A Mathematical Analysis of Computational Trust Models with The Introduction of Con-man Agents*

Amirali Salehi-Abari and Tony White
School of Computer Science, Carleton University
{*asabari, arpwhite*}*@scs.carleton.ca*

Abstract

Recent work has demonstrated that several trust and reputation models can be exploited by malicious agents with cyclical behaviour if the details of the trust model are known to the malicious agents. In each cycle, a malicious agent with cyclical behaviour first regains a high trust value with a number of cooperations and then misuses its gained trust by engaging in a bad transaction. Using a game theoretic formulation, Salehi-Abari and White have proposed an adaptive trust and reputation model, called AER, that is resistant to exploitation by cyclical behaviour. Their simulation results imply that FIRE, Regret, and a model due to Yu and Singh, can always be exploited with an appropriate value for the period of the cyclical behaviour. Furthermore, their results demonstrate that this is not so for AER. This paper provides a mathematical analysis of the properties of the Yu and Singh scheme, Regret, FIRE, probabilistic trust models, and the AER scheme when faced with cyclical behaviour by malicious agents. Three main results are proven. First, malicious agents can always select a cycle period that allows them to exploit the Yu and Singh model, Regret, FIRE, and probabilistic trust models indefinitely. Second, malicious agents cannot select a single, finite cycle period that allows them to exploit AER forever; their exploitation time is bounded. Finally, the number of cooperations required to achieve a given trust value increases monotonically with each cycle in AER.

1 Introduction

In human and artificial societies trust is a extremely important concept that drives decision making and the lifecycle of relationships. According to Jarvenpaa et al. [7], trust is an essential aspect of any relationship in which the trustor does not have control over the actions of a trustee, the decision is important, and the environment is uncertain. As a consequence of reputation and trust receiving considerable attention in many diverse domains – such as distributed artificial intelligence, computational economics, evolutionary biology, psychology, and sociology – there are many definitions of trust available, each having aspects unique to those domains. Mui et al. define trust as "a subjective expectation an agent has about another's future behavior based on the history of their encounters" [10]. In contrast to trust definitions focussing more on the history of agents' encounters, reputation is based on the aggregated information from other individuals. As an example, Sabater and Sierra [12] declared that "reputation is the opinion or view of someone about something".

Sabater and Sierra [13] created a taxonomy of computational trust and reputation models using several intrinsic features. From their perspective, an important distinction is whether a trust and reputation model is cognitive or game-theoretical in terms of its conceptual model. Furthermore, trust and reputation models might use different sources of information such as direct experiences, witness information, sociological information and prejudice. Witness information is the information that comes from other members of the community whereas sociological information is extracted from the social relations between individuals and their roles in the community. Prejudice is connected to identifying characteristics of individuals (e.g., religious beliefs). Trust and reputation of an individual can be seen either as a global property available to all members of a society (centralized models) or as a subjective property assessed by each individual (decentralized models). Trust and reputation models vary in terms of individual behavior assumptions; in some models, cheating behaviors and malicious individuals are not considered at all whereas in others possible cheating behaviors are taken into account. There are many models of trust, a review of which can be found in [11] and [13].

*Version: Jan 19, 2010.

Regret [12] is a decentralized trust and reputation system oriented to e-commerce environments. The system takes into account three different sources of information: direct experiences, information from third party agents and social structures. Yu and Singh [19] developed an approach for social reputation management, in which they represented an agent's ratings regarding another agent as a scalar and combined them with testimonies using combination schemes similar to certainty factors. Huynh et al. [5, 6] introduced a trust and reputation model called FIRE that incorporates interaction trust, role-based trust, witness reputation, and certified reputation to provide a trust metric.

Considerable progress has recently been made in the development of probabilistic trust models, the Beta Reputation System (BRS) and TRAVOS being two examples [8, 10, 16, 17, 4]. BRS [8] is a probabilistic trust model that works based on the beta distribution. The system is centralized and designed to meet the requirements of online communities. Mui et. al [10] calculates the probability of an agent being trustworthy on the next interaction by considering the frequency of positive and negative direct impressions gathered from the social network. TRAVOS [16, 17] calculates trust using probability theory and takes into account the past interactions and reputation information gathered from third parties while coping with inaccurate reputations. Hang et al. proposed an adaptive probabilistic trust model that combines probability and certainty and offers a trust update mechanism to estimate the trustworthiness of referrers [4].

Fullam et al. [3] have defined the following set of criteria to evaluate trust and reputation models: (1) The model should be multi-dimensional and multi-faceted; (2) Converge quickly; (3) It should precisely model the agent's behavior; (4) Adaptive: the trust value should be adapted if the target's behavior changes; (5) Efficient in terms of computation. Salehi-Abari and White [14] added the exploitation resistance feature to this set. They declared that exploitation resistance "implies that adversaries cannot take advantage of the trust model and its associated systems parameters even when they are known or partially known to adversaries." This paper refines the definition provided in [14] by defining vulnerability classes for a trust model.

Most recently, researchers have identified the existence of cheaters (exploitation) in artificial societies employing trust and reputation models [9, 14, 15], and the existence of inaccurate witnesses [18, 2, 20], and [15]. Kerr and Cohen [9] examined the security of several e-commerce marketplaces employing a trust and reputation system. To this end, they proposed several attacks and examined their effects on each marketplace. Salehi-Abari and White [14] introduced and formally modeled the con-man attack. In the con-man attack, the con-man has cyclical behaviour such that in each cycle he first regains a high trust value with a number of cooperations and then misuses the trust gained by engaging in a bad transaction. Salehi-Abari and White [14] empirically demonstrated the vulnerability of several trust models (i.e., the model due to Yu and Singh, Regret, and FIRE) against this attack. Moreover, their proposed adaptive updating scheme, AER, prevented such exploitation as supported by empirical simulation evidence. However, Salehi-Abari and White could not demonstrate that their results are valid for all parameter settings. More specifically, it was not clear whether the success and failure of the con-man attack against the examined trust models is the result of specific parameter settings or the design (and nature) of those models.

This paper is motivated by the need to develop trust and reputation schemes that have provable properties. While simulation can often provide insights into average case trust and reputation model performance, analytical results based upon known or potential attacks are important to increase confidence in the true utility of such models. It is the authors' belief that widespread deployment of sophisticated trust and reputation models will only occur when such analytical results are forthcoming. To this end, this paper provides mathematical analysis of the con-man attack against prominent trust models such as Yu and Singh's model, Regret, FIRE, probabilistic trust models, and AER.

There are two types of contributions in this paper. To begin, we define what is meant by an attack on a trust and reputation model and what it is meant for such models to be vulnerable to an attack or exhibit exploitation resistance to the attack. The paper includes a description of 3 classes of vulnerability, characterized by model specification. Our principal contributions are analytical and consist of 5 results. First, we prove that the Yu and Singh model can be exploited indefinitely if malicious agents are aware of the model's parameter settings. Second, Regret, FIRE and probabilistic trust models can be exploited indefinitely by malicious agents mounting a con-man attack even when malicious agents are not aware of the model's parameter. Third, malicious agents can not indefinitely exploit AER. Fourth, the number of cooperations required to achieve a given trust value increases monotonically without any upper bound in AER, while this is not true for the other models. Fifth, as forgiveness is a frequently noted aspect of trust and reputation theory [12, 1], it is proven than the AER scheme is forgiving but that forgiveness is slower when several defections have happened.

The remainder of this paper proceeds as follows. Section 2 provides background material and briefly describes four trust and reputation models whose properties in the face of the con-man attack are analyzed in this paper. Section 3 introduces definitions for vulnerability and exploitation resistance and provides a formal model of the con-man attack. We describe our hypotheses and conjectures in Section 4. Sections 5, 6, 7, 8, and 9 present the proofs for the claims. Finally, concluding remarks and future work are explained in Section 10.

2 Background and Terminology

2.1 Direct Interaction Components

Direct interaction is the most popular source of information for trust and reputation models [11]. Different fields have their own interpretation and understanding of direct interaction. In the context of e-commerce, direct interaction might be considered as buying or selling a product, whereas in peer-to-peer systems (e.g., file sharing systems) direct interaction is uploading or downloading files.

Trust and reputation models usually have a direct interaction trust variable that indicates the level of an agent's trustworthiness. This trust value is calculated based on previous direct interactions. We discuss the direct interaction trust components of Yu and Singh's model, Regret, FIRE, and probabilistic trust models in the following subsections.

2.1.1 Yu and Singh

Yu and Singh's [19] trust variable is defined by $T_{i,j}(t)$ indicating the trust rating assigned by agent i to agent j after t interactions between agent i and agent j, with $T_{i,j}(t) \in [-1, +1]$ and $T_{i,j}(0) = 0$.

An agent will update this variable based on the perception of cooperation/defection. Cooperation by the other agents generates positive evidence of α, with $1 > \alpha > 0$ and defection generates negative evidence of β, with $-1 < \beta < 0$. The following trust updating scheme is proposed by [19]:

$$T_{i,j}(t) = \begin{cases} T_{i,j}(t) + \alpha(1 - T_{i,j}(t)) & T_{i,j}(t) \geq 0, \, Cooperation \\ (T_{i,j}(t) + \alpha)/(1 - min(|T_{i,j}(t)|, |\alpha|)) & T_{i,j}(t) < 0, \, Cooperation \\ (T_{i,j}(t) + \beta)/(1 - min(|T_{i,j}(t)|, |\beta|)) & T_{i,j}(t) > 0, \, Defection \\ T_{i,j}(t) + \beta(1 + T_{i,j}(t)) & T_{i,j}(t) \leq 0, \, Defection \end{cases} \tag{1}$$

2.1.2 Regret

Regret uses the term subjective reputation (direct trust) to talk about the trust calculated directly from an agent's impressions. Regret defines an impression as the subjective evaluation made by an agent on a certain aspect of an outcome. The variable $r_{i,j}(t)$, with $r_{i,j}(t) \in [-1, 1]$, is the rating associated with the impression of agent i about agent j as a consequence of specific outcome at time t. $R_{i,j}$ is the set of all $r_{i,j}(t)$ for all possible t. A subjective reputation at time t from agent i's point of view regarding agent j is noted as $T_{i,j}(t)$[1]. To calculate $T_{i,j}(t)$, Regret uses a weighted mean of the impressions' rating factors, giving more importance to recent impressions. Intuitively, a more recent rating is weighted more than those that are less recent. The formula to calculate $T_{i,j}(t)$ is:

$$T_{i,j}(t) = \sum_{w_k \in R_{i,j}} \rho(t, t_k).w_k \tag{2}$$

where t_k is the time that w_k is recorded, t is the current time, $\rho(t, t_k) = \frac{f(t_k,t)}{\sum_{r_l \in W_{i,j}} f(t_l,t)}$, and $f(t_k, t) = \frac{t_k}{t}$ which is called the rating recency function.

2.1.3 FIRE

FIRE utilizes the direct trust component of Regret but does not use the rating recency function of Regret, the method used to calculate the weights for each rating. The rating recency function of Regret does not actually reflect a rating's recency. Therefore, FIRE introduced a new rating recency function based on the time difference between current time and the rating time. The parameter λ is introduced into the rating recency function to scale time values. As a result, this parameter makes the rating

[1]For the purpose of simplification, we have changed the original notations from [12].

recency function adjustable to suit the time granularity in different applications. FIRE's rating recency function is given by the following formula:

$$f(t_k, t) = e^{-\frac{t-t_k}{\lambda}} \tag{3}$$

2.1.4 Probabilistic Trust Models

Probabilistic trust models are built based on observations of past interactions between agents. Trust is calculated using probability theory and takes into account the past interactions. They usually simplify the outcome of an interaction by providing a binary rating, where 1 and 0 represent successful and unsuccessful interactions (cooperation and defection) respectively.

In probabilistic trust models, the probability that agent j satisfies its obligations for agent i is expressed by $B_{i,j}$. The trust value of agent i for agent j at time t, denoted by $T_{i,j}(t)$, is the expected value of $B_{i,j}$ given the set of outcomes $O_{i,j}(t)$ at time t. Note that, each outcome can be either cooperation or defection.

$$T_{i,j}(t) = E[B_{i,j}|O_{i,j}(t)] \tag{4}$$

The expected value of a continuous random variable is dependent on the probability density function (pdf) used to model. Probabilistic trust models usually utilize the beta family of probability density functions (PDF) to model the probability of having a successful interaction with a particular given agent. The general formula of beta distributions is presented in Equation 5. The values of α and β define the shape of the density function.

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du}, \;\; where \;\; \alpha, \beta > 0 \tag{5}$$

As the standard equation for the expected value of a beta distribution is $E[B|\alpha, \beta] = \frac{\alpha}{\alpha+\beta}$, the trust value $T_{i,j}(t)$ after t interactions will be calculated by Equation 6.

$$T_{i,j}(t) = E[B_{i,j}|\alpha, \beta] = \frac{\alpha}{\alpha+\beta} \tag{6}$$

where $\alpha = n_c(t)+1$ and $\beta = n_d(t)+1$. $n_c(t)$ and $n_d(t)$ denote the number of cooperations (successful interactions) and the number of defections (unsuccessful interactions)[2].

3 Definitions

In this paper we consider an agent's trust and reputation model, M, to be characterized by two attributes, S and P; S is the trust and reputation strategy being employed and P is the set of parameter values that are used to operate it. This paper deals with the concept of vulnerability. We define more precisely *vulnerability* and the levels of vulnerability of trust models against an attack as follows:

Definition Attack. An attack, A, is a sequence of cooperations and defections used by a malicious agent, *ma*, to achieve or maintain a trustworthy status as maintained by an agent, *ta*, with which it is interacting.

Definition Vunerability. A trust model, M, is vulnerable to an attack, A, if a malicious agent, *ma*, adopting some strategy and with full or partial knowledge of an agent, *ta*, and its associated trust model, M, can be trustworthy as determined by *ta*.

We define the following levels of vulnerability in this paper:

Definition Low-level. A trust model, M, is vulnerable to an attack, A, with low-level risk, if it is vulnerable only for some specific model parameter settings and *ma* needs to be aware of the parameter values used by *ta* to mount a successful attack.

Definition Medium-level. A trust model, M, is vulnerable to an attack, A, with medium-level risk, if it is vulnerable for any parameter settings and *ma* needs to be aware of the value of parameters used by *ta* to successfully mount an attack.

[2]It is worth mentioning that the trust value in probabilistic models is in the range of [0, 1] as opposed to Yu and Singh, Regret, and FIRE models in which trust is in the range of [-1, 1].

Definition High-level. A trust model, M, is vulnerable to an attack, A, with high-level risk, if ma is able to successfully mount an attack under any conditions even when ma is not aware of the values of parameters.

Finally, we say that a trust and reputation model, M, exhibits exploitation resistance to an attack, A, if it is not vulnerable to that attack. We also refer to a trust and reputation model, M, as being exploitation resistant when faced with an attack, A.

3.1 Con-man Attack and Terminology

The con-man attack introduced and modeled by Salehi-Abari and White [14] has been applied to direct trust components of several trust models and are shown to be vulnerable to it. In the con-man attack, a con-man usually takes advantage of someone else and attempts to defraud that person by gaining their confidence. The con-man attack is modeled by introducing the parameter θ. The con-man will defect after cooperating θ times. After each defection, the con-man will again cooperate θ times. The con-man will repeat this interaction pattern several times (maybe, forever).

In this paper, there is a slight modification in the con-man interaction pattern when compared to [14]. Here, the con-man has a higher level of intelligence such that it will defect in an interaction with the victim agent whenever its trust value is equal to or greater than a threshold, denoted by T_c. In other words, the con-man will cooperate until its trust value reaches T_c. We formally model this interaction pattern with Equation 7:

$$L = \left\{(C^{\theta_i}D)^{+} | i = 0...n, \theta_i \in \mathbb{N}\right\} \tag{7}$$

Where C and D represent cooperation and defection respectively. The main difference in this interaction pattern when compared with that presented in [14] is that θ_i is subject to change for each cycle of cooperation and defection instead of being a constant. The value of θ_i is determined by the number of the cooperations which the con-man needs in order to increase its trust value above T_c.

For the purpose of simplification in the proofs which follow, we rewrite the interaction pattern in such a way that the i^{th} cycle of interactions starts with a defection and followed by θ_i cooperations. In this sense, the first cooperations of con-man, θ_0, which result in an increment of trust from T_0 to T_c are not modeled. In other words, we consider the con-man has already built up its trust to T_c from T_0 by θ_0 cooperations. The modified interaction pattern of the con-man is presented in Equation 8.

$$L = \left\{(DC^{\theta_i})^{+} | i = 1...n, \theta_i \in \mathbb{N}\right\} \tag{8}$$

More precisely, we herein highlight the terminology that is used in this paper and is illustrated in Figure 1. The variable θ_i is the number of cooperations that the con-man will have in the i^{th} cycle of defection-cooperations. The i^{th} cycle includes the i^{th} defection and θ_i cooperations. The trust value at the end of the i^{th} cycle is T_c or greater; i.e., T_c defines the criterion for ending the i^{th} cycle. The trust value of the con-man before the i^{th} defection is denoted by $T_b(i)$. The trust value of the con-man after the i^{th} defection is denoted by $T_d(i)$.

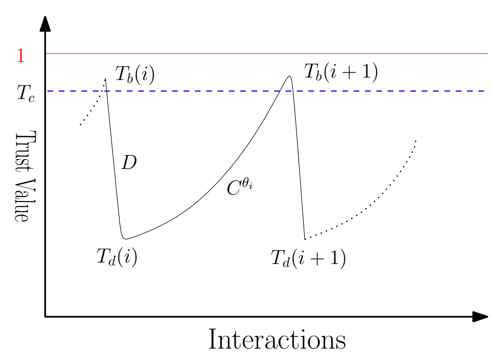

Figure 1: Trust value changes in the i^{th} cycle of defection-cooperations

4 Hypothesis and Conjectures

This paper intends to prove that:

- The model due to Yu and Singh is vulnerable to the con-man attack with medium-level risk.
- Regret is vulnerable to the con-man attack with high-level risk.
- FIRE is vulnerable to the con-man attack with medium-level risk.
- The probabilistic trust models are Regret is vulnerable to the con-man attack with high-level risk.
- AER is not vulnerable to the con-man attack (i.e., it will not let the con-man regain a high trust value easily with the same or a smaller number of cooperations). Moreover, it motivates the con-man to be more cooperative (i.e., $\theta_{i+1} > \theta_i$). In other words, the con-man needs a higher number of cooperations in each cycle of defection-cooperations in order to reach a specific trust threshold, T_c.

5 Yu and Singh Model

In this section, we will prove that Yu and Singh's model [19] as presented in Equation 1 is vulnerable to the con-man attack. To simplify our analysis and our proofs, we will rewrite the update scheme as presented in Equation 9.

$$T_{n+1} = \begin{cases} T_n + \alpha(1 - T_n) & T_n \geq 0 \quad , Cooperation \\ (T_n + \alpha)/(1 + T_n) & -\alpha < T_n < 0 \ , Cooperation \\ (T_n + \alpha)/(1 - \alpha) & T_n \leq -\alpha \quad , Cooperation \\ (T_n + \beta)/(1 - T_n) & 0 < T_n < -\beta \ , Defection \\ (T_n + \beta)/(1 + \beta) & T_n \geq |\beta| \quad , Defection \\ T_n + \beta(1 + T_n) & T_n \leq 0 \quad , Defection \end{cases} \tag{9}$$

Our proof strategy is to find θ_i for any parameter setting in any defection-cooperations cycle and show that $\theta_{i+1} \leq \theta_i$. It is not straightforward to calculate θ_i for the Yu and Singh model since it includes several recurrent formulae as shown in Equation 9. The proof is broken down into a series of cases that mirror the distinct forms of Equation 9. We find the explicit, or closed, form for the first form of Equation 9 by using Lemma 11.1 provided in the Appendix. Lemmas, 5.6, 5.7 and 5.8 present results for the cases shown in Figures 2, 3 and 4 respectively. Theorem 5.1 provides a closed form solution for the number of cooperations that are required in a con-man attack to reach a given level of trust that is simply a function of α and β. This results in Theorem 5.2 that proves the middle-level risk vulnerability of this trust model.

Lemma 5.1 *Let $T_s \geq 0$ be the starting trust value, the trust value after n cooperations will be calculated by $T_n = (T_s - 1) \times (1 - \alpha)^n + 1$ while using Yu and Singh Model. Note that, $\alpha > 0$ is constant over n cooperations.*

Proof As $T_s \geq 0$, the trust value for each cooperation will be updated by the first recurrent formula of Equation 1 (i.e., $T_{i+1} = T_i + \alpha(1 - T_i)$). We will rearrange this formula as follows:

$$T_{i+1} = T_i + \alpha(1 - T_i) \tag{10}$$
$$= T_i + \alpha - \alpha T_i \tag{11}$$
$$= (1 - \alpha)T_i + \alpha \tag{12}$$

Equation 12 is the recurrence equation with the same form as Equation 26 when b and d are substituted by $(1 - \alpha)$ and α respectively. By using Lemma 11.1, the closed form solution for T_n using Equation 12 is as follows:

$$\begin{aligned} T_n &= \left(T_s + \frac{\alpha}{(1-\alpha) - 1}\right) \times (1-\alpha)^n + \frac{-\alpha}{(1-\alpha) - 1} \\ &= \left(T_s + \frac{\alpha}{-\alpha}\right) \times (1-\alpha)^n + \frac{-\alpha}{-\alpha} \\ &= (T_s - 1) \times (1-\alpha)^n + 1 \quad \blacksquare \end{aligned}$$

Lemma 5.2 *Let $-\alpha < T_s < 0$ be the starting trust value, the trust value after 1 cooperation will be $0 < T_{I^+} < \alpha$, given by $T_{I^+} = (T_s + \alpha)/(1 + T_s)$.*

Proof As $-\alpha < T_s < 0$, the trust value after one cooperation will be updated based on the third formula in Equation 9. Therefore, the value of $T_{I^+} = (T_s+\alpha)/(1+T_s)$. We can consider that T_{I^+} is the function of T_s since only T_s is a variable changing in the range of $-\alpha < T_s < 0$ and the value of α will not be changed by this cooperation. we can consider the following function for the calculation of T_{I^+}:

$$f(x) = \frac{x+\alpha}{1+x} \quad , \quad -\alpha < x < 0 \tag{13}$$

The first derivation of $f(x)$ denoted by $\acute{f}(x)$ can be calculated by as follows:

$$\acute{f}(x) = \frac{(1+x)-(x+\alpha)}{(1+x)^2} = \frac{1-\alpha}{(1+x)^2} > 0 \tag{14}$$

As $\acute{f}(x) > 0$, $\acute{f}(x)$ is ever-increasing and we have $f(-\alpha) < f(x) < f(0)$ which is equal to, $0 < f(x) < \alpha$. Therefore, $0 < T_{I^+} < \alpha$ for any value of $-\alpha < T_s < 0$. ∎

Lemma 5.3 *Let $T_s \leq -\alpha$ be the starting trust value, the trust value after n cooperations can be calculated by $T_n = (T_s+1) \times (\frac{1}{1-\alpha})^n - 1$, if $T_n < 0$. Note that, $\alpha > 0$ is constant over n cooperations.*

Proof As $T_s \leq -\alpha$, for the consecutive cooperations, the trust value will be updated by the third recurrent formula of Equation 1 (i.e., $T_{i+1} = (T_i+\alpha)/(1-\alpha)$) until the trust value reaches the value $-\alpha$. We will rearrange this formula as follows:

$$\begin{aligned} T_{i+1} &= \frac{T_i+\alpha}{1-\alpha} \\ &= \frac{1}{1-\alpha}T_i + \frac{\alpha}{1-\alpha} \end{aligned}$$

$$T_{i+1} = \frac{1}{1-\alpha}T_i + \frac{\alpha}{1-\alpha} \tag{15}$$

Equation 15 is the recurrence equation with the same form as Equation 26 when b and d are substituted by $\frac{1}{1-\alpha}$ and $\frac{\alpha}{1-\alpha}$ respectively. By using Lemma 11.1, the explicit form of the Equation 15 can be calculated as follows:

$$\begin{aligned} T_n &= \left(T_s + \frac{\frac{\alpha}{1-\alpha}}{\frac{1}{1-\alpha}-1}\right) \times (\frac{1}{1-\alpha})^n + \frac{-\frac{\alpha}{1-\alpha}}{\frac{1}{1-\alpha}-1} \\ &= \left(T_s + \frac{\frac{\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha}}\right) \times (\frac{1}{1-\alpha})^n + \frac{-\frac{\alpha}{1-\alpha}}{\frac{\alpha}{1-\alpha}} \\ &= (T_s+1) \times (\frac{1}{1-\alpha})^n - 1 \end{aligned}$$

Note that, the updated trust value found by using the third recurrent formula of Equation 1 should be in the range of $(-1, 0]$ (i.e., $T_n < 0$). Therefore, this explicit form can be replaced with the recurrent one as long as $T_n < 0$. In other words, for those n and T_s that $T_n < 0$, this explicit formula is valid. ∎

Definition $\theta_i^+(T_s, T_t)$ specifies the number of cooperations for incrementing the trust value from T_s to the target trust value, denoted by T_t, where $T_s \geq 0$, $T_t > T_s$, and the agent is in the i^{th} cycle of defection-cooperations.

Definition $\theta_i^-(T_s, T_t)$ specifies the number of cooperations for incrementing the trust value from T_s to the target trust value, denoted by T_t, where $T_s < -\alpha$, $T_s < T_t$, $T_t < 0$, and the agent is in the i^{th} cycle of defection-cooperations.

Lemma 5.4 *Let $T_s \geq 0$ and $T_t > T_s$ be the starting trust value and target trust value respectively, then $\theta_i^+(T_s, T_t) = \frac{\ln\left(\frac{T_t-1}{T_s-1}\right)}{\ln(1-\alpha)}$. Note that, $\alpha > 0$ is a constant over $\theta_i^+(T_s, T_t)$ cooperations.*

Proof As $T_s > 0$, We can find $\theta_i^+(T_s, T_t)$ using Lemma 5.1, where T_n should be replaced by T_t.

$$
\begin{aligned}
&T_t = (T_s - 1) \times (1-\alpha)^{\theta_i^+} + 1 \Longrightarrow \\
&T_t - 1 = (T_s - 1) \times (1-\alpha)^{\theta_i^+} \Longrightarrow \\
&\frac{T_t - 1}{T_s - 1} = (1-\alpha)^{\theta_i^+} \Longrightarrow \\
&\ln\left(\frac{T_t - 1}{T_s - 1}\right) = \ln((1-\alpha)^{\theta_i^+}) \Longrightarrow \\
&\ln\left(\frac{T_t - 1}{T_s - 1}\right) = \theta_i^+ \times \ln(1-\alpha) \Longrightarrow
\end{aligned}
$$

$$
\theta_i^+(T_s, T_t) = \frac{\ln\left(\frac{T_t-1}{T_s-1}\right)}{\ln(1-\alpha)} \quad \blacksquare \tag{16}
$$

Lemma 5.5 *Let $T_s < -\alpha$ and $T_t > T_s$ be the starting trust value and target trust value respectively. If $T_t < 0$ then $\theta_i^-(T_s, T_t) = \frac{\ln\left(\frac{T_s+1}{1+T_t}\right)}{\ln(1-\alpha)}$. Note that, $\alpha > 0$ is a constant over $\theta_i^-(T_s, T_t)$ cooperations.*

Proof As $T_s < -\alpha$ and $T_t < 0$, We can find $\theta_i^-(T_s, T_t)$ using Lemma 5.3, where T_n should be replaced by T_t.

$$
\begin{aligned}
&T_t = (T_s + 1) \times (\frac{1}{1-\alpha})^{\theta_i^-} - 1 \Longrightarrow \\
&1 + T_t = (T_s + 1) \times (\frac{1}{1-\alpha})^{\theta_i^-} \Longrightarrow \\
&\frac{1 + T_t}{T_s + 1} = (\frac{1}{1-\alpha})^{\theta_i^-} \Longrightarrow \\
&\frac{T_s + 1}{1 + T_t} = (1-\alpha)^{\theta_i^-} \Longrightarrow \\
&\ln\left(\frac{T_s + 1}{1 + T_t}\right) = \theta_i^- \times \ln(1-\alpha) \Longrightarrow \\
&\theta_i^-(T_s, T_t) = \frac{\ln\left(\frac{T_s+1}{1+T_t}\right)}{\ln(1-\alpha)} \quad \blacksquare
\end{aligned}
$$

Lemma 5.6 *Given that $T_c > 0$ and $T_c \geq |\beta|$, then $\theta_i = \frac{\ln(1+\beta)}{\ln(1-\alpha)}$.*

Proof When the con-man does the i^{th} defection, its trust is updated by $T_d(i) = (T_b(i) + \beta)/(1 - min(T_b(i), |\beta|))$. As $T_b(i) \geq T_c > 0$, We simplify the formula by replacing $T_b(i)$ with its lower bound T_c. In this sense, we have $T_d(i) = (T_c + \beta)/(1 - min(T_c, |\beta|))$. As $T_c \geq |\beta|$ based on the assumption of this lemma, $T_d(i)$ will be calculated based on $T_d(i) = (T_c + \beta)/(1 + \beta)$ and $T_d(i) > 0$ because $T_c + \beta > 0$ and $1 + \beta > 0$ as illustrated in Figure 2.

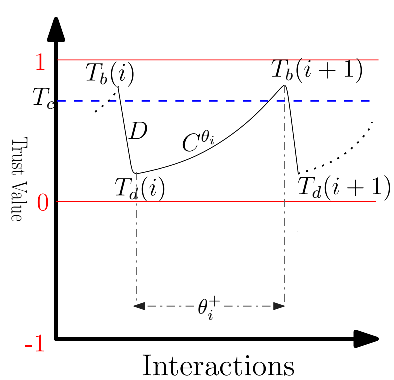

Figure 2: $T_c \geq |\beta|$

As $T_d(i) > 0$, $\theta_i = \theta_i^+(T_s, T_t)$ using Lemma 5.4 where T_s and T_t should be substituted by $T_d(i)$ and T_c respectively.

$$\theta_i = \theta_i^+(T_d(i), T_c) \Longrightarrow$$
$$\theta_i = \frac{\ln\left(\frac{T_c - 1}{T_d(i) - 1}\right)}{\ln(1-\alpha)}$$

By replacing $T_d(i)$ with $(T_c + \beta)/(1+\beta)$, and simplifying we have:

$$\begin{aligned}
\theta_i &= \frac{\ln\left(\frac{T_c - 1}{(T_c+\beta)/(1+\beta) - 1}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{T_c - 1}{(T_c-1)/(1+\beta)}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{(T_c-1)(1+\beta)}{(T_c-1)}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln(1+\beta)}{\ln(1-\alpha)} \quad \blacksquare
\end{aligned}$$

Lemma 5.7 *Given that* $0 < T_c < |\beta|$ *and* $-\alpha < T_d(i)$*, then* $\theta_i = \frac{\ln(1+\beta)}{\ln(1-\alpha)}$.

Proof We herein attempt to find θ_i when $T_c \leq |\beta|$. In this case, according to Equation 9, $T_d(i)$ will be calculated based on $T_d(i) = (T_c + \beta)/(1 - T_c)$ and $T_d(i) \leq 0$ because $T_c + \beta \leq 0$ and $1 - T_c > 0$. Moreover, we consider that $T_d(i) > -\alpha$ as illustrated in Figure 3.

Using Lemma 5.2, since $-\alpha < T_d(i) < 0$, the trust value after 1 cooperation will be $0 < T_{I^+} < \alpha$ where T_{I^+} is calculated by $T_{I^+} = (T_d(i) + \alpha)/(1 + T_d(i))$. Afterward, $\theta_i^+(T_{I^+}, T_c)$ cooperations are needed to reach the trust value of T_c. As $T_{I^+} > 0$, $\theta_i^+(T_{I^+}, T_c)$ can be calculated by using Lemma 5.4 where T_s and T_t should be substituted by T_{I^+} and T_c respectively.

$$\theta_i^+(T_{I^+}, T_c) = \frac{\ln\left(\frac{T_c - 1}{T_{I^+} - 1}\right)}{\ln(1-\alpha)}$$

We will calculate the exact value of T_{I^+} by considering the fact that $T_d(i) = (T_c+\beta)/(1-T_c)$ as follows:

$$\begin{aligned}
T_{I^+} &= \frac{T_d(i) + \alpha}{1 + T_d(i)} \\
&= \frac{\frac{T_c+\beta}{1-T_c} + \alpha}{1 + \frac{T_c+\beta}{1-T_c}} \\
&= \frac{T_c + \beta + \alpha(1 - T_c)}{1 - T_c + T_c + \beta} \\
&= \frac{T_c + \beta + \alpha(1 - T_c)}{1 + \beta}
\end{aligned}$$

By replacing the value of T_{I^+} in the formula for $\theta_i^+(T_{I^+}, T_c)$, we have:

$$\begin{aligned}
\theta_i^+(T_{I^+}, T_c) &= \frac{\ln\left(\frac{T_c - 1}{T_{I^+} - 1}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{T_c - 1}{\frac{T_c+\beta+\alpha(1-T_c)}{1+\beta} - 1}\right)}{\ln(1-\alpha)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\ln\left(\frac{(T_c-1)(1+\beta)}{T_c+\beta+\alpha(1-T_c)-1-\beta}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{(T_c-1)(1+\beta)}{T_c+\alpha(1-T_c)-1}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{(T_c-1)(1+\beta)}{(T_c-1)-\alpha(T_c-1)}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{(T_c-1)(1+\beta)}{(T_c-1)(1-\alpha)}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(\frac{1+\beta}{1-\alpha}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln(1+\beta)-\ln(1-\alpha)}{\ln(1-\alpha)} \\
&= \frac{\ln(1+\beta)}{\ln(1-\alpha)} - 1
\end{aligned}
$$

As $\theta_i = 1 + \theta_i^+(T_{I^+}, T_c)$, we have :

$$
\begin{aligned}
\theta_i &= 1 + \frac{\ln(1+\beta)}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(1+\beta)}{\ln(1-\alpha)} \quad \blacksquare
\end{aligned}
$$

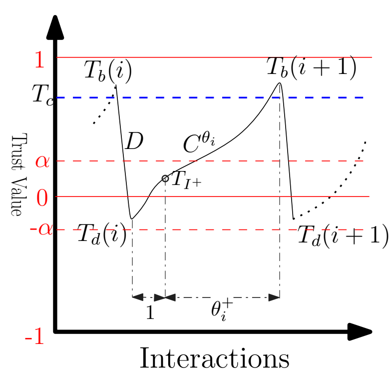

Figure 3: $T_c < |\beta|$ and $T_d(i) > -\alpha$

Lemma 5.8 *Given that* $0 < T_c < |\beta|$ *and* $-\alpha \geq T_d(i)$*, then* $\theta_i = \frac{\ln(1+\beta)}{\ln(1-\alpha)}$.

Proof As $T_c \leq |\beta|$, $T_d(i)$ will be calculated based on $T_d(i) = (T_c + \beta)/(1 - T_c)$ and $T_d(i) \leq 0$ because $T_c + \beta \leq 0$ and $1 - T_c > 0$. Moreover, we consider that $T_d(i) \leq -\alpha$. As shown in Figure 4, the trust value will be increased to T_{I^-} after $\theta_i^-(T_d(i), -\alpha)$ cooperations, then with a cooperation the trust value be increased to T_{I^+}. Afterward, $\theta_i^+(T_{I^+}, T_c)$ cooperations are needed to reach the trust value of T_c or above. Therefore, $\theta_i = \theta_i^-(T_d(i), -\alpha) + 1 + \theta_i^+(T_{I^+}, T_c)$.

As $T_d(i) < -\alpha$ and $-\alpha < 0$, $\theta_i^-(T_d(i), -\alpha)$ can be calculated by using Lemma 5.5 where T_s and T_t should be substituted by $T_d(i)$ and $-\alpha$ respectively in the original formula. Dropping the $T_d(i)$ and $-\alpha$ arguments from $\theta_i^-(T_d(i), -\alpha)$, we see that:

$$
\theta_i^- = \frac{\ln(T_d(i)+1)}{\ln(1-\alpha)} - 1
$$

By replacing $T_d(i)$ with $(T_c + \beta)/(1 - T_c)$ and simplifying we have:

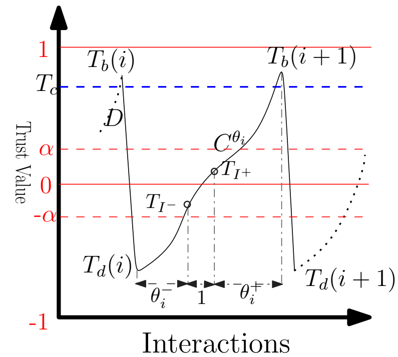

Figure 4: $T_c < |\beta|$ and $T_d(i) \leq -\alpha$

$$
\begin{aligned}
\theta_i^- &= \frac{\ln(T_d(i)+1)}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(\frac{T_c+\beta}{1-T_c}+1)}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(\frac{T_c+\beta+1-T_c}{1-T_c})}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(\frac{\beta+1}{1-T_c})}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(\beta+1) - \ln(1-T_c)}{\ln(1-\alpha)} - 1 \\
&= \frac{\ln(\beta+1)}{\ln(1-\alpha)} - \frac{\ln(1-T_c)}{\ln(1-\alpha)} - 1
\end{aligned}
$$

The trust value after θ_i^- cooperations will be T_I^-. T_I^- is in the range of $[-\alpha, 0]$. Using Lemma 5.2, the trust value after 1 cooperation will be $0 < T_{I^+} < \alpha$ where T_{I^+} is calculated by $T_{I^+} = (T_{I^-}+\alpha)/(1+T_{I^-})$. We consider the worst case by considering that $T_I^+ = 0$. Afterward, $\theta_i^+(0, T_c)$ cooperations are needed to reach the trust value of T_c. $\theta_i^+(T_{I^+}, T_c)$ can be calculated by using Lemma 5.4 where T_s and T_t should be substituted by 0 and T_c respectively in the original formula.

$$
\begin{aligned}
\theta_i^+(0, T_c) &= \frac{\ln\left(\frac{T_c-1}{0-1}\right)}{\ln(1-\alpha)} \\
&= \frac{\ln\left(1-T_c\right)}{\ln(1-\alpha)}
\end{aligned}
$$

Now, we can calculate θ_i as follows:

$$
\begin{aligned}
\theta_i &= \theta_i^-(T_d(i), T_{I^-}) + 1 + \theta_i^+(T_{I^-}, T_{I^+}) \\
&= \left(\frac{\ln(\beta+1)}{\ln(1-\alpha)} - \frac{\ln(1-T_c)}{\ln(1-\alpha)} - 1\right) + 1 + \frac{\ln\left(1-T_c\right)}{\ln(1-\alpha)} \\
&= \frac{\ln(\beta+1)}{\ln(1-\alpha)} \quad \blacksquare
\end{aligned}
$$

Theorem 5.1 *If $1 > T_c > 0$, θ_i will be calculated by $\frac{\ln(\beta+1)}{\ln(1-\alpha)}$ for Yu and Singh's model.*

Proof Using both Lemma 5.7 and Lemma 5.8, $\theta_i = \frac{\ln(\beta+1)}{\ln(1-\alpha)}$ for $T_c < |\beta|$. Moreover, Lemma 5.6 demonstrates that $\theta_i = \frac{\ln(\beta+1)}{\ln(1-\alpha)}$ for $T_c \geq |\beta|$. Therefore, it is proved that θ_i for any values of $T_c > 0$, α, and β is calculated by Equation 17.

$$\theta_i = \frac{\ln(\beta+1)}{\ln(1-\alpha)} \quad \blacksquare \tag{17}$$

Theorem 5.2 *Yu and Singh model is vulnerable to the con-man attack with a medium-level risk for any values of α and β.*

Proof Referring to Lemma 5.1, θ_i for any values of T_c, α, and β is calculated by Equation 17. The value of θ_i only depends on the value of α and β. As α and β are constants in Yu and Singh's model, so $\theta_i = \theta_{i+1}$ for $i = 1, 2, ..., N$. Therefore, there is a constant $\theta_{critical}$ that the con-man can select in order to be recognized as trustworthy despite being a con-man. Moreover, the con-man can regain the lost trust value with the same number of cooperations in each cycle of defection-cooperations. As a result, Yu and Singh's model is vulnerable to the con-man attack for any values of $T_c > 0$, α and β. $\blacksquare$

6 The Regret Model

We here will prove that the Regret model [12] as presented in Equation 2 is vulnerable to the con-man attack. Our proof strategy is to find θ_i for any parameter setting in any defection-cooperations cycle and show that $\theta_{i+1} \leq \theta_i$. Throughout our analysis and proofs, cooperation and defection are mapped to 1 and -1 respectively for the Regret model and the appropriate value is used as an input of the trust model. To simplify our analysis and proofs, we will rewrite the update scheme as presented in Equation 18.

$$T(t) = \sum_{i=1}^{t} \rho(t,i) \times w_i \tag{18}$$

where i is the time that w_i is recorded, t is the current time, $\rho(t,i) = \frac{f(i,t)}{\sum_{i=1}^{t} f(i,t)}$, and $f(i,t) = \frac{i}{t}$.

Lemma 6.1 *Let be $\rho(t,i) = \frac{f(i,t)}{\sum_{j=1}^{t} f(j,t)}$ and $f(i,t) = \frac{i}{t}$ in the Regret model, then $\rho(t,i) = \frac{2i}{t(t+1)}$*

Proof

$$\begin{aligned}
\rho(t,i) &= \frac{f(i,t)}{\sum_{j=1}^{t} f(j,t)} \\
&= \frac{\frac{i}{t}}{\sum_{j=1}^{t} \frac{j}{t}} \\
&= \frac{i}{\sum_{j=1}^{t} j} \\
&= \frac{i}{\frac{t(t+1)}{2}} \\
&= \frac{2i}{t(t+1)} \quad \blacksquare
\end{aligned}$$

Lemma 6.2 *The Regret model can be updated by $T(t) = \frac{2}{t(t+1)} \sum_{i=1}^{t} i \times w_i$*

Proof Using Equation 18 and Lemma 6.1, the updating schema for Regret can be written as follows:

$$\begin{aligned}
T(t) &= \sum_{i=1}^{t} \rho(t,i) \times w_i \\
&= \sum_{i=1}^{t} \frac{2i}{t(t+1)} \times w_i \\
&= \frac{2}{t(t+1)} \sum_{i=1}^{t} i \times w_i \quad \blacksquare
\end{aligned}$$

Lemma 6.3 *Let $T(t-1)$ be the trust value after the $t-1^{th}$ interactions and w_t be the t^{th} rating then $T(t) = \frac{1}{t+1}\left((t-1)T(t-1) + 2w_t\right)$.*

Proof Using Lemma 6.2, we can show that $\frac{T(t-1)\times(t-1)\times t}{2} = \sum_{i=1}^{t-1} i \times w_i$. Moreover, we can rewrite $T(t)$ as follows:

$$
\begin{aligned}
T(t) &= \frac{2}{t(t+1)}\sum_{i=1}^{t} i \times w_i \\
&= \frac{2}{t(t+1)}\left(\sum_{i=1}^{t-1} i \times w_i + t.w_t\right) \\
&= \frac{2}{t(t+1)}\left(\frac{T(t-1)\times(t-1)\times t}{2} + t.w_t\right) \\
&= \frac{1}{t+1}\left(T(t-1)\times(t-1) + 2w_t\right) \quad \blacksquare
\end{aligned}
$$

Lemma 6.4 *Let $T(t)$ be the trust value after the t^{th} interactions, the trust value after θ cooperations will be calculated by $T(t+\theta) = \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times (T(t)-1) + 1$.*

Proof Using Lemma 6.2, we have:

$$
\begin{aligned}
T(t+\theta) &= \frac{2}{(t+\theta)(t+\theta+1)}\sum_{i=1}^{t+\theta} i \times w_i \\
&= \frac{2}{(t+\theta)(t+\theta+1)}\left(\sum_{i=1}^{t} i \times w_i + \sum_{i=t+1}^{t+\theta} i \times w_i\right)
\end{aligned}
$$

Using Lemma 6.2, we know that $\frac{T(t)\times(t+1)\times t}{2} = \sum_{i=1}^{t} i \times w_i$. Therefore, we have:

$$
T(t+\theta) = \frac{2}{(t+\theta)(t+\theta+1)}\left(\frac{T(t)\times(t+1)\times t}{2} + \sum_{i=t+1}^{t+\theta} i \times w_i\right)
$$

As the last θ interactions were cooperations, therefore $w_i = 1$ for $t < i \leq t+\theta$. Therefore we have:

$$
\begin{aligned}
T(t+\theta) &= \frac{2}{(t+\theta)(t+\theta+1)}\left(\frac{T(t)\times(t+1)\times t}{2} + \sum_{i=t+1}^{t+\theta} i\right) \\
&= \frac{2}{(t+\theta)(t+\theta+1)}\left(\frac{T(t)\times(t+1)\times t}{2} + \frac{\theta(2t+\theta+1)}{2}\right) \\
&= \left(\frac{T(t)\times(t+1)\times t}{(t+\theta)(t+\theta+1)} + \frac{\theta(2t+\theta+1)}{(t+\theta)(t+\theta+1)}\right) \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + \frac{\theta(2t+\theta+1)}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + \frac{(\theta+t-t)(t+t+\theta+1)}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + \frac{((t+\theta)-t)((t+\theta+1)+t)}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + \frac{(t+\theta)(t+\theta+1)+t(t+\theta)-t(t+\theta+1)-t^2}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + \frac{(t+\theta)(t+\theta+1)-t-t^2}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times T(t) + 1 + \frac{-t(t+1)}{(t+\theta)(t+\theta+1)} \\
&= \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times (T(t)-1) + 1 \quad \blacksquare
\end{aligned}
$$

Lemma 6.5 *Let t_i be the total number of interactions as of the beginning of the i^{th} cycle of defection-cooperations, $t_1 = \theta_0$ and $t_{i+1} = t_i + \theta_i + 1$.*

Proof As mentioned before, θ_0 interactions are necessary in order that the trust value increases from T_0 (initial trust value) to T_c. Therefore, $t_1 = \theta_0$. Moreover, the i^{th} cycle of defection-cooperations includes 1 defection and θ_i cooperations. Therefore, at the beginning of the $i+1^{th}$ cycle, the number of interactions is equal to $t_{i+1} = t_i + \theta_i + 1$. $\blacksquare$

Lemma 6.6 *Let $T(t)$ and T_{ta} be the starting trust value and the target trust value as a result of θ cooperations respectively, where t is the number of interactions used for calculation of $T(t)$. The value of θ is calculated by $\theta = \frac{-1-2t+\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2}$.*

Proof We can find θ using Lemma 6.4, where T_{ta} should be replaced by $T(t+\theta)$, so we have:

$$T_{ta} = \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times (T(t)-1) + 1 \Longrightarrow$$
$$T_{ta} - 1 = \frac{(t+1)t}{(t+\theta)(t+\theta+1)} \times (T(t)-1) \Longrightarrow$$
$$(t+\theta)(t+\theta+1) = \frac{(t+1)t}{T_{ta}-1} \times (T(t)-1)$$

Let $t+\theta = x$, then we have:

$$x(x+1) = \frac{(t+1)t}{T_{ta}-1} \times (T(t)-1)$$
$$x^2 + x - \frac{(t+1)t}{T_{ta}-1} \times (T(t)-1) = 0$$

By solving this quadratic formula, we have:

$$x_1 = \frac{-1+\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2} \qquad x_2 = \frac{-1-\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2}$$

As $t+\theta > 0$, x_2 is an invalid answer and only x_1 is acceptable. By substituting $t+\theta$ for x we have:

$$\begin{aligned} t+\theta &= \frac{-1+\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2} \Longrightarrow \\ \theta &= \frac{-1+\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2} - t \Longrightarrow \\ \theta &= \frac{-1-2t+\sqrt{1+\frac{4t(t+1)(T(t)-1)}{T_{ta}-1}}}{2} \end{aligned} \quad \blacksquare$$

Lemma 6.7 *Let t_i be the number of interactions at the beginning of the i^{th} cycle, the trust value after the i^{th} defection, $T_d(i)$ will be calculated by $T_d(i) = \frac{1}{t_i+2}(T_b(i) \times T_b(i) - 2)$ where $T_b(i)$ is the trust value before the i^{th} defection.*

Proof The i^{th} defection is the $t_i + 1^{th}$ interactions so $T_d(i) = T(t_i+1)$. Using Lemma 6.3, $T(t_i+1) = \frac{1}{t_i+2}(t_i \times T(t_i) + 2w_{t_i+1})$ using lemme 6.3 where t_i+1 is replaced by t. As t_i+1 is a defection, $w_{t_i+1} = -1$. Therefore, $T_d(i) = T(t_i+1) = \frac{1}{t_i+2}(t_i \times T(t_i) - 2)$. Moreover, we know that $T(t_i) = T_b(i)$. Therefore, $T_d(i) = T(t_i+1) = \frac{1}{t_i+2}(t_i \times T_b(i) - 2)$ $\blacksquare$

Lemma 6.8 *Given t_i is the number of interactions at the beginning of the i^{th} cycle, θ_i for the Regret model will be calculated by $\theta_i = \frac{-3-2t_i+\sqrt{(2t_i+\frac{T_c-5}{T_c-1})^2-\frac{8(T_c+1)}{(T_c-1)^2}}}{2}$*

Proof When the con-man does the i^{th} defection, its trust value will be updated by $T_d(i) = T(t_i+1) = \frac{1}{t_i+2}(t_i \times T_b(i) - 2)$ using Lemma 6.3. As $T_b(i) \geq T_c > 0$, we simplify the formula by replacing $T_b(i)$ with its lower bound T_c [3]. In this sense, we have $T_d(i) = \frac{1}{t_i+2}(t_i \times T_c - 2)$.

Using Lemma 6.6, θ_i can be calculated as demonstrated below, where t_i+1 and T_c should be replaced by t and T_{ta} respectively.

$$\begin{aligned}
\theta_i &= \frac{-1-2(t_i+1)+\sqrt{1+\frac{4(t_i+1)(t_i+1+1)(T(t_i+1)-1)}{T_c-1}}}{2} \Longrightarrow \\
\theta_i &= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i+2)(T(t_i+1)-1)}{T_c-1}}}{2}
\end{aligned}$$

We will replace $T(t_i+1)$ with the value of $T_d(i)$, so we have:

$$\begin{aligned}
\theta_i &= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i+2)(T_d(i)-1)}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i+2)(\frac{1}{t_i+2}(t_i\times T_c-2)-1)}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i+2)(\frac{t_i\times T_c-2-t_i-2}{t_i+2})}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i+2)(\frac{t_i(T_c-1)-4}{t_i+2})}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{1+\frac{4(t_i+1)(t_i(T_c-1)-4)}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{1+4t_i(t_i+1)+\frac{-16(t_i+1))}{T_c-1}}}{2} \\
&= \frac{-3-2t_i+\sqrt{4t_i^2+4t_i+\frac{-16t_i}{T_c-1}+\frac{-16}{T_c-1}+1}}{2} \\
&= \frac{-3-2t_i+\sqrt{4t_i^2+(4-\frac{16}{T_c-1})t_i+\frac{-16}{T_c-1}+1}}{2} \\
&= \frac{-3-2t_i+\sqrt{4t_i^2+(\frac{4T_c-20}{T_c-1})t_i+\frac{-16}{T_c-1}+1}}{2} \\
&= \frac{-3-2t_i+\sqrt{(2t_i+\frac{T_c-5}{T_c-1})^2-\frac{8(T_c+1)}{(T_c-1)^2}}}{2} \quad \blacksquare
\end{aligned}$$

Theorem 6.1 *Let $1 > T_c > 0$, $\theta_i < \frac{T_c+1}{1-T_c}$ for any $i \in \mathbb{N}$ in the Regret model.*

Proof

$$\begin{aligned}
&0 < T_c < 1 \Longrightarrow \\
&0 < (T_c-1)^2 < 1 \Longrightarrow \\
&1 < \frac{1}{(T_c-1)^2} \Longrightarrow \\
&0 > -\frac{8(T_c+1)}{(T_c-1)^2} \Longrightarrow \\
&(2t_i+\frac{T_c-5}{T_c-1})^2 > (2t_i+\frac{T_c-5}{T_c-1})^2 - \frac{8(T_c+1)}{(T_c-1)^2} \Longrightarrow
\end{aligned}$$

[3]This simplification should be investigated more in the future.

$$
\begin{aligned}
&\sqrt{(2t_i + \frac{T_c - 5}{T_c - 1})^2} > \sqrt{(2t_i + \frac{T_c - 5}{T_c - 1})^2 - \frac{8(T_c + 1)}{(T_c - 1)^2}} \Longrightarrow \\
&-3 - 2t_i + 2t_i + \frac{T_c - 5}{T_c - 1} > -3 - 2t_i + \sqrt{(2t_i + \frac{T_c - 5}{T_c - 1})^2 - \frac{8(T_c + 1)}{(T_c - 1)^2}} \Longrightarrow \\
&-3 + \frac{T_c - 5}{T_c - 1} > -3 - 2t_i + \sqrt{(2t_i + \frac{T_c - 5}{T_c - 1})^2 - \frac{8(T_c + 1)}{(T_c - 1)^2}} \Longrightarrow \\
&\frac{-3 + \frac{T_c - 5}{T_c - 1}}{2} > \frac{-3 - 2t_i + \sqrt{(2t_i + \frac{T_c - 5}{T_c - 1})^2 - \frac{8(T_c + 1)}{(T_c - 1)^2}}}{2} \Longrightarrow \\
&\frac{-3(T_c - 1) + T_c - 5}{2(T_c - 1)} > \theta_i \Longrightarrow \\
&\frac{-2T_c - 2}{2(T_c - 1)} > \theta_i \Longrightarrow \\
&\frac{-2(T_c + 1)}{2(T_c - 1)} > \theta_i \Longrightarrow \\
&\theta_i < \frac{T_c + 1}{1 - T_c} \quad \blacksquare
\end{aligned}
$$

Theorem 6.2 *The Regret model is vulnerable to the con-man attack with high-level risk.*

Proof Referring to Theorem 6.1, $\theta_i < \frac{T_c+1}{1-T_c}$ for any $i \in \mathbb{N}$ and any values of T_c. The upper-bound value of θ_i only depends on the value of T_c. The con-man, by selecting $\theta = \frac{T_c+1}{1-T_c}$, can make sure that in each cycle of cooperations-defection it can reach the trust value of T_c. As a result, there is a constant θ, $\theta_{critical}$, that the con-man can select in order to be recognized as trustworthy despite being a con-man. Moreover, the con-man can regain the lost trust value with the same number of cooperations in each cycle of defection-cooperations by choosing $\theta_{critical} = \frac{T_c+1}{1-T_c}$. Therefore, the Regret model is vulnerable to the con-man attack for any value of $1 > T_c > 0$. $\blacksquare$

7 The FIRE Model

We here will prove that the FIRE model [6] is vulnerable to the con-man attack. Our proof strategy is to find θ_i for any parameter setting in any defection-cooperations cycle and show that $\theta_i \leq \theta_c$ where θ_c is a constant. Throughout our analysis and proofs, the cooperation and defection is mapped to 1 and -1 respectively for the FIRE model and the value is used as an input of the trust model. To simplify our analysis and proofs, we will rewrite the updating scheme as presented in Equation 19.

$$T(t) = \sum_{i=1}^{t} \rho(t,i) \times w_i \tag{19}$$

where i is the time that w_i is recorded, t is the current time, $\rho(t,i) = \frac{f(i,t)}{\sum_{j=1}^{t} f(j,t)}$, and $f(i,t) = e^{-\frac{t-i}{\lambda}}$.

Lemma 7.1 *Let be $\rho(t,i) = \frac{f(i,t)}{\sum_{j=1}^{t} f(j,t)}$ and $f(i,t) = e^{-\frac{t-i}{\lambda}}$ in the FIRE model, then $\rho(t,i) = \frac{e^{\frac{i}{\lambda}}.(e^{\frac{-1}{\lambda}}-1)}{1-e^{\frac{t}{\lambda}}}$*

Proof

$$
\begin{aligned}
\rho(t,i) &= \frac{f(i,t)}{\sum_{j=1}^{t} f(j,t)} \\
&= \frac{e^{-\frac{t-i}{\lambda}}}{\sum_{j=1}^{t} e^{-\frac{t-j}{\lambda}}} \\
&= \frac{e^{\frac{-t}{\lambda}}.e^{\frac{i}{\lambda}}}{\sum_{j=1}^{t} e^{\frac{-t}{\lambda}}.e^{\frac{j}{\lambda}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{e^{\frac{-t}{\lambda}}.e^{\frac{i}{\lambda}}}{e^{\frac{-t}{\lambda}}\sum_{j=1}^{t} e^{\frac{j}{\lambda}}} \\
&= \frac{e^{\frac{i}{\lambda}}}{\sum_{j=1}^{t} e^{\frac{j}{\lambda}}}
\end{aligned}
$$

$\sum_{j=1}^{t} e^{\frac{j}{\lambda}} = \sum_{j=0}^{t-1} e^{\frac{1}{\lambda}}.e^{\frac{j}{\lambda}}$ is a geometric series where $a = e^{\frac{1}{\lambda}}$ is the first term of the series, and $r = e^{\frac{1}{\lambda}}$ is the common ratio. The result of the series is:

$$
\sum_{j=1}^{t} e^{\frac{j}{\lambda}} = \frac{a(1-r^t)}{1-r} = \frac{e^{\frac{1}{\lambda}}(1-e^{\frac{t}{\lambda}})}{1-e^{\frac{1}{\lambda}}} = \frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}
$$

Therefore, we have:

$$
\begin{aligned}
\rho(t,i) &= \frac{e^{\frac{i}{\lambda}}}{\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}} \\
&= \frac{e^{\frac{i}{\lambda}}.(e^{\frac{-1}{\lambda}}-1)}{1-e^{\frac{t}{\lambda}}}
\end{aligned}
$$

Lemma 7.2 *The FIRE model can be updated by* $T(t) = \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t}{\lambda}}} \sum_{i=1}^{t} e^{\frac{i}{\lambda}} \times w_i$

Proof Using Equation 19 and Lemma 7.1, the updating schema for FIRE can be written as follows:

$$
\begin{aligned}
T(t) &= \sum_{i=1}^{t} \rho(t,i) \times w_i \\
&= \sum_{i=1}^{t} \frac{e^{\frac{i}{\lambda}}.(e^{\frac{-1}{\lambda}}-1)}{1-e^{\frac{t}{\lambda}}} \times w_i \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t}{\lambda}}} \sum_{i=1}^{t} e^{\frac{i}{\lambda}} \times w_i \qquad \blacksquare
\end{aligned}
$$

Lemma 7.3 *Let $T(t-1)$ be the trust value after the $t-1^{th}$ interactions and w_t be the t^{th} rating then* $T(t) = \frac{1-e^{\frac{t-1}{\lambda}}}{1-e^{\frac{t}{\lambda}}}(T(t-1) - w_t) + w_t$.

Proof Using Lemma 7.2, we can show that $\frac{1-e^{\frac{t-1}{\lambda}}}{e^{\frac{-1}{\lambda}}-1} T(t-1) = \sum_{i=1}^{t-1} e^{\frac{i}{\lambda}} \times w_i$. Moreover, we can rewrite $T(t)$ as follows:

$$
\begin{aligned}
T(t) &= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t}{\lambda}}} \sum_{i=1}^{t} e^{\frac{i}{\lambda}} \times w_i \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t}{\lambda}}} \left(\sum_{i=1}^{t-1} e^{\frac{i}{\lambda}} \times w_i + e^{\frac{t}{\lambda}}.w_t \right) \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t}{\lambda}}} \left(\frac{1-e^{\frac{t-1}{\lambda}}}{e^{\frac{-1}{\lambda}}-1} T(t-1) + e^{\frac{t}{\lambda}}.w_t \right) \\
&= \frac{1}{1-e^{\frac{t}{\lambda}}} \left((1-e^{\frac{t-1}{\lambda}})T(t-1) + (e^{\frac{t-1}{\lambda}} - e^{\frac{t}{\lambda}}).w_t \right) \\
&= \frac{1}{1-e^{\frac{t}{\lambda}}} \left((1-e^{\frac{t-1}{\lambda}})T(t-1) - w_t(-e^{\frac{t-1}{\lambda}} + e^{\frac{t}{\lambda}}) \right) \\
&= \frac{1}{1-e^{\frac{t}{\lambda}}} \left((1-e^{\frac{t-1}{\lambda}})T(t-1) - w_t(1-e^{\frac{t-1}{\lambda}} - 1 + e^{\frac{t}{\lambda}}) \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{1-e^{\frac{t}{\lambda}}}\left((1-e^{\frac{t-1}{\lambda}})(T(t-1)-w_t)-w_t(-1+e^{\frac{t}{\lambda}})\right) \\
&= \frac{1-e^{\frac{t-1}{\lambda}}}{1-e^{\frac{t}{\lambda}}}(T(t-1)-w_t)+w_t
\end{aligned}
$$

Lemma 7.4 *Let $T(t)$ be the trust value after the t^{th} interactions, the trust value after θ cooperations will be calculated by $T(t+\theta)=\frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}\times(T(t)-1)+1$.*

Proof Using Lemma 7.2, we have:

$$
\begin{aligned}
T(t+\theta) &= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\sum_{i=1}^{t+\theta} e^{\frac{i}{\lambda}}\times w_i \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\left(\sum_{i=1}^{t} e^{\frac{i}{\lambda}}\times w_i+\sum_{i=t+1}^{t+\theta} e^{\frac{i}{\lambda}}\times w_i\right)
\end{aligned}
$$

Using Lemma 7.2, we know that $\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}T(t)=\sum_{i=1}^{t} i\times w_i$. Therefore, we have:

$$
T(t+\theta) = \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\left(\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}T(t)+\sum_{i=t+1}^{t+\theta} e^{\frac{i}{\lambda}}\times w_i\right)
$$

As the last θ interactions were cooperations, therefore $w_i=1$ for $t<i\leq t+\theta$. Therefore we have:

$$
\begin{aligned}
T(t+\theta) &= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\left(\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}T(t)+\sum_{i=t+1}^{t+\theta} e^{\frac{i}{\lambda}}\right) \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\left(\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}T(t)+\frac{e^{\frac{t+1}{\lambda}}.(1-e^{\frac{\theta}{\lambda}})}{1-e^{\frac{1}{\lambda}}}\right) \\
&= \frac{e^{\frac{-1}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}\left(\frac{1-e^{\frac{t}{\lambda}}}{e^{\frac{-1}{\lambda}}-1}T(t)+\frac{e^{\frac{t}{\lambda}}.(1-e^{\frac{\theta}{\lambda}})}{e^{\frac{-1}{\lambda}}-1}\right) \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}T(t)+\frac{e^{\frac{t}{\lambda}}.(1-e^{\frac{\theta}{\lambda}})}{1-e^{\frac{t+\theta}{\lambda}}} \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}T(t)+\frac{e^{\frac{t}{\lambda}}-e^{\frac{t+\theta}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}} \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}T(t)+\frac{e^{\frac{t}{\lambda}}-1+1-e^{\frac{t+\theta}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}} \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}T(t)+\frac{e^{\frac{t}{\lambda}}-1}{1-e^{\frac{t+\theta}{\lambda}}}+1 \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}T(t)-\frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}+1 \\
&= \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}}\times(T(t)-1)+1 \quad \blacksquare
\end{aligned}
$$

Lemma 7.5 *Let $T(t)$ and T_{ta} be the starting trust value and the target trust value as a result of θ cooperation respectively, where t is the number of interactions used for calculation of $T(t)$. θ is calculated by $\theta=\lambda\ln\left(1-\frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1}\times(T(t)-1)\right)-t$.*

Proof We can find θ using Lemma 7.4, where T_{ta} should be replaced by $T(t+\theta)$, so we have:

$$T_{ta} = \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}} \times (T(t)-1)+1 \Longrightarrow$$

$$T_{ta}-1 = \frac{1-e^{\frac{t}{\lambda}}}{1-e^{\frac{t+\theta}{\lambda}}} \times (T(t)-1) \Longrightarrow$$

$$1-e^{\frac{t+\theta}{\lambda}} = \frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1} \times (T(t)-1) \Longrightarrow$$

$$-e^{\frac{t+\theta}{\lambda}} = \frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1} \times (T(t)-1)-1 \Longrightarrow$$

$$e^{\frac{t+\theta}{\lambda}} = 1-\frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1} \times (T(t)-1) \Longrightarrow$$

$$\frac{t+\theta}{\lambda} = \ln\left(1-\frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1} \times (T(t)-1)\right) \Longrightarrow$$

$$\theta = \lambda \ln\left(1-\frac{1-e^{\frac{t}{\lambda}}}{T_{ta}-1} \times (T(t)-1)\right)-t \quad \blacksquare$$

Lemma 7.6 *Let t_i be the number of interactions at the beginning of the i^{th} cycle, the trust value after the i^{th} defection, $T_d(i)$ will be calculated by $T_d(i) = \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T_b(i)+1)-1$ where $T_b(i)$ is the trust value before the i^{th} defection.*

Proof The i^{th} defection is the t_i+1^{th} interactions so $T_d(i) = T(t_i+1)$. Using lemme 7.3, we have:

$$\begin{aligned} T(t_i+1) &= \frac{1-e^{\frac{t_i+1-1}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}}(T(t_i+1-1)-w_{t_i+1})+w_{t_i+1} \\ &= \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}}(T(t_i)-w_{t_i+1})+w_{t_i+1} \end{aligned}$$

As t_i+1 is a defection, $w_{t_i+1} = -1$. Therefore, $T_d(i) = T(t_i+1)$ and we have:

$$T_d(i) = \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T(t_i)+1)-1$$

we know that $T(t_i) = T_b(i)$. Therefore, we have:

$$T_d(i) = \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T_b(i)+1)-1 \quad \blacksquare$$

Theorem 7.1 *Given t_i be the number of interactions at the beginning of the i^{th} cycle, θ_i for the FIRE model will be calculated by:*

$$\theta_i = \lambda \ln\left(\frac{T_c+1-2e^{\frac{1}{\lambda}}}{T_c-1}\right)-1 \tag{20}$$

Proof When the con-man does the i^{th} defection, its trust value will be updated by $T_d(i) = T(t_i+1) = \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T_b(i)+1)-1$ using Lemma 7.3. As $T_b(i) \geq T_c > 0$, We simplify the formula by replacing $T_b(i)$ with its lower bound T_c [4]. In this sense, we have $T_d(i) = \frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T_c+1)-1$.

[4] This simplification should be investigated more in the future.

Using Lemma 7.5, θ_i can be calculated as demonstrated below, where t_i+1 and T_c should be replaced by t and T_{ta} respectively.

$$
\begin{aligned}
\theta_i &= \lambda \ln\left(1 - \frac{1-e^{\frac{t_i+1}{\lambda}}}{T_c-1} \times (T(t_i+1)-1)\right) - (t_i+1) \Longrightarrow \\
\theta_i &= \lambda \ln\left(1 - \frac{1-e^{\frac{t_i+1}{\lambda}}}{T_c-1} \times (T(t_i+1)-1)\right) - t_i - 1
\end{aligned}
$$

We will replace $T(t_i+1)$ with the value of $T_d(i)$, so we have:

$$
\begin{aligned}
\theta_i &= \lambda \ln\left(1 - \frac{1-e^{\frac{t_i+1}{\lambda}}}{T_c-1} \times (T_d(i)-1)\right) - t_i - 1 \\
&= \lambda \ln\left(1 - \frac{1-e^{\frac{t_i+1}{\lambda}}}{T_c-1} \times (\frac{1-e^{\frac{t_i}{\lambda}}}{1-e^{\frac{t_i+1}{\lambda}}} \times (T_c+1)-1-1)\right) - t_i - 1 \\
&= \lambda \ln\left(1 - \frac{1-e^{\frac{t_i}{\lambda}}}{T_c-1} \times (T_c+1) + 2 \times \frac{1-e^{\frac{t_i+1}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c-1-((T_c+1)(1-e^{\frac{t_i}{\lambda}}))+2-2e^{\frac{t_i+1}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c-1-(T_c-T_c e^{\frac{t_i}{\lambda}}+1-e^{\frac{t_i}{\lambda}})+2-2e^{\frac{t_i+1}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c-1-T_c+T_c e^{\frac{t_i}{\lambda}}-1+e^{\frac{t_i}{\lambda}}+2-2e^{\frac{t_i+1}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c e^{\frac{t_i}{\lambda}}+e^{\frac{t_i}{\lambda}}-2e^{\frac{t_i+1}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{(T_c+1-2e^{\frac{1}{\lambda}})e^{\frac{t_i}{\lambda}}}{T_c-1}\right) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c+1-2e^{\frac{1}{\lambda}}}{T_c-1}\right) + \lambda \ln(e^{\frac{t_i}{\lambda}}) - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c+1-2e^{\frac{1}{\lambda}}}{T_c-1}\right) + \lambda \times \frac{t_i}{\lambda} - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c+1-2e^{\frac{1}{\lambda}}}{T_c-1}\right) + t_i - t_i - 1 \\
&= \lambda \ln\left(\frac{T_c+1-2e^{\frac{1}{\lambda}}}{T_c-1}\right) - 1 \quad \blacksquare
\end{aligned}
$$

Theorem 7.2 *The FIRE model is vulnerable to the con-man attack with medium-level risk for any value of λ.*

Proof Referring to Theorem 7.1, θ_i for any values of T_c, and λ, is calculated by Equation 20. The value of θ_i only depends on the value of λ and T_c where T_c is constant and determined by the con-man. As λ is constant in the FIRE model over the course of interactions, so $\theta_i = \theta_{i+1}$ for $i = 1, 2, ..., N$. Therefore, there is a constant θ that the con-man can select in order to be recognized as trustworthy despite being a con-man. This constant θ can be calculated by Equation 20. Moreover, the con-man can regain lost trust with the same number of cooperations in each cycle of defection-cooperations. As a result, the FIRE model is vulnerable to the con-man attack for any values of $T_c > 0$ and λ.

8 Probabilistic Trust Models

We here will prove that probabilistic trust models as presented in Equation 6 are vulnerable to the con-man attack. Using the fact that $n_c(t) + n_d(t) = t$, we will rewrite Equation 6 as follows:

$$T(t) = \frac{n_c(t) + 1}{t + 2} \tag{21}$$

where $T(t)$ and $n_c(t)$ are the trust value and the number of cooperations after t interactions respectively. As before, our proof strategy is to show that the value of θ is a simple function of the trust threshold, T_c, thereby implying that the trust model is vulnerable to a con-man attack.

Lemma 8.1 *Let $n_c(t)$ and $T(t)$ be the number of cooperations and trust value after t interactions respectively, then $n_c(t) = T(t)(t + 2) - 1$*

Proof We will find $n_c(t)$ from Equation 21 as follows:

$$\begin{aligned}
&T(t) = \frac{n_c(t) + 1}{t + 2} \Longrightarrow \\
&T(t)(t + 2) = n_c(t) + 1 \Longrightarrow \\
&n_c(t) = T(t)(t + 2) - 1 \quad \blacksquare
\end{aligned}$$

Lemma 8.2 *Let $T(t)$ be the trust value after t interactions, the trust value after θ cooperations will be calculated by $T(t + \theta) = \frac{(t+2)T(t)+\theta}{t+\theta+2}$*

Proof Given $n_c(t)$, the total number of cooperations after θ cooperative interactions, $n_c(t + \theta) = n_c(t) + \theta$. Using this fact and Equation 21, we have:

$$\begin{aligned}
T(t + \theta) &= \frac{n_c(t + \theta) + 1}{(t + \theta) + 2} \Longrightarrow \\
T(t + \theta) &= \frac{n_c(t) + \theta + 1}{t + \theta + 2} \Longrightarrow
\end{aligned}$$

using Lemma 8.1, we will replace $n_c(t)$ by $T(t)(t + 2) - 1$, so we have:

$$\begin{aligned}
T(t + \theta) &= \frac{T(t)(t + 2) - 1 + \theta + 1}{t + \theta + 2} \Longrightarrow \\
T(t + \theta) &= \frac{(t + 2)T(t) + \theta}{t + \theta + 2} \quad \blacksquare
\end{aligned}$$

Lemma 8.3 *Let $T(t)$ be the trust value after t interactions, if the $t + 1$-th interaction is a defection, $T(t + 1) = \frac{T(t)(t+2)}{t+3}$.*

Proof According to Equation 21, we have:

$$T(t + 1) = \frac{n_c(t + 1) + 1}{t + 1 + 2}$$

As the $t + 1$-th interaction is a defection, $n_c(t) = n_c(t + 1)$. So, we have $T(t + 1) = \frac{n_c(t)+1}{t+3}$. Using Lemma 8.1, we have:

$$\begin{aligned}
T(t + 1) &= \frac{n_c(t) + 1}{t + 3} \\
&= \frac{T(t)(t + 2) - 1 + 1}{t + 3} \\
&= \frac{T(t)(t + 2)}{t + 3} \quad \blacksquare
\end{aligned}$$

Lemma 8.4 *Let $T(t)$ and T_{ta} be the starting trust value and the target trust value as a result of θ cooperations respectively, where t is the number of interactions used for calculation of $T(t)$. θ is calculated by $\theta = \frac{(t+2)(T_{ta}-T(t))}{1-T_{ta}}$.*

Proof We can find θ using Lemma 8.2, where T_{ta} should be replaced by $T(t+\theta)$, so we have:

$$\begin{aligned}
&T_{ta} = \frac{(t+2)T(t)+\theta}{t+\theta+2} \Longrightarrow \\
&T_{ta}(t+\theta+2) = (t+2)T(t)+\theta \Longrightarrow \\
&\theta T_{ta} + (t+2)T_{ta} = (t+2)T(t)+\theta \Longrightarrow \\
&\theta T_{ta} - \theta = (t+2)T(t) - (t+2)T_{ta} \Longrightarrow \\
&\theta(T_{ta}-1) = (t+2)(T(t)-T_{ta}) \Longrightarrow \\
&\theta = \frac{(t+2)(T(t)-T_{ta})}{T_{ta}-1} \Longrightarrow \\
&\theta = \frac{(t+2)(T_{ta}-T(t))}{1-T_{ta}} \qquad \blacksquare
\end{aligned}$$

Theorem 8.1 *Given t_i as the number of interactions at the beginning of the i^{th} cycle, θ_i for a probabilistic trust model will be calculated by $\theta_i = \frac{T_c}{1-T_c}$*

Proof When the con-man does the i^{th} defection, its trust value will be updated by $T_d(i) = T(t_i+1) == \frac{T_b(i)(t_i+2)}{t_i+3}$ using Lemma 8.3. As $T_b(i) \geq T_c > 0$, We simplify the formula by replacing $T_b(i)$ with its lower bound T_c [5]. In this sense, we have $T_d(i) = \frac{T_c(t_i+2)}{t_i+3}$.

Using Lemma 8.4, θ_i can be calculated as demonstrated below, where t_i+1 and T_c should be replaced by t and T_{ta} respectively.

$$\begin{aligned}
\theta_i &= \frac{(t_i+1+2)(T_c-T(t_i+1))}{1-T_c} \Longrightarrow \\
&= \frac{(t_i+3)(T_c-T(t_i+1))}{1-T_c}
\end{aligned}$$

We will replace $T(t_i+1)$ with the value of $T_d(i)$, so we have:

$$\begin{aligned}
\theta_i &= \frac{(t_i+3)(T_c-T_d(i))}{1-T_c} \\
&= \frac{(t_i+3)(T_c-\frac{T_c(t_i+2)}{t_i+3})}{1-T_c} \\
&= \frac{(t_i+3)T_c-T_c(t_i+2)}{1-T_c} \\
&= \frac{T_c}{1-T_c} \qquad \blacksquare
\end{aligned}$$

Theorem 8.2 *Probabilistic models are vulnerable to the con-man attack with high-level risk.*

Proof Referring to Theorem 8.1, θ_i for any values of T_c is calculated by $\frac{T_c}{1-T_c}$. The value of θ_i only depends on the value of T_c which can be determined by the con-man. Note that T_c is constant in all cycles of cooperations-defection. Therefore, $\theta_i = \theta_{i+1} = \frac{T_c}{1-T_c}$ for $i = 1, 2, ..., N$. Therefore, there is a constant θ, $\theta_{critical}$, that the con-man can select in order to be recognized as trustworthy despite being a con-man. Moreover, the con-man can regain the lost trust value with the same number of cooperations in each cycle of defection-cooperations. As a result, probabilistic models are vulnerable to the con-man attack for any values of $1 > T_c > 0$. $\blacksquare$

[5] This simplification should be investigated more in the future.

9 AER

We will prove here that the AER update scheme for α and β (Equations 22 and 23) is not vulnerable to the con-man attack in such a way that the con-man requires more cooperations in each cycle of defection-cooperations when compared to the previous cycle in order to reach to T_c. In other words, $\theta_i < \theta_{i+1}$.

$$\alpha := \alpha \times (1 - |\beta|) \tag{22}$$

$$\beta := \beta - \gamma_d \times (1 + \beta) \tag{23}$$

Where γ_d is the discounting factor and is in the range of $[0, 1]$. Note that, α and β will be updated when a defection occurs. This is a crucial point, as the use of results from the Yu and Singh analysis rely upon it. From now on, the value of α and β after the i^{th} defection are denoted by $\alpha(i)$ and $\beta(i)$ respectively. Therefore, we can rewrite Equations 22 and 23 as follows:

$$\begin{aligned} \alpha(i) &= \alpha(i-1) \times (1 - |\beta(i-1)|) \\ \beta(i) &= \beta(i-1) - \gamma_d \times (1 + \beta(i-1)) \end{aligned}$$

Lemma 9.1 *$\beta(i)$ is ever-decreasing (i.e., $\beta(i) > \beta(i+1)$).*

Proof If we rewrite the formula for $\beta(i)$, we have:

$$\begin{aligned} \beta(i) &= \beta(i-1) - \gamma_d \times (1 + \beta(i-1)) \\ &= \beta(i-1) - \gamma_d - \gamma_d \times \beta(i-1)) \\ &= (1 - \gamma_d) \times \beta(i-1) - \gamma_d \end{aligned}$$

The proof is by using induction:

$$\begin{aligned} \beta(1) &= (1 - \gamma_d) \times \beta(0) - \gamma_d \Longrightarrow \\ \beta(1) + \gamma_d &= (1 - \gamma_d) \times \beta(0) \overset{\gamma_d \geq 0}{\Longrightarrow} \\ \beta(1) &< (1 - \gamma_d) \times \beta(0) \Longrightarrow \\ \beta(1) &< \beta(0) \end{aligned}$$

Now we will assume that $\beta(i-1) > \beta(i)$ and we prove that $\beta(i) > \beta(i+1)$.

$$\begin{aligned} &\beta(i-1) > \beta(i) \Longrightarrow \\ &(1 - \gamma_d) \times \beta(i-1) > (1 - \gamma_d) \times \beta(i) \Longrightarrow \\ &(1 - \gamma_d) \times \beta(i-1) - \gamma_d > (1 - \gamma_d) \times \beta(i) - \gamma_d \Longrightarrow \\ &\beta(i) > \beta(i+1) \end{aligned}$$

Therefore, $\beta(i)$ is ever-decreasing. ∎

Lemma 9.2 *Given that $-1 < \beta(0) < 0$, $\beta(i)$ is in the range of (0,1) for any $i \in \mathbb{N}$.*

Proof We can rewrite the formula for β as follows:

$$\beta(i) = (1 - \gamma_d) \times \beta(i-1) - \gamma_d \tag{24}$$

Equation 24 is the recurrence equation with the same form as Equation 26 when b and d are substituted by $(1-\gamma_d)$ and $-\gamma$ respectively. Therefore, using Lemma 11.1, the explicit form of Equation 24 is as follows:

$$\begin{aligned} \beta(n) &= \left(\beta(0) + \frac{-\gamma_d}{1 - \gamma_d - 1}\right) \times (1 - \gamma_d)^n + \frac{\gamma_d}{1 - \gamma_d - 1} \\ &= (\beta(0) + 1) \times (1 - \gamma_d)^n - 1 \end{aligned}$$

As $-1 < \beta(0) < 0$, we have:

$$
\begin{aligned}
&-1 < \beta(0) < 0 \Longrightarrow \\
&0 < \beta(0) + 1 < 1 \Longrightarrow \\
&0 < (\beta(0) + 1) \times (1 - \gamma_d)^i < (1 - \gamma_d)^i
\end{aligned}
$$

As $(1 - \gamma_d)^i < 1$ for any value of i, then we have:

$$
\begin{aligned}
&0 < (\beta(0) + 1) \times (1 - \gamma_d)^i < 1 \Longrightarrow \\
&-1 < (\beta(0) + 1) \times (1 - \gamma_d)^i - 1 < 0 \Longrightarrow \\
&-1 < \beta(i) < 0 \quad \blacksquare
\end{aligned}
$$

Lemma 9.3 *Given that* $-1 < \beta(0) < 0$ *then* $0 < 1 - |\beta(i)| < 1$ *for any* $i \in \mathbb{N}$.

Proof Using Lemma 9.2, we know that $-1 < \beta(i) < 0$. Therefore, we have:

$$
\begin{aligned}
&0 < |\beta(i)| < 1 \Longrightarrow \\
&-1 < -|\beta(i)| < 0 \Longrightarrow \\
&0 < 1 - |\beta(i)| < 1 \quad \blacksquare
\end{aligned}
$$

Lemma 9.4 $\alpha(i)$ *is ever-decreasing (i.e.* $\alpha(i) > \alpha(i+1)$*).*

Proof We will prove by induction. $\alpha(1) = \alpha(0) \times (1 - |\beta(0)|)$ and using Lemma 9.3 we know that $0 < 1 - |\beta(0)| < 1$, therefore $\alpha(1) < \alpha(0)$.

As $\beta(i) > \beta(i+1)$ according to Lemma 9.1 and $-1 < \beta(i) < 0$ according to Lemma 9.2, therefore $|\beta(i)| < |\beta(i+1)|$. As a result, $1 - |\beta(i)| > 1 - |\beta(i+1)|$

$$
\begin{aligned}
&\alpha(i) < \alpha(i-1) \Longrightarrow \\
&\alpha(i) \times (1 - |\beta(i)|) < \alpha(i-1) \times (1 - |\beta(i-1)|) \Longrightarrow \\
&\alpha(i+1) < \alpha(i)
\end{aligned}
$$

Therefore, $\alpha(i)$ is ever-decreasing. $\blacksquare$

Lemma 9.5 *Given that* $0 < \alpha(0) < 1$, $\alpha(i)$ *is in the range of (0,1) for any* $i \in \mathbb{N}$.

Proof The proof is by induction. As $\beta(i) > \beta(i+1)$ according to Lemma 9.1 and $-1 < \beta(i) < 0$ according to Lemma 9.2, therefore $|\beta(i)| < |\beta(i+1)|$. As a result, $1 - |\beta(i)| > 1 - |\beta(i+1)|$

As $0 < \alpha(0) < 1$ and $0 < 1 - |\beta(i)| < 1$, we have:

$$
\begin{aligned}
&0 < \alpha(0) < 1 \Longrightarrow \\
&0 < \alpha(0) \times (1 - |\beta(0)|) < 1 \Longrightarrow \\
&0 < \alpha(1) < 1
\end{aligned}
$$

We have the following for any $i \in \mathbb{N}$:

$$
\begin{aligned}
&0 < \alpha(i) < 1 \Longrightarrow \\
&0 < \alpha(i) \times (1 - |\beta(i)|) < 1 \Longrightarrow \\
&0 < \alpha(i+1) < 1 \quad \blacksquare
\end{aligned}
$$

Lemma 9.6 *Let* $T_c > 0$, θ_i *will be calculated by* $\frac{\ln(\beta(i)+1)}{\ln(1-\alpha(i))}$ *for the con-resistance extension of Yu and Singh's model.*

Proof In Lemma 5.1, we proved that for any constant value of $0 < \alpha < 1$ and $-1 < \beta < 0$, θ_i will be calculated based on the formula presented in Equation 17. In all of the proofs associated with Lemma 5.1, we assume that the value of α and β does not change during each cycle of defection-cooperations. This assumption is still valid when the value of α and β will be changed only after each defection and before the trust value is changed. Therefore, we can rewrite the Equation 17 as follows:

$$
\theta_i = \frac{\ln(\beta(i) + 1)}{\ln(1 - \alpha(i))} \quad \blacksquare \tag{25}
$$

Lemma 9.7 *Given that* $-1 < \beta(0) < 0$ *then* $0 > \ln(1+\beta(i)) > \ln(1+\beta(i+1)) > -\infty$ *for any* $i \in \mathbb{N}$.

Proof In Lemma 9.1, we have shown that $\beta(i) > \beta(i+1)$ for any $i \in \mathbb{N}$. Moreover, it is proved that $0 > \beta(i) > -1$, for any $i \in \mathbb{N}$, in Lemma 9.2. Therefore, we have:

$$\begin{aligned}
&0 > \beta(i) > \beta(i+1) > -1 \Longrightarrow \\
&1 > 1+\beta(i) > 1+\beta(i+1) > 0 \Longrightarrow \\
&\ln(1) > \ln(1+\beta(i)) > \ln(1+\beta(i+1)) > -\infty \Longrightarrow \\
&0 > \ln(1+\beta(i)) > \ln(1+\beta(i+1)) > -\infty \qquad \blacksquare
\end{aligned}$$

Lemma 9.8 *Given that* $0 < \alpha(0) < 1$ *then* $0 > \frac{1}{\ln(1-\alpha(i))} > \frac{1}{\ln(1-\alpha(i+1))} > -\infty$ *for any* $i \in \mathbb{N}$.

Proof In Lemma 9.4, we have shown that $\alpha(i) > \alpha(i+1)$ for any $i \in \mathbb{N}$. Moreover, it is proved that $0 > \alpha(i) > -1$, for any $i \in \mathbb{N}$, in Lemma 9.5. Therefore, we have:

$$\begin{aligned}
&0 < \alpha(i+1) < \alpha(i) < 1 \Longrightarrow \\
&0 < 1-\alpha(i) < 1-\alpha(i+1) < 1 \Longrightarrow \\
&-\infty < \ln(1-\alpha(i)) < \ln(1-\alpha(i+1)) < 0 \Longrightarrow \\
&0 > \frac{1}{\ln(1-\alpha(i))} > \frac{1}{\ln(1-\alpha(i+1))} > -\infty \qquad \blacksquare
\end{aligned}$$

Theorem 9.1 *AER will not let the con-man regain a high trust value easily with the same or smaller number of cooperations (i.e.,* $\theta_i < \theta_{i+1}$*). AER is not vulnerable (i.e., is exploitation resistant) to the con-man attack.*

Proof Using both Lemma 9.7 and Lemma 9.8, we have:

$$\begin{aligned}
&0 < \frac{\ln(1+\beta(i))}{\ln(1-\alpha(i))} < \frac{\ln(1+\beta(i+1))}{\ln(1-\alpha(i+1))} < +\infty \Longrightarrow \\
&0 < \theta_i < \theta_{i+1} < \infty
\end{aligned}$$

In AER, the con-man can not regain a high trust value easily with the same number of cooperations with each cycle. This motivates the con-man to be more cooperative in its interactions after each defection. AER is not vulnerable to the con-man attack. $\blacksquare$

Corollary 9.1 *AER is forgiving in any cycle of defection-cooperations but is more strict after each defection.*

Proof Lemma 9.6 demonstrates that in any cycle of defection-cooperations there exists a θ_i, calculated by Equation 25, for which the trust threshold, T_c, can be achieved. This is the evidence of that the con-man can be forgiven (gain lost trust) by θ_i cooperations. Moreover, Theorem 9.1, by showing that $\theta_i < \theta_{i+1}$, indicates that AER is more strict in terms of forgiveness (regaining lost trust) after each defection.

10 Conclusions and Future work

This paper is motivated by the dire need to develop trust and reputation schemes that have provable properties for artificial societies, especially e-commerce. While simulation can often provide insights into trust and reputation model performance, analytical results based upon known or potential attacks are important to increase confidence in the true utility of such models. Moreover, they can draw our attention to some vulnerability or interesting features of the model which might be neglected or not discovered in agent-based models. It is the position of this paper that widespread deployment of sophisticated trust and reputation models will only occur when such analytical results are forthcoming.

This paper has proven that simple malicious agents with cyclical behaviour (con-man agents) can exploit the Yu and Singh's trust model, Regret, FIRE, and probabilistic trust models regardless of the model's parameters. However, AER has been shown to be exploitation resistant. Furthermore, malicious agents with cyclical behaviour will have to increase the number of cooperations in each and every cycle with AER in order to achieve a specific trust value. It is proven that AER is forgiving but that the rate of forgiveness slows with every defection.

Future work will design adaptive schemes similar to AER for Regret, FIRE, and probabilistic trust models and their exploitation resistance to the con-man attack proven. Finally, new attacks on adaptive schemes will be constructed using formal models as described in Section 3.1 for the con-man attack and analyzed empirically through simulation and analytically using techniques similar to those used in this paper.

References

[1] Robert Axelrod. *The Evolution of Cooperation.* New York: Basic Books, 1984.

[2] Chrysanthos Dellarocas. Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems. In *In ICIS*, pages 520–525, 2000.

[3] Karen K. Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater, Andreas Schlosser, Zvi Topol, K. Suzanne Barber, Jeffrey S. Rosenschein, Laurent Vercouter, and Marco Voss. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 512–518, New York, NY, USA, 2005. ACM.

[4] Chung-Wei Hang, Yonghong Wang, and Munindar P. Singh. An adaptive probabilistic trust model and its evaluation. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 1485–1488, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[5] T.D. Huynh, N. R. Jennings, and N. Shadbolt. Developing an integrated trust and reputation model for open multi-agent systems. In *7th International Workshop on Trust in Agent Societies*, pages 65–74, 2004.

[6] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[7] Sirkka L. Jarvenpaa, Noam Tractinsky, and Michael Vitale. Consumer trust in an internet store. *Inf. Technol. and Management*, 1(1-2):45–71, 2000.

[8] Audun Josang and Roslan Ismail. The beta reputation system. In *In Proceedings of the 15th Bled Electronic Commerce Conference*, 2002.

[9] Reid Kerr and Robin Cohen. Smart cheaters do prosper: Defeating trust and reputation systems. In *AAMAS '09: Proceedings of the eighth international joint conference on Autonomous agents and multiagent systems*, Budapest, Hunagry, 2009. ACM.

[10] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation for e-businesses. In *HICSS '02: Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7*, page 188, Washington, DC, USA, 2002. IEEE Computer Society.

[11] Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *Knowl. Eng. Rev.*, 19(1):1–25, 2004.

[12] Jordi Sabater and Carles Sierra. Regret: A reputation model for gregarious societies. In *Fourth Workshop on Deception Fraud and Trust in Agent Societies*, pages 61–70, 2001.

[13] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60, 2005.

[14] Amirali Salehi-Abari and Tony White. Towards con-resistant trust models for distributed agent systems. In *IJCAI '09: Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence*, pages 272–277, Pasadena, CA, USA, 2009.

[15] Amirali Salehi-Abari and Tony White. Witness-based collusion and trust-aware societies. In *Proceedings of the IEEE International Conference on Computational Science and Engineering*, volume 4, pages 1008–1014, Vancouver, Canada, 2009. IEEE Computer Society.

[16] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. Coping with inaccurate reputation sources: experimental analysis of a probabilistic trust model. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 997–1004, New York, NY, USA, 2005. ACM.

[17] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Journal of Autonomous Agents and Multi-Agent Systems*, 12:2006, 2006.

[18] Andrew Whitby, Audun Josang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of 7th International Workshop on Trust in Agent Societies*, 2004.

[19] Bin Yu and Munindar P. Singh. A social mechanism of reputation management in electronic communities. In *CIA '00: Proceedings of the 4th International Workshop on Cooperative Information Agents IV, The Future of Information Agents in Cyberspace*, pages 154–165, London, UK, 2000. Springer-Verlag.

[20] Bin Yu and Munindar P. Singh. Detecting deception in reputation management. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 73–80, New York, NY, USA, 2003. ACM.

11 Appendix

Lemma 11.1 *Let A_n be the recurrent formula as presented in Equation 26, where $b \neq 0$ and d are constants. The closed form for A_n is $A_n = \left(A_0 + \frac{d}{b-1}\right) \times b^n + \frac{-d}{b-1}$.*

$$A_n = b \times A_{n-1} + d \tag{26}$$

Proof Let $A_n = b \times A_{n-1} + d$ and $A_{n-1} = b \times A_{n-2} + d$, by subtracting these two equations, we have:

$$\begin{aligned} A_n - A_{n-1} &= b \times A_{n-1} + d - b \times A_{n-2} - d \\ &= b \times (A_{n-1} - A_{n-2}) \Longrightarrow \\ A_n &= b \times (A_{n-1} - A_{n-2}) + A_{n-1} \\ &= (b+1) \times A_{n-1} - b \times A_{n-2} \end{aligned}$$

If we guess $A_n = cx^n$

$$\begin{aligned} A_n &= (b+1) \times A_{n-1} - b \times A_{n-2} \Longrightarrow \\ cx^n &= (b+1) \times cx^{n-1} - b \times cx^{n-2} \Longrightarrow \\ \frac{1}{cx^{n-2}} \times cx^n &= \frac{1}{cx^{n-2}} \times ((b+1)cx^{n-1} - bcx^{n-2}) \Longrightarrow \\ x^2 &= (b+1) \times x - b \end{aligned}$$

By rearranging the terms, we have the following equation $x^2 - (b+1) \times x + b = 0$ and we will find x with the quadratic formula.

$$\begin{aligned} x &= \frac{b+1 \pm \sqrt{(b+1)^2 - 4b}}{2} \\ &= \frac{b+1 \pm \sqrt{(b-1)^2}}{2} \\ &= \frac{b+1 \pm |b-1|}{2} \\ &= b \quad or \quad 1 \end{aligned}$$

Therefore, there are two solutions to the recurrence equation:

$$A_n = c \times b^n \qquad or \qquad A_n = c \times 1^n = c$$

As any linear combination of these two solutions is also a solution, the general solution is as follows:

$$A_n = c_1 \times b^n + c_2 \tag{27}$$

Equation 27 is a solution to the recurrence without boundary conditions for all constants c_1 and c_2. At this point, we attempt to find constants c_1 and c_2 to obtain a solution consistent with the boundary conditions, $A_0 = A_0$ and $A_1 = b \times A_0 + d$. From the first condition, we know:

$$A_0 = c_1 \times b^0 + c_2 \Longrightarrow A_0 = c_1 + c_2$$

From the second condition, we know:

$$A_1 = c_1 \times b^1 + c_2 = b \times A_0 + d$$

We now have two linear equations in two unknowns, so there is a unique solution:

$$c_1 = A_0 + \frac{d}{b-1} \qquad\qquad c_2 = \frac{-d}{b-1}$$

Therefore, we have a complete solution to the recurrence with boundary conditions:

$$A_n \quad = \quad \left(A_0 + \frac{d}{b-1}\right) \times b^n + \frac{-d}{b-1} \qquad \blacksquare$$