Towards Con-Resistant Trust Models for Distributed Agent Systems

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

Abstract

Artificial societies – distributed systems of autonomous agents – are becoming increasingly important in e-commerce. Agents base their decisions on trust and reputation in ways analogous to human societies. Many different definitions for trust and reputation have been proposed that incorporate many sources of information; however, system designs have tended to focus much of their attention on direct interactions. Furthermore, trust updating schemes for direct interactions have tended to uncouple updates for positive and negative feedback. Consequently, behaviour in which cycles of positive feedback followed by a single negative feedback results in untrustworthy agents remaining undetected. This con-man style of behaviour is formally described and desirable characteristics of con-resistant trust schemes proposed. A con-resistant scheme is proposed and compared with FIRE, Regret and Yu and Singh's model [Yu and Singh, 2000]. Simulation experiments demonstrate the utility of the con-resistant scheme.

1 Introduction

With the growth of open distributed systems - especially network services through the internet - artificial societies have been formed in these environments. As a consequence, real-world assumptions and the whole range of possible social behaviors need to be taken into account in these artificial societies. By analogy to human societies in which trust is one of the most crucial concepts driving decision making and relationships, *trust* is indispensable for any interactions among individuals in these artificial societies.

As reputation and trust have recently received considerable attention in diverse domains such as distributed artificial intelligence, computational economics, evolutionary biology, psychology, and sociology, there are many diverse definitions of trust available in these 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" [Mui *et al.*, 2002]. While trust definitions focus more on the history of agents' encounters, reputation is based on the aggregated information from other individuals. For instance, Sabater and Sierra [Sabater and Sierra, 2001] declared that "reputation is the opinion or view of someone about something".

[Sabater and Sierra, 2005] categorized computational trust and reputation models based on various intrinsic features. From their perspective, a trust and reputation model can be cognitive or game-theoretical in terms of its conceptual model. A cognitive model works based on beliefs and the mental states of individuals as opposed to game-theoretical models that rely on the result of pragmatic games and numerical aggregation of past interactions; the latter is used in this paper. 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., skin color or 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 computational models of trust, a review of which can be found in [Ramchurn *et al.*, 2004] and [Sabater and Sierra, 2005].

One approach to building a trust or reputation model is to have a central agency that keeps records of the recent activity of the users on the system. Amazon and eBay are important practical examples of centralized reputation management systems. EBay is an online auction and shopping website in which people and businesses buy and sell goods and services worldwide. In eBay, sellers receive feedback (+1, 0, -1) in each auction and their reputation is calculated as the sum of those ratings over the last six months.

Regret [Sabater and Sierra, 2001] 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. Direct trust (subjective reputation), witness reputation, neighborhood reputation and sys-

tem reputation are introduced in Regret where each trust and reputation value has an associated reliability measure.

[Yu and Singh, 2000] 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. The drawbacks of this combination model led them to consider alternate approaches; specifically, an evidential model of reputation management based on Dempster-Shafer theory [Yu and Singh, 2002]. 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 [Hang *et al.*, 2008]. [Huynh *et al.*, 2006] 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.

[Fullam *et al.*, 2005] 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.

We believe that in addition to the above criteria, *exploitation resistance* is a crucial feature of trust models. Exploitation resistance reflects the ability of a trust model to be impervious to agents who try to manipulate the trust model and who aim to abuse the presumption of trust. More precisely, 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.

There are few trust models which consider the existence of an adversary in providing witness information and present solutions for dealing with inaccurate reputation. TRAVOS [Teacy *et al.*, 2005] models an agent's trust in an interaction partner. Trust is calculated using probability theory that takes account of past interactions and reputation information gathered from third parties while coping with inaccurate reputations. [Yu and Singh, 2003] is similar to TRAVOS, in that it rates opinion source accuracy based on a subset of observations of trustee behavior.

As discussed earlier, direct experience (direct interaction) has been widely used by trust and reputation models as a source of information to judge whether the interacting partner is trustworthy; the work reported here does too. Trust models usually possess the components dedicated to direct interaction. For instance, Regret and FIRE employ direct trust (subjective reputation) and interaction trust components respectively. Unfortunately, to our knowledge, most of the direct interaction trust models and formulas are designed without the assumption of the persistence of con-men. A con-man (confidence man) is someone who takes advantage of someone else usually for financial gain using what is known as a confidence trick. The lack of an ability to detect con-man behaviors motivates the work reported in this paper.

Our contributions include modeling the con-man attack, demonstration of the vulnerability of three well-known trust models against the con-man attack, a proposal for desirable characteristics of con-resistant trust models and the introduction of a con-resistant extension to [Yu and Singh, 2000].

The remainder of this paper proceeds as follows. Section 2 discusses the direct interaction trust components of [Yu and Singh, 2000], FIRE, and Regret. We describe our proposed con-man attack in Section 3. Desirable characteristics of con-resistant trust schemes and a con-resistant scheme are proposed in Section 4. Evaluation of the proposed scheme through simulation experiments and concluding remarks are explained in Sections 5 and 6 respectively.

2 Direct Interaction Trust

Direct interaction is the most popular source of information for trust and reputation models [Ramchurn *et al.*, 2004]. 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 demonstrating the level of an opponent's trustworthiness. This trust value is calculated based on previous direct interactions. We discuss the direct interaction trust components of three well-known trust models (Yu and Singh, Regret and FIRE) in the following subsections.

2.1 Yu and Singh

[Yu and Singh, 2000]'s 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, while $T_{i,j}(t) \in [-1, +1]$ and $T_{i,j}(0) = 0$. One agent in the view of the other agent can have one of the following levels of trustworthiness: *Trustworthy*, *Not Yet Known*, or *Untrustworthy*.

An agent will update this variable based on the perception of cooperation/defection. Cooperation by the other agents generates positive evidence of $\alpha > 0$ and defection generates negative evidence of $\beta < 0$. The following trust updating scheme is proposed by [Yu and Singh, 2000]:
If $T_{i,j}(t) > 0$ **and** $Cooperation$ **then**
$T_{i,j}(t+1) := T_{i,j}(t) + \alpha(1 - T_{i,j}(t))$
If $T_{i,j}(t) < 0$ **and** $Cooperation$ **then**
$T_{i,j}(t+1) := (T_{i,j}(t) + \alpha)/(1 - min(|T_{i,j}(t))|, |\alpha|)$
If $T_{i,j}(t) > 0$ **and** $Defection$ **then**
$T_{i,j}(t+1) := (T_{i,j}(t) + \beta)/(1 - min(|T_{i,j}(t))|, |\beta|)$
If $T_{i,j}(t) < 0$ **and** $Defection$ **then**
$T_{i,j}(t+1) := T_{i,j}(t) + \beta(1 + T_{i,j}(t))$

2.2 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. $w_{i,j}(t) \in [-1, 1]$ is the rating associated with the impression of agent i about agent j as a consequence of specific outcome at time t. $W_{i,j}$ is the set of all $w_{i,j}(t)$ 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 $T_{i,j}(t)$, Re-

[1]For the purpose of simplification, we have changed the original notations from [Sabater and Sierra, 2001].

gret 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:

$$T_{i,j}(t) = \sum_{r_k \in W_{i,j}} \rho(t, t_k).r_k \quad (1)$$

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.

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

$$f(t_k, t) = e^{-\frac{t-t_k}{\lambda}} \quad (2)$$

3 Con-man Attack

A con-man, also known as a "confidence man", is someone who takes advantage of someone else – usually for financial gain – using what is known as a confidence trick, where a confidence trick or confidence game is an attempt to defraud a person or group by gaining their confidence.

To model the con-man attack, we use the terms *cooperation* and *defection* from the language of game theory. The level of trust of an agent towards another agent can be changed based on the evaluation of an interaction. If an agent perceives the other agent was cooperative during the specific interaction, its trust in the other agent will be increased. In contrast, if the agent perceives that the other agent has defected for a specific interaction, it will decrease its trust in that agent.

Cooperation and defection in direct interactions have different interpretations depending on the context. In the context of e-commerce, defection in an interaction can be interpreted as the agent failing to satisfy the terms of a contract, selling poor quality goods, late delivery, or does not pay the requested amount of money to a seller depending on the role of the agent [Ramchurn *et al.*, 2004].

What the con-man does is to build up trust from the victim's view point by cooperating with him/her several times. Then, when it comes to a high risk interaction, the con-man will defect. After the con-man has defrauded the victim, he/she has two choices: never interact again with the victim or regain the lost trust with some subsequent cooperative behavior. The con-man, by regaining the victim's trust, can again con (defect) the victim.

In our view, it is hard to understand the intention of a cooperative person and to make sure he/she will continue cooperating forever and will never be tempted to con. Therefore, this work does not plan to identify the con-man before the con happens. Our work is aimed at identifying the repetition of the confidence trick and not let the con-man regain a high trust value easily.

We model the repetition of a confidence trick by introducing the parameter θ. The con-man will defect after θ times of cooperation. After each defection, the con-man will again cooperate θ times. The con-man will repeat this interaction pattern several times (maybe, forever). The formal language (natural language) L over the alphabet $\Sigma = \{C, D\}$ demonstrates the interaction pattern of the con-man:

$$L = \{(C^\theta D)^+ | \theta \geq 1\} \quad (3)$$

where C and D stand for cooperation and defection respectively.

In Section 5, we will demonstrate how three well-known trust models fail to identify the repetition of a confidence trick and the con-man still will have/can gain a high trust value. We have observed this type of attack in reputation management systems used by eBay, for example. In eBay, sellers with good reputations can take advantage of their good reputations to sell a few faulty and low-quality products among plenty of high-quality products that they are selling. For instance, a microphone seller with a good reputation might sell 980 high-quality microphones and 20 faulty microphones every month. Despite his/her defection for selling 20 damaged microphones, he/she can still have a high reputation value (above 90%) since the reputation value is calculated as the sum of all ratings over the last six months.

It should be observed that agents with time-varying behavior have been previously studied in other works to test the adaptability of trust models. For instance, [Hang *et al.*, 2008] introduced damping and capricious agents to analyze the adaptability of its trust scheme. Capricious agents change their behavior between cooperation and defection every two cycles and damping agents have cooperative behavior for several cycles before defecting for the remainder.

4 Con-resistant Trust Models

4.1 Characteristics

To explore the features of con-resistant trust models, we provide a hypothetical example. Alice and Carol are the owners of two separate bakeries. Alice can identify con-men but Carol can not. Bob is the manager of a flour mill. Bob offers to provide high quality flour to each bakery; both accept.

Carol initially accepts daily shipments of 50kg. After 10 satisfactory shipments (cooperations), Carol increases her trust in Bob by doubling her daily order to 100kg. The next day Bob sends low-quality flour at an unchanged price to Carol (a defection). Carol understands the defection and reduces her order to its initial size (50kg). Bob realizes that Carol detected the defection and so cooperates by providing high-quality flour. Bob and Carol continue this cyclical interaction pattern (10 days cooperation then one day defection) for a long time; Carol never understands that Bob is playing a confidence trick on her.

Alice also accepts 50kg daily from Bob who attempts the same confidence trick. Alice doubles her order after 10 satisfactory shipments. However, when Bob defects, Alice realizing the defection reduces her order to 40kg (less than its initial size) and doubles the number of shipments required before increasing her order. When Bob repeats this cycle, Alice remembers the previous defections and reduces her order by 10kg when compared to the starting cycle order. After 5 cycles, Alice cancels her contract with Bob.

Alice detects the confidence trick by doing two things: reducing trust more severely than the previous reduction and decreasing the rate at which trust accumulates with each cooperation. She does this by remembering defections, which Carol does not.

We propose the following characteristics of con-resistant trust models:

- **Cautiously increment trust after defection:** The more the agent perceives defection, the corresponding trust value should be increased more slowly by perceiving the consecutive cooperations.
- **Larger punishment after each defection:** The more the agent perceives defection, the corresponding trust value should be dropped more sharply by perceiving each defection.

If α is the rate of trust increment and β is the rate of trust decrement (referring to Section 2.1), then defection should decrease α but increase the absolute value of β based on the above characteristics. The above characteristics will not remove forgiveness from trust models, which is a frequently noted aspect of trust and reputation theory [Sabater and Sierra, 2001; Axelrod, 1984]. The above characteristics are mainly motivated by the facts that forgiveness is slower when several defections have happened, and punishments are bigger for those who defect more.

4.2 A Con-resistant Extension

We extend the direct trust of [Yu and Singh, 2000] to be con-resistant as defined above. We introduce the following update schema for a positive evidence weighting coefficient of $\alpha > 0$ and a negative evidence weighting coefficient $\beta < 0$ when the agent perceives defection:

$$\alpha := \alpha \times (1 - |\beta|) \tag{4}$$

$$\beta := \beta - \gamma_d \times (1 + \beta) \tag{5}$$

Where γ_d is the discounting factor , and can be calculated based on following formula:

$$\gamma_d = C \times |T_{i,j}| \tag{6}$$

Based on the presented formulae[2], α is decreased with the rate of $1 - |\beta|$ which results in a large decrement of α for a high value of $|\beta|$ and a small decrement of α for a low value of $|\beta|$. We have chosen this rate of decrement because in our view after several defections (when $|\beta|$ is high), making up for a defection should be harder and require more cooperation. As presented in Formula 6, the discounting factor γ_d for the β update is proportional to the absolute value of trust of agent i in agent j, $|T_{i,j}|$. We hypothesize that the discounting factor should be high when the target agent is either trustworthy ($T_{i,j}$ is close to 1) or untrustworthy ($T_{i,j}$ is close to -1). This hypothesis is motivated by the well-known fact that "Trust is hard to earn but easy to lose". $0 < C \leq 1$ is a constant in Formula 6 and is set to $\frac{1}{e}$ in our experiments.

Furthermore, we introduce the following update formula for α when the agents observe cooperation from the other agents:

$$\alpha := \alpha + \gamma_c \times (\alpha_0 - \alpha) \tag{7}$$

$$\alpha := Min(\alpha_0, \alpha) \tag{8}$$

This update results in the increment of α while α will never exceed its initial value, α_0. Therefore, an agent which previously had a decrement in α as a consequence of defection can compensate for it and gradually increase α to the initial value of α_0 by cooperating for some time. γ_c is the learning rate (discounting factor). We believe that if an agent has a high value of β because of its previous defections, its α should be increased more slowly when it is cooperating. Therefore, γ_c should decrease as the magnitude of β increases and we propose the following formula:

$$\gamma_c = 1 - |\beta| \tag{9}$$

5 Experiments

All simulations were run with two agents, one trust-aware agent (TAA) which utilizes a specific trust model (e.g., Regret, FIRE, and Yu and Singh) and a con-man agent (CA). The interaction of agents with each other can be either cooperation or defection. If the trust-aware agent uses Regret or FIRE as a trust model, the cooperation and defection is mapped to 1 and -1 respectively and the value is used as an input of the trust model. In the case of using the Yu and Singh trust model, cooperation and defection will be used directly for the updates of trust value. The interaction strategy of TAAs is tit-for-tat which starts by cooperation and then imitates the opponent's last move. The interaction strategy of CAs follows the formal language presented in Section 3 which is solely dependent on the parameter θ. The strategy of a CA is denoted by SCA(θ). Each agent has 400 interactions with its opponent in one simulation.

5.1 Con-man Attack Vulnerability Demonstration

We continue by demonstrating the vulnerability of three trust models (FIRE, Regret and [Yu and Singh, 2000]) against the con-man attack presented in Section 3.

We ran 5 simulations in each of which a trust-aware agent using the Yu and Singh trust model interacts with a con-man agent with an interaction strategy of SCA(5), SCA(10), SCA(20), SCA(30), or SCA(40). The values of α and β for the Yu and Singh model were set to 0.05 and -0.5 respectively. These values are conservative, leading to trust being built up slowly and reduced quickly. Figure 1 demonstrates the variation of the trust value of TAA over the simulation. It is interesting that the con-man agent with $\theta > 10$ is eventually determined to be trustworthy from the perspective of the trust-aware agent. Although the magnitude of β is set at

[2]Technically, α, β and γ_d should have the subscripts of i and j but they are omitted in the interest of clarity.

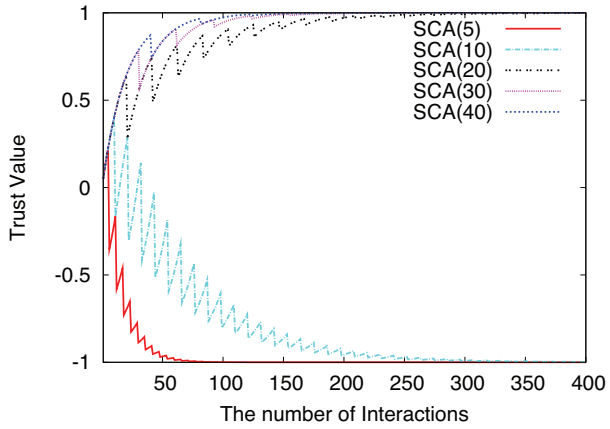

Figure 1: Exploitation of Yu & Singh model by a con-man.

ten times that of α, which leads to a small improvement for a cooperation and a big drop for a defection, the con-man by choosing $\theta > 10$ is known as trustworthy in this trust model with this parameter setting. It is straightforward to show that for each α and β, there is a θ_t that the con-man by choosing its SCA ($\theta > \theta_t$) will still be recognized as trustworthy despite being a con-man.

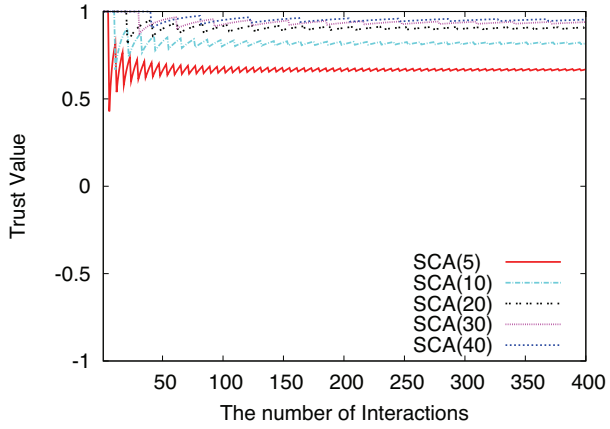

Figure 2: Exploitation of Regret model by a con-man.

We repeated the previous experiment where the trust-aware agent employs the Regret model as explained in Section 2.2. Figure 2 shows the trust value variation of TAA over the 400 interactions. It is clear that the con-man with SCA(5) can stabilize its trust value at 0.66. Moreover, by increasing θ to 10, 20, 30 and 40, the con-man agent can reach a trust value of 0.81, 0.90, 0.93, and 0.95, which are high values of trust for a con-man; i.e., the agent is considered trustworthy.

The previous experiments were repeated with the trust-aware agent employing the FIRE model as explained in Section 2.3. We set $\lambda = \frac{-5}{Ln(0.5)}$ as proposed in the original research. Figure 3 depicts the variation of the trust value of TAA over the simulation (The larger gray box magnifies part of the graph for clarity). Although FIRE is more sensitive to

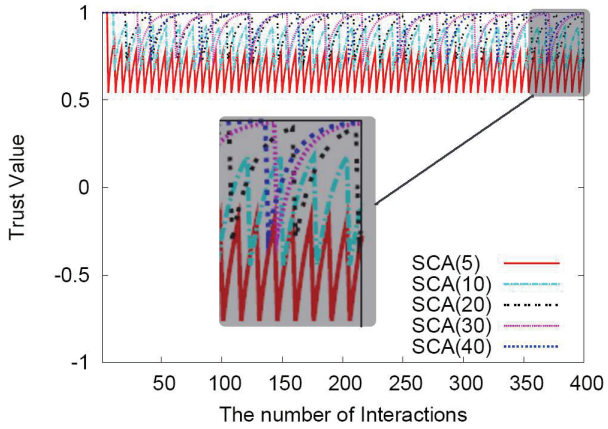

Figure 3: Exploitation of FIRE model by a con-man.

defection when compared to Regret as a result of its enhanced rating recency function, it is still vulnerable to the con-man attack. As shown, the con-man with SCA(5) can have trust value in the range of 0.56 to 0.73. Moreover, by increasing θ to 10, 20, 30 and 40, the con-man agent can ensure that its trust value will not fall below that of 0.67, 0.72, 0.73, and 0.74 respectively, while the maximum value is close to 1.

5.2 Results For A Con-resistant Trust Model

We ran the simulations with the same settings as explained in Section 5.1 with the difference that the trust-aware agent used our con-resistant trust updates as presented in Section 4.2. The initial values of α and β (α_0 and β_0) were set to 0.05 and -0.5 respectively. Figure 4 shows the trust value variation of the TAA over the 400 interactions. Interestingly, regardless of the value of θ for SCA(θ), the con-man was recognized by the trust model and achieved a low value of trust. It is worth noting that the con-man still has a chance to be forgiven but with a very large number of cooperations and a change in its pattern of interaction. Figure 4 also shows that the speed of detection of the con-man is inversely proportional to θ.

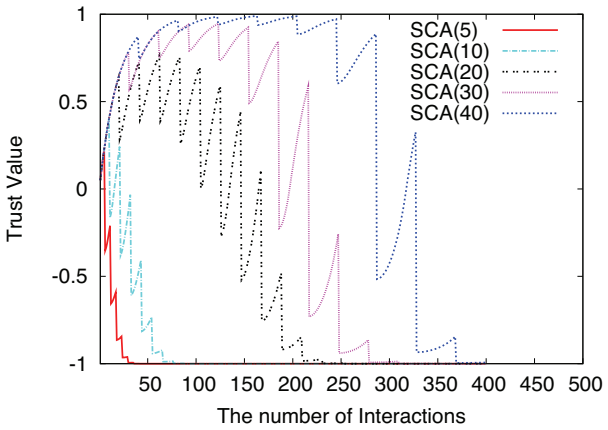

Figure 4: The reaction of con-resistant extension of Yu & Singh to the con-man attack.

To understand the effect of α_0 and β_0, we ran a similar simulation experiment when the con-man only uses SCA(20) but the trust-aware agent had different initialization values of α_0 and β_0 for each simulation. Not only did the trust-aware agent recognize the CA as an untrustworthy agent during the 400 interactions but also the final values of α and β for different initializations were close to each other as presented in Table 1. Similar final values for α and β support the hypothesis that α and β update formulae are insensitive to the values of α_0 and β_0.

	$\alpha_0 = 0.20$ $\beta_0 = -0.2$	$\alpha_0 = 0.15$ $\beta_0 = -0.3$	$\alpha_0 = 0.10$ $\beta_0 = -0.4$	$\alpha_0 = 0.05$ $\beta_0 = -0.5$
α	0.00003	0.00002	0.00002	0.00005
β	-0.99983	-0.99984	-0.99981	-0.99893

Table 1: Final Values of α and β after 400 interactions of the trust-aware agent with the con-man with SCA(20).

6 Concluding Remarks and Future Work

This paper is motivated by the dire need for trust and reputation models in artificial societies, especially e-commerce. While reviewing important existing trust and reputation models from the literature, embracing centralized and decentralized models, we have noted a tendency to focus on trust variation rather than the identification of trustworthy or untrustworthy agents and a reliance on information derived from direct interactions. As a result, we have noted the exposure of such models to repeated cycles of exploitation. This vulnerability reinforces the need for new criteria for trust and reputation models called attack resistance which reflects the ability of a trust model to be unaffected by agents who try to manipulate the trust model.

The con-man attack introduced and modeled in this work has been applied to direct trust components of trust models. In the con-man attack, a con-man usually takes advantage of someone else and attempts to defraud that person by gaining their confidence. We have demonstrated how a con-man can exploit three well-known trust models [Yu and Singh, 2000], Regret, and FIRE such that he/she is still known as trustworthy after repeated cycles of interaction while conning others. Therefore, we have introduced two characteristics of con-resistant trust models: first, cautiously increment trust after having seen any defection and second, larger punishments after each defection. Based on the proposed features, we proposed a con-resistant scheme and empirically demonstrated its utility.

We plan to design a con-resistant extension for Regret that can also be used for FIRE. With the advent of probabilistic trust models, future work will include the design of con-resistant probabilistic trust and reputation models.

The con-man attack can be extended to more complicated attacks in which the con-man observes the behavior of his/her opponents and change his/her interaction patterns based on those observations (e.g., θ can be adaptive over the interactions of a con-man). Design of these attacks might provide more comprehensive insight in characteristics of con-resistant trust models.

Furthermore, it would be interesting to observe the effect of con-man agents in a society of agents that employ social mechanisms (e.g., the use of witness information) which helps agents avoid encounters with con-man agents and thereby reduce exposure to confidence tricks.

References

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

[Fullam *et al.*, 2005] 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*, pages 512–518, New York, NY, USA, 2005. ACM.

[Hang *et al.*, 2008] Chung-Wei Hang, Yonghong Wang, and Munindar P. Singh. An adaptive probabilistic trust model and its evaluation. In *AAMAS (3)*, pages 1485–1488, 2008.

[Huynh *et al.*, 2006] 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.

[Mui *et al.*, 2002] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation for e-businesses. In *HICSS '02*, page 188, Washington, DC, USA, 2002. IEEE Computer Society.

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

[Sabater and Sierra, 2001] 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.

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

[Teacy *et al.*, 2005] 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*, pages 997–1004, New York, NY, USA, 2005. ACM.

[Yu and Singh, 2000] Bin Yu and Munindar P. Singh. A social mechanism of reputation management in electronic communities. In *CIA '00*, pages 154–165, London, UK, 2000. Springer-Verlag.

[Yu and Singh, 2002] Bin Yu and Munindar P. Singh. An evidential model of distributed reputation management. In *AAMAS '02*, pages 294–301, New York, NY, USA, 2002. ACM.

[Yu and Singh, 2003] Bin Yu and Munindar P. Singh. Detecting deception in reputation management. In *AAMAS '03*, pages 73–80, New York, NY, USA, 2003. ACM.