Towards robust reputation system based on clustering approach

Xin Zhou
Department of Social Informatics
Kyoto University
Kyoto, Japan
xin@ai.soc.i.kyoto-u.ac.jp

Shigeo Matsubara
Department of Social Informatics
Kyoto University
Kyoto, Japan
matsubara@i.kyoto-u.ac.jp

***Abstract*—Service computing is playing a more and more important role in current Internet activities, especially with the rapid adoption of electric markets, more and more individuals are engaging with commercial services. As the potential profit of service computing is becoming clear, malicious users are ramping up unfair rating attacks that can mislead honest service consumers into transacting with dishonest service providers. Moreover, some dishonest service providers may collude with dishonest service consumers to damage the reputation of service rivals. In this paper, we proposed a clustering-based reputation system that is robust to various unfair rating attacks. The model categorizes consumers as either honest or dishonest according to their rating ratio. It utilizes the Dirichlet distribution in determining reputation values. We analyze the profits and costs attained by the attacker and elucidate the conditions under which an attack is profitable. Experiments demonstrate that our clustering-based reputation model is more robust than the state-of-art model against currently successful attacks.**

***Keywords*-Web service; Feedback rating; Reputation; Clustering**

I. Introduction

The rapid adoption of electric markets such as eBay, Amazon and Taobao has throw into strong relief two major problems: 1). Even though large number of service consumers have no or little past experience with the services, each service consumer must attempt assess and identify reliable interaction partners by themselves. 2). Unfair rating attacks from dishonest service consumers can mislead honest service consumers to transact with dishonest service providers. Moreover, some dishonest service providers behave differently toward different consumers, and can collude with dishonest service consumers to boost their reputation in the e-commerce system. These problems are currently countered, to mitigate the potential risk to the consumers, by adopting some form of reputation system [1], [2]. Such systems not only aggregate and filter information for service consumers, but also act as an incentive for service providers to improve their service quality. Reputation is defined as a subjective assessment of service quality and is typically determined by collecting ratings or feedback from service consumers. It acts as a global and public value that can be observed by all the participators in the system. Hence, new comers who have no experience can utilize the reputation system to mitigate the risks of selecting a partner.

Various approaches [3], [4], [5] have been proposed for building robust reputation systems for service providers, and most use personalized similarity-based credibility to evaluate the reputation of a service provider. However, those techniques are usually unreliable since the rating distributions are marginally effective. Not only are novices exposed to significant risks, but also experts are unable to exploit the rating information efficiently even if they accumulate a lot of data. Others models [6], [7], [8] use statistical theory to handle unfair ratings; they are designed to filter out the ratings that deviate in some way from the mainstream ratings. TRAVOS, however, evaluates rating accuracy against the consumers past opinions, and hence can avoid the Sybil attack. When considering the timeliness of a rating and the dynamic changes possible in service quality, the estimates of rating accuracy may be inaccurate.

Actually, in a system where service providers can collude with consumers, there is no way to remove the impact of unfair rating completely if the malicious consumers can rapidly modify their attack strategies. An interesting solution is to analyze the cost of performing an effective unfair rating attack against a specific reputation model and negate the incentive that drives unfair rating attacks.

The clustering method has been proven to be effective in immunizing reputation systems against unfair ratings [9]. However, different from previous proposals that cluster the rater based on similarity among raters, the model proposed in this paper uses the rating ratio information to detect unfair raters. This information reflects the inherent behavior of customers; 50%$\sim$60% of eBay customers do not leave ratings for various reasons [10]. This means that the transaction volume is much larger than the number of received ratings. Further, we update the trustworthiness of each rater by applying a fitness function. Service provider reputations are aggregated by using a Dirichlet distribution. Based on this model, we analyze the costs and profits associated with effective attacks and reveal the conditions under which such attacks become attractive.

The main contributions of this paper include:

- A clustering-based reputation system is proposed that is

978-1-4673-7281-7/15 $31.00 © 2015 IEEE
DOI 10.1109/SCC.2015.15

robust to various attacks. It uses rating ratio information to separate honest raters from dishonest ones.
- We describe the conditions under which attacks make economic sense.

The rest of our paper is organized as follows. We discuss related work in Section II and introduce the clustering-based reputation model in Section III. Section IV evaluates the performance of the model by comparing it to a state-of-art model. Our analysis of the potential profit of performing an unfair rating attack is presented in Section V. The research is rounded off with a conclusion in the last section.

II. Related Works

Since Resnick et al. [11] pointed out the issues posed by web site reputation, various reputation systems based on feedback have been published. However, various forms of unfair attacks have been observed and are being studied by the trust and reputation community [12], [13]. The key unfair rating attacks are listed here:

- **Constant**: An individual dishonest rater gives constant and unfair ratings to a service provider.
- **Camouflage**: Dishonest raters gain and then abuse the trust of providers.
- **Whitewashing**: Dishonest rates try to escape their reputations by using new accounts that have the default value of trust.
- **Sybil**: A group of dishonest raters gives constant and unfair ratings to a service provider.
- **Sybil Camouflage**: A group of dishonest raters act together in mounting a Camouflage attack.
- **Sybil Whitewashing**: A group of dishonest raters act together in conducting a Camouflage attack.

Jøsang and Ismail proposed a Bayesian reputation system based the beta probability distribution [6]. The beta distribution is used to determine and update the reputation value from behavior data. The beta reputation model may fail under coalition attack when a group of malicious users try to modify the reputation value deliberately with fake feedback. The TRAVOS model, which also uses the Beta distribution to update the new received information, is not sensitive to changes in reputation value [7].

Some reputation systems robust to malicious behavior have been proposed, such as the hierarchical Bayesian inferred trust model [14] and robust linear Markov [8]. The first model gains its robustness by learning from all observed information, including direct experience or third-parties. The latter model updates the reputation in a hidden Markov model based on new observations but is vulnerable to Sybil attacks. Other various reputation models [3], [4], [5] have been proposed to deal with unfair attacks. Those methods use personalized similarity-based credibility to evaluate the reputation of a service provider. They compute reputation for individuals, so consumers who have little or no experience can not benefit from those systems. To summarize, those systems manage the reputation of individual peers in a decentralized manner based on the behavior of each individual. The reputation value is different from one to another, and the underlying differences are obscure.

Different from previous research, the model proposed in this paper uses a centralized scheme to generate reputation values that can be understood by all consumers. The clustering approach is adopted to tag raters as either honest or dishonest. Previous research that considered clustering, such as the multiagent evolutionary trust model (MET) [5], mainly focused on categorizing the raters according to rating history. In MET, raters with similar rating histories are grouped into trust networks. The drawback is that the resulting schemes react slowly to the changes in reputation value. In our model, raters are classified by their ratings given in the current period, and the reputation of a service provider is immediately updated when changes in the reputation value the honest cluster are detected. The trustworthiness of each consumer is maintained and contributes to the reputation calculation via the Dirichlet distribution.

III. The clustering-based reputation model

The key point of a reputation model is how to detect the dishonest service raters and decrease their trustworthiness when calculating the reputation of a service provider.

A. Clustering the rating vector

We assume a service computing environment with M service providers $S = \{S_i | i = 1, 2, ..., M\}$ each having one functional equivalently service with different quality, and N service consumers $C = \{C_j | j = 1, 2, ..., N\}$. The rating given by consumer C_j to provider S_i after the transaction at time t denoted as $r_{t,C_j,S_i} \in [0, 1]$. As the ratings accumulate, rating vector R_{t,S_i} received in time period t by S_i can be expressed as:

$$\overrightarrow{R}_t^{S_i} = [r_{t,C_1,S_i}, ..., r_{t,C_j,S_i}, ...] \tag{1}$$

The number of element in $\overrightarrow{R}_t^{S_i}$ is no more than N, because not all the consumers will transact with S_i in time t and not all consumers will leave their ratings after the transaction. The rating ratio of consumer C_j is given by:

$$\rho(t)_{C_j} = \gamma_{C_j} / \eta_{C_j} \tag{2}$$

where γ_{C_j} is the number of rating for C_j, and η_{C_j} is the number of total transaction for C_j. Rating vector $\overrightarrow{R}_t^{S_i}$ is used adopted by the clustering algorithm when classifying the ratings into Z clusters $T_1, T2, ..., T_Z$. A fast and robust cluster algorithm is applied on the rating vector to generate a set of clusters [15]. For large data sets, the clustering algorithm can find the density peaks that are robust with respect to the choice of cutoff distance between points. Take figure 1 as an example, in day 90, two clusters occur simultaneously. In this situation, an additional mechanism is needed to detect the dishonest cluster.

B. Detecting the dishonest clusters

As we have derived Z clusters $T_1, T_2, ..., T_Z$, in each cluster, the set of ratings are denoted as $T_k = \{C_j | C_j \in C\}$. For each consumer C_j, if their rating on S_i at time t is classified into the honest cluster, the trustworthiness value $\phi(t)_{C_j}^{S_i}$ of C_j on S_i will updated by the fitness function $h(t, C_j, S_i)$. The trustworthiness value of fair raters will be increased by weighting them when calculating the reputation of S_i, unfair raters are deweighted. To detect dishonest clusters, there must be some property that can distinguish unfair from fair raters. The underlying statistics of eBay usage showed that normal customers, C_j , exhibit a rating ratio, $\rho(t)_{C_j}$, around 0.5∼0.6. The perpetrators of each unfair rating attack naturally attempt to minimize their cost in performing the attack and they do so by increasing their rating ratios. As a result, a ρ-ϕ graph can be plotted for rating vector $\overrightarrow{R}_t^{S_i}$. Hence, for Sybil attack, the unfair consumers will be categorized as high ρ, low ϕ group and detected as such. The reason is that the unfair raters will seize every opportunity to enter malicious ratings. That is, for every transaction, they will give an unfair rating to decrease or increase the reputation of the provider deliberately. However, for Sybil Camouflage attack, the unfair raters must secure a relative high degree of trust ϕ and high ρ because they initially pretend to be fair raters. Rating ratio ρ plays a key role in distinguishing the dishonest clusters. We define rating ratio for a cluster T_k as:

$$\rho(t)_{T_k} = \frac{1}{K} \sum_{j=1}^{K} \rho(t)_{C_j}, and\ C_j \in T_k \tag{3}$$

where K is the number of raters in cluster T_k. The above equation reveals that the rating ratio of a cluster is the average rating ratio of all its members. For example, in figure 1, the rating ratio of cluster $C1$ and cluster $C2$ is calculated to detect the dishonest cluster. A threshold value TH is defined to distinguish the dishonest clusters in Algorithm 1.

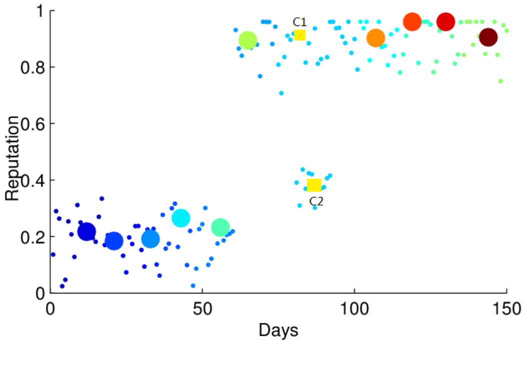


Figure 1. An example of clustering. At day period (80, 90), two clusters $C1$ and $C2$ occur simultaneously.

C. Calculating the reputation based on honest cluster

For each rater C_j in the honest cluster at time t, the trust value is updated by the fitness function. As each honest rater C_j has their direct experience with the quality of S_i, variable $\phi(t)_{C_j}^{S_i}$ indicates how much trust should be ascribed to rating r_{t,C_j,S_i}. Suppose the number of members in honest cluster T_k is K, in the case of discrete distributions parameterized by $\phi(t)_{C_j}^{S_i}$, the probability density of C_j usually means assigning a K-dimensional Dirichlet distribution:

$$Dir(C_l|\alpha) = \frac{1}{Beta(\alpha)} \prod_{l=1}^{K} (C_l)^{\alpha_l} \tag{4}$$

And the $Beta(\alpha)$ is defined in terms of the gamma function as:

$$Beta(\alpha) = \frac{\prod_{l=1}^{K} \Gamma(\alpha_l)}{\Gamma(\Sigma_{l=1}^{K} \alpha_l)} \tag{5}$$

where $\alpha =< \alpha_1, \alpha_2, ..., \alpha_K >$, and $\alpha_l = \phi(t)_{C_l}^{S_i}$, C_l is the l-th member in cluster T_k. Given that the expected value of the Dirichlet distribution is defined as $E[C_l|\alpha] = \alpha_l / \sum_{l=1}^{K} \alpha_l$. The reputation of S_i can be derived as:

$$\begin{aligned} \mathbf{R}_{t,S_i} = E[r_{t,C_j,S_i}|\alpha] &= \sum_{l=1}^{K} r_{t,C_l,S_i} \cdot p(C_j = C_l|\alpha) \\ &= \sum_{l=1}^{K} r_{t,C_l,S_i} \cdot E[C_l|\alpha] \\ &= \sum_{l=1}^{K} r_{t,C_l,S_i} \cdot \frac{\alpha_l}{\sum_{n=1}^{K} \alpha_n} \end{aligned} \tag{6}$$

where r_{t,C_l,S_i} is the rating value here.

D. Fitness function

The fitness function acts here as a sanction mechanism to reward the honest rater, while devaluing the ratings of dishonest raters. Given this background, we define the admissible function as: A fitness function is said to be **admissible** if it always rewards the honest rater and punishes the dishonest rater at time t. That is, for honest rater C_h and dishonest rater C_d at time t, the following equations hold for the admissible function $h(t, C_j, S_i)$:

$$\begin{aligned} \phi(t)_{C_h}^{S_i} &= h(t, C_h, S_i) \\ &> h(t-1, C_h, S_i) = \phi(t-1)_{C_h}^{S_i} \end{aligned} \tag{7}$$

$$\begin{aligned} \phi(t)_{C_d}^{S_i} &= h(t, C_d, S_i) \\ &< h(t-1, C_d, S_i) = \phi(t-1)_{C_d}^{S_i} \end{aligned} \tag{8}$$

A fitness function sensitive to dishonest raters can effectively mitigate their impact. However, an honest rater may be erroneously devalued especially when the service requests

Table I
ROBUSTNESS OF REPUTATION MODELS VS. ATTACKS ON CONSTANT REPUTATION.

Models	Constant	Camouflage	Whitewashing
MET	0.960±0.025	**0.966±0.020**	0.966±0.021
CRM	**0.995±0.019**	**0.966±0.020**	**0.994±0.019**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.926±0.039	0.920±0.037	0.934±0.039
CRM	**0.979±0.030**	**0.995±0.033**	**0.990±0.032**

* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

are stacked. For simple balance, the fitness function adopted in this paper uses the admissible function defined as:

$$h(t, C_h, S_i) = h(t-1, C_h, S_i) + 1 \tag{9}$$

$$h(t, C_d, S_i) = h(t-1, C_d, S_i) - 1 \tag{10}$$

where $h(t-1, C_h, S_i) > 0$ in equation 10. For honest and dishonest rater, their trust values are updated by equation 9 and 10 respectively.

The pseudo-code summary of the clustering-based reputation model is given in Algorithm 1.

Algorithm 1 Clustering-based reputation model

```
procedure CRM($S_i$, $\overrightarrow{R}_t^{S_i}$)
    Inputs: $S_i$, evaluated service provider;
            $\overrightarrow{R}_t^{S_i}$, rating vector received by $S_i$ at time $t$;
    Output: $\mathbf{R}_{t,S_i}$, the reputation of $S_i$ at time $t$.
    $T_1, T_2, ..., T_Z$ = CLUSTERING($\overrightarrow{R}_t^{S_i}$)
    $\forall T_k, (1 \leq k \leq Z)$
        Calculate $\rho(t)_{T_k}$ by equation (3)
    $\rho(t)_{T_m} = MIN(\rho(t)_{T_k})$
    Tag cluster $T_k$ as dishonest if it satisfied:
        $\rho(t)_{T_k} - \rho(t)_{T_m} > TH$
    $\forall T_k, (1 \leq k \leq Z)$
        Update trustworthiness by fitness function (9) and (10)
    $\forall$ honest clusters
        Calculate the reputation $\mathbf{R}_{t,S_i}$ by equation (6)
    return $\mathbf{R}_{t,S_i}$
```

When service provider S_i receives the rating vector, the CRM procedure will respond by reflecting the latest reputation of S_i. It first categorizes the ratings into several clusters, and detects unfair clusters by their rating ratios. This clustering approach makes CRM preferable with large data set as its computational complexity is only sensitive to the number of recently received ratings. The threshold value of $\rho(t)_{T_k}$ delineating honest from dishonest clusters can be learnt by a supervised machine learning algorithm and thus updated over time. The rest of the code is straightforward; rater trust is updated and the final reputation value is derived.

Table II
MEAN ABSOLUTE ERROR (MAE) OF REPUTATION ESTIMATION FOR HONEST DUOPOLY SERVICE ON CONSTANT REPUTATION.

Models	Constant	Camouflage	Whitewashing
MET	0.014±0.005	0.013±0.005	0.014±0.004
CRM	**0.009±0.002**	**0.008±0.002**	**0.009±0.001**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.027±0.018	0.027±0.009	0.027±0.009
CRM	**0.014±0.003**	**0.012±0.003**	**0.014±0.003**

* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

Table III
MEAN ABSOLUTE ERROR (MAE) OF REPUTATION ESTIMATION FOR DISHONEST DUOPOLY SERVICE ON CONSTANT REPUTATION.

Models	Constant	Camouflage	Whitewashing
MET	0.057±0.018	0.054±0.018	0.052±0.015
CRM	**0.030±0.025**	**0.038±0.026**	**0.029±0.021**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.087±0.030	0.088±0.032	0.090±0.029
CRM	**0.030±0.025**	**0.034±0.026**	**0.032±0.021**

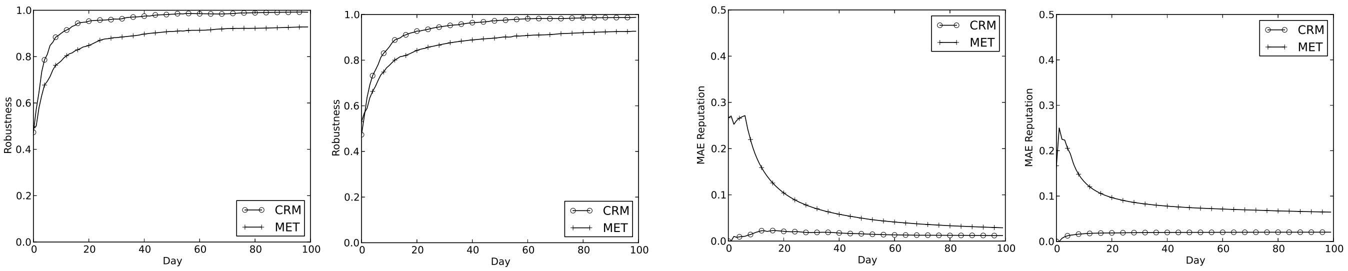
* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

IV. EXPERIMENTATION

To evaluate the proposed model, we first introduce the duopoly service providers testbed used in paper [5]. Different attack strategies will be simulated to assess the robustness and accuracy of the proposed model CRM relative to MET as the paper concludes it is more robust and effective than the state-of-art models against typical attacks.

A. Simulation Setup

As the papers on different reputation models used their own evaluation method, despite a comprehensive testbed is proposed in [16], the data used in the testbed is lack of rating ratio information. Hence, we reuse the e-market testbed designed for simulating ”Duopoly Market” where two service provider occupy a large proportion of the transaction volume. The dishonest duopoly provider may collude with the dishonest consumers to perform various attack to the damage the reputation of honest provider. The simulation assumes that of the 198 common service providers, half are honest and the other half are dishonest. Furthermore, for the non-Sybil based attack case, there are 12 dishonest consumers (attackers) and 24 honest consumers. The number is switched in the Sybil attack case, that is, 24 attackers and 12 honest consumers. Each consumer interacts with one provider each day, so the consumer has the probability of 0.5 of interacting with a common service provider. When choosing which provider to access for service, the honest consumer tends to select the provider randomly. In the duopoly case, the honest consumer uses the reputation model to decide which one should be accessed. The attacker will



(a) Robustness vs. Sybil Cam (b) Robustness vs. Sybil WW (c) Honest duopoly provider (d) Dishonest duopoly provider

Figure 2. Constant Reputation (a-b): Robustness of Reputation Model under Attack; (c-d): MAE of Duopoly Provider Reputation under Sybil WW attack.

choose the duopoly seller according to the attack mode. After each transaction, each consumer rates the service provider with probability P_r. The rating scores given by honest consumers following Normal distribution $N(\mu, \sigma)$, where μ is the actually reputation of a service and $\sigma = 0.05$. For fairness, we use all the parameters and environment in MET [5] except giving each consumer a probability to rate the service. The threshold value TH is determined by the supervised machine learning algorithm initially, and can be updated as the ratings accumulated. The robustness of reputation model(M) against attack model (Atk) is defined as:

$$\mathcal{R}(M, Atk) = \frac{Tran(S_H)}{C_H \times Days \times Ratio} \tag{11}$$

where $Tran(S_H)$ is the transaction volume of the honest duopoly provider by honest consumers, and C_H is the number of honest consumers. The value of $\mathcal{R}(M, Atk)$ normally is in [0, 1], where 0 indicates the model is completely vulnerable to attack type Atk; while 1 denotes the model is complete proof against the attack. The accuracy of the model is evaluated by mean absolute error (MAE):

$$MAE(S_i) = \frac{\sum_t |'R_{t,S_i} - \mathbf{R}_{t,S_i}|}{Days} \tag{12}$$

where $'R_{t,S_i}$ is the actual reputation value of S_i, and $\mathbf{R}_{t,S_i}$ is the reputation as estimated by the reputation model. As in the MET model, the reputation value of S_i is calculated from the ratings of all honest consumers, hence, the above equation can be transformed into:

$$MAE(S_i) = \frac{\sum_t \sum_{C_j} |'R_{t,S_i} - \mathbf{R}_{t,C_j,S_i}|}{C_H \times Days} \tag{13}$$

Small MAE values indicate that the model is more accurate.

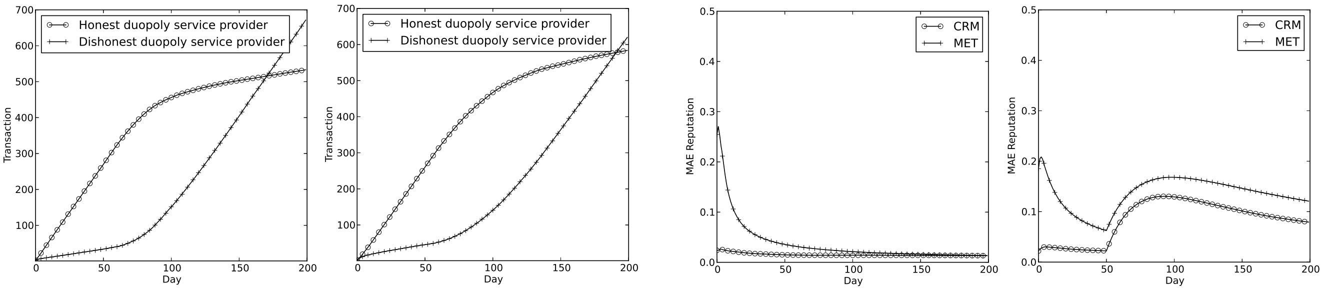
Each attack is carried out 50 times to reduce the randomness. The mean and standard deviation values are shown in Table II and III, and the best results are in bold font.

B. Experiment on robustness on constant reputation

Experiments were carried out to evaluate the robustness of the reputation model. In Table I, the two models have almost the same results with CRM slight outperforming the MET model. All the results are consistent with the authors results [5]. However, when CRM results are observed more carefully, it is may be thought strange that $\mathcal{R}(CRM, Sybil\ Cam)$ is more robust than $\mathcal{R}(CRM, Cam)$. The reason is that in performing the Camouflage attack, all attackers must first establish their trust before day 20, and then submit unfair ratings. Luckily, the dishonest consumer give fair rating at the beginning of the attack, the fair rating are used to facilitate the reveal of the actually reputation of honest service provider. Consequently, in the next few days, honest consumers will choose the provider with high reputation. This result is consistent with Figure 2(a), in which the robustness value increases faster than under Sybil Whitewashing attack 2(b). Note that the two models yield the same robustness value on the final day. We plot the daily robustness value for Sybil Whitewashing attack in Figure 2(b). The CRM curves in Figures 2(a) and 2(b) show that it converges to the excepted robustness value faster than the MET model. This rapid gain property can help the model to resist attacks performed at the very beginning stage. The underlying reason why CRM gains fast robustness value at the beginning is that CRM generates a public reputation value that can be observed by all consumers, while only the consumer in the trust network can sense service provider quality.

C. Comparison of MAE

In Table II and III, for both honest and dishonest duopoly sellers, the clustering-based reputation model attains the best results. CRM shows significant improvement considering the deviation of rating scores is set as $\sigma = 0.05$ in simulation setup. Both models can mitigate the influence of malicious raters. The MAE reputation value for dishonest service providers is much higher than that of honest providers, and his result is consistent with the MAE result shown in [5]. The MAE reputation for non-Sybil attack is generally lower than that of the corresponding Sybil attack; the main reason is that the number of fair ratings on honest and dishonest providers is decreased. For MET, this makes the trust network become a sparse network, hence, it is hard to build up the level of trust in the advisor. Sybil Camouflage is an exception



(a) CRM vs. Sybil Cam (b) MET vs. Sybil Cam (c) Honest duopoly provider (d) Dishonest duopoly provider

Figure 3. Dynamic Reputation. (a-b): Transaction of Reputation Model under Attack; (c-d): MAE of Duopoly Provider Reputation under Sybil WW.

Table IV
Robustness of Reputation Models vs. Attacks on Dynamic Reputation.

Models	Constant	Camouflage	Whitewashing
MET	0.521±0.067	0.528±0.086	0.523±0.062
CRM	**0.564±0.065**	**0.561±0.052**	**0.575±0.052**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.517±0.089	0.485±0.144	0.485±0.130
CRM	**0.570±0.082**	**0.557±0.085**	**0.561±0.104**

* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

Table V
Mean Absolute Error (MAE) of Reputation Estimation for Honest Duopoly Service on Dynamic Reputation.

Models	Constant	Camouflage	Whitewashing
MET	0.007±0.002	0.007±0.002	0.008±0.002
CRM	**0.007±0.002**	**0.008±0.003**	**0.007±0.003**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.012±0.005	0.014±0.006	0.012±0.005
CRM	**0.013±0.005**	**0.012±0.005**	**0.013±0.005**

* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

Table VI
Mean Absolute Error (MAE) of Reputation Estimation for Dishonest Duopoly Service on Dynamic Reputation.

Models	Constant	Camouflage	Whitewashing
MET	0.103±0.024	0.101±0.031	0.102±0.024
CRM	**0.063±0.021**	**0.064±0.018**	**0.060±0.018**
Models	Sybil	Sybil Cam*	Sybil WW*
MET	0.122±0.034	0.136±0.050	0.132±0.047
CRM	**0.068±0.029**	**0.073±0.029**	**0.074±0.039**

* Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

as the malicious rater would like to subvert the reputation of a provider. It first pretends to be a fair rater and tries to gain a high level of trust, and the fair ratings help the other consumers to select the fair service provider. Hence, the MAE reputation is not sensitive to Sybil attack.

The daily MAE reputation shown in Figures 2(c) and 2(d) reveals that the proposed model converges to the true reputation value faster than MET. When no honest consumers interact with the provider, the estimated reputation value can not be calculated. Therefore, at the very beginning, the MEA reputation value remains at 0.0.

D. Robustness comparison on dynamic reputation

The reputation of a service provider is always dynamic and changes with time. At some point, the provider may update its service quality by offering a better service to consumers. A popular adopted dynamic reputation changes model is pairwise model. The service providers change their strategies in halfway, either from high reputation value to low in order to rip off the attained reputation [3], [17]. Or they learn from their previous mistakes and ameliorate their behavior accordingly [18], [19], [20]. In this subsection, we conduct further experiments to compare the dynamic adaption ability of reputation models. The actual reputation value for a dishonest service provider is updated to 0.9 at day 50. One problem is that all honest consumers know that the honest provider has the higher reputation before day 50. Consequently, all of them will select the honest provider for interaction, there is no chance of discovering the emergence of a potentially good provider. We force the consumer to interact with other service providers by setting a random service selection probability, e_i, of 0.1, it represents a balance between exploration and exploitation. In order to give more time for the model to adapt to the changes, the experiments were extended to 200 days. As the robustness function defined before can not be directly applied to determine the actual reputation value of the dishonest provider (updated to 0.9), we update the definition of robustness as follows:

$$\mathcal{R}(M, Atk) = \frac{Tran(S_i)}{C_H \times Days \times Ratio} \tag{14}$$

where $Tran(S_i)$ is the transaction volume of duopoly provider S_i with higher actually reputation value. Each experiment was conducted 50 times, the averaged results are listed in Table IV. Bold font indicates the best value. They show that CRM outperforms MET on all attacks. As in Figure 2, for static reputation, the robustness of CRM increases rapidly. However, when the dishonest provider updates its quality, according to equation 14, the robustness value of CRM should be smaller than that of MET because robustness is proportional to the transaction volume of duopoly provider with higher reputation value. Further detailed experiments examined the daily change in robustness. Figure 4 shows that CRM gain rapidly increases and when the dishonest duopoly updates its reputation value, its (CRM) robustness value is smaller than that of MET. These CRM characteristics are reasonable since at first more honest consumers interact with the honest duopoly provider, and fewer interact with the dishonest duopoly provider as the actual reputation value can be observed publicly. This process is confirmed in Figures 3(a) and 3(b). In the figure, we can observe that CRM adapts faster than MET as at day 70 the transaction slope of the higher reputation provider (previously dishonest duopoly provider) is larger than that of the lower reputation provider (honest duopoly provider). The fast gain metric makes the model more accurate than the state-of-art model according to the reputation evaluation model MACAU [21].

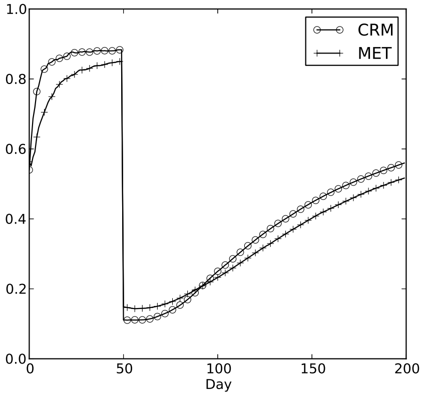


Figure 4. Robustness of Reputation Models on Dynamic Reputation.

E. MAE reputation comparison on dynamic reputation

As the reputation value of the dishonest duopoly provider is updated halfway, its MAE reputation value is obviously increased, which can be observed in Figure 3(d). This is made clear by comparing Table VI with III, the quality improvement made by the dishonest duopoly provider nearly doubles its MAE reputation value compared it the value in the static reputation situation. While CRM can reduce the MAE reputation value to nearly half that of the MET model, the actual reputation value detected by the honest cluster is learnt faster than with the MET model. This makes CRM adapt to the changes more quickly.

For honest consumers, the two models generate almost the same results. Given a 200 day period, the two models both converge to the actual reputation value but with different convergence rates. MET is slower than CRM because the different trustworthiness networks must communicate with each other and learn from the information shared, while in CRM, the information is public and every consumer can use the estimated reputation value to make rational decisions. The time evolved MAE Figure 3(c) for the honest duopoly provider is consistent with the data in Table V.

V. Analysis

All attacks are performed to benefit the dishonest provider. The main objective is to increase transaction volume. We analyze the potential profit and the corresponding cost of each attack. Assume we have a service platform with M service providers and N service consumers, each consumer must pay μ_{S_i} to access the service from provider S_i. Service provider S_j intends to launch the unfair attack on S_i. The corresponding profit from an interaction with $S_j (j \neq i)$ is denoted τ_{S_j}, where $\tau_{S_j} < \mu_{S_i}$. For CRM, the most effective attack is Sybil Camouflage attack. For whitewashing attack, it is hard to establish the trust of a service provider as the fitness function is admissible. In order to launch a successful attack in the clustering-based model CRM, the dishonest raters should keep their rating ratio ρ the same as the honest consumers. That is, transactions are not rated at the rate of $1 - \rho$. The total loss for the attack is $\tau_{S_i} \times (1 - \rho)$. Assume new consumers following the Poisson distribution with dynamic rate $\lambda(t)$ at time t [22], then the number of new consumers is $\int_1^t \lambda(x)dx$. Assume all the new consumers are misled by the dishonest provider S_j because of the successful attack. Profit is attained if the following inequalities are satisfied:

$$\int_1^t \lambda(x)dx \times \tau_{S_j} > \mu_{S_i} \times (N_d/\rho_d) \tag{15}$$

$$N_d > N_h, and\ \rho_d \approx \rho_h \tag{16}$$

where N_d, ρ_d is the number of dishonest consumers and their rating ratio, while N_h, ρ_h is for honest consumers. Inequality 16 holds because of Sybil attack. Inequality 15 is to ensure the provider achieves profit from the attack. Conclusions can be made as follows:

- Decreasing rating ratio ρ increases attack cost and thus suppresses attack likelihood.
- The minimized number of new interactions required to ensure profit is: $(\mu_{S_i} \times (N_d/\rho_d))/\tau_{S_j}$. If value τ_{S_j}/μ_{S_i} can be viewed as the profit ratio of one type of service, we can derive the more profitable of one type of service (larger τ_{S_j}/μ_{S_i}), the more worthy to launch the attack. The conclusion is consistent with economic phenomenon.

VI. Conclusion

In this paper, we proposed a clustering-based reputation model that can resist various types of attacks. The clustering approach is based on the rating ratios of consumers, the honest consumers have no incentive to rate each transaction. Dishonest consumers, however, will utilize every transaction to give an unfair rating to subvert the rating-based reputation system. The proposed model first classifies the ratings into clusters and detects the honest cluster based on the rating ratio of the cluster. It aggregates the reputation values from the honest customers by harnessing the Dirichlet distribution. Simulations showed that our model is more robust than the state-of-art model and its reputation estimates have low mean absolute error. Finally, as it is impossible to totally prevent unfair attacks, we conducted a preliminary analysis of the conditions under which it is worthwhile to attack our proposed reputation model.

Acknowledgment

This research was partly supported by a Grant-in-Aid for Scientific Research (S) (24220002, 2012-2016) from Japan Society for Promotion of Science (JSPS).

References

[1] S. Goto, Y. Murakami, and T. Ishida., "Reputation-based selection of language services." in *Services Computing (SCC), 2011 IEEE International Conference on*, 2011, pp. 330–337.

[2] W. Qiu, Z. Zheng, X. Wang, X. Yang, and M. R. Lyu, "Reputation-aware qos value prediction of web services," in *Services Computing (SCC), 2013 IEEE International Conference on*, 2013, pp. 41–48.

[3] S. Wang, Z. Zheng, Q. Sun, H. Zou, and F. Yang, "Evaluating feedback ratings for measuring reputation of web services," in *Services Computing (SCC), 2011 IEEE International Conference on*, 2011, pp. 192–199.

[4] S. Liu, J. Zhang, C. Miao, Y.-L. Theng, and A. C. Kot, "i-club: An integrated clustering-based approach to improve the robustness of reputation systems," in *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 1151–1152.

[5] S. Jiang, J. Zhang, and Y.-S. Ong, "An evolutionary model for constructing robust trust networks," in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 813–820.

[6] A. Jøsang and R. Ismail, "The beta reputation system." in *Proceedings of the 15th bled electronic commerce conference*, 2002, pp. 41–55.

[7] W. Teacy, J. Patel, N. Jennings, and M. Luck, "Travos: Trust and reputation in the context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183–198, 2006.

[8] X. Wang, L. Liu, and J. Su, "Rlm: A general model for trust representation and aggregation," *Services Computing, IEEE Transactions on*, vol. 5, no. 1, pp. 131–143, 2012.

[9] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in *Proceedings of the 2nd ACM conference on Electronic commerce*. ACM, 2000, pp. 150–157.

[10] L. Cabral and A. Horiaçsu, "The dynamics of seller reputation: Evidence from ebay*," *The Journal of Industrial Economics*, vol. 58, no. 1, pp. 54–78, 2010.

[11] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, "Reputation systems," *Communications of the ACM*, vol. 43, no. 12, pp. 45–48, 2000.

[12] A. Jøsang, "Robustness of trust and reputation systems: Does it matter?" in *Trust Management VI*. Springer, 2012, pp. 253–262.

[13] L. Zhang, S. Jiang, J. Zhang, and W. K. Ng, "Robustness of trust models and combinations for handling unfair ratings," in *Trust Management VI*. Springer, 2012, pp. 36–51.

[14] W. Teacy, M. Luck, A. Rogers, and N. R. Jennings, "An efficient and versatile approach to trust and reputation using hierarchical bayesian modelling," *Artificial Intelligence*, vol. 193, pp. 149–185, 2012.

[15] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[16] A. A. Irissappane, S. Jiang, and J. Zhang, "Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attack." in *UMAP Workshops*, vol. 12, 2012.

[17] L. Xiong and L. Liu., "Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities." *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, no. 7, pp. 843–857, 2004.

[18] M. Zaki and A. Bouguettaya, "Rateweb: Reputation assessment for trust establishment among web services." *The VLDB Journal – The International Journal on Very Large Data Bases*, vol. 18, no. 4, pp. 885 – 911, 2009.

[19] G. Vogiatzis, I. MacGillivray, and M. Chli., "A probabilistic model for trust and reputation." in *In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, vol. 1-Vol.1. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 225 – 232.

[20] Y. Wu, C. Yan, Z. Ding, G. Liu, P. Wang, C. Jiang, and M. Zhou., "A novel method for calculating service reputation," *Automation Science and Engineering, IEEE Transactions on*, vol. 10, no. 3, pp. 634–642, 2013.

[21] C. J. Hazard and M. P. Singh, "Macau: a basis for evaluating reputation systems," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. AAAI Press, 2013, pp. 191–197.

[22] B. Khosravifar, J. Bentahar, and A. Moazin, "Analyzing the relationships between some parameters of web services reputation," in *Web Services (ICWS), 2010 IEEE International Conference on*. IEEE, 2010, pp. 329–336.