An efficient algorithm for implementing consensus in the DAG-based distributed ledger for IoT system

Huisu Jang
Seoul National University
Seoul, Korea 08826
Email: gmltn7798@snu.ac.kr

***Abstract*—The DAG-based tangle model proposed by the IOTA Foundation aims to remove the transaction fee by employing a different protocol from that used in the blockchain. They proposed the MCMC algorithm for tip selection in the tangle; however, concerns about centralization by the coordinator nodes remain. Additionally, the economic incentive to choose the algorithm is insufficient. The present study proposes a light and efficient tip selection algorithm that considers only the subtangle of each step by considering the Bayesian inference. Experimental results have confirmed that the proposed methodology has considerable performance similar to the existing methodology and has faster computation time. The proposed methodology has the same resistance as the MCMC algorithm for possible attacks on the same reason as the MCMC algorithm.**

I. Introduction

To address the high transaction fee problem in micropayments, [1] has proposed the tangle, which is a directed acyclic graph (DAG)-based distributed ledger, in which the group issuing transactions is identical to the group confirming transactions. In the tangle, all participants in the system have the same demand (purpose) for issuing a transaction. The tangle presents a conditional cost of mandatory transaction confirmations to achieve the purpose of issuing a transaction. In other words, the issuance of the transaction replaces the incentive for transaction approval so that the demand and supply are matched within the system without exogenous supply, such as the miner provided. Thus, he tried to maintain the distributed ledger without transaction fee while discarding the blockchain-type ledger adopted in the bitcoin or ethereum.

Currently, a light node in the IOTA receives the results of the MCMC-based tip selection algorithm from a node operated by the Foundation, known as a "coordinator," by calling the tip selection API to select the tips to approve. To choose the tips, a coordinator node considers a reliable transaction (a "milestone") issued by a coordinator node to implement a MCMC algorithm. However, transactions with a trustworthy status granted by the coordinator node of the Foundation bring a centralization issue to IOTA. Additionally, there is no reason to enforce a suggested tip selection method for general users. [2] separated nodes complying with the default tip selection algorithm with the selfish nodes pursuing their own profits. Rather, it sounds a bit more logically plausible that every node behaves "selfishly" in a way that minimizes their own cost. A general participant wants their transactions cumulative confirmation to become sufficiently large in the tangle. Therefore, they may want to take a strategy that allows their transaction weight to accumulate quickly. Hence, at this point, the only strategy they can take is to choose tips that can quickly increase their own cumulative weight.

In what follows, we propose a more efficient and light tip selection algorithm that allows users to achieve their objective of maximizing their probability of confirmation by selecting adequate tips. In addition, we discuss the resistance of the proposed algorithm against possible attack scenarios, and finally, present recommendations for future works.

II. Related work

In the field of cryptocurrency, DAG-based ideas have been considered to circumvent several drawbacks of the blockchain[3], [4], [5], [6]. The greedy heaviest-observed subtree (GHOST) protocol suggests that the whole chain maintains several uncle blocks that fail to generate a new block because they take a longer time do so compared with the first block maker. The GHOST protocol considers the blockchain as a tree. It can be used as a solution to the network security problem, which is caused by the faster block generation time leading to a higher stale rate [4]. [6] proposed the DAG-based cryptocurrency model in which considers the miner and the block. In this model, the block contains only one transaction and is called the site.

III. Bayesian inference

Before we make a statement of the proposed method, we would give a brief description of Bayesian inference. Bayesian inference, a kind of statistical inference, uses the Bayes' rule to update the unknown objective probability density as more evidence or information becomes available. Apart from being widely applied in the fields of science, engineering, and philosophy, Bayesian inference has been employed especially in the dynamic analysis of data sequences [7], [8], [9].

In Bayesian inference, the posterior distribution can be deduced from two antecedent probabilities, a prior and likelihood, according to Bayes' rule expressed as

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}, \tag{1}$$

where H means any hypothesis affected by the data, E, which is the evidence corresponding to new data; $P(H|E)$ is a

probability that we ultimately want to know after the evidence E is observed; $P(E|H)$ is the probability of observing the evidence when the hypothesis or the model is given; and $P(H)$ is a prior density, that is, a probability of the objective hypothesis or model before the evidence is given.

In this study, we apply the Bayes rule to study the data sequence of a specific type, that is, a time series of subtangles in the IOTA ecosystem. To select tips in the IOTA ecosystem, we want to precisely estimate the t-th probability distribution for the nodes, the posterior distribution, $P(H_t|E_t)$ given the evidence of t-th subtangle. The posterior density can be approximated in proportion to the product of the $(t-1)$-th prior density, $P(H_{t-1})$, and a likelihood, $P(E_t|H_{t-1})$, which refers to the probability of the t-th subtangle given the $(t-1)$-th approximated density for the nodes.

$$P(H_t|E_t) \propto P(E_t|H_{t-1})P(H_{t-1}) \tag{2}$$

A detailed description of Bayesian inference in the proposed algorithm is presented in the next section.

IV. Light tip selection algorithm

We also restrict the network structure as a tangle of IOTA and develop the argument by accepting the basic assumptions and related definition in the white paper of IOTA. In this work, we approximate a discrete probability density for the existing tip set as the original tip selection algorithm does [1]. The difference from the conventional tip selection algorithm is that the probabilistic distribution of the tip set is projected into the space of the node set consisting of nodes issued on a site in the subtangle. The reason why the tip set is selected by the cumulative weight of the node is that a node with a high cumulative weight would want to maintain such high cumulative weight by carrying out confirmations quickly. The proposed tip selection algorithm is described below.

1 If the node participates in the network for the first time, two of the currently available tips are randomly selected.

2 In subsequent tip selections, each node selects a new tip set from the prior density based on the precedence subtangle, wherein all sites are directly or indirectly confirmed by the tip set recently confirmed by each node.

3 Based on the updated subtangle, the discrete likelihood distribution can be suggested for the nodes issuing a transaction in the updated subtangle. The value of the likelihood distribution of each node should be approximated in order to reflect the principle that malicious nodes have a smaller probability than typical participant nodes based on the already known information of the preceding subtangle.

4 The posterior distribution is updated given the likelihood and a prior distribution.

While the existing MCMC methodology discusses the probability density over the tips itself, the current study deals with the probability density over the nodes that issued sites. Hereafter, we shall update the probability density only for the moments when the user should select the tips and the moment at which a transaction is issued. For the sake of brevity, the moment of issuing the t-th transaction and selecting the t-th tips is referred as the t-th phase. We addresses the modeling of the discrete posterior density transition over the time by deploying the Bayes' rule. In the other word, a discrete posterior density at the t-th phase is determined by the t-th prior and the t-th likelihood distribution, which is created based on the t-th subtangle consisting of approved sites by the selected tips at the t-th phase.

the posterior distribution obtained in the immediately preceding phase becomes the prior distribution at the present phase. Using the Markov chain structure, it is assumed that the present prior distribution contains cumulative information according to the subtangle at each precedent phase. Before becoming the t-th state, a node has a discrete posterior distribution for the nodes and participants in the network based on the $(t-1)$-th subtangle. Assuming that the set of participating nodes known by this node at the $(t-1)$-th phase is K, then the equation below is established,

$$\sum_{i \in K} p(X=i) = 1, \quad i \in K, \tag{3}$$

where p_i is the prior density of the i-th node.

Before updating the posterior distribution, we first propose a probability distribution for choosing tips in the t-th phase based on the t-th prior distribution. We defined additional set M and K' at the t-th phase. Set M contains new nodes, which are not contained in the set K and newly appear in the t-th subtangle. Set K has a subset K' which contains the nodes excluded from the tip issuing node set at the t-th phase. Each k, k' and m means the number of elements in the set K, K' and M.

Based on the above assumptions and equation 3, we can derive a probability distribution for tip selection as follows,

$$p_{tip}(X=x) = \begin{cases} \frac{k-k'}{k-k'+m} \frac{p(X=x)}{\sum_{i \in K \setminus K'} p(X=i)}, & x \in K \setminus K' \\ \frac{1}{k-k'+m}, & x \in M \end{cases} \tag{4}$$

In the t-th phase, new information we can obtain is the subtangle created after selecting a set of tips based on a given above probability distribution. We need an appropriate likelihood distribution to estimate the posterior probability density for the union of the known set of nodes K and the set of new nodes L from the subtangle. For the sake of argument, we classify the entire node set into three subset to propose a proper likelihood distribution. The characteristics of each subset and the estimates of the likelihood distribution are described in each subsection.

A. ***Set A**: only included in the prior distribution*

Since the set A consists of nodes that are included in the prior distribution but are excluded when forming the subtangle, the new information relevant to set A can not be obtained from the subtangle. Therefore, the likelihood distribution is obtained from the average cumulative weight held in the prior density in a manner similar to the existing tangle. The nodes included in set A can be considered as two cases as follows: the number of issued transactions and the average cumulative weight are

small or large nodes. The second case is more likely to be a malicious intentional node. Despite the fact that a large number of transactions have been issued by a node, if a node can only be observed in the restricted subtangle, it is likely that the node intentionally attempted. Therefore, it is reasonable to give a likelihood distribution in inverse proportion to the average cumulative weight. We have proposed the following likelihood distribution for the set A, taking into account the case of other subsets.

$$p_A(X=x) = \frac{N_A}{N_A+N_B+N_C}\frac{exp(-\alpha_1 w_x)}{\sum_{i\in A} exp(-\alpha_1 w_i)}, \quad x \in A, \tag{5}$$

where N_S is the number of elements in set S, w_x is the average cumulative weight of node i based on the $(t-1)$-th subtangle, and α_1 is the parameter for the distribution. Set B and C are defined by each subsequent subsection.

B. ***Set B**: both included in the prior distribution and the t-th subtangle*

Set B covers nodes that are observed in both consecutive phases. When a normal node is observed in successive phases, it is assumed that transactions are issued on average λ times between the two phases depending on the Poisson process assumption. It can be expected that the average cumulative weight of a typical node will increase by about λ between the two consecutive phase.

Consider now the case where the average cumulative weight change is noticeably larger than lambda. Regardless of the direction of change, this means that the corresponding node is not usual. For example, an abnormal average cumulative weight increase may be a signal that the node has started a malicious attack. An extraordinary reduction in the average cumulative weight may also be a signal that the $(t-1)$-th subtangle contained the parasite tangle of malicious node. In other words, it is reasonable that a likelihood distribution is presented based on the extreme change in the average cumulative weight. We employ a step function for implementing the corresponding likelihood distribution. If the absolute value of the difference of the node's average cumulative weight between the $(t-1)$ and t-th phases is greater than the sum of λ and the parameter α_3, then the likelihood that a transaction generated by the node is selected as a tip is reduced. Therefore, the likelihood distribution of set B can be approcimated as follows,

$$p_B(X=x) = \begin{cases} \frac{n_1}{n_1+n_2}p, & \textit{if} \quad |w_x^{(t-1)} - w_x^t| \geq \lambda + \alpha_3 \\ \frac{n_2}{n_1+n_2}(1-p), & \textit{if} \quad |w_x^{(t-1)} - w_x^t| < \lambda + \alpha_3, \end{cases} \tag{6}$$

where n_1 is the number of nodes satisfying the first condition, and n_2 is the number of nodes satisfying the second condition in set B, p is the threshold probability, α_3 is the parameter for the model.

C. ***Set C**: only included in the t-th subtangle*

In the case of set C, it consists of newly observed nodes in the t-th subtangle. In other words, it is possible to estimate the likelihood distribution with the same principle as in the case of the set A. The only difference between the two cases is that set C uses node information of the t-th subtangle unlike set A, which investigates node information of the $(t-1)$-th subtangle.

$$p_C(X=x) = \frac{N_C}{N_A+N_B+N_C}\frac{exp(-\alpha_3 w_x)}{\sum_{i\in C} exp(-\alpha_1 w_i)}, \quad x \in C, \tag{7}$$

where N_S is the number of elements in set S, w_x is the average cumulative weight of node i based on the t-th subtangle, and α_3 is the parameter for the distribution.

Assuming that a sincere node and a malicious node each issued a total of n transactions, the transaction issued by the sincere one would cumulate its own weight independently from other sincere participants. After the adaptation period, the cumulative weight would increase linearly. After the adaptation period, a cumulative weight of each transaction issued by a sincere node would increase linearly and assume that the value is w. Under the same conditions, all the n transactions issued by a malicious node would inevitably have different cumulative weights that are dependent on each anothers transactions. In order to hide its malicious intent, a node must approve its own transactionand not by a node of the main tangle. In this case, the cumulative weight of the first issued transaction is w, the cumulative weight of the following transactions would be bound to $w-1, w-2, w-3, ...$

V. Empirical study

In order to evaluate the stability of the system for each method, we performed simulation studies where λ is 20 or 50, and the number of iteration is 1000, and 3000, respectively. Figure 1 shows that the cumulative weight increases with the slope of the lambda without any difference for each methodology. Figure 1 also confirmed that each methodology reliably increases the cumulative weight of any transaction under the usual circumstance, and that the slope depends on the volume of transactions issued on average per unit time. There exists a remarkable difference in average cumulative weight between the empirical result and original paper [1]. While [1] have mentioned that there is an adaptation period in which a cumulative weight is exponentially increased, this simulation study confirms that the cumulative weight is linearly increased according to the number of iterations.

Results of the simulation shows that the number of tips of each methodology is concentrated around a λ for each methodology, which is slightly different from the [1]'s assumption the number of tips remains roughly stationary in time and in concentrated around a number $L_0 = 2\lambda h$. We have also found that the number of tips of the proposed algorithm achieves roughly stationary in time around the λ as in other algorithms. Figure 2 presents the number of tips according to λ.

Figure 3 shows the elapsed time of each iteration for the MCMC and proposed algorithm, respectively. Unlike the MCMC algorithm of extracting particles from the main tangle and performing MCMC simulations, the proposed method considers only the subtangle calculated at each time step, so that the elapsed time of each step is shorter than the MCMC algorithm.

VI. Discussion

The proposed and MCMC algorithms have the same principle to prevent malicious node attacks. Both methods attempt to

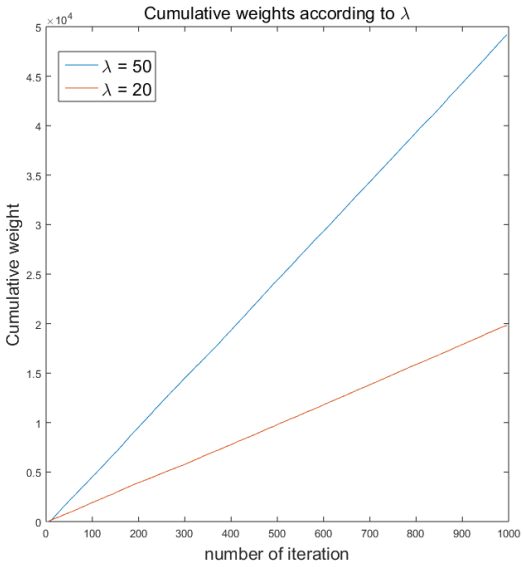

Fig. 1. Cumulative weights according to λ with random selection algorithm

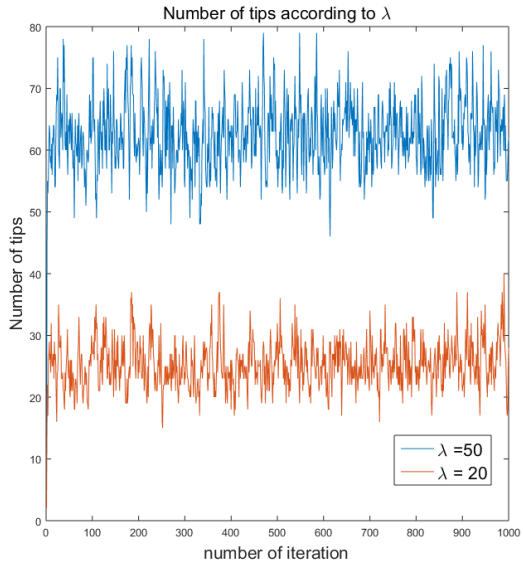

Fig. 2. Cumulative weights according to λ with random selection algorithm

minimize the probability of attack by reducing the probability that a malicious node is selected as a tip. Therefore, the proposed algorithm is also resistant to possible attack scenarios for the same reaons discussed in [1].

Major difference between the two algorithms is revealed in the way to determine whether a node is malicious. Two algorithms consider the different criteria to determine whether a node is malicious. In particular, the proposed methodology evaluates the probability the node is malicious based on the sudden increase or decrease in the cumulative weight of any node, that is, the large variability of the cumulative weight. Given the fact that many types of network attacks are made through the sudden appearance of specific parasite subtangle,

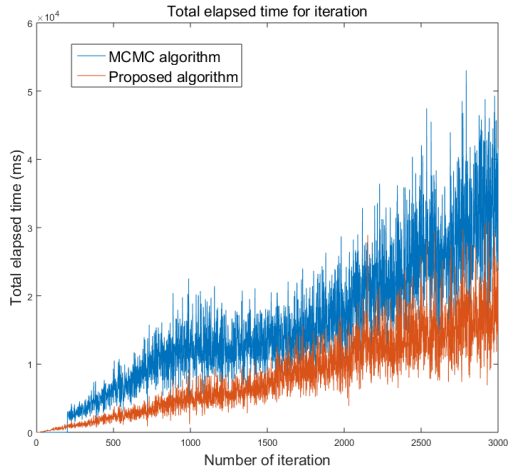

Fig. 3. Cumulative weights according to λ with random selection algorithm

the proposed algorithm attempts to give a penalty by using Bayesian inference for the sudden volatility of cumulative weight over time. For instance, if a paracite subtangle appears and disappears between consecutive states, the probability of selecting that node as a tip will decrease exponentially by the proposed algorithm in the previous section.

The main advantage of the proposed methodology is that it is light and efficient. Each node is supposed to keep only the own subtangle at each time step and the advanced posterior (or prior) information according to the time. To issue a transaction, each node needs to refer the information of subtangle to confirm the validity of the transactions directly or indirectly approved by the selected tips. In other words, the information of subtangle is essential for every time step, irrespective of the proposed algorithm. Thus, the proposed algorithm is expected to help mitigate the problem of centralization issues by ”coordinators” present in the tangle because the proposed algorithm is very light and efficient to utilize information. At the same time, the proposed algorithm protects against possible attacks with the same principle as MCMC algorithm.

VII. Conclusion and Future Work

In this paper, we have experimentally confirmed that a lightweight and efficient tip selection algorithm has the same safety as the existing MCMC algorithm in tangles and enables faster tip selection. There are other important branches to support the strengths of the proposed method as a future work. An empirical simulation study involving malicious nodes will be able to investigate the response patterns of each algorithm in an abnormal situation and expand the related discussion. The theoretical support based on the Markov chain and Bayes' rules will also strengthen the argument of the proposed methodology.

References

[1] S. Popov, “The tangle,” *cit. on*, p. 131, 2016.

[2] S. Popov, O. Saa, and P. Finardi, “Equilibria in the tangle,” *arXiv preprint arXiv:1712.05385*, 2017.

[3] S. D. Lerner, “Dagcoin: a cryptocurrency without blocks,” 2015.

[4] Y. Sompolinsky and A. Zohar, "Accelerating bitcoins transaction processing," *Fast Money Grows on Trees, Not Chains*, 2013.

[5] J. Poon and T. Dryja, "The bitcoin lightning network," 2015.

[6] Y. Lewenberg, Y. Sompolinsky, and A. Zohar, "Inclusive block chain protocols," in *International Conference on Financial Cryptography and Data Security*. Springer, 2015, pp. 528–547.

[7] L. Bardwell, P. Fearnhead *et al.*, "Bayesian detection of abnormal segments in multiple time series," *Bayesian Analysis*, vol. 12, no. 1, pp. 193–218, 2017.

[8] M. Sanchez-Castillo, D. Blanco, I. Tienda-Luna, M. Carrion, and Y. Huang, "A bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data," *Bioinformatics*, vol. 1, p. 7, 2017.

[9] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, S. L. Scott *et al.*, "Inferring causal impact using bayesian structural time-series models," *The Annals of Applied Statistics*, vol. 9, no. 1, pp. 247–274, 2015.