Trust Dynamics Analysis of CTR Scheme Subversion under Virtual Anonymity and Trust-Unaware Partner Selection

Jerzy Konorski

Faculty of ETI, Gdańsk University of Technology
ul. Narutowicza 11/12, 80-233 Gdańsk, Poland
jekon@eti.pg.gda.pl

Abstract

We propose a framework to study Markovian trust value dynamics in a centralized *Computational Trust and Reputation* (CTR) scheme under trust-unaware partner selection using a mean-value approximation. Analytically founded answers are sought to questions like: Can dishonest agents subvert the CTR scheme (*i.e.*, acquire higher trust values than honest agents)? Is indirect reciprocity incentivized? Is there a qualitative impact of a growing proportion off dishonest agents and collusion among them?

1 Introduction

A number of tenets of *Computational Trust and Reputation* (CTR) scheme design are recognized, but have yet to be captured analytically. In this paper we develop an analytical framework to study the trust dynamics for a centralized CTR scheme under trust-unaware partner selection. Our CTR scheme features a single *Reputation Aggregation Engine* (RAE). Agents occasionally interact to exchange *service*. Each interaction involves two *partners*, a *service provider* and a *service recipient*; the latter decide on service providers according to a *partner selection policy* and after an interaction report to RAE the amount of received service (called *reputation data* or *reported service*). RAE aggregates reputation data into agents' trust values, which it subsequently disseminates. Agents enjoy virtual anonymity [Del00], [Sei04], i.e., use time-varying pseudonyms that only RAE can map to agents' permanent identities. A *service policy* dictates the *treatment* of the partner–the amount of provided or reported service subject to *available service*. If treatments depend on partners' trust values, the service policy is called *trust-aware*, otherwise it is *trust-unaware*. In the former case, a closed loop containing the RAE to service policy link is created in Figure 1a (symbols are explained in Sec. 2). By *goodwill* we mean the sensitivity of treatments to partners' trust values. Favorable treatments disregarding trust values signify utmost goodwill, whereas nonzero treatments only of high-trust partners signify scant goodwill. Better treatment of agents with higher trust values rewards their favorable past treatments of third-party agents; such a service policy exhibits *indirect reciprocity*. We view a trust value as a reputation-based decision-support variable governing who to interact with and how–hence close to "standing" [Now98], [Oht04].

Likewise, a partner selection policy can be trust-aware or trust-unaware. In the former, a closed loop containing the RAE to partner selection policy link is created in Figure 1a, and a distinction between *pooled* and *partitioned*

In: J. Zhang, R. Cohen, and M. Sensoy (eds.): Proceedings of the 18th International Workshop on Trust in Agent Societies, Singapore, 09-MAY-2016, published at http://ceur-ws.org

service availability is relevant. With partitioned service availability, separate resources are reserved for each service recipient being interacted with. With pooled service availability all recipients draw on a common resource pool and high reputation can backfire on an agent, who can then experience *reputation damage* [Yu13]. Figure 1b summarizes the main goals of CTR schemes assuming agents divide into intrinsically good and bad. Under trust-unaware partner selection and service policies a CTR scheme is open-loop, and only aims at raising red flags on bad agents. Trust-aware partner selection enables ostracizing bad agents subject to reputation damage control. Trust-aware service policy enables differentiation of provided service (also of reported service, which is not important under trust-aware partner selection).

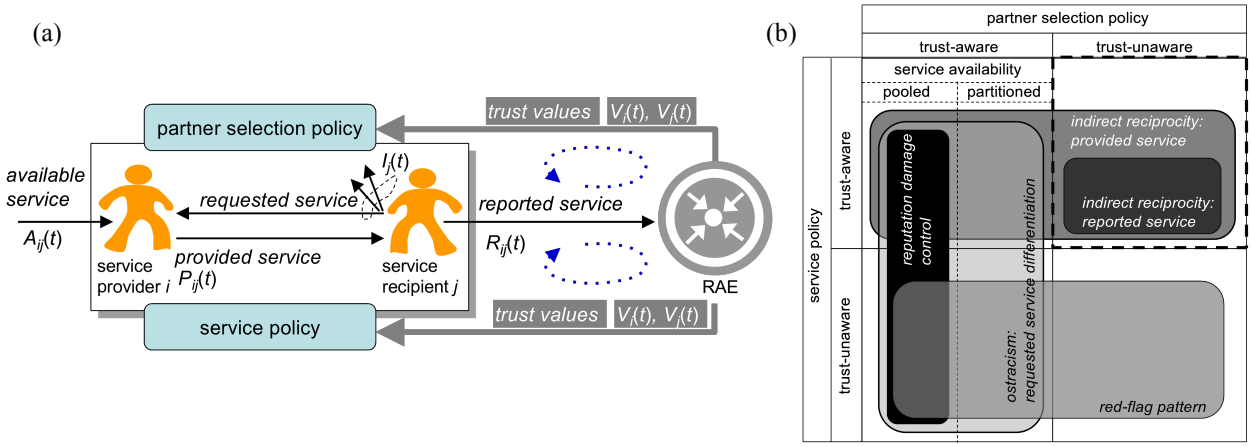

Figure 1: (a) Multi-agent CTR scheme with closed loops; (b) Goals of CTR schemes in open- and closed-loop models

We assume trust-unaware, *e.g.*, random partner selection, as appropriate in systems like mobile ad hoc networks, where interactions are restricted by connectivity rather than trustworthiness, or communities from which bad agents cannot be entirely removed since they possess some scarce resource [Hen15]. Compared to trust-aware partner selection this is a more challenging setting, since bad agents never cease to receive and report service the way they want, which can easily confuse RAE. Good agents' defense then lies in trust-aware service policy. We call intrinsically good and bad agents *honest* and *skimp* & *slander* (s&s), respectively; we also speak of an *honest* or *s&s service policy*. The latter entails (selfish or malicious) *skimp* and/or *slander* attacks, *i.e.*, providing and/or reporting less service than honest service policy dictates. Selfish s&s agents strive to maximize own trust values, whereas malicious s&s agents strive to minimize those of honest agents. In a *collusion* scenario, s&s agents recognize one another despite virtual anonymity and apply towards fellow colluders *cronyism* and/or *undue appraisal* attacks, *i.e.*, provide and/or report more service than honest service policy dictates. RAE aggregates reputation data weighted by current trust values and runs a clustering algorithm producing two levels of trust values. Ideally, the higher level will be acquired by honest agents and the lower by s&s agents. In the opposite case we say that a *CTR scheme subversion* occurs. We only allow time-invariant service policies, reflecting the notion of *constant* attacks in [Zha12]. The proposed simple framework for Markovian trust value dynamics gives analytically founded answers to questions like:

- Can s&s agents subvert the CTR scheme and under what conditions?
- Is indirect reciprocity incentivized, *i.e.*, will honest agents condition treatments upon trust values?
- Is there a qualitative impact of a growing proportion of s&s agents and collusion among them?

Our analysis should also capture or somehow relate to well-known tenets of CTR scheme design; several of them, listed below as T1 through T6, can be deduced from the above discussion and literature.

T1. *A CTR scheme should bring honest agents better treatments than they would get without it*, assuming all agents self-optimize. Since treatments are not differentiated in a trust-unaware system, tenet T1 simply postulates no CTR scheme subversion.

T2. *Trust values may fail to reflect agents' intrinsic qualities*, CTR scheme subversion being an extreme case.

T3. *Goodwill prescribed by an honest service policy should be scant enough to defend against selfish s&s agents, but enough to defend against malicious ones.*

T4. Good reputation is a fiat currency: it only increases an agent's fitness if reciprocity norms are observed [Now98]. *A CTR scheme should incentivize indirect reciprocity among honest agents*, cf. [Wan12].

T5. Increased and more coordinated input from s&s agents to RAE makes it hard to differentiate between honest and s&s agents: *as the proportion of s&s agents grows and collusion among them sets in, a CTR scheme becomes less effective*, cf. [Zha12].

T6. A CTR scheme relies upon reporting received service, for which endogenous incentives should be sought: *reporting received service should help honest agents acquire higher trust values.*

The remainder of the paper is organized as follows. Section 2 details the model of the multi-agent system and CTR scheme. Section 3 presents the trust value dynamics for weighted averaging of reputation data and a class of service policies. In Section 4 we investigate steady-state trust values and give analytical foundation for some tenets of CTR scheme design. A beneficial model extension is discussed in Section 5. Section 6 briefly discusses selected related works. Section 7 concludes and outlines directions of future work.

2 Model Specifics

Consider a large set N of agents and a single RAE. We make the following assumptions, cf. Figure 1a:

(i) Time is divided into *cycles* $t = 1, 2, \ldots$; in each one service recipients select partners to request service from. Received service is reported to RAE and aggregated into agents' trust values, denoted $V_i(t) \in [0, 1]$.

(ii) In cycle t agent j selects a random set of partners $I_j(t) \subseteq N \backslash \{j\}$.

(iii) $A_{ij}(t) \in [0, 1]$ is the amount of service available at a service provider i for a service recipient j in cycle t, relative to the amount requested by j. Due to temporary resource shortages, variable requested service etc., $A_{ij}(t)$ are modeled as exogenous random variables with the iid property across $t = 1, 2, \ldots$, and public-knowledge probability distributions $F_{ij}(a) = Prob(A_{ij}(t) < a)$. They are also independent across $j \in N \backslash \{i\}$ (partitioned service availability).

(iv) $P_{ij}(t) \in [0, 1]$ and $R_{ij}(t) \in [0, 1]$ denote treatments in cycle t, respectively the amount of service provided by i to j and reported by j as received from i. In general, $P_{ij}(t) \neq A_{ij}(t)$ as prescribed by i's service policy, (*e.g.*, may depend on $V_j(t)$), and $R_{ij}(t) \neq P_{ij}(t)$ as prescribed by j's service policy (*e.g.*, may depend on $V_i(t)$).

(v) $N = S \cup H$, where S and H are disjoint sets of s&s (selfish or malicious) and honest agents; let $\xi = |S|/|N| > 0$.

(vi) In a collusion scenario s&s agents can tell fellow s&s agents and selectively apply cronyism and/or undue appraisal; in a non-collusion scenario they cannot differentiate treatments other than based on trust values.

Note that RAE cannot reliably guess agents' intrinsic qualities (s&s or honest): deciding if a record of $R_{ij}(t), t = 1, 2, \ldots$ "looks honest" given $F_{ij}(\cdot)$ would need the knowledge of both agents' intrinsic qualities.

3 Analytical Framework

In this section we construct trust value dynamics for a class of service policies; asymptotic trust values can then be obtained as fixed points of a corresponding nonlinear mapping.

3.1 Reputation Aggregation and Trust Value Dynamics

RAE uses in each cycle an eigenvector-type reputation aggregation algorithm that obtains a weighted average of reputation data regarding agent i:

$$R_{i,avg}(t) = \sum_{j \in N \setminus \{i\}} V_j(t) \delta^{\Delta_{ij}(t)} R_{ij}(t - \Delta_{ij}(t)) \ , \tag{1}$$

where $\delta \in [0,1]$ is a decay factor and $\Delta_{ij}(t)$ is the number of cycles since i last provided service to j ($\Delta_{ij}(t) = 0$ for $i \in I_j(t)$). RAE distinguishes just two levels of trust values and assigns them to subsets $N_{\text{high}}(t)$ and $N_{\text{low}}(t)$ of agents with "high" and "low" $R_{i,\text{avg}}(t)$ values. To determine these subsets, RAE applies a clustering algorithm such that if $i \in N_{\text{high}}(t)$ and $j \in N_{\text{low}}(t)$ then $R_{i,\text{avg}}(t) \geq R_{j,\text{avg}}(t)$. Normalization is applied so that agents in $N_{\text{high}}(t)$ acquire trust values equal to 1. This yields (with 0/0 defined to be 0):

$$V_i(t+1) = \frac{\sum_{l \in N_i(t)} R_l, avg(t)/|N_i(t)|}{\sum_{l \in N high(t)} R_l, avg(t)/|N_{high}(t)|} \ , \tag{2}$$

where $N_i(t)$ is the subset containing i. If s&s and honest service policies differ substantially then $N_{\text{high}}(t) = S$ and $N_{\text{low}}(t) = H$ or *vice versa*, *i.e.*, s&s and honest agents are accurately partitioned; otherwise the partition is not very relevant. Due to possible CTR scheme subversion, RAE cannot pinpoint s&s agents with certainty. Given $V_i(0)$, (1) and (2) produce explicit trust value dynamics in the form of a multidimensional homogeneous discrete-time Markov chain on $[0,1]^{|N|}$. There may be absorbing states: e.g., zero trust values may tend to spread among agents, reflecting a denial of service. It is not clear whether such an absorbing state will ever be reached; even so, one can speak of a *wide-sense steady state* of (2) if, for all $i \in N$ and a long enough time, the variability range of $V_i(t)$ is much narrower than the interval [0, 1]. (An absorbing state is a special case.) For large $|N|$, a wide-sense steady state can be observed almost indefinitely, as illustrated in Sec. 3.4.

3.2 Service Policy

A service provider i interacting with a service recipient j is limited by $P_{ij}(t) \leq A_{ij}(t)$ and its service policy; assume that the latter moreover implies $P_{ij}(t) \leq p_{ij}$, where $p_{ij} \in [0,1]$ is a threshold. The above two constraints can be met, *e.g.*, by $P_{ij}(t) = \sigma_i(A_{ij}(t), p_{ij})$, where $\sigma_i(a,p) = min(a,p)$ or ap (a *discriminative* or *multiplicative* policy, respectively). For $R_{ij}(t)$, we assume $R_{ij}(t) \leq P_{ij}(t)$ and let j's service policy moreover imply $R_{ij}(t) \leq r_{ij}$, where $r_{ij} \in [0,1]$ is a threshold. Thus:

$$R_{ij}(t) = \sigma_j(P_{ij}(t), r_{ij}) = \sigma_j(\sigma_i(A_{ij}(t), p_{ij}), r_{ij}) \ . \tag{3}$$

The thresholds p_{ij} and r_{ij} quantify agents' goodwill (propensity for skimp, cronyism, slander, or undue appraisal). It is natural that p_{ij} increases with $V_j(t)$ and r_{ij} increases with $V_i(t)$. If both i and j are s&s and collude then i applies cronyism and j applies undue appraisal–respectively, $p_{ij} = r_{ij} = 1$. Finally, towards an honest partner, or any partner in a non-collusion scenario, an s&s service policy applies discretionary thresholds.

3.3 Symmetric System Model

Assume in addition that (vii) service policy only distinguishes between s&s and honest partners, and (viii) $F_{ij}(\cdot) \equiv F(\cdot)$. Assumption (vii) simplifies the description of a service policy, since now only $\sigma s(\cdot,\cdot\cdot)$ and $\sigma h(\cdot,\cdot\cdot)$, as well as pqq' and rqq', need to be distinguished, where q,q' $\in$ {s, h}. All $V_i(t), i \in S$, are now represented as $Vs(t)$, and all $V_i(t), i \in H$, as $Vh(t)$. We take

$$p_{\text{qq'}} = \begin{cases} \tilde{y} & \text{q = s and q' = s} \\ y & \text{q = s and q' = h} \\ wL(V_{\text{q'}}(t), x) & \text{q = h} \end{cases} \qquad r_{\text{qq'}} = \begin{cases} \tilde{z} & \text{q = s and q' = s} \\ z & \text{q = h and q' = s} \\ wL(V_{\text{q}}(t), x) & \text{q' = h} \end{cases} \tag{4}$$

where $\tilde{y} = y$ and $\tilde{z} = z$ in a non-collusion scenario (s&s agents cannot tell s&s partners from honest), and $\tilde{y} = \tilde{z} = 1$ in a collusion scenario; $y, z \in [0,1]$ are discretionary thresholds imposed by s&s agents ($y = z = 0$ and $y = z = 1$ signify scant and utmost goodwill, respectively). Honest agents offer goodwill (limit the amount of provided or reported service) depending on the partner's current trust value through a function $L : [0,1]^2 \to [0,1]$; $L(v,x)$ is returned for the trust value v given the *goodwill shape factor* $x \in [0,1]$, whereas $w \in [0,1]$ is the

goodwill downscale factor. Goodwill is calibrated from utmost, with $x = w = 1$ (trust values are disregarded and no downscaling applies), to scant, with $x = 0$ and $w \approx 0$ (only highest-trust agents receive nonzero service w). We take $L(v, x)$ to be continuous and nondecreasing in v and $x > 0$, converging point-wise to $\mathbf{1}(v \leq 1)$ as $x \to 1$ and to $\mathbf{1}(v = 1)$ as $x \to 0$, where $\mathbf{1}(\cdot)$ is the indicator function. Examples are *unit-step* functions, which produce an intuitive "binary" honest service policy only offering goodwill if the partner has a large enough trust value: $P_{ij}(t) = A_{ij}(t) \cdot w \cdot \mathbf{1}(V_j(t) \geq 1 - x)$ and $R_{ij}(t) = P_{ij}(t) \cdot w \cdot \mathbf{1}(V_i(t) \geq 1 - x)$; the type of honest service policy (discriminative or multiplicative) is then irrelevant. Note that indirect reciprocity occurs for $x < 1$.

3.4 Mean-Value Dynamics

We note that in our model, (1) is statistically indistinguishable from $\sum_{j \in N \setminus \{i\}} V_j(t) \delta^{\Delta_{ij}(t)} R_{ij}(t)$. We substitute the latter into (2) and use the law of large numbers to obtain approximate deterministic trust value dynamics:

$$V_i(t+1) \approx \frac{\sum_{j \in N} V_j(t) \mathsf{E} \delta^{\Delta_{ij}(t)} \mathsf{E} R_{lj}(t)|_{l \in N_i(t)}}{\sum_{j \in N} V_j(t) \mathsf{E} \delta^{\Delta_{ij}(t)} \mathsf{E} R_{lj}(t)|_{l \in N_{\text{high}}(t)}} = \frac{\sum_{j \in N} V_j(t) \mathsf{E} R_{lj}(t)|_{l \in N_i(t)}}{\sum_{j \in N} V_j(t) \mathsf{E} R_{lj}(t)|_{l \in N_{\text{high}}(t)}}, i \in N, \tag{5}$$

where E denotes the statistical expectation of $F(\cdot)$. We group the summands assuming that s&s and honest agents are accurately partitioned, and replace $V_j(t)$ by $V_{\text{q}}(t)$ and $R_{ij}(t)$ by $R_{\text{qq'}}(t)$, where q,q' $\in \{\text{s}, \text{h}\}$. Then (5) becomes:

$$V_{\text{q}}(t+1) = \frac{\xi V_{\text{s}}(t) \mathsf{E} R_{\text{qs}}(t) + (1-\xi) V_{\text{h}}(t) \mathsf{E} R_{\text{qh}}(t)}{M(t)}, \text{q} = \text{s}, \text{h}, \tag{6}$$

where $M(t)$ is the larger of the two numerators. Thus $\max\{V_{\text{s}}(t), V_{\text{h}}(t)\} = 1$.

To account for the s&s and honest service policies introduce a continuous, nondecreasing and quasi-concave function $\Psi : [0,1] \to [0, \mathsf{E}A]$, where $\Psi(p) = \mathsf{E}\min\{A, p\} = \int_0^p (1 - F(a))da$ and A $\sim F(\cdot)$; we have $\Psi(0) = 0$, $\Psi(1) = \mathsf{E}A$, and $\Psi(pr) \geq p\Psi(r)$ for any $p, r \in [0,1]$. Let mean reported service be represented by $\Omega_{\text{qq'}} : [0,1]^2 \to [0, \mathsf{E}A]$, with $\Omega_{\text{qq'}}(p,r) = \mathsf{E} R_{\text{qq'}}(t)|_{p_{\text{qq'}}=p, r_{\text{qq'}}=r}$. From (3),

$$\Omega_{\text{qq'}}(p,r) = \begin{cases} \Psi(\min\{p,r\}), & \text{discriminative } \sigma q(\cdot,\cdot\cdot) \text{ and } \sigma q'(\cdot,\cdot\cdot) \\ r\Psi(p), & \text{discriminative } \sigma q(\cdot,\cdot\cdot) \text{ , multiplicative } \sigma q'(\cdot,\cdot\cdot) \\ p\Psi(\min\{1, r/p\}), & \text{multiplicative } \sigma q(\cdot,\cdot\cdot) \text{ , discriminative } \sigma q'(\cdot,\cdot\cdot) \\ pr\Psi(1), & \text{multiplicative } \sigma q(\cdot,\cdot\cdot) \text{ and } \sigma q'(\cdot,\cdot\cdot). \end{cases} \tag{7}$$

Hence $\Omega_{\text{qq'}}(\cdot,\cdot\cdot)$ is continuous and nondecreasing in both variables, with $\Omega_{\text{qq'}}(0,r) = \Omega_{\text{qq'}}(p,0) = 0$ and $\Omega_{\text{qq'}}(1,1) = \Psi(1) = \mathsf{E}A$. Based on (4), (6) can be rewritten as

$$\begin{aligned} V_{\text{s}}(t+1) &= \frac{\xi V_{\text{s}}(t) \Omega_{\text{ss}}(\tilde{y}, \tilde{z}) + (1-\xi) V_{\text{h}}(t) \Omega_{\text{sh}}(y, wL(V_{\text{s}}(t), x))}{M(t)}, \\ V_{\text{h}}(t+1) &= \frac{\xi V_{\text{s}}(t) \Omega_{\text{hs}}(wL(V_{\text{s}}(t), x), z) + (1-\xi) V_{\text{h}}(t) \Omega_{\text{hh}}(wL(V_{\text{h}}(t), x), wL(V_{\text{h}}(t), x))}{M(t)}, \end{aligned} \tag{8}$$

Equation (8) defines a deterministic fixed-point iteration process for a nonlinear continuous mapping of $(V_{\text{s}}, V_{\text{h}})$, where the compact convex set $[0,1]^2$ maps to itself. By Brouwer's theorem, a fixed point exists. Since RAE is unable to distinguish s&s and honest agents, necessarily $V_{\text{s}}(0) = V_{\text{h}}(0) = 1$. If iterations converge to a fixed point, the limits $V_{\text{so}} = V_{\text{s}}(\infty)$ and $V_{\text{ho}} = V_{\text{h}}(\infty)$ approximate the wide-sense steady-state of (2); possible other fixed points are then irrelevant.

In Figure 2 we compare the mean-value trajectories (8) subject to $V_{\text{s}}(0) = V_{\text{h}}(0) = 1$, and the averages $\sum_{i \in S} V_i(t)/|S|$ and $\sum_{i \in H} V_i(t)/|H|$ of the Markovian trajectories (2) subject to $V_i(0) = 1$. RAE adopts a simple clustering algorithm: having ordered the averages of reputation data (1) so that $R_{i_1,\text{avg}} \geq \ldots \geq R_{i_{|N|},\text{avg}}$, minimize the sum of the "high" and "low" subset diameters, *i.e.*, find $\mu^* = \text{argmin}_{1+m<\mu<|N|-m}(R_{i_{1+m},\text{avg}} - R_{i_\mu,\text{avg}} + R_{i_{\mu+1},\text{avg}} - R_{i_{|N|-m},\text{avg}})$, where $m \geq 0$ eliminates possible outliers (we take $m = 1$). Then $N_{\text{high}}(t) = \{i_1, \ldots, i_{\mu^*}\}$.

To warrant the law of large numbers, there should be enough nonnegligible summands in (2), hence δ should be large enough. When $\delta = 1$ and $x = 0.25$ (Figure 2a,b), the clustering is perfect, *i.e.*, $N_i(t) = S$ iff $i \in S$, hence (8) quickly converges to a wide-sense steady state of (2) (with honest agents absorbed at $V_i(t) = 0$). For $y = 0.89$, $V_{\text{h}}(t) = 0$ is approximated rather inaccurately as it converges, yet its steady-state behavior is captured.

Decreasing δ to 0.8 (Figure 2c,d) has little effect if $y = 0.89$, but for $y = 0.51$ causes a CTR scheme subversion and a discrepancy between the Markovian and mean-value trajectories; this is because the clustering algorithm fixes upon incorrect clusters after $t = 32$. A further decrease of δ to 0.6 (Figure 2e) confuses the clustering algorithm altogether and drives the Markovian trajectories to a denial of service. In Figure 2f, $x = 0.65$, $\delta = 1$, and s&s agents offer high goodwill: $y = z = 0.9$. The clustering algorithm is again confused and no wide-sense steady state occurs; nevertheless the discrepancies between the Markovian and mean-value trajectories are mild.

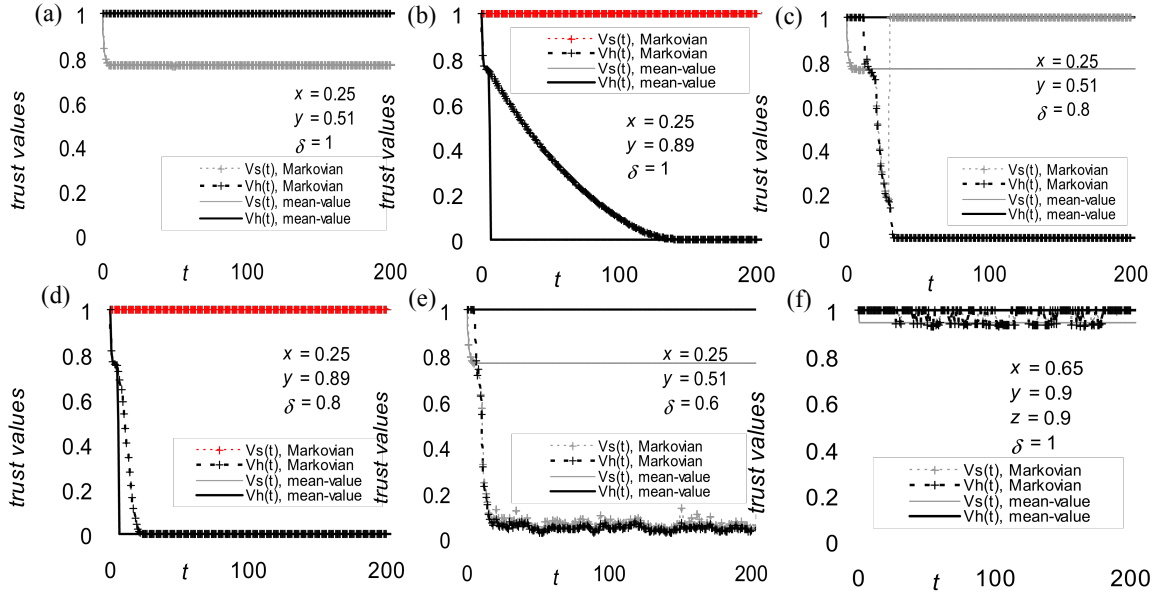

Figure 2: Markovian and mean-value trust value trajectories in a collusion scenario; $\xi = 25\%$, unit-step $L(\cdot,\cdot)$, multiplicative $\sigma_s(\cdot,\cdot)$

4 Mean-Value Analysis

The results in Figure 2 encourage the use of (8) instead of (2) for investigation of some qualitative properties of agents' trust values. First we verify that our CTR scheme is in a sense "fair" to s&s agents.

Proposition 1. If s&s agents do not apply skimp attacks then there exists an s&s service policy (namely, extreme slander, $z = 0$) that yields them $V\text{so} = 1$ regardless of the honest service policy (x and w). If they do apply skimp attacks with a small enough y while being in the minority ($\xi < 1/2$) then there exists a honest service policy (namely, a large enough w with any x) that drives $V\text{so}$ arbitrarily close to 0.

Proof: To satisfy the premise of the first part, one substitutes $wL(V_\text{h}(t), x)$ for y in (8), moreover, it must be that either (a) $\sigma_\text{s}(\cdot,\cdot)$ and $\sigma_\text{h}(\cdot,\cdot)$ are of the same type (discriminative or multiplicative), or (b) $\sigma_\text{s}(\cdot,\cdot)$ is discriminative and $\sigma_\text{h}(\cdot,\cdot)$ is multiplicative. Let $z = 0$. Recall that $V_\text{s}(0) = V_\text{h}(0) = 1$ and assume $V_\text{h}(t) \le V_\text{s}(t) = 1$ for some $t \ge 0$. For brevity denote $L_\text{s} = L(V_\text{s}(t), x)$ and $L_\text{h} = L(V_\text{h}(t), x)$ ($L_\text{h} \le L_\text{s} = 1$ by the properties of $L(\cdot,\cdot)$). Then $\Omega_\text{hh}(wL_\text{h}, wL_\text{h}) \le \Omega_\text{sh}(wL_\text{h}, wL_\text{s})$: in case (a) by the properties of $\Omega_{\text{qq}'}(\cdot,\cdot)$, and in case (b) by (7) and the properties of $\Psi(\cdot)$. Consequently, $V_\text{h}(t+1) \le V_\text{s}(t+1)$ and by induction, $V_\text{so} = 1$. For the second part observe that for a small enough y and a non-collusion scenario, $V_\text{s}(t)$ is arbitrarily close to 0 for all t. For a collusion scenario assume $V_\text{s}(t) \le V_\text{h}(t) = 1$ for some t. Then from the first equation of (8),

$$V_\text{s}(t+1) \le \frac{c\Psi(1)}{\Omega_\text{hh}(w,w)} V_\text{s}(t) + \frac{\Omega_\text{hh}(y, wL_\text{s})}{\Omega_\text{hh}(w,w)} \tag{9}$$

where $c = \xi/(1-\xi) < 1$. Let $\alpha = c\Psi(1)/\Omega_\text{hh}(w,w)$. In a generic equation $v = \alpha v + \Omega_\text{sh}(y, wL(v,x))/\Omega_\text{hh}(w,w)$, the right-hand side can be made arbitrarily close to αv with a small enough y, cf. (7). On the other hand, taking w close enough to 1 makes α arbitrarily close to c, hence $V_\text{s}(t+1) < V_\text{s}(t), V_\text{h}(t+1) = 1$, and the largest root of the equation is arbitrarily close to 0 depending on the behavior of the last term at $v = 0$. Thus if the fixed-point iteration process (9) for $t' > t$ converges to some V_so then V_so is arbitrarily close to 0. ■

Analyzing (8) similarly one sees that lack of indirect reciprocity ($x = 1$) enables s&s agents to acquire trust values equal to honest agents', which illustrates tenets T1 and T4. The following two propositions show a qualitative difference between a collusion and a non-collusion scenario as regards CTR scheme subversions.

Proposition 2. In a non-collusion scenario with $\xi < 1/2$, taking $w = 1$ prevents CTR scheme subversion, *i.e.*, $V_{\rm so} \leq V_{\rm ho} = 1$ for any x, y, and z.

Proof: To use the inductive argument assume $V_{\rm s}(t) \leq V_{\rm h}(t) = 1$ for some $t \leq 0$. Rewrite (8) substituting $\tilde{y} = y, \tilde{z} = z$ and using shorthand $L_{\rm s}$ for $L(V_{\rm s}(t), x)$:

$$\begin{aligned} M(t)V_{\rm s}(t+1) &= \xi V_{\rm s}(t)\Omega_{\rm ss}(y,z) + (1-\xi)\Omega_{\rm sh}(y, wL_{\rm s}) \ , \\ M(t)V_{\rm h}(t+1) &= \xi V_{\rm s}(t)\Omega_{\rm hs}(wL_{\rm s}, z) + (1-\xi)\Omega_{\rm hh}(w,w) \ . \end{aligned} \tag{10}$$

Define $\mathit{diff} = 2M(t)(V_{\rm h}(t+1) - V_{\rm s}(t+1))$, and take a large enough w such that $\Omega_{\rm hh}(w,w) \geq \Omega_{\rm sh}(1, wL_{\rm s})$ (by (7), this is guaranteed if $\sigma_{\rm h}(\cdot,\cdot\cdot)$ is discriminative and requires $w \geq L_{\rm s}$ if $\sigma_{\rm h}(\cdot,\cdot\cdot)$ is multiplicative). Then

$$\begin{aligned} \mathit{diff} &= 2(1-\xi)(\Omega_{\rm hh}(w,w) - \Omega_{\rm sh}(y, wL_{\rm s})) - 2\xi V_{\rm s}(t)(\Omega_{\rm ss}(y,z) - \Omega_{\rm hs}(wL_{\rm s}, z)) \\ &\geq \Omega_{\rm hh}(w,w) - \Omega_{\rm sh}(1, wL_{\rm s}) - (\Omega_{\rm ss}(1,z) - \Omega_{\rm hs}(wL_{\rm s}, z)) = RHS \ . \end{aligned} \tag{11}$$

We can show that $w = 1$ guarantees $RHS \geq 0$ regardless of x, y, and z, but $w < 1$ does not. Indeed, in the case where $\sigma_{\rm s}(\cdot,\cdot\cdot)$ and $\sigma_{\rm h}(\cdot,\cdot\cdot)$ are both multiplicative, we have $RHS = w^2\Psi(1) - wL_{\rm s}\Psi(1) - (z\Psi(1) - zwL_{\rm s}\Psi(1)) = (1-z)(1-wL_{\rm s})\Psi(1) - (1-w^2)\Psi(1)$. The other three cases can be examined analogously. Hence, $V_{\rm s}(t+1) \leq V_{\rm h}(t+1) = 1$ and the proposition follows. ■

Proposition 3. In a collusion scenario, CTR scheme subversion is always possible, *i.e.*, for any x and w, and regardless of ξ, there exist y and z such that $V_{\rm ho} < V_{\rm so} = 1$; if x is close enough to 0 then $V_{\rm ho} = 0$ (denial of service) is also observed.

Proof: First we will show that subversion is indeed possible when $z = 0$ and y is large enough. With $z = 0$, $\tilde{y} = \tilde{z} = 1$ and $V_{\rm s}(0) = V_{\rm h}(0) = 1$, (8) implies

$$\begin{aligned} M(t)V_{\rm s}(1) &= \xi\Psi(1) + (1-\xi)\Omega_{\rm sh}(y,w), \\ M(t)V_{\rm h}(1) &= (1-\xi)\Omega_{\rm hh}(w,w). \end{aligned} \tag{12}$$

Pick a y that fulfils

$$\Omega_{\rm sh}(y,w) > \frac{(1-\xi)\Omega_{\rm hh}(w,w) - \xi\Psi(1)}{1-\xi} = \Omega_{\rm hh}(w,w) - c\Psi(1), \tag{13}$$

where c is defined in (9) (we omit the argument that such a y exists). Then (12) implies $V_{\rm h}(1) < V_{\rm s}(1) = 1$. Suppose $V_{\rm h}(t) < V_{\rm s}(t) = 1$ for some $t > 0$, then using shorthand $L_{\rm h} = L(V_{\rm h}(t), x)$ and $L_{\rm s} = L(V_{\rm s}(t), x) = 1$ we have from (8):

$$\begin{aligned} M(t)V_{\rm s}(t+1) &= \xi\Psi(1) + (1-\xi)V_{\rm h}(t)\Omega_{\rm sh}(y,w), \\ M(t)V_{\rm h}(t+1) &= (1-\xi)V_{\rm h}(t)\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h}). \end{aligned} \tag{14}$$

The condition for $V_{\rm h}(t+1) < V_{\rm s}(t+1) = 1$ is $\Omega_{\rm sh}(y,w) > \Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h}) - c\Psi(1)/V_{\rm h}(1)$, which (13) ensures, hence the first part of the proposition follows by induction. To investigate x close to 0, notice from (8) that

$$V_{\rm h}(t+1) = \frac{(1-\xi)V_{\rm h}(t)\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})}{M(t)} = \frac{V_{\rm h}(t)\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})}{c\Psi(1) + V_{\rm h}(t)\Omega_{\rm sh}(y,w)}, \tag{15}$$

where c is defined in (9). The right-hand side of a generic equation $v = v\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})/(c\Psi(1) + v\Omega_{\rm sh}(y,w))$ at $v = 0$ equals 0, at $v = 1$ equals $\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})/(c\Psi(1) + \Omega_{\rm hh}(y,w)) < 1$ by (13), has a derivative 0 at $v = 0$ by the properties of $L(\cdot,\cdot\cdot)$ and $\Omega_{\rm qq'}(\cdot,\cdot\cdot)$, and by choosing x close enough to 0 can be made less than v for all $v < 1$. Therefore the fixed-point iteration process (15) starting from $V_{\rm h}(t) < 1$ produces $V_{\rm h}(t') < 1$ for all $t' > t$ and converges to $V_{\rm ho} = 0$. ■

One conclusion from Proposition 3 is that the acquired trust values may contrast with agents' intrinsic qualities, as tenet T2 states. Another is that the qualitative difference between a collusion and a non-collusion scenario regarding possible CTR scheme subversions occurs irrespective of ξ. This challenges the intuition tenets T2 and T5 convey that s&s agents must be numerous enough to subvert a CTR scheme. Furthermore, if Proposition 3, second part, holds for some x', it also does for all $x < x'$. Hence in a collusion scenario honest agents' best

response to malicious s&s agents is $x > x_{\rm h}$ (while small x punish selfish s&s agents, which illustrates tenet T3). The $x_{\rm h}$ that separates "too scant" and "enough" goodwill depends on whether $\sigma_{\rm s}(\cdot,\cdot\cdot)$ and $\sigma_{\rm s}(\cdot,\cdot\cdot)$ are discriminative or multiplicative, and decreases in ξ, since so does the right-hand side of (9). For a unit-step $L(\cdot,\cdot\cdot), x_{\rm h}$ can be found analytically (details are omitted): if $w \le c$ then $x_{\rm h} = 1$, otherwise

$$x_{\rm h} = \begin{cases} 1 - \frac{(w-c)\Psi(1)}{\Psi(w)}, & \text{discriminative } \ \sigma_{\rm h}(\cdot,\cdot\cdot), \\ c/w, & \text{multiplicative } \ \sigma_{\rm h}(\cdot,\cdot\cdot). \end{cases} \tag{16}$$

We now show that "enough" goodwill protects honest agents from a denial of service caused by malicious s&s agents that are in the minority.

Proposition 4. In a collusion scenario with $\xi < 1/2$, and x, w large enough, $V_{\rm ho} > 0$ regardless of y and z.

Proof: The second equation of (8) implies

$$V_{\rm h}(t+1) \ge \min\left\{ \frac{(1-\xi)V_{\rm h}(t)\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})}{\xi V_{\rm s}(t)\Psi(1) + (1-\xi)V_{\rm h}(t)\Omega_{\rm sh}(y, wL_{\rm s})}, 1 \right\} \ge \frac{\Omega_{\rm hh}(wL_{\rm h}, wL_{\rm h})}{\Psi(1)} \cdot \frac{V_{\rm h}(t)}{c + V_{\rm h}(t)}, \tag{17}$$

where c is defined in (9). If x and w are large enough then, by the continuity of $\Omega_{\rm hh}(\cdot,\cdot\cdot)$, the first fraction of the right-hand side can be made arbitrarily close to 1 for any $V_{\rm h}(t) > 0$. Hence

$$V_{\rm h}(t+1) \ge (1-\epsilon) \cdot \frac{V_{\rm h}(t)}{c + V_{\rm h}(t)}, \tag{18}$$

where $\epsilon > 0$ is arbitrarily small. Let x be such that $\epsilon < 1 - c$. The right-hand side of a generic equation $v = (1-\epsilon)v/(c+v)$ is an increasing concave function of v, less than 1 at $v = 1$, whose derivative at $v = 0$ is $(1-\epsilon)/c > 1$. Therefore if (8) converges to some $V_{\rm ho}$ then $V_{\rm ho}$ is greater than or equal to the unique root of the equation, *i.e.*, $V_{\rm ho} \ge 1 - c - \epsilon > 0$. (Note that failure to report by honest agents is tantamount to a larger ξ, hence results in a smaller $V_{\rm ho}$, cf. tenet T6.) ■

Figure 3 plots $V_{\rm so}$ and $V_{\rm ho}$ against y and z in two sample settings with uniformly distributed $A_{ij}(t)$ *i.e.*, $F(a) = a$ for $a \in [0,1]$, implying $\Psi(p) = p(1-p/2)$, to illustrate tenet T2–trust values may contrast with agents' behavior. One sees that larger y and z may, but need not mean larger trust values for s&s agents: the effect of getting more favorable reputation data from honest agents competes with that of producing more favorable reputation data about them. Note that $V_{\rm so} = 1$ for large enough y (and small enough z) signifies a CTR scheme subversion, and $V_{\rm so} = 0$ for a range of small y, cf. Proposition 1. Qualitatively similar results were obtained for other settings.

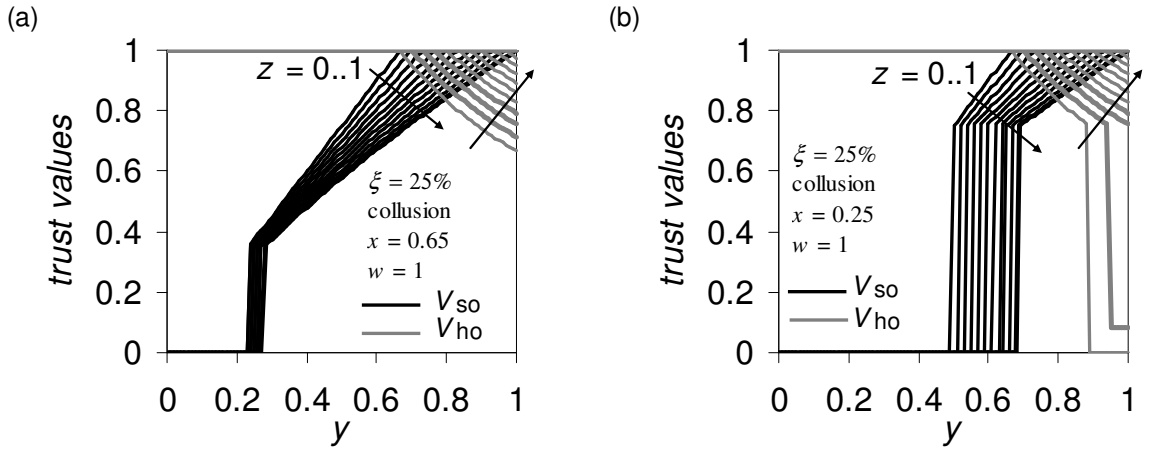

Figure 3: Trust values for various s&s service policies in a collusion scenario; $V_s(0) = V_h(0) = 1$, unit-step $L(\cdot,\cdot\cdot)$, multiplicative $\sigma_s(\cdot,\cdot\cdot), w = 1$; (a) $x = 0.65$, (b) $x = 0.25$

5 Bilateral Contribution

In the reputation aggregation algorithm governed by (1), agent j contributes to the new trust value of agent i through $R_{ij}(t)$, service reported as received from i. Such a *unilateral* type of contribution is fairly common in existing CTR schemes. Consider now a *bilateral* type of contribution, so that (1) becomes

$$R_{i,avg}(t) = \sum_{j \in N \setminus \{i\}} V_j(t) \delta^{\max\{\Delta_{ij}(t), \Delta_{ji}(t)} \Gamma\big(R_{ij}(t - \Delta_{ij}(t)), R_{ji}(t - \Delta_{ji}(t))\big) \ , \tag{19}$$

where $\Gamma : [0,1]^2 \to [0,1]$ is a bivariate function. The intuition behind bilateral contribution is that an s&s agent's trust value should be lowered not only by skimp attacks (providing too little service), but also by slander attacks (underreporting received service). A similar approach is taken in [Jia06], where $\Gamma(u,v) = u + v$ can be inferred; traces of it can also be found in some opinion portals where voting down someone else's opinion causes one to lose points, and in [Saa2010], where agents' scores are based both on provided and received service. Since it is reasonable to demand that $\Gamma(u,v) \leq \min\{u,v\}$, let $\Gamma(u,v) = \min\{u,v\}$. Recalling (3) we replace $\mathsf{E}R_{ij}(t)$ by

$$R_{i \leftrightarrow j}(t) = \mathsf{E}\min\{R_{ij}(t), R_{ji}(t)\} = \mathsf{E}\min\{\sigma_j(\sigma_i(A_{ij}(t), p_{ij}), r_{ij}), \sigma_i(\sigma_j(A_{ji}(t), p_{ji}), r_{ji}\} \ . \tag{20}$$

To adopt the mean-value approximation of Sec. 3.4 let us introduce generalized functions

$$\Psi_\beta(p) = \mathsf{E}\min\{A', \beta A'', p\} = \int_0^1 \left(1 - F(\frac{a}{\beta})\right)(1 - F(a))da \ , \tag{21}$$

$$\Omega_{\mathrm{q} \leftrightarrow \mathrm{q'}}(p, r, \pi, \rho) = \mathsf{E}R_{i \leftrightarrow j}(t)|_{p_{\mathrm{qq'}}=p, r_{\mathrm{qq'}}=r, p_{\mathrm{q'q}}=\pi, r_{\mathrm{q'q}}=\rho} \ . \tag{22}$$

where $\beta \in [0,1]$, $A', A'' \sim F(\cdot)$, and $\mathrm{q}, \mathrm{q'} \in \{\mathrm{s}, \mathrm{h}\}$. Analogously to (7) and with 0/0 taken to be 0 we have:

$$\Omega_{\mathrm{q} \leftrightarrow \mathrm{q'}}(p, r, \pi, \rho) = \begin{cases} \Psi_1(\min\{p, r, \pi, \rho\}), & \text{discriminative } \sigma q(\cdot,\cdot) \text{ and } \sigma q'(\cdot,\cdot) \\ \Psi_{\pi\rho}(\min\{p, r\}), & \text{discriminative } \sigma q(\cdot,\cdot) \text{ , multiplicative } \sigma q'(\cdot,\cdot) \\ \Psi_{pr}(\min\{\pi, \rho\}), & \text{multiplicative } \sigma q(\cdot,\cdot) \text{ , discriminative } \sigma q'(\cdot,\cdot) \\ pr\Psi_{\pi\rho/pr}(1), & \text{multiplicative } \sigma q(\cdot,\cdot) \text{ and } \sigma q'(\cdot,\cdot). \end{cases} \tag{23}$$

Using a succinct notation $L_\mathrm{s} = L(V_\mathrm{s}(t), x)$ and $L_\mathrm{h} = L(V_\mathrm{h}(t), x)$ we turn (8) into

$$\begin{aligned} V_\mathrm{s}(t+1) &= \frac{\xi V_\mathrm{s}(t) \Omega_{\mathrm{s} \leftrightarrow \mathrm{s}}(\tilde{y}, \tilde{z}, \tilde{y}, \tilde{z}) + (1-\xi) V_\mathrm{h}(t) \Omega_{\mathrm{s} \leftrightarrow \mathrm{h}}(y, wL_\mathrm{s}, wL_\mathrm{s}, z)}{M(t)} \ , \\ V_\mathrm{h}(t+1) &= \frac{\xi V_\mathrm{s}(t) \Omega_{\mathrm{h} \leftrightarrow \mathrm{s}}(wL_\mathrm{s}, z, y, wL_\mathrm{s}) + (1-\xi) V_\mathrm{h}(t) \Omega_{\mathrm{h} \leftrightarrow \mathrm{h}}(wL_\mathrm{h}, wL_\mathrm{h}, wL_\mathrm{h}, wL_\mathrm{h})}{M(t)} \ . \end{aligned} \tag{24}$$

Bilateral contribution proves highly beneficial, since it prevents CTR scheme subversion by a minority of s&s agents in both a non-collusion and a collusion scenario, thus in effect neutralizes s&s agents' collusion.

Proposition 5. With $\xi < 1/2$ and $w = 1$, CTR scheme subversion is not possible under bilateral contribution, *i.e.*, $V_\mathrm{so} \leq V_\mathrm{ho} = 1$ for any $x, y, z \in [0,1]$.

Proof: Define *diff* similarly as in the proof of Proposition 2 and assume $V_\mathrm{s}(t) \leq V_\mathrm{h}(t) = 1$ for some $t > 0$. Then $L\mathrm{h} = 1$ and, since $\xi < 1/2$, we have:

$$\begin{aligned} \mathit{diff} &= 2(1-\xi)(\Omega_{\mathrm{h} \leftrightarrow \mathrm{h}}(w,w,w,w) - \Omega_{\mathrm{s} \leftrightarrow \mathrm{h}}(y, wL_\mathrm{s}, wL_\mathrm{s}, z)) - 2\xi V_\mathrm{s}(t)(\psi - \Omega_{\mathrm{h} \leftrightarrow \mathrm{s}}(wL_\mathrm{s}, z, y, wL_\mathrm{s})) \\ &\geq \Omega_{\mathrm{h} \leftrightarrow \mathrm{h}}(w,w,w,w) - \Omega_{\mathrm{s} \leftrightarrow \mathrm{h}}(y, wL_\mathrm{s}, wL_\mathrm{s}, z)) - (\psi - \Omega_{\mathrm{h} \leftrightarrow \mathrm{s}}(wL_\mathrm{s}, z, y, wL_\mathrm{s})), \end{aligned} \tag{25}$$

where $\psi = \Psi_1(\min(\tilde{y}, \tilde{z})$ or $\tilde{y}\tilde{z}\Psi_1(1)$ if the s&s service police is discriminative or multiplicative, respectively; in both cases $\psi \leq \Psi_1(1)$. From (23), $\Omega_{\mathrm{h} \leftrightarrow \mathrm{s}}(wL_\mathrm{s}, z, y, wL_\mathrm{s}) = \Omega_{\mathrm{s} \leftrightarrow \mathrm{h}}(y, wL_\mathrm{s}, wL_\mathrm{s}, z)$ and $\Omega_{\mathrm{h} \leftrightarrow \mathrm{h}}(1,1,1,1) = \Psi_1(1)$. Thus if $w = 1$ then $\mathit{diff} \geq 0$, *i.e.*, $V_\mathrm{s}(t+1) \leq V_\mathrm{s}(t+1) = 1$ and the proof follows by induction. (Note that if $w < 1$, the proposition may not hold.) ■

Having two reputation aggregation algorithms to choose from (*i.e.*, unilateral and bilateral contribution), RAE can try to confuse s&s agents by concealing the algorithm in use. However, in agreement with [Ker09], such "security by obscurity" does not bring any improvement; a broader discussion is omitted for lack of space.

6 Related Work

Works that hint at trust value dynamics induced by one or both closed-loops in Figure 1b are many, but relatively few explicitly construct and analyze them. The model of [Mui02] recognizes a closed loop between and an agent's reputation, trust, and utility resulting from reciprocity (close in meaning to a trust-aware service policy). It is postulated that a more honest service policy foster higher reputation and more trust. We observe to the contrary that CTR scheme subversions are possible if s&s agents collude. In [Jia06], an agent's service policy is modeled as a probability of positive or negative rating of an interacting agent conditioned on both agents' intrinsic goodness or badness. It is shown that non-colluding attackers cannot subvert the scheme regardless of their number. Surprisingly, colluding attackers whose ratings favor fellow colluders cannot either. Our model links ratings to current trust values rather than agents' intrinsic qualities, which gives rise to a trust-aware service policy. We show that colluding attackers always can subvert the scheme unless bilateral contribution is in use. Our Markovian analysis takes advantage of a mean-value approximation; a mean field approximation similar in spirit to ours is applied in [Hub04], albeit in a very different, economic setting. Deterministic reputation dynamics are studied in [Cao06], but reported service, agents' misbehavior, or trust-weighted reputation aggregation are not explicitly accounted for. In [Chk10], each agent forms opinions cycle by cycle based on other agents' opinions weighted by their current reputations; the latter depend on previous-cycle reputations. In our context, an agent's trust value cannot influence other agents' other than through reported service–this precludes various manipulations which [Chk10] overcomes using control- and game-theoretic methods. Under trust-aware partner selection, reputation dynamics arise because more frequent partners become even more selectable. Existing analyses account for various ratings collection models [Huy06], [Has14]. With pooled service availability one expects oscillatory trajectories of reputation scores [Yu13]. Our model prevents reputation damage owing to trust-unaware partner selection and partitioned service availability. Game-theoretic analyses of reputation models are usually tied to a simple interaction game, e.g., social dilemma, donation, or product choice. Evolutionary models of random pairwise interactions "natively" give rise to reputation dynamics under trust-unaware partner selection and trust-aware service policy [Now98] [Oht04]. Evolutionary reputation dynamics under trust-aware partner selection are investigated, e.g., in [Fu08], [Wan12], [Pel14]. Non-evolutionary studies of reputation dynamics in repeated games often have one long-lived player interact sequentially with many short-lived ones [Ekm06], [Liu11]. Agents' intrinsic qualities, collusion, service policy, or service availability are not directly reflected.

7 Conclusion And Future Work

The analyzed trust value dynamics capture some intuitions and tenets of CTR design, and run counter some others. For example,

- a CTR scheme with trust-weighted aggregation of reputation data can be made invulnerable to subversion in a non-collusion scenario if s&s agents are in the minority; however, subversion is possible in a collusion scenario with an arbitrarily small proportion of s&s agents,
- defenses against selfish and malicious s&s agents can be combined by offering a moderate degree of goodwill (and indirect reciprocity) towards low-trust agents,
- bilateral-contribution reputation aggregation at RAE in effect neutralizes s&s agents' collusion, and finally,
- concealment of the reputation aggregation algorithm from agents does not bring honest agents better treatments "security by obscurity" does not work.

Further research will focus on other reputation aggregation algorithms, multiple- rather than two-level trust values, a multi-faceted notion of service, and extensions to quasi-static settings with adaptive service policies adjusted based on received service. Group rather than pairwise interactions are another natural generalization.

7.0.1 Acknowledgements

Preliminary ideas of the paper were developed during the author's participation in the Future Internet Engineering project supported by the European Regional Development Fund under Grant POIG.01.01.02-00-045/09-00.

References

[Cao06] J. Cao, W. Yu, and Y. Qu. A new complex network model and convergence dynamics for reputation computation in virtual organizations. *Physics Letters A*, 356(6): 414–425, Aug. 2006.

[Chk10] A. G. Chkhartishvili, D. A. Gubanov, N. A. Korgin, and D. A. Novikov. Models of reputation dynamics in expertise by social networks. Proc. UKACC Int. Conf. on Control, Coventry, UK, Sept. 2010.

[Del00] C. Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. Proc. 2nd ACM Conf. on Electronic Commerce, Minneapolis, MN, Oct. 2000.

[Ekm06] M. Ekmekci. Sustainable reputations with rating systems. Mimeo, Princeton University 2006.

[Fu08] F. Fu, C. Hauert, M. A. Nowak, and L. Wang. Reputation-based partner choice promotes cooperation in social networks. *Phys. Rev. E*, 78(026117), 2008.

[Has14] M. R. Hassan, G. Karmakar, and J. Kamruzzaman. Reputation and user requirement based price modeling for dynamic spectrum access. *IEEE Trans. Mobile Comput.*, 13(9): 2128–2140, Sept. 2014.

[Hen15] F. Hendrikx, K. Bubendorfer, and R. Chard. Reputation systems: A survey and taxonomy. *J. Parallel and Distrib. Comput.*, 75(1): 184–197, Jan. 2015.

[Hub04] B. A Huberman and Fang Wu. The dynamics of reputations. *J. Stat. Mechanics*, P04006, April 2004.

[Huy06] Trung Dong Huynh, N. R. Jennings, and N. R. Shadbolt. Certified reputation: how an agent can trust a stranger. Proc. AAMAS'06, Hakodate, Japan, May 2006.

[Jia06] T. Jiang and J. S. Baras. Trust evaluation in anarchy: A case study on autonomous networks. Proc. 25th IEEE INFOCOM, Barcelona, Spain, April 2006.

[Ker09] R. Kerr and R. Cohen. Smart cheaters do prosper: Defeating trust and reputation systems. Proc. AAMS'09, Budapest, Hungary, May 2009.

[Liu11] Q. Liu. Information acquisition and reputation dynamics. *Rev. Econ. Studies*, 78(4): 1400–1425, 2011.

[Mui02] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. Proc. 35th Annual Hawaii International Conf. on System Sciences, 2002.

[Now98] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393: 573–577, June 1998.

[Oht04] H. Ohtsuki and Y. Iwasa. How should we define goodness?–reputation dynamics in indirect reciprocity. *J. Theor. Biol.*, 231(1): 107–120, Nov. 2004.

[Pel14] A. Peleteiro, J. C. Burguillo, and Siang Yew Chong. Exploring indirect reciprocity in complex networks using coalitions and rewiring. Proc. AAMAS'14, Paris, France, May 2014.

[Ram04] S. D. Ramchurn, Dong Huynh, and N. R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1): 1–25, March 2004.

[Saa2010] S. Saavedra, D. Smith, and F. Reed-Tsochas. Cooperation under indirect reciprocity and imitative trust," PLoS ONE, vol. 5, 2010.

[Sei04] J.-M. Seigneur, S. Farrell, C. Damsgaard Jensen, E. Gray, and Y. Chen. End-to-end trust starts with recognition. LNCS vol. 2802, Berlin, Heidelberg: Springer, 2004.

[Wan12] Z. Wang, L.Wang, Z.-Y.Yin, and C.-Y. Xia. Inferring reputation promotes the evolution of cooperation in spatial social dilemma games. PLoS ONE, 7(7): e40218, 2012.

[Yu13] Han Yu, Zhiqi Shen, C. Leung, C. Miao, and V. R. Lesser. A survey of multi-agent trust management systems. *IEEE Access*, 1: 35–50, 2013.

[Zha12] L. Zhang, S. Jiang, Jie Zhang, and Wee Keong Ng. Robustness of trust models and combinations for handling unfair ratings. IFIP Advances in Information and Communication Technology, vol. 374, Berlin, Heidelberg: Springer, 2012.