Filtering Trust Opinions through Reinforcement Learning

Han Yu[a], Zhiqi Shen[a], Chunyan Miao[a], Bo An[a], Cyril Leung[b]

[a]*School of Computer Engineering, Nanyang Technological University, Singapore 637659*
[b]*Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, V6T1Z4, Canada*

Abstract

In open dynamic online communities such as e-commerce, participants need to rely on services provided by others in order to thrive. Accurately estimating the trustworthiness of a potential interaction partner is vital to a participant's well-being. It is generally recognized in the research community that third-party testimony sharing is an effective way for participants to gain knowledge about the trustworthiness of potential interaction partners without having to incur the risk of actually interacting with them. However, the presence of biased testimonies adversely affects a participant's long term well-being. Existing trust computational models often require complicated manual tuning of key parameters to combat biased testimonies. Such an approach heavily involves subjective judgements and adapts poorly to changes in environment. In this study, we propose the Actor-Critic Trust (ACT) model, which is an adaptive trust evidence aggregation model based on the principles of reinforcement learning. The proposed method dynamically adjusts the selection of credible witnesses as well as the key parameters associated with the direct and indirect trust evidence sources based on the observed benefits received by the trusting entity. Extensive simulations have shown that the ACT approach significantly outperforms existing approaches in terms of mitigating the adverse effect of biased testimonies. Such a performance is due to the proposed accountability mechanism that enables ACT to attribute the outcome of an interaction to individual witnesses and sources of trust evidence, and adjust future evidence aggregation decisions without the need for human intervention. The advantage of the proposed model is particularly significant when service providers and witnesses strategically collude to improve their chances of being selected for interaction by service consumers.

Keywords: Trust, reputation, credibility, collusion.

1. Introduction

In open and highly dynamic online communities where users are from diverse backgrounds and may have conflicting goals, distributed social control is needed to sustain long term interactions among them. Nowadays, such systems are quite common (e.g., service oriented computing systems [1], e-commerce systems [2], wireless communication networks [3], and multi-agent systems [4] etc.). In such environments in which services and devices usually have limited capabilities, users often have to interact with each other in order to achieve their goals. These interactions usually involve an exchange of services, information, or goods with value. Selfish users may renege on their commitments, thereby breaching the trust placed in them by others. Therefore, trust and reputation management mechanisms are often used to minimize the negative impact of selfish users.

Generally, users in an open online community that can be modeled as multi-agent systems (MASs) may play two types of roles [1]:

- *service providers* (SPs), who provide services, goods or information requested by others and do not need to rely on others to perform these services; and

- *service consumers* (SCs), who need to rely on service providers to accomplish certain tasks.

The main objective of evidence-based trust models is to estimate the trustworthiness of a potential interaction partner which represents its true behavior pattern. Evidences about a service provider from the perspective of a service consumer are usually from two sources:

- *direct trust evidence*: which consists of a service consumer's direct interaction experience with the service provider; and

- *indirect trust evidence*: which consists of third-party testimonies about the service provider from other service providers in the system.

In practical systems, it is not possible to definitively know the trustworthiness of a service provider. Therefore, it is often estimated using trust evidences. The estimation of a service provider's trustworthiness derived from the direct trust evidence of a service consumer alone is called *direct trust*, while that derived from the indirect trust evidence is called *indirect trust*. An estimation derived from both sources of trust evidence is commonly known as the *reputation* of a service provider. In the eyes of a service consumer, other service consumers who provide it with indirect trust evidence (i.e. testimonies) about a service provider are regarded as *witnesses*. A witness's reliability in terms of providing useful testimonies is referred to as its *credibility*.

Since such systems tend to be very large in practice, service consumers often have to interact with service providers with whom they may not be very familiar (i.e. have little or no prior interaction experience with) [5]. Thus, it is both necessary and advantageous to allow service consumers to act as witnesses to provide their own first-hand interaction experience as testimonies to other service consumers who lack such information.

However, such an approach is not without its perils. A witness might fabricate or hide information to promote the reputations of service providers who are related to it in some way, due to self-interest; the behavior patterns of service providers might change over time, thus rendering some service consumers' existing record of past interaction experience obsolete; service consumers may use different criteria to define the success and failure of an interaction, thereby making it difficult to use testimonies in a uniform way. Biased testimonies resulting from one or more of these factors can degrade the accuracy of trust decisions [1]. Therefore, testimonies from witnesses need to be filtered before being used to evaluate a service consumer's reputation.

To this end, a number of evidence-based trust and reputation management (TRM) models have been proposed over the years. The general flow for a service consumer to decide which service provider to select for interaction is illustrated in Fig. 1. Each service consumer continuously records its direct interaction experience with service providers over time. When a service provider's trustworthiness needs to be evaluated, the service consumer may request third-party testimonies from witnesses, depending on the service consumer's confidence on its own direct trust evidence. These testimonies are preprocessed in an attempt to filter out unfair ratings. The resulting direct and indirect trust evidences are then aggregated to form a trustworthiness evaluation for that particular service provider. At the end of this process, the service consumer decides which service provider to interact with based on their trustworthiness evaluations.

Existing approaches for third-party testimony filtering and aggregation fall into three main categories:

- filtering witness testimonies before aggregating them without recording the

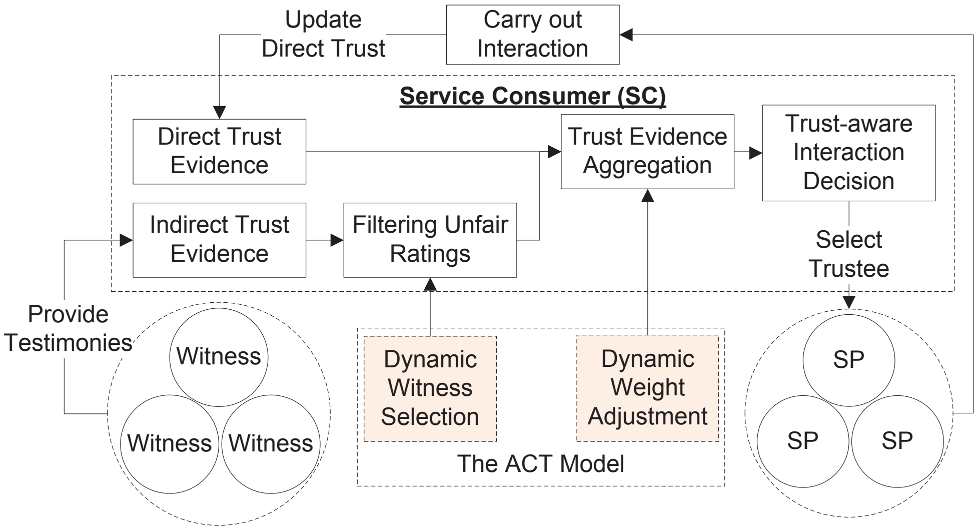


Figure 1: The general flow of trust-aware interaction decision making for evidence-based trust and reputation management models, and the contributions by the proposed ACT approach.

witnesses' credibility in terms of providing useful testimonies [6, 7, 8, 9];

- aggregating testimonies according to witnesses' credibility evaluations [10, 11, 12]; and
- incorporating incentive mechanisms into existing trust models to induce witnesses to provide fair testimonies [13, 14, 15].

However, these approaches typically require manual tuning of key parameters in their models which heavily involves subjective judgements and adapts poorly to changes in environment.

In this paper, we address these limitations by proposing the Actor-Critic Trust (ACT) model based on the principles of the Actor-Critic Learning Method [16]. The ACT approach makes the following two key technical contributions. It enables existing evidence-based trust models to dynamically make two important decisions when presented with third-party testimonies for a service provider: 1) how much weight to give to its own personal direct trust evidence and the collective opinions from witnesses, and 2) how much weight to assign to the testimonies of each witness. The reward and penalty strategy design in the ACT approach enables the trust models to adapt based on the actual outcomes of past interactions between a service consumer and other service providers. As a result, its performance is less affected by the changes in the composition of the witness population.

It also does not require additional infrastructure support (e.g., social relationship information) for its operation. Experimental results, presented in Section V, show that the ACT approach is more effective in mitigating the adverse effects of unfair testimonies than existing approaches, especially when witnesses strategically collude with malicious service providers.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents the basic notations used in this paper. Details of the proposed ACT approach are presented in Section 4. Section 5 analyzes the impact of the value chosen for a key parameter affects the expected performance. Section 6 describes the simulation test-bed and analyzes the results. Section 7 presents a summary of our contributions and possible future work.

2. Related Work

It is widely recognized within the research community that the importance of incorporating mechanisms to mitigate the adverse effects of biased testimonies. In this section, we discuss some recent research work on aggregating trust evidence from different sources and filtering out biased testimonies. For a more comprehensive review of this field, readers may refer to [1, 2, 3].

2.1. Trust Evidence Aggregation Approaches

Evidence-based trust models often make use of two distinct sources of information to evaluate the trustworthiness of a service provider: 1) *direct trust evidence*: a service consumer's personal interaction experience with a service provider, and 2) *indirect trust evidence*: third-party testimonies about the service provider. The majority of existing trust models adopt a weighted average approach when aggregating these two sources of trust evidence [3]. Direct trust evidence is often assigned a weight of $0 \leq \gamma \leq 1$, and indirect evidence is assigned a corresponding weight of $(1 - \gamma)$. Existing approaches for aggregating direct and indirect trust evidence can be divided into two broad categories: 1) *static approaches*, where the value of γ is pre-defined; and 2) *dynamic approaches*, in which the value of γ is continually adjusted by the service consumer.

In many papers, static γ values for trust evidence aggregation. The majority of them tend to take a balanced approach by assigning a value of 0.5 to γ [8, 12, 9, 17, 18]. In some studies, the authors assign the value 0 [19, 20] or 1 [21] to γ to exclusively use only one source of trust information. Barber and Kim [22] have empirically shown, without considering the presence of biased testimonies, that direct trust evidence is the most useful to a service consumer over the long term while indirect trust evidence gives an accurate picture more quickly. Thus, approaches that discard one source or the other, forfeit some of the advantages provided by evidence based trust models. However, using a static value for γ is also not always a good strategy.

Some researchers have explored adjusting the value of γ dynamically based on different rationales. In [23], the value of γ is varied according to the number of direct observations on the behavior of a service provider s_j available to a service consumer c_i. It is assumed that every service consumer starts with no prior interaction experience with a service provider and gradually accumulates direct trust evidence over time. Initially, the service consumer relies completely on indirect trust evidence (i.e. $\gamma = 0$) to select service providers for interaction. As the number of its interactions with a service provider s_j increases, the value of γ also increases according to the formula

$$\gamma = \begin{cases} \frac{N_j^i}{N_{min}}, & \text{if } N_j^i < N_{min} \\ 1 & \text{, Otherwise} \end{cases} \tag{1}$$

where N_j^i is the total number of direct observations of s_j's behavior by c_i, and N_{min} is the minimum number of direct observations required in order to achieve a pre-determined acceptable level of error rate ε and confidence level ϑ. N_{min} is calculated following the *Chernoff Bound Theorem*:

$$N_{min} = -\frac{1}{2\varepsilon^2} \ln(\frac{1-\vartheta}{2}). \tag{2}$$

This approach is not concerned with filtering potentially biased third-party testimonies. Rather, its aim is to accumulate enough direct trust evidence so that a service consumer can make a statistically accurate estimation on the trustworthiness of a service provider without relying on indirect trust evidence. Since the value of γ increases to 1, this approach implicitly assumes that agent behaviors do not change with time. This may not always be true and limits the applicability of the approach under more dynamic scenarios. On the other hand, the ACT approach does not make this assumption and continuously make adjustments as the situation changes.

In [5], an approach based the Q-learning technique [24] to select an appropriate γ value from a predetermined static set of values Γ has been proposed. In order to select appropriate values for the set Γ, expert opinions about the underlying system characteristics are assumed to be available. Based on the reward accumulated by a service consumer under different γ values, Q-learning selects the γ value associated with the highest accumulated reward at each time step. This work provided the first step towards using interaction outcomes to enable the service consumer to weight the two sources of trust evidence. However, as this method uses a predetermined set of γ values, its performance is affected by the quality of the expert opinions used to form the set of permissible γ values. In contrast, the ACT model adjust both the γ value as well as the weight values for individual witnesses in finely grained steps so that it does not have to rely on the subjective opinions of the designer.

2.2. Testimony Filtering Approaches

Over the years, many models for filtering potentially biased third-party testimonies have been proposed. However, these models are usually based on assumptions of the presence of some infrastructure support or special characteristics in the environment. In this section, some representative models in this sub-field are discussed.

The ReGreT model [25] makes use of the social relationships among the members of a community to determine the credibility of witnesses. Pre-determined fuzzy rules are used to estimate the credibility of each witness which, in turn, is used as the weight of its testimony for a service provider when aggregating all the testimonies. This model relies on the availability of social network information among the agents which may not be present in many systems.

In [7], unfair testimonies are assumed to exhibit certain characteristics. The proposed approach is closely coupled with the Beta Reputation System (BRS) [26] which records testimonies in the form of counts of successful and unsuccessful interactions with a service provider. The received testimonies are aggregated with equal weights to form a majority opinion and then, each testimony is tested to see if it is outside the q quartile and $(1 - q)$ quartile of the majority opinion. If so, the testimony is discarded and the majority opinion updated. This model assumes that the majority opinion is always correct. Thus, it is not effective in highly hostile environments where the majority of witnesses are malicious.

In [8], it is assumed that the direct experience of the service consumer is the most reliable source of belief about the trustworthiness of a particular service provider, and it is used as the basis for filtering testimonies before aggregating them to form a reputation evaluation. An entropy-based approach is proposed to measure how much a testimony deviates from the current belief of the service consumer before deciding whether to incorporate it into the current belief. However, by depending on having sufficient direct interaction experience with a service provider, this assumption conflicts with the purpose for relying on third-party testimonies, which is to help service consumers make better interaction decisions when they lack direct trust evidence.

The model in [9] supports interaction outcomes recorded in multi-dimensional forms. It applies two rounds of clustering of the received testimonies to identify testimonies which are extremely positive or extremely negative about a trustee. If neither the extremely positive opinion cluster nor the extremely negative opinion cluster forms a clear majority, they are both discarded as unfair testimonies and the remaining testimonies are used to estimate the reputation of a service provider. Otherwise, the majority cluster is considered as the reliable testimonies. Due to its iterative nature, the computational complexity of this method is high, with a time complexity of $O(mn^2)$ where m is the number of candidate service providers whose reputations need to be evaluated and n is the number of testimonies received for each candidate service provider. The method is also not robust in hostile environments where the majority of the witnesses are malicious.

3. System Model

Before discussing details of the proposed model, we introduce the the system model under which the ACT approach is designed to operate. At each time step t, a service consumer c_i will interact with at most one service provider s_j in our target system. For each interaction, c_i chooses a service provider from among several candidates based on their estimated trustworthiness values. Whenever c_i needs to assess the trustworthiness of s_j, it draws upon both its own direct trust evidence about s_j (if there is any) as well as testimonies from a list of witnesses $W_{i,j}(t)$ which are known by c_i at time t to have previously interacted with s_j. A witness w_k may reply to c_i's request at time step t with a testimony $d_j^k(t)$. In this study, a malicious w_k may distort its testimonies before sharing them with others. The service provider chosen for interaction by c_i at time step t is affected by the selection of witnesses as well as the weights given to direct and indirect trust evidence.

each interaction with s_j, c_i incurs a utility cost of C. If s_j successfully completes the task assigned to it by c_i, c_i receives a utility gain of G. We assume that the outcome of the interaction $O_{i \to j}(t)$ can be observed by c_i within the same time step in which the interaction occurs. We further assume that the interaction outcome is either successful ($O_{i \to j}(t) = 1$) or unsuccessful ($O_{i \to j}(t) = 0$). By comparing the recommendation $d_j^k(t)$ by each $w_k \in W_{i,j}(t)$ about s_j at time t with $O_{i \to j}(t)$, c_i can learn the ranking of each w_k in $W_{i,j}(t)$. New witnesses for s_j discovered by

Table 1: Symbols used in this Paper

Symbol	Meaning
c_i	A service consumer.
s_j	A service provider.
w_k	A witness.
$W_{i,j}(t)$	A list of witnesses for s_j known to c_i at time t.
$O_{i\to j}(t)$	The outcome of an interaction between c_i and s_j at time t.
$test_j^k(t)$	A testimony from w_k with regard to s_j at time t.
$d_j^k(t)$	The interaction decision as suggested by $test_j^k(t)$.
$D_{i\to j}^d(t)$	The decision by c_i on whether to interact s_j with at time t based on direct trust evidence only.
$D_{i\to j}^{ind}(t)$	The decision by c_i on whether to interact s_j with at time t based on indirect trust evidence only.
$D_{i\to j}(t)$	The overall decision by c_i on whether to interact s_j with at time t based on both direct and indirect trust evidence.
C	The cost incurred by c_i when engaging the service of s_j.
G	The utility derived from a successful interaction.
R	The reward assigned to a source of trust evidence.
P	The penalty assigned to a source of trust evidence.
$\tau_{i\to j}^d(t)$	The direct trust placed on s_j by c_i at time t.
$\tau_{i\to j}^{ind}(t)$	The indirect trust placed on s_j by c_i at time t.
$r_j(t)$	The reputation of s_j at time t.
$\pi_{k,j}$	The credibility of w_k for s_j in c_i's local record.
π_d	The weight assigned to the direct source of trust evidence by a consumer.

c_i over time are added into $W_{i,j}$. The interaction outcome value, $O_{i\to j}(t)$, is further compared with the recommended interaction decision value, $D_{i\to j}^d(t)$, based on direct trust evidence and the value, $D_{i\to j}^{ind}(t)$, based on indirect trust evidence from the testimonies of selected witnesses. Reward and penalty values are assigned to these two sources of trust evidence by c_i in its local record to determine how much to rely on either source in the future.

The objective of an individual service consumer is to maximize its utility over its lifetime in the presence of malicious service providers and malicious witnesses. For convenience, the main symbols used in this paper are listed in Table 1.

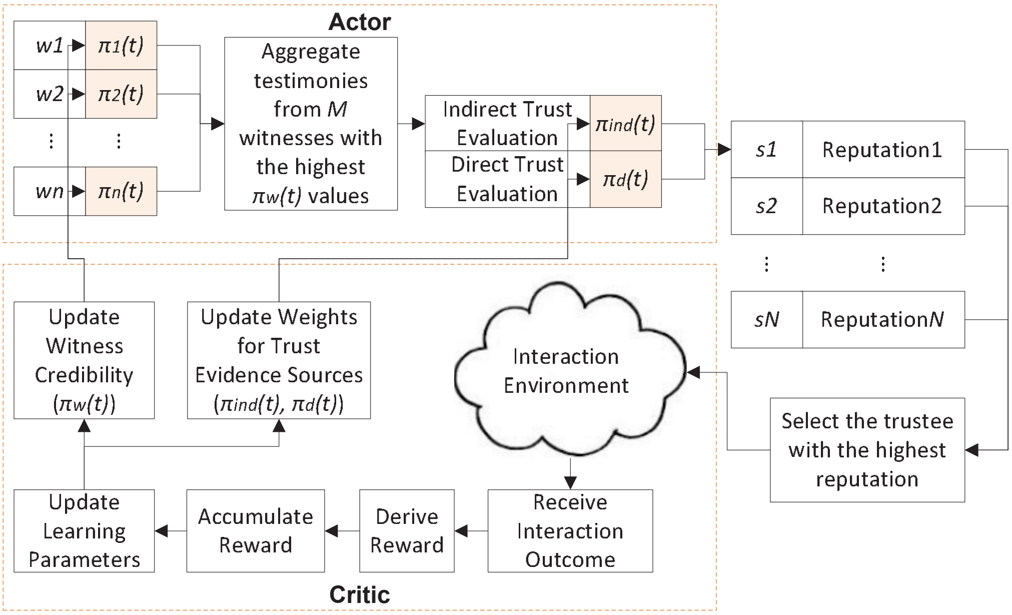

Actor
w1
π1(t)
w2
π2(t)
wn
πn(t)
Aggregate testimonies from M witnesses with the highest πw(t) values
Indirect Trust Evaluation
πind(t)
Direct Trust Evaluation
πd(t)
s1
Reputation1
s2
Reputation2
sN
ReputationN
Update Witness Credibility (πw(t))
Update Weights for Trust Evidence Sources (πind(t), πd(t))
Interaction Environment
Select the trustee with the highest reputation
Update Learning Parameters
Accumulate Reward
Derive Reward
Receive Interaction Outcome
Critic

Figure 2: The general framework of the ACT approach based on reinforcement learning.

4. The ACT Approach

We now propose the ACT approach to assist service consumers make trust-aware interaction decisions in the presence of potentially unreliable third-party testimonies. The general framework of the proposed ACT approach is presented in Fig. 2. Each service consumer c_i keeps two local lists: 1) a *list of known witnesses*, and 2) a *list of known service providers*. Since a witness may only have interacted with a few service providers, the list of known witnesses organizes the witnesses into sub-lists indexed according to known service providers. The list of known service providers stores the direct trust evidence c_i has for each known service provider and the weight assigned to the direct trust evidence $\gamma_{i,j}$ in the case of that service provider. These two lists grow as c_i acquires more interaction experience with these two types of system participants.

The ACT approach is designed based on a variant of the reinforcement learning (RL) approach - the actor-critic method [27]. The actor-critic method requires minimal computation when selecting an action. The actor module represents the policy used to choose which witnesses' testimonies should be selected and how much weight each of them should have when aggregating them together to form the indirect trust evidence. The policy also determines how much weight should be given to the direct and indirect trust evidence in order to evaluate the service provider's trustworthiness. The critic module represents the value function that determines whether the service provider c_i is better off or worse off after each interaction with a selected service provider s_j. Overtime, the learning parameters of the ACT approach are updated in such a way that more preference is given to witnesses and the source of trust evidence that enhance c_i's well-being.

Although the ACT approach can be used together with many possible trust evaluation models, to be specific, we assume that the popular Beta Reputation System (BRS) [26] is used as the underlying trust evaluation method. The direct trust for s_j by c_i at t can be calculated using the BRS as:

$$\tau^d_{i\to j}(t) \triangleq \frac{\alpha + 1}{\alpha + \beta + 2} \tag{3}$$

where α is the total number of successful interactions between s_j and c_i, while β is the total number of unsuccessful interactions between s_j and c_i up to t.

4.1. Learning Witness Credibility Ranking

In the critic module, the reward function for c_i is defined as:

$$r_{i,j} \triangleq \frac{1}{T_{i,j}} \sum_{t=1}^{T_{i,j}} [\mu^i_j(t) \cdot (G - C) - (1 - \mu^i_j(t)) \cdot C]. \tag{4}$$

$r_{i,j}$ is computed at the end of each interaction between c_i and s_j. It reflects the level of success achieved by c_i following the current interaction decision policy. $T_{i,j}$ is the number of interactions between c_i and s_j, and the parameter $\mu^i_j(t)$ is defined as:

$$\mu^i_j(t) = \begin{cases} 0, \text{ if } O_{i\to j}(t) = 0 | D_{i\to j}(t) = 1 \\ 1, \text{ if } O_{i\to j}(t) = 1 | D_{i\to j}(t) = 1 \end{cases} \tag{5}$$

$D_{i\to j}(t)$ denotes the overall decision by c_i to interact with s_j at time t based on both the direct and indirect trust evidence currently available. Here, we only consider the case when the decision is to interact with a service provider (i.e. $D_{i\to j}(t) = 1$), because in order for a service consumer c_i to be able to observe the actual interaction outcome with a service provider s_j at the end of time t, s_j must be selected by c_i for interaction in that time step. When $D_{i\to j}(t) = 0$, it implies that c_i deems s_j untrustworthy based on its reputation value. Thus, in these cases, no interaction between them will take place at that time and no $O_{i\to j}(t)$ value can be observed. In this paper, we assume that the agents' direct trust values and indirect trust values are normalized to a range of [0, 1]. A testimony $test_j^k(t)$ is simply w_k's direct trust value for s_j based on its own direct trust evidence up to time t. Thus, its value is also within the range [0, 1].

Once the latest interaction outcome is known, a reward correction value $\theta_{k,j}$ is computed for each of the M selected witnesses whose testimonies have been used to calculate the reputation of s_j namely:

$$\theta_{k,j} = \frac{1}{T_{k,j}} \sum_{t=1}^{T_{k,j}} [d_j^k(t) \cdot (1 - O_{i \to j}(t))]. \tag{6}$$

$T_{k,j}$ denotes the total number of times that w_k's testimonies about s_j has been used by c_i, and $d_j^k(t)$ represents the interaction recommendation implied by w_k's testimony, $test_j^k(t)$, on s_j at time step t and is given by:

$$d_j^k(t) = \begin{cases} 0, \text{ if } test_j^k(t) < Th \\ 1, \text{ otherwise} \end{cases} \tag{7}$$

where $Th \in [0, 1]$ is a predefined threshold value. $\theta_{k,j}$ increases with the number of times that w_k has given testimonies suggesting a service provider is trustworthy but the actual interaction outcome is unsuccessful. It is used to penalize the act of unfairly praising a service provider, which is the most common form of collusion between service providers and witnesses.

After the interaction outcome with a service provider is known, the critic process is carried out by updating the learning parameter $p_{k,j}$ for each of the M witnesses whose testimonies resulted in the selection of s_j by c_i at t as:

$$p_{k,j} \leftarrow p_{k,j} + \rho \cdot (r_{i,j} - \widetilde{r}_{i,j} - \delta \cdot \theta_{k,j})(1 - \pi_{k,j}). \tag{8}$$

The constant $(0 < \rho \leq 1)$ denotes the learning rate. As ρ increases, the learning parameter $p_{k,j}$ changes more rapidly as new interaction outcomes become available. In this paper, we choose a ρ value close to 0 to make $p_{k,j}$ vary more smoothly. The constant $(0 < \delta \leq 1)$ represents the bias towards penalizing collusion when updating the learning parameter; we select its value to be significantly smaller than 1 to avoid drastic changes in the value of $p_{k,j}$.

The credibility ranking value $\pi_{k,j}$ of each known w_k with regard to a service provider s_j is calculated using the *Gibbs softmax method* [24] as:

$$\pi_{k,j} = \frac{e^{p_{k,j}}}{\sum_{l=1}^{M} e^{p_{l,j}}}. \tag{9}$$

The resulting values of $\pi_{k,j}$ is used to rank the witnesses known to c_i to facilitate subsequent witness selections. The sum of all $\pi_{k,j}$ values always equals to 1. Thus, it can be regarded as the probability of soliciting testimonies from each of the witnesses known to c_i at time t.

After the credibility ranking values are calculated, the total accumulated reward $\widetilde{r}_{i,j}$ is updated. It is used as a reference in the process of evaluating the well-being of c_i resulted from interactions with s_j. It is updated as:

$$\widetilde{r}_{i,j} \leftarrow \varphi \cdot \widetilde{r}_{i,j} + (1 - \varphi) \cdot r_{i,j} \tag{10}$$

where the constant $0 < \varphi \leq 1)$ determines the influence of the latest rewards in the smoothed baseline reward $\widetilde{r}_{i,j}$. When $\varphi = 1$, only the current reward is used to evaluate the credibility of each witness.

The indirect trust for s_j by c_i can be computed as:

$$\tau^{ind}_{i\to j}(t) \triangleq \frac{\sum_{k=1}^{M}(\pi_{k,j} \cdot test^k_j(t))}{\sum_{k=1}^{M}\pi_{k,j}}. \tag{11}$$

4.2. Learning the Weights for Sources of Trust Evidence

With the values of $\tau^{d}_{i\to j}(t)$ and $\tau^{ind}_{i\to j}(t)$ calculated using Eq. (3) and Eq. (11) respectively, the next step is to aggregate them to compute the reputation of s_j. In the ACT approach, for each s_j known to c_i, two critic modules are are used to learn the weights for the two sources of trust evidence and one actor module is used for estimating the trustworthiness of s_j. The critic module in the proposed method determines the relative merit of each source of trust evidence through reward accumulation. The learning process is similar to that presented in Section IV.A. Since the two critic modules are essentially the same but only use different sources of trust evidence as input data, in the following, we only discuss the critic module for the direct trust evidence source.

The value function of the critic module is designed as:

$$r_d = \frac{1}{T_{i,j}}\sum_{t=1}^{T_{i,j}}[\widetilde{\mu}(t) \cdot R + (1 - \widetilde{\mu}(t)) \cdot P] \tag{12}$$

$$\widetilde{\mu}(t) = \begin{cases} 1, \text{ if } O_{i\to j}(t) = D^d_{i\to j}(t) \\ 0, \text{ otherwise} \end{cases} \tag{13}$$

$$D^d_{i\to j}(t) = \begin{cases} 1, \text{ if } \tau^d_{i\to j}(t) \geq Th \\ 0, \text{ otherwise} \end{cases} \tag{14}$$

r_d can be considered as the time averaged per interaction reward achieved by c_i through relying on its direct trust evidence source about s_j with the current weight value $\gamma_{i,j}$. R and P are predetermined constant values for reward and penalty, based on the consequences of the interaction decision. The ratio of R to P, rather than their absolute values, is important to the learning process. A small $R : P$ ratio means that trust is hard for a service provider to gain, but easy to lose. The variable $\widetilde{\mu}(t)$) determines whether this trust evidence source should be rewarded or penalized at time t. Its value toggles between 0 and 1 according to the relationship between the interaction decision $D^d_{i\to j}(t)$, which is related the direct trustworthiness evaluation ($0 \leq \tau^d_{i\to j}(t) \leq 1$), and the actual interaction outcome $O_{i\to j}(t)$. As $D^d_{i\to j}(t)$ is only one component of the overall interaction, it is possible that even as $D^d_{i\to j}(t)$ suggests not to interact with s_j, the overall decision is otherwise.

Once the latest r_d is calculated, it is compared with the baseline reward $\widetilde{r_d}$ accumulated by this trust evidence source to update the learning parameter p_d according to:

$$p_d \leftarrow p_d + \rho \cdot (r_d - \widetilde{r_d}) \cdot (1 - \pi_d). \tag{15}$$

After p_d is updated, $\widetilde{r_d}$ is updated to incorporate the latest reward r_d:

$$\widetilde{r_d} \leftarrow \varphi \cdot \widetilde{r_d} + (1 - \varphi) \cdot r_d. \tag{16}$$

$\widetilde{r_d}$ can be treated as a basis for comparing whether c_i is better off or worse off by aggregating the direct trust evidence into the estimation for the trustworthiness of s_j using the latest $\gamma_{i,j}$ value.

Similarly, the learning parameter p_{ind} for the indirect source of trust evidence can be obtained. When both p_d and p_{ind} are obtained, the learning parameters π_d and π_{ind} are updated as:

$$\pi_d \triangleq \frac{e^{p_d}}{e^{p_d} + e^{p_{ind}}} \tag{17}$$

$$\pi_{ind} \triangleq \frac{e^{p_{ind}}}{e^{p_d} + e^{p_{ind}}}. \tag{18}$$

π_d and π_{ind} can be treated as the probability of selecting each source of trust evidence $\pi_d + \pi_{ind} = 1$. In the ACT approach, $\gamma_{i,j} = \pi_d$.

4.3. Exploration v.s. Exploitation

While the strategy for exploiting known witnesses with high credibility is relatively straightforward (i.e. selecting the top M most credible witnesses to request testimonies from), balancing it with exploration for addition witnesses requires careful design. In the ACT approach, the exploration process is controlled by two parameters: 1) an exploration probability $Pr(Exp)$, and 2) the magnitude of M. The value of $Pr(Exp)$ is initialized to 1 at the start of a service consumer c_i's life time to enable c_i to explore when the list of known witnesses is empty. The value of $Pr(Exp)$ is gradually decreased over time until it reaches a pre-defined minimum value, Pr_{min}. Testimonies returned by previously unknown witnesses are given the benefit of the doubt and included in the calculation of the service provider's reputation with weight values equal to the lowest $\pi_{k,j}$ among that of the selected known witnesses. This is to ensure that c_i will always have some opportunity to discover new witnesses.

Algorithm 1 The ACT Testimony Aggregation Algorithm

```
Require: $\tau^{d}_{i \to j}(t)$ for all $s_j$ with who $c_i$ had prior interactions.
if $c_i$ needs to select an SP for interaction then
    $explorationProbability = random(0, 1)$
    if $explorationProbability \leq Pr(Exp)$ then
        Randomly select an unknown SP with for interaction
    else
        Rank known SPs in descending order of their $\tau^{d}_{i \to j}(t)$ values
        for each candidate known SP $s_j$ do
            $c_i$ asks the top $M$ ranked witnesses known to $c_i$ for testimonies on $s_j$
        end for
        Evaluate $\tau^{ind}_{i \to j}$ for each known candidate SP following Eq.(11)
        Evaluate $r_j(t)$ for each known SP following Eq.(19)
        Delegate the task to the SP with the highest $r_j(t)$ value
        Observe the interaction outcome $O_{i \to j}(t)$ with the selected SP
        Update $\tau^{d}_{i \to j}(t)$ following Eq.(3)
        Update the $p_{k,j}$ values according to Eq.(8) for all witnesses $w_k$ who provided testimonies for the selected SP in the last time step
        Update their $\pi_{k,j}$ values following Eq.(9)
        Re-rank known witnesses according to their new $\pi_{k,j}$ values
        Update the $\gamma_{i,j}$ and $(1 - \gamma_{i,j})$ values according to Eq.(17) and Eq.(18) respectively
    end if
end if
```

A service provider s_j's reputation is calculated as:

$$r_j(t) \triangleq \gamma_{i,j} \cdot \tau^{d}_{i\to j}(t) + (1 - \gamma_{i,j}) \cdot \tau^{ind}_{i\to j}(t). \tag{19}$$

$r_j(t)$ represents the overall reputation of s_j and is used by c_i to estimate s_j's trustworthiness. At each time step, c_i might have more than one candidate service providers to choose from. In this study, we assume that c_i always selects the service provider with the highest overall reputation for interaction. The working process of the ACT approach is shown in Algorithm 1.

5. Analysis

Biased testimonies from witnesses can influence the probability of occurrence of two types of errors: 1) unfairly negative reputation assessment for a trustworthy service provider (ε_1); and 2) unfairly positive reputation assessment for an untrustworthy service provider (ε_2). Let $H_0(j)$ denotes the unknown hypothesis that the service providers s_j is untrustworthy and $H_1(j)$ denotes the unknown hypothesis that s_j is trustworthy. The probabilities that the biased testimonies will influence the aggregate *indirect* trust decision by c_i on whether to interact with s_j at time t, i.e., erroneous non-interaction ($P_{\varepsilon_1}(Th)$) and erroneous interaction ($P_{\varepsilon_2}(Th)$) are:

$$\begin{aligned} P_{\varepsilon_1}(Th) &\triangleq Pr\{D^{ind}_{i\to j}(t) = 0 | H_1(j)\} \\ &= 1 - \int_{Th}^{1} Pr\{\tau^{ind}_{i\to j}(t) = l | H_1(j)\} dl, \end{aligned}$$

$$\begin{aligned} P_{\varepsilon_2}(Th) &\triangleq Pr\{D^{ind}_{i\to j}(t) = 1 | H_0(j)\} \\ &= \int_{Th}^{1} Pr\{\tau^{ind}_{i\to j}(t) = l | H_0(j)\} dl. \end{aligned}$$

As an erroneous interaction may result in the service consumer losing utility, preventing ε_2 is more important than preventing ε_1.

To simplify the notations used in the analysis, we define functions $X(Th, \Delta Th)$ and $Y(Th, \Delta Th)$ as:

$$X(Th, \Delta Th) \triangleq \int_{Th}^{Th+\Delta Th} Pr\{\tau^{ind}_{i\to j}(t) = l | H_1(j)\} dl,$$

$$Y(Th, \Delta Th) \triangleq \int_{Th}^{Th+\Delta Th} Pr\{\tau^{ind}_{i\to j}(t) = l | H_0(j)\} dl.$$

If the decision threshold Th is increased to $(Th + \Delta Th)$ where ΔTh is a positive value that ensures $0 \leq (Th + \Delta Th) \leq 1$:

$$\begin{aligned}
P_{\varepsilon_1}(Th + \Delta Th) &= 1 - \int_{Th+\Delta Th}^{1} Pr\{\tau_{i \to j}^{ind}(t) = l | H_1(j)\} dl \\
&= 1 - (\int_{Th}^{1} Pr\{\tau_{i \to j}^{ind}(t) = l | H_1(j)\} dl \\
&\quad - \int_{Th}^{Th+\Delta Th} Pr\{\tau_{i \to j}^{ind}(t) = l | H_1(j)\} dl) \\
&= P_{\varepsilon_1}(Th) + X(Th, \Delta Th),
\end{aligned}$$

$$\begin{aligned}
P_{\varepsilon_2}(Th + \Delta Th) &= \int_{Th+\Delta Th}^{1} Pr\{\tau_{i \to j}^{ind}(t) = l | H_0(j)\} dl \\
&= \int_{Th}^{1} Pr\{\tau_{i \to j}^{ind}(t) = l | H_0(j)\} dl \\
&\quad - \int_{Th}^{Th+\Delta Th} Pr\{\tau_{i \to j}^{ind}(t) = l | H_0(j)\} dl \\
&= P_{\varepsilon_2}(Th) - Y(Th, \Delta Th).
\end{aligned}$$

Since $X(Th, \Delta Th) \geq 0$ and $Y(Th, \Delta Th) \geq 0$, it can be deduced that in general:

$$P_{\varepsilon_1}(Th^{'}) \geq P_{\varepsilon_1}(Th), \tag{20}$$

$$P_{\varepsilon_2}(Th^{'}) \leq P_{\varepsilon_2}(Th), \tag{21}$$

whenever $Th^{'} > Th$. By increasing the threshold Th, a service consumer c_i takes a risk-averse approach that reduces the probability of ε_2 at the expense of increasing the probability of ε_1. However, by doing so, c_i's long term utility gain may be affected by reduced number of interactions.

6. Experimental Evaluations

In order to comprehensively evaluate the performance of the ACT model under different witness behavior conditions, we have designed a test-bed which allows the well-being of service consumers adopting different approaches to be gauged. Through extensive simulations, it has been shown that the ACT approach significantly outperforms existing approaches in terms of the reduction in normalized average utility loss and, in the case of colluding witnesses, the reduction in their collusion power.

6.1. Simulation Test-bed

The test-bed simulates a scenario where a number of service consumers need the services offered by service providers. A service consumer incurs a cost of C in order to utilize the service of a service provider. If the service provider acts honestly, i.e. satisfies the service consumer's request, the service consumer gains an amount of utility of G after the interaction; otherwise, it gains zero utility. Therefore, the maximum average utility gain a service consumer can achieve is $G-C$, corresponding to all its interactions with service providers being successful; the minimum of this value is $-C$, if all its interactions are unsuccessful.

The main purpose of this test-bed is to investigate the effectiveness of the proposed ACT approach in mitigating the adverse effects of unfair testimonies relative to existing approaches. Although there are multiple ways of modeling the malicious behavior of service providers in a system, it is impractical to investigate the proposed model for all possible service provider population configurations. In our experiments, we adopt one of the common modeling approaches used by previous studies such as [12]. The service provider population is hostile to the service consumers and consists of

- 10% honest service providers (which renege randomly with a probability of 10%);
- 10% Type I dishonest service providers (which renege randomly with an initial probability of 40%);
- 40% Type II dishonest service providers (which renege randomly with an initial probability of 60%); and
- 40% Type III dishonest service providers (which renege randomly with an initial probability of 80%).

Except for the honest service provider group, the behavior patterns of all other groups changes gradually during the simulation. A service provider's behavior can change according to three different profiles: 1) increasing reneging probability, 2) decreasing reneging probability, or 3) unchanging reneging probability. The magnitude of each change is randomly chosen from the interval [0, 0.01]. Each dishonest service provider chooses one of the three profiles in each interaction with equal probability (i.e. $\frac{1}{3}$). The test-bed environment consists of 1000 service providers with different behavior patterns. During each round of simulation, each service consumer attempts to solve a total of N_m problems. The service consumers select service providers for interaction based on their reputation. The outcome of the interaction is assumed to be binary, namely $successful$ or $unsuccessful$, depending on whether the service provider provides the requested service.

There are 100 witnesses who accumulate direct trust evidence about the service providers and respond to service consumers requesting testimonies. When a request for testimony is received by a witness it will return a testimony to the requester if it has prior interaction experience with the particular service provider in the request; otherwise, it will decline the request. Two categories of malicious testimony sharing strategies are studied: 1) *random lying*, and 2) *collusive lying*.

In the case of random lying, a malicious witness does not collude with any other service provider. It either positively distorts a testimony (*ballot-stuffing*) or negatively distorts a testimony (*badmouthing*) following a preset lying probability. In the case of collusive lying, a number of service providers collude with lying witnesses to inflate their reputation in the eyes of service consumers (*ballot-stuffing*). The colluding witnesses do not give unfair testimonies about service providers who are outside the collusion ring. This is because, relative to a large online community, the sizes of collusion rings tend to be small. The costs for ballot-stuffing within collusion rings are significantly less than the costs for badmouthing a large number of competitors. The actual situation observed on e-commerce systems such as eBay.com supports this assumption [12]. In both random lying and collusive lying cases, the distortions are implemented as offset values added to or subtracted from the original testimony. Two types of unfair testimonies are supported in the test-bed:

- *Moderately Unfair Testimonies (MUT)*: the magnitude of the offset is randomly chosen in the range [0.1, 0.4];

- *Highly Unfair Testimonies (HUT)*: the magnitude of the offset is randomly chosen in the range [0.8, 1.0].

The values of the distorted testimonies are always kept within the range [0, 1] by hard-limiting to 1 (or 0) if the distorted testimonies after adding (or subtracting) exceeds 1 (or falls below 0).

In the proposed ACT approach, we use BRS as the trust evaluation model in this study. The values selected for the parameters in the ACT approach are listed in Table 2. These values are selected based on the experience reported from previous studies in reinforcement learning [16, 24]. They can achieve a good balance between learning speed and smooth changes in learning results.

Table 2: Parameter Values used in the Simulations

Parameter	Value
Th	0.5
φ	0.6
δ	0.05
ρ	0.4
M	10
N_m	200
G	5
C	1
R	1
P	-10
Pr_{min}	0.1

6.2. Evaluation Metrics

Two evaluation metrics from [12] are adopted to facilitate comparisons with state-of-the-art methods:

1. *Normalized Average Utility Leftover (NAUL)*: the normalized average utility ($0 \leq \sigma \leq 1$) measures the average per time step utility gain as a percentage of the maximum possible utility gain for each service consumer over its lifetime. It is calculated as:

$$\sigma = \frac{\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} g_i(t) - g_{min}}{g_{max} - g_{min}}. \tag{22}$$

   T is the total number of times a service consumer c_i has interacted with the service providers, N is the number of service consumers adopting the same approach as c_i does in the test-bed and $g_{max} = G - C$, $g_{min} = -C$. $g_i(t)$ is the actual utility gain of each c_i after each interaction at time t. If the interaction is successful $g_i(t) = g_{max}$; otherwise, $g_i(t) = g_{min}$. *NAUL* is then $(1 - \sigma)$. With perfect foresight, $(1 - \sigma) = 0$. It measures the percentage difference between the actual utility gain and the maximum possible utility gain per service consumer per time step (i.e., the *leftover* utility that the service consumers following a trust-aware interaction approach are not able to gain). The closer $(1 - \sigma)$ is to 0, the better the performance of a given model.

2. *Collusion Power*: the Collusion Power, cp, is a measure of the effectiveness of different models in the face of collusion [12]. It is defined as:

$$cp = \frac{\sum_{c_i \in A_{nc}} \#try(c_i)}{|A_{nc}| \cdot N_m} \tag{23}$$

   where A_{nc} denotes the set of non-colluding service consumers, c_i is a service consumer in this set, and $\#try(c_i)$ is the number of times c_i interacted with any colluding service provider during the simulation. In essence, cp represents the percentage of all tasks delegated to any of the colluding service providers in the simulated community.

6.3. Experiment Design

For each experiment, the composition of the common witness population is altered to simulate different scenarios. In the following sections, Hon denotes a population consisting entirely of honest common witnesses. BMn denotes a population consisting of $n\%$ badmouthing witnesses and $(100 - n)\%$ honest witnesses. BSn denotes a population consisting of $n\%$ ballot-stuffing witnesses and $(100 - n)\%$ honest witnesses. The malicious witness populations consist of half giving out MUTs and half giving out HUTs.

The experiments conducted in this study include two parts: 1) verifying the effectiveness of the adaptive trust evidence aggregation module of the ACT approach (labeled as ACT''), and 2) verifying the effectiveness of the ACT approach as a whole (labeled as ACT). In Part 1 of the experiment, five groups of service consumers are simulated for comparison. They are:

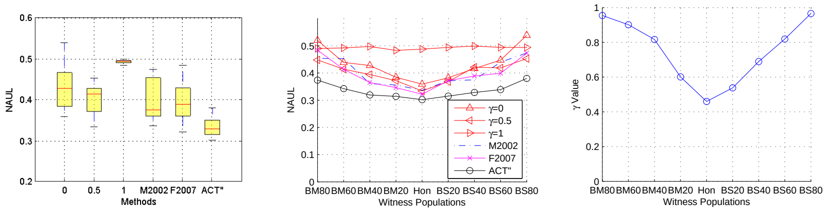
- Group $\gamma = 0$: service consumers who completely rely on indirect trust evidence;

- Group $\gamma = 0.5$: service consumers who rely on a balanced mix of direct and indirect trust evidence;

- Group $\gamma = 1$: service consumers who completely rely on direct trust evidence;

- Group $M2002$: service consumers who use the method described in [23] to set the γ value;

- Group $F2007$: service consumers who use the method described in [5] to set the γ value.



(a) Ranges of variation of *NAUL* by service consumer groups under non-collusive conditions.

(b) Performance of various service consumer groups under different non-collusive common witness populations.

(c) The variation of the γ value from the record of a service consumer in Group ACT'' with respect to an honest service provider under different non-collusive common witness populations .

Figure 3: Results for Experiment Part 1.

The group of service consumers equipped with the ACT approach is labeled as Group ACT''. Each group consists of 100 agents. All competing groups only request for testimonies from the common witness group.

In Part 2 of this experiment, we compare the performance of the complete ACT approach against:

- Group $W2010$: service consumers which employ an existing state-of-the-art method [12];
- Group $Y2003$: service consumers which employ a classic method [11];
- Group $B2002$: service consumers who only rely on their direct interaction experience to evaluate a service provider's trustworthiness using BRS [26].

The group of service consumers equipped with the ACT method is labeled as Group ACT. Each group also consists of 100 agents. All groups only request for testimonies from the common witness group same as in Part 1 of the experiment.

6.4. Experiment Results - Part 1

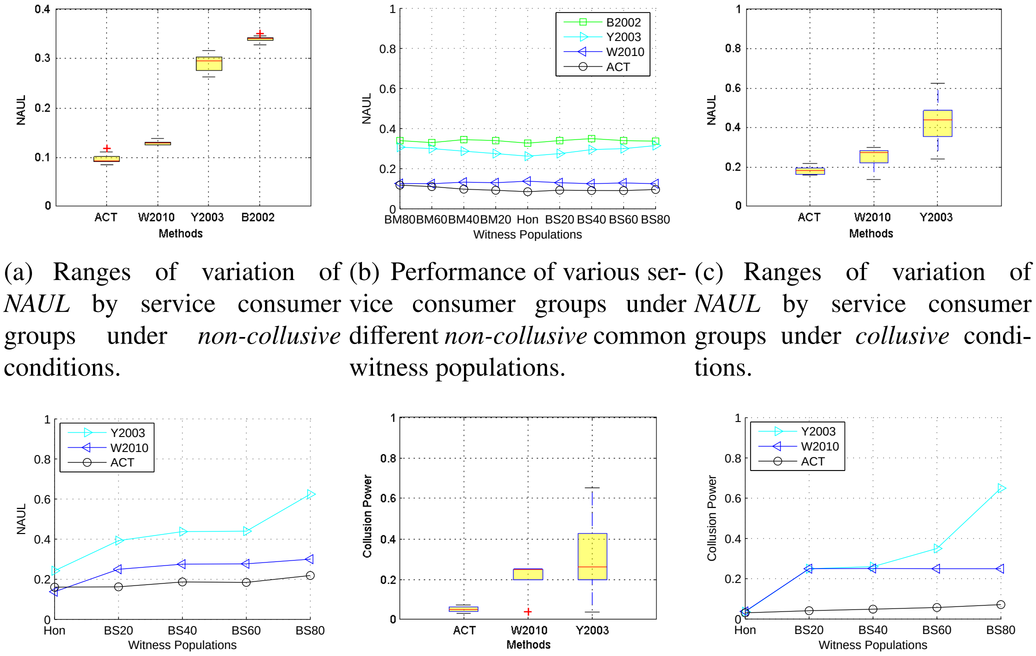
6.4.1. The Effect of Adaptive γ Values

Part 1 of this study is conducted assuming non-collusive common witnesses. The common witness population composition is altered from $BM80$ to Hon and then to $BS80$ to test the performance of service consumers employing different testimony aggregation methods. The results are summarized in Figure 6.3. It can be observed that Group $\gamma = 1$ achieves the highest *NAUL* values as they need more exploration to identify trustworthy service providers. Its performance is not affected by the changes in the common witness population composition. Completely relying on indirect trust evidence is also not a good strategy as the performance of Group $\gamma = 0$ is heavily affected by the presence of unreliable witnesses of both *BM* and *BS* types. However, the saving in exploration from completely relying on third party testimonies allows Group $\gamma = 0$ to achieve lower *NAUL* values than Group $\gamma = 1$. Nevertheless, the advantage drops with number of misbehaving witnesses as shown in Figure 6.3. The performance of the Group $\gamma = 0.5$ is the best among the three groups using static γ values. Group $F2007$'s performance is similar to that of Group $M2002$. As $F2007$ tries to learn which static strategy (γ=0,0.5,or 1) is the best under different conditions, its performance more or less tracks that of Group $\gamma = 0.5$ in our experiments. Group ACT'' outperforms all other methods under all testing conditions by an average of 20.79% in terms of the reduction in *NAUL*. A detailed breakdown of the comparisons is shown in Table 3.

The performance achieved by the proposed ACT'' service consumers can be attributed to their ability to adapt the values of γ for each service provider as the environment conditions change in a continuous manner. Figure 6.3 shows a snap-shot of the γ value from a service consumer in Group ACT'' with respect to an honest service provider in its local record. It can be seen that as the witness population becomes increasingly hostile, the reliance on third-party testimonies is reduced to mitigate their negative influence on the service consumer's interaction decisions.

Table 3: Improvement of Group ACT'' over other Groups

Group	Improvement
$\gamma = 0$	23.00%
$\gamma = 0.5$	16.82%
$\gamma = 1$	31.99%
$M2002$	16.73%
$F2007$	15.41%
Average	20.79%



(a) Ranges of variation of *NAUL* by service consumer groups under *non-collusive* conditions.

(b) Performance of various service consumer groups under different *non-collusive* common witness populations.

(c) Ranges of variation of *NAUL* by service consumer groups under *collusive* conditions.

(d) Performance of various service consumer groups under different *collusive* common witness populations.

(e) Ranges of variation of Collusion Power by service consumer groups under *collusive* conditions.

(f) Performance of various service consumer groups under different *collusive* common witness populations.

Figure 4: Results for Experiment Part 2.

6.5. *Experiment Results - Part 2*

6.5.1. *Performance of ACT under non-Collusive Lying*

In Part 2 of this study, the performance of the complete ACT approach is investigated. The distributions of the *NAUL* achieved by all five models in this study are shown in Figure 4(a). Group ACT has achieved significantly lower level of *NAUL* than existing models. As shown in Figure 4(b), when the percentage of malicious witnesses increases, the performance of Group *B*2002 is relatively stable as it does not take into account testimonies from witnesses when making trustworthiness evaluations. However, the *NAUL* of Group *Y*2003 deteriorates significantly. The performance of groups *W*2010 and *ACT* are relatively consistent across different witness population configurations. The consistent performance achieved by the ACT approach is due to that fact that it uses the interaction outcomes with the service providers rather than the majority opinion of the witnesses to update the credibility ranking of known witnesses, as well as its ability to adjust

Table 4: Improvement of Group *ACT* over other Groups

Group	Badmouthing	Ballot-stuffing	Overall
*W*2010	18.44%	29.91%	25.16%
*Y*2003	64.18%	70.20%	66.98%
*B*2002	69.07%	74.18%	71.66%

its preference of the two trust evidence sources dynamically. As can be seen from Table 4, overall, Group *ACT* outperforms all other groups in terms of reduction in *NAUL* by significant margins. The advantage is more significant under ballot-stuffing conditions due to the addition of the reward correction value $\theta_{k,j}$ in Eq. (6) that penalizes positively biased testimonies.

6.5.2. Performance of ACT under Collusive Lying

In our test-bed, the collusive witnesses always form collusion rings with Type III malicious service providers to try to promote their reputation. The proportion of collusive witnesses in the total common witness population is varied from *Hon* to *BS80*. From Figure 4(c), it can be seen that the presence of colluding witnesses tricks the *Y*2003 group into interacting more often with collusive service providers than other groups. In addition, by comparing Figure 4(c) with Figure 4(a), we find that the negative impact of collusion is more powerful than that of non-collusive random lying. The most adversely affected group is still the *Y*2003 group. The highest *NAUL* of this group is about 0.3 under *BS80* without collusion. However, under *BS80* with collusion, this value increases to around 0.6 (as shown in Figure 4(d)). This is due to the fact that colluding witnesses do not give unfair testimonies about non-colluding service providers, so that their testimonies are considered accurate in these cases. Thus, they are strategically building up their credibility with the service consumers in order to mislead them into interacting with collusive service providers later.

The performance of all the models studied in our test-bed deteriorated under the influence of collusion as shown in Table 5. Although Group *ACT* and Group *W*2010 managed to maintain the witness agents' collusion power at relatively low levels compared to other groups as illustrated in Figure 4(e), their performances in terms of *NAUL* still deteriorated under collusion. It is observed, from Table 6, that the ACT approach significantly outperforms all other approaches in terms of mitigating the adverse effect of collusion. The over performance in terms of reduction in collusion power is the most significant when the majority of the witness

Table 5: Performance Deterioration due to Collusion (*NAUL*)

Group	without Collusion	with Collusion
ACT	0.0890	0.1825
*W*2010	0.1283	0.2479
*Y*2003	0.2895	0.4145

Table 6: Improvement of Group *ACT* over other Groups

Group	Collusion Power	*NAUL*
*W*2010	77.60%	26.37%
*Y*2003	85.94%	55.97%

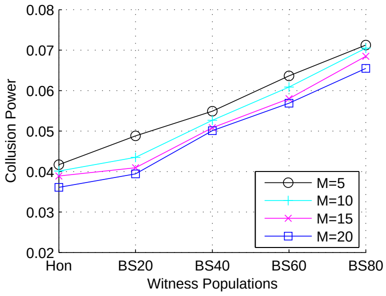


Figure 5: The influence of the parameter M on the performance of ACT under different witness population compositions.

population consists of collusive witnesses, as can be seen from Figure 4(f).

6.6. Sensitivity Analysis

To study the influence of M on the proposed ACT approach, we alter the value of M and re-run the experiments. The value of M is varied to be equivalent to between 5% to 20% of the common witness population. The experiments are re-run only for the cases where collusion exists since collusive testimonies are more powerful in affecting the credibility models. From Fig. 5, it can be seen that generally, collusion power increases with the fraction of colluding witness agents. However, the value of collusion power is maintained at a relatively low level by the ACT approach. This trend is true for the different values of M.

It is expected that the effectiveness of the ACT approach improves with M. However, the value of M also determines the storage capacity required at each individual service consumer as well as the time taken to estimate the reputation of a service provider. Therefore, a service consumer needs to balance the trade-off between potentially more accurate interaction decisions and the extra effort required to gather testimonies from more witnesses.

6.7. Analysis of Results

Several reasons contribute to the superior performance of Group ACT model over Groups $W2010$ and $Y2003$:

- $Y2003$ uses the number of past interactions between a service consumer c_i and the service provider of interest s_j to determine whether third-party testimonies are required. If the number of past interactions between c_i and s_j exceeds a predefined threshold, c_i will not ask for testimonies when estimating s_j's trustworthiness. However, since the behavior of the witnesses are changing in the experiments, c_i's direct trust evidence may become outdated. This increases c_i's risk exposure in the long run.

- $W2010$ applies an adaptive strategy in aggregating third-party testimonies. However, it also uses a service consumer c_i's own evaluation of a service provider s_j's trustworthiness as a baseline to determine which testimonies are potentially unfair. It proposed a measure of uncertainty induced by additional testimonies. If a new testimony contradicts c_i's current belief about the trustworthiness of s_j, it would be regarded as increasing c_i's uncertainty and discarded. While this approach is more dynamic than $Y2003$, it still suffers from the effect of changing service provider behavior to some degree.

- In contrast, the ACT approach always seeks testimonies from witnesses when estimating a service provider's reputation. By learning the weights assigned to different witnesses' testimonies based on the outcomes after each interaction, the ACT approach dynamically decides which witnesses to keep in the top M list for each service provider based on their contributions to the well-being of the service consumer. Even in the face of highly hostile witness populations, the ACT approach still can maintain a relatively good performance by relying more on the direct trust evidence source. This mechanism also helps the service consumers when the behavior of a service provider changes. If this change is reflected first in the testimonies, the service consumer can increase the weight given to the indirect trust evidence source to reduce the need for trial and error; if this change is detected first by the service consumer itself, it can increase the weight given to the direct trust evidence source to reduce its chance of being misled by outdated opinions from others.

7. Conclusions

A trust evidence aggregation model, based on the principles of the actor-critic learning, was proposed to mitigate the adverse effects of biased testimonies. It dynamically adjusts the weights given to selected testimonies as well as the relative emphasis given to the direct and indirect trust evidence sources to reduce a service consumer's risk of being misled by biased third-party testimonies or outdated direct past experience. The ACT approach can be applied to most existing trust models as long as their trust evaluations can be normalized to a range of [0, 1] and the interaction outcomes can be represented as either successful or unsuccessful. Experimental results show that the ACT approach is more effective than existing models in mitigating the adverse effects of biased testimonies.

In the computational trust research literature, the most popular metrics used to determine the relative merits of trust models are individually rational in nature. Such measures include various forms of long term average monetary gain for service consumers, and the deviation of estimated trustworthiness from ground truth. Other means of assessing trust decisions on the social welfare of an entire system are rarely considered. It is our belief that apart from the utility enhancement goals, trust models must take into consideration the fair treatment of trustworthy service providers during their decision making processes. This is a crucial consideration that may hold the key to ensuring the long term sustainable operation of a system built on trust. We will investigate this topic in subsequent work.

Acknowledgment

This research is supported in part by Interactive and Digital Media Programme Office (IDMPO), National Research Foundation (NRF) hosted at Media Development Authority (MDA) of Singapore under Grant No.: MDA/IDM/2012/8/8-2 VOL 01, and the National Natural Science Foundation of China (No.61202212).

References

[1] A. Jøang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2) (2007) 618–644.

[2] Z. Noorian, M. Ulieru, The state of the art in trust and reputation systems: A framework for comparison, Journal of Theoretical and Applied Electronic Commerce Research 5 (2) (2010) 97–117.

[3] H. Yu, Z. Shen, C. Miao, C. Leung, D. Niyato, A survey of trust and reputation management systems in wireless communications, Proceedings of the IEEE 98 (10) (2010) 1755–1772.

[4] H. Yu, Z. Shen, C. Leung, C. Miao, V. R. Lesser, A survey of multi-agent trust management systems, IEEE Access 1 (1) (2013) 35–50.

[5] K. K. Fullam, K. S. Barber, Dynamically learning sources of trust information: Experience vs. reputation, in: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07), 2007, pp. 1055–1060.

[6] S. Ba, P. Pavlou, Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior, MIS Quarterly 26 (3) (2002) 243–268.

[7] A. Whitby, A. Jøsang, J. Indulska, Filtering out unfair ratings in bayesian reputation systems, in: Workshop on Trust in Agent Societies at the 4rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05), 2005.

[8] J. Weng, C. Miao, A. Goh, An entropy-based approach to protecting rating systems from unfair testimonies, IEICE - Transactions on Information and Systems E89-D (9) (2006) 2502–2511.

[9] S. Liu, J. Zhang, C. Miao, Y.-L. Theng, A. C. Kot, iclub: An integrated clustering-based approach to improve the robustness of reputation systems, in: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11), 2011, pp. 1151–1152.

[10] R. Jurca, B. Faltings, Towards Incentive-Compatible Reputation Management, Vol. 2631, 2003, pp. 13–24.

[11] B. Yu, M. P. Singh, Detecting deception in reputation management, in: Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'03), 2003, pp. 73–80.

[12] J. Weng, Z. Shen, C. Miao, A. Goh, C. Leung, Credibility: How agents can handle unfair third-party testimonies in computational trust models, IEEE Transactions on Knowledge and Data Engineering (TKDE) 22 (9) (2010) 1286–1298.

[13] N. Miller, P. Resnick, R. Zeckhauser, Eliciting informative feedback: The peer-prediction method, Management Science 51 (9) (2005) 1359–1373.

[14] J. Zhang, R. Cohen, K. Larson, A Trust-based Incentive Mechanism for E-Marketplaces, Vol. 5396, 2008, pp. 135–161.

[15] S. Marsh, P. Briggs, Examining Trust, Forgiveness and Regret as Computational Concepts, Human-Computer Interaction Series, Part I, 2009, pp. 9–43.

[16] G. Tesauro, Temporal difference learning and td-gammon, Communications of the ACM 38 (3) (1995) 58–68.

[17] Z. Shen, H. Yu, C. Miao, J. Weng, Trust-based web-service selection in virtual communities, Journal for Web Intelligence and Agent Systems (WIAS) 9 (3) (2011) 227–238.

[18] H. Yu, S. Liu, A. C. Kot, C. Miao, C. Leung, Dynamic witness selection for trustworthy distributed cooperative sensing in cognitive radio networks, in: IEEE 13th International Conference onCommunication Technology (ICCT), 2011, pp. 1–6.

[19] C. M. Jonker, J. Treur, Formal analysis of models for the dynamics of trust based on experiences, in: Proceedings of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World: MultiAgent System Engineering (MAAMAW'99), 1999, pp. 221–231.

[20] M. Schillo, P. Funk, I. Stadtwald, M. Rovatsos, Using trust for detecting deceitful agents in artificial societies, Journal of Applied Artificial Intelligence 14 (8) (2000) 825–848.

[21] J. Shi, G. V. Bochmann, C. Adams, Dealing with recommendations in a statistical trust model, in: Workshop on Trust in Agent Societies in conjunction with the 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'05), 2005, pp. 144–155.

[22] K. S. Barber, J. Kim, Soft Security: Isolating Unreliable Agents from Society, Vol. 2631, 2003, pp. 224–233.

[23] L. Mui, M. Mohtashemi, A computational model of trust and reputation, in: 35th Annual Hawaii International Conference on System Sciences (HICSS'02), Vol. 7, 2002, pp. 188–197.

[24] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[25] J. Sabater, C. Sierra, Reputation and social network analysis in multi-agent systems, in: Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02), 2002, pp. 475–482.

[26] A. Jøang, R. Ismail, The beta reputation system, in: Proceedings of the 15th Bled Electronic Commerce Conference, 2002, pp. 41–55.

[27] V. R. Konda, J. N. Tsitsiklis, On actor-critic algorithms, SIAM Journal on Control and Optimization 42 (4) (2002) 1143–1166.