacks are taking place in their pools or not. From the pool managers' point of view, the condition of withholding a block is considered defection and the negation of this condition is considered cooperation.

A. Simulation Setup

In our implementation, the number of mining rounds for each cycle is a random variable that is uniformly distributed within a defined range. Since the block withholding detection method relies on a statistical test, the quantity of samples is crucial to the accuracy of this test. Therefore, we select a fairly large number as the lower bound for the probability distribution range. In the previous section, we defined i as the number

of rounds for simplicity. However, in our implementation, the value of i represents the detection cycle that can contain any number of rounds within the probability distribution range. Therefore, the reputation value for all miners from pool p will be updated by the pool manager once a new detection cycle takes place. The pool manager simply compares the number of actual proof-of-work (POW) against the number of expected POW for all member miners and then the reputation value for each miner is updated accordingly.

In other words, the pool manager compares the actual POW and the expected POW for all member miners for the period between the detection cycle $i-1$ and i. If $x_j < E[x_j]$ and $x_i \notin CI$ for miner M_j, where x_j denotes the actual POW, $E[x_j]$ denotes the expected POW since the last detection cycle for miner M_j, and CI is the confidence interval for the attack detection, then the pool manager identifies miner M_j as an attacker and updates the reputation of M_j considering M_j has defected in round i. Likewise, if the condition is false for miner M_j, then the pool manager updates the reputation of M_j considering M_j has cooperated in round i.

B. Performance in Simulation and Technical Discussion

To observe the performance of our trust model in the reputation-based mining paradigm, we perform our simulation for $250,000$ rounds of mining. The scatter plot in Figure 4 shows a summary of the performance of our trust model after round $250,000$. In this plot, x axis represents the reputation value and y axis represents the total number of block withholding attacks. As we explained in the previous section, the detection method in our simulation is based on the confidence interval statistical test with the confidence of 98%. Therefore a percentage of attacks is expected to remain undetected.

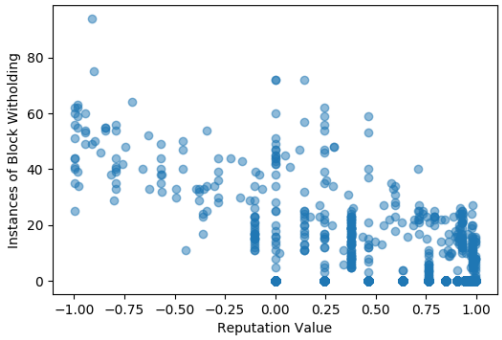

Fig. 4: Reputation value vs block withholding attack.

Also, our mining simulation is dynamic. It means that miners may enter and exit the system at any time. The scatter plot shown in Figure **??** is expected to demonstrate the realistic correlation between the number of block withholding attacks and the reputation. As it is evident, a large percentage of miners have conducted block withholding attacks but their reputation remains relatively high. This outcome is significantly different from the examples we showed in Figures 1 through 4. This difference is mostly due to the fact that a large number of block withholding attacks remain undetected. The statistical test with the confidence interval of 98% will not be able to detect a significant number of attacks. This situation becomes worse if the attack instances are scattered over a large period of time. From this experiment, we can observe that the performance of the detection method is the key to the success rate of any trust measurement scheme.

V. Conclusion

We proposed a trust model that is designed for the cryptocurrency mining reputation-based paradigm. Our trust model is designed to be defection sensitive and preserve the history of the players with only three parameters. Our trust function also takes into account the two stages of trust, i.e., initial trust and continuous trust, and incentivizes participants to commit to honest mining strategies. The method of detection of all block withholding needs to be further developed as the success of the trust model relies on the performance of attacks' detection, which is reasonable and a critical observation.

VI. Acknowledgment

Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-18-1-0483. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd annual Hawaii international conference on system sciences*, pages 9–pp. IEEE, 2000.

[2] Mauro Conti, E Sandeep Kumar, Chhagan Lal, and Sushmita Ruj. A survey on security and privacy issues of bitcoin. *IEEE Communications Surveys & Tutorials*, 20(4):3416–3452, 2018.

[3] Partha Dasgupta. Trust as a commodity. *Trust: Making and breaking cooperative relations*, 4:49–72, 2000.

[4] Diego Gambetta et al. Can we trust trust. *Trust: Making and breaking cooperative relations*, 13:213–237, 2000.

[5] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A computational model of trust and reputation. In *35th Annual Hawaii International Conference on System Sciences*, pages 2431–2439. IEEE, 2002.

[6] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of reputation in multi-agents systems: a review. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, pages 280–287, 2002.

[7] Mehrdad Nojoumian. Trust, influence and reputation management based on human reasoning. In *4th AAAI Workshop on Incentives and Trust in E-Communities*, pages 21–24, 2015.

[8] Mehrdad Nojoumian. Rational trust modeling. In *International Conference on Decision and Game Theory for Security (GameSec)*, pages 418–431. Springer, 2018.

[9] Mehrdad Nojoumian, Arash Golchubian, Laurent Njilla, Kevin Kwiat, and Charles Kamhoua. Incentivizing blockchain miners to avoid dishonest mining strategies by a reputation-based paradigm. In *Science and Information Conference*, pages 1118–1134. Springer, 2018.

[10] Mehrdad Nojoumian and Timothy C Lethbridge. A new approach for the trust calculation in social networks. In *International Conf. on E-Business and Telecommunication Networks*, pages 64–77. Springer, 2006.

[11] Florian Tschorsch and Björn Scheuermann. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys & Tutorials*, 18(3):2084–2123, 2016.