A Blockchain-Enabled Quantitative Approach to Trust and Reputation Management with Sparse Evidence

Extended Abstract

Leonit Zeynalvand
Nanyang Technological University
Singapore
leonit001@e.ntu.edu.sg

Tie Luo
Missouri University of Science and Technology
Rolla, MO, USA
tluo@mst.edu

Ewa Andrejczuk
Nanyang Technological University
Singapore
ewaa@ntu.edu.sg

Dusit Niyato
Nanyang Technological University
Singapore
dniyato@ntu.edu.sg

Sin G. Teo
Institute for Infocomm Research
Singapore
teosg@i2r.a-star.edu.sg

Jie Zhang
Nanyang Technological University
Singapore
zhangj@ntu.edu.sg

ABSTRACT

The prevalence of e-commerce applications poses new trust challenges that render traditional Trust and Reputation Management (TRM) approaches inadequate. The first challenge is that TRM is built on evidence (direct or indirect observations) but evidence is becoming increasingly sparse because nowadays users have many more venues to share information. This makes it hard to derive trust models that are robust to attacks such as whitewashing and Sybil attacks. Second, the cost of attacks has reduced significantly due to the widespread presence of bots in e-commerce applications, which tends to invalidate the traditional assumption that majority users are honest. In this paper, we propose a new TRM framework called BEQA, which uses Blockchain to transform multiple disjoint and sparse sets of evidence into a single and dense evidence set. To address the second challenge, we introduce and formulate the cost of Sybil attacks using Blockchain transaction fees. In addition, we make a key observation that existing trust models have overlooked publicity (evidence originating from influencers) that exist in e-commerce applications. Thus, we formulate publicity as a whitewashing deposit such that a higher level of publicity will impose higher cost on Sybil attacks.

KEYWORDS

Blockchain; trust management; whitewashing attack; Sybil attack

ACM Reference Format:
Leonit Zeynalvand, Tie Luo, Ewa Andrejczuk, Dusit Niyato, Sin G. Teo, and Jie Zhang. 2021. A Blockchain-Enabled Quantitative Approach to Trust and Reputation Management with Sparse Evidence: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 2 pages.

1 DESIGN OF BEQA

BEQA quantifies the cost of whitewashing and Sybil attacks using Blockchain transaction fees. For an arbitrary *entity* (whose trustworthiness is to be assessed; e.g., a seller in an e-marketplace), BEQA first uses a heuristic method to identify the publicity of the entity, which is the set of feedback given by influencers. Second, BEQA establishes a mathematical relationship between the weight of a submitted feedback and the cost incurred by the user who submitted the feedback, where the weight grows over time with an unknown growth rate (that is determined subsequently). Third, BEQA quantifies the cost of a Sybil attack as a function of the growth rate, and calculates the growth rate in such a way that the cost of a Sybil attack is determined by the aggregate cost incurred by the influencers of the entity. BEQA leverages this aggregate cost as a whitewashing deposit for that entity to discourage both whitewashing and Sybil attacks.

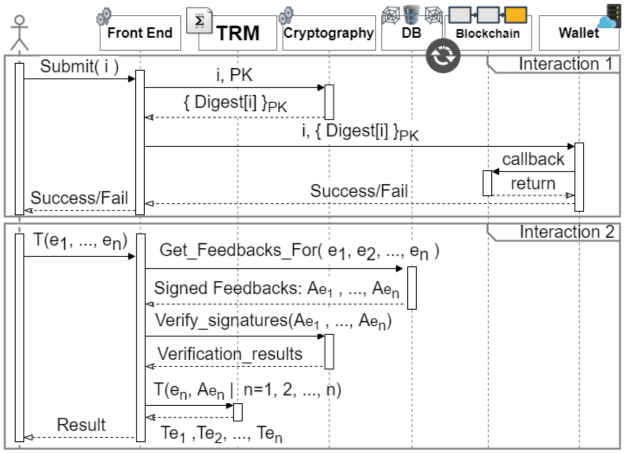


Figure 1: Sequence diagram of the processes in BEQA. Interaction 1 illustrates the assessment data submission process. Interaction 2 illustrates trustworthiness enquiry process.

1.1 Process Overview

Feedback. We use the term *feedback* to refer to the information shared by a user about an entity. A feedback i contains quantitative information (such as user rating) which we denote by v_i.

Blockchain transaction fee. Blockchain is a tamper-resistant record-keeping commodity, which means the records are tamper-resistant regardless of who keep them. This, has been time-tested the longest in Bitcoin [3–5] for more than 10 years. Adding records will incur a *transaction fee* to be paid to the record-keepers.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

We illustrate the process overview of BEQA in Fig. 1. As illustrated in interaction 1 (upper portion of figure), to submit feedback $i = (e, v_i, pub)$ about an entity e, a user with a public-key pub first signs the feedback with its own private key PK. Thus, each user has a randomly generated key pair (not that of the Blockchain), which makes BEQA robust against *repudiation attacks*. Next, the user submits the signed feedback to Blockchain using a wallet (which is a software used in Blockchain to perform transactions). This will incur the user a transaction fee f_i; when the transaction is confirmed on Blockchain, the feedback is assigned a time-stamp t_i.

As illustrated in interaction 2 (lower portion of Fig. 1), when a user enquires about the trustworthiness of some entities $e_1, e_2, ..., e_n$, BEQA fetches from a database all the existing feedbacks for those entities as well as the corresponding fees and timestamps. This database is synchronized with the Blockchain (only on transactions relevant to the BEQA framework). Examples of database technologies, which can perform targeted crawling and indexing of Blockchain, can be found in [1, 2]. Next, BEQA verifies the user signatures of these fetched feedbacks. This prevents *identity attacks* in which an attacker associates its malicious feedback with a legitimate user's identity. Finally, BEQA's trust model, which is explained in the next section, uses Blockchain transaction fees to quantify the cost of whitewashing and Sybil attacks and provide a robust assessment of the trustworthiness of entities $e_1, e_2, ..., e_n$.

1.2 Details of TRM

Weighting feedbacks. BEQA assigns a weight of zero to feedbacks from publicity influencers, and hence filters out their opinions; but BEQA uses their total spend on transaction fees as a whitewashing deposit. BEQA weights the remaining of the feedbacks proportionately to the transaction costs incurred by the submitter. This costs are grown over time using an exponential growth function with an unknown growth rate. This way, BEQA quantifies the weight of a feedback as a function of this growth rate.

Formally, let f_j be the transaction fee of feedback j, A_e be the set of all feedbacks about entity e, and B_i be the set of all feedbacks submitted not later than feedback i by the submitter of i about all entities including but not limited to e. We obtain the publicity of an entity e from (1) where μ_x, σ_x^2, $\mathcal{S}_x$, and $\mathcal{K}_x$ are the first four statistical moments of the set $\{x_i = \sum_{j\in B_i} f_j \mid i \in A_e\}$ as follows:

$$P_e = \left\{ i \in A_e \,\middle|\, x_i \geq \mu_x + \sigma_x(1 - \frac{1}{\mathcal{K}_x})\left(\frac{2}{\pi}\tan^{-1}(-|\mathcal{S}_x|) + 3\right)\right\} \quad (1)$$

We define the weight of a feedback i about an entity e as w_i:

$$w_i = \begin{cases} \sum_{j\in B_i} f_j (1 + r_e)^{t_i - t_j} & i \notin P_e \\ 0 & i \in P_e \end{cases} \quad (2)$$

where t_i and t_j are the timestamps of feedback i and feedback j respectively, P_e is the set containing the *publicity* of entity e, and r_e in (2) is a growth rate that BEQA calculates for each entity dynamically.

Computing the growth rate r_e. Formally, given a set of entities E and publicity set P_e, where $e \in E$, we calculate r_e from (4) where R_0^{max} is obtained using (3).

$$R_0^{max} = \min\left\{ \underset{r_e}{\operatorname{argmax}} \left(\sum_{i\in A_e} w_i < \sum_{i\in P_e} \sum_{j\in B_i} f_j \right) \,\middle|\, e \in E \right\} \quad (3)$$

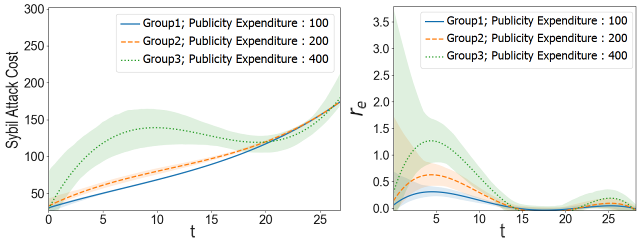
$$r_e = \frac{\sum_{i\in P_e} \sum_{j\in B_i} f_j}{\max\left\{\sum_{i\in P_{e'}} \sum_{j\in B_i} f_j \,\middle|\, e' \in E\right\}} \times R_0^{max} \quad (4)$$

Trust computation. Having obtained r_e for $e \in E$ from (4), we can compute the trustworthiness of each entity using (5).

$$T_e = \sum_{i\in A_e} v_i \cdot w_i \Big/ \sum_{i\in A_e} w_i \quad (5)$$

2 EVALUATION

We simulate a TRM scenario with three groups of entities. The only difference between the groups is their publicity expenditure which we define as the total spend of their publicity influencers. That is, the number of the total transactions of the influencers of an entity is a random variable with a mixture of uniform distributions with different means of 100, 200, and 400 and equal mixture weights. At each time-step of the simulation, each entity interacts with 3 percent of randomly chosen honest users. Then, these users each submits a feedback which incurs the user a transaction fee that we treat as the cost unit. We say a Sybil attack is successful if the feedbacks of the attacker about the target entity outweigh that of the honest users.



(a) (b)

Figure 2: Simulation results: (a) Cost of a successful Sybil attack for entities with different whitewashing deposits at each time step; (b) r_e for entities with different whitewashing deposits at each time step.

We present the experiment results in Fig. 2(a)-(b). The results show that in general a higher publicity expenditure leads to a higher Sybil attack cost. This is a desired property that discourages Sybil attacks, and also encourages higher publicity expenditure as a deposit against white washing.

REFERENCES

[1] Planaria corp. 2019. Planaria: crawl and index the Bitcoin blockchain. Retrieved November 7, 2019 from https://github.com/interplanaria/planaria
[2] Matashard. 2018. Easily access data stored on Blockchain. https://metashard.com/. (Accessed on 6 Feb 2021).
[3] Satoshi Nakamoto. 2019. *Bitcoin: A peer-to-peer electronic cash system.* Technical Report. Manubot.
[4] Michael Bedford Taylor. 2017. The evolution of bitcoin hardware. *Computer* 50, 9 (2017), 58–66.
[5] Craig S Wright. 2008. Bitcoin: A peer-to-peer electronic cash system. *Available at SSRN* 3440802 (2008).