A whitelist and blacklist-based co-evolutionary strategy for defending against multifarious trust attacks

Shujuan Ji^{1,2} · Haiyan Ma¹ · Yongquan Liang¹ · Hofung Leung³ · Chunjin Zhang⁴

Published online: 13 May 2017

© The Author(s) 2017, corrected publication May/2018

Abstract With electronic commerce becoming increasingly popular, the problems of trust have become one of the main challenges in the development of electronic commerce. Although various mechanisms have been adopted to guarantee trust between customers and sellers (or platforms), trust and reputation systems are still frequently attacked by deceptive, collusive, or strategic agents. Therefore, it is difficult to keep these systems robust. It has been mentioned that a combined usage of both trust and distrust propagation can lead to better results. However, little work has been known to realize this insight successfully. Besides, literatures either use a social network with trust/distrust information or use one advisor list in evaluating all sellers, which leads to the lack of pertinence and inaccuracy of evaluation. This paper proposes a defending strategy called *WBCEA*, in which, each buyer agent is modeled with two attributes (i.e., the trustworthy facet and the untrustworthy facet) and two lists (i.e., the whitelist and the

blacklist). Based on the social network that are constructed and maintained according to its whitelist and blacklist, the honest buyer agent can find trustable buyers and evaluate the candidate sellers according to its own experience and ratings of trustable buyers. Experiments are designed and implemented to verify the accuracy and robustness of this strategy. Results show that our strategy outperforms existing ones, especially when majority of buyers are dishonest in the electronic market.

Keywords Trust and reputation systems · Attack defense · Whitelist and blacklist · Co-evolutionary algorithm

1 Introduction

As electronic commerce becomes increasingly popular, more and more people are involved. The problems of trust have become the main challenges in the development of electronic commerce. In multi-agent based electronic commerce, self-interested agents may be deceptive, collusive or strategic. Unfair rating attacks (such as the collusive unfair ratings, *Sybil*, *Camouflage*, *Whitewashing*, and discrimination attacks [7, 23]) from dishonest reviewer render reputation systems' vulnerable to mislead buyers to transact with dishonest sellers. Dishonest reviewers may also employ sophisticated attacking strategies (such as a combination of various unfair rating strategies) to avoid being detected. Lots of trust models have been proposed to cope with unfair/false ratings. However, these models are not completely robust against various strategic attacks, i.e., they have limitations in defending against some kinds of strategic attacks. To address such problems, we design a new robust algorithm called *WBCEA* for improving intelligent agents' capabilities in accurately estimating the trustworthiness of

✉ Shujuan Ji
jane_ji2003@aliyun.com

¹ Shandong Provincial Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science and Technology, Qingdao, China

² Shandong Provincial Key Laboratory of Novel Distributed Computer Software Technology, Shandong Normal University, Jinan, China

³ Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

⁴ Center of Network, Shandong University of Science and Technology, Qingdao, China

sellers under various types of attacks and further reducing the risk of purchase.

The main steps of *WBCEA* are as follows. First, based on historical ratings, the defending buyer agent should evaluate the trustworthiness of reviewers who rated these recommended sellers according to its own experiences or the trustable buyers' experiences (the trustable buyers are evolved based on the buyer agent's whitelist and blacklist). And then, the defending buyer generates a list of most trustworthy reviewers as advisors (based on their trustworthiness) for each recommended seller agent. Thirdly, it evaluates each seller's trustworthiness according to its own experience and advisors' ratings, and selects the most trustworthy seller as trading partner. Finally, the defending buyer agent updates its own whitelist and blacklist.

In contrast to existing strategy, the novel features of this strategy are as follows. First, considering that each buyer agent has both a trustworthy facet and an untrustworthy facet, *WBCEA* considers both trustworthy facet and untrustworthy facet of each reviewer. Moreover, each buyer agent maintains two lists (i.e., a whitelist and a blacklist) to keep track of the most trustworthy advisors (i.e. reviewers) and most untrustworthy advisors that are evolved according to its own experience. Secondly, based on blacklist and whitelist, a customized optimal advisors list is generated for evaluating each recommended seller, which is similar to *PEALGA* while different from *MET* (which adopts one advisor list for evaluating all sellers). However, the agents in each optimal advisor lists of *PEALGA* are selected from all the buyers in the system, while the agents in each optimal advisor lists of *WBCEA* are selected from those buyers who have traded with the recommended seller. Therefore, the latter optimal advisor lists are targeted more. Thirdly, *WBCEA* considers both trust and distrust information in social network of buyers. That is different from *PEALGA*, which only considers trust information.

The rest of this paper is developed as follows. Section 2 reviews related literature. Section 3 gives a framework for multi-agent based electronic commerce platform. Section 4 illustrates the *WBCEA* strategy in detail. Section 5 verifies the performance of our approach using experiments. Section 6 concludes this paper with future work directions.

2 Literature review

The aim of this section is to review the models which are designed for detecting and defending against malicious agents. Since our work is based on a local (buyer's) view, not on the global (electronic commerce platform) view, we only review models designed from buyers' viewpoints. In general, the defending models can be divided into two categories, i.e., trust-based approaches, and the trust

and distrust-based approaches. The following paragraphs illustrate these two kinds of models in detail.

It is widely agreed that trust means the confidence that one or many entities behave as they expected [16]. Based on trust information, many defending models such as *BRS* [19], *iCLUB* [11, 12], *TRAVOS* [18], *ReferralChain* [20, 21], *Personalized* [22], *MET* [6], *PEALGA* [5] have been designed. These models can effectively defend against some kinds of attacks. However, *BRS* becomes inefficient and *iCLUB* is unstable when the majority of buyers are dishonest because both of them employ the "majority-rule". When dishonest advisors adopt shifty attacks, *TRAVOS* does not work well because it assumes that each advisor's rating behavior is consistent. *ReferralChain* sets 1 to the initial trust of each new buyer (advisor) which provides a chance for dishonest advisors to abuse the initial trust (i.e., *Whitewashing*). *Personalized* model is vulnerable when buyers have insufficient experiences with advisors and the majority of advisors are dishonest (i.e., combination of *Sybil* and *Whitewashing*). *MET* evolves one advisor list, which is not necessarily suitable enough for estimating the trustworthiness of all the sellers. *PEALGA* pre-evolves a customized advisor list for evaluating each candidate seller. However, this algorithm still only considers trust information between buyers.

Distrust is recognized to play an equally important role as trust [3]. Though some attack defending models such as *GBR* [13] and *Multi-faceted* [4] are designed, the investigation of utilizing distrust is still in its infancy [14, 15]. *GBR* is proposed to combat web spam employing the "majority-rule". Therefore, it has strong bias towards seed pages if a small seed set is used. However, it is time consuming to get a large seed set artificially. Therefore, it is difficult for researchers to tradeoff between the number of seed pages and time complexity. *Multi-faceted* considers both interpersonal and impersonal aspects that may bring in redundant and even noisy information. Moreover, it cannot be very prominent in the fight against most kinds of attacks, especially *Sybil* and *Whitewashing* attack. Both *GBR* and *Multi-faceted* only generate one advisor list for evaluating all sellers, which leads to problems such as lack of pertinence or inaccuracy of evaluation.

As links among users in a social network are consist of trust as well as distrust connections, both of these information can be used in the design of defending strategies. However, existing models have some shortcomings. For example, the performance of *GBR* depends on training data (i.e., the number of seed pages). In contrast to *GBR*, our strategy *WBCEA* also relies on data as its performance can only keep stable when enough transaction experience is accumulated. Of course, we can use the rolling incremental updated transaction and rating data as input of our strategy, which are easier to acquire compared to the seed pages in *GBR*.

Fig. 1 A framework for electronic commerce platform with advisor mechanism

3 A framework for multi-agent-based electronic commerce platform whitelist and blacklist mechanism

We propose a framework for multi-agent-based electronic commerce platform (see Fig. 1). In this framework, there are three kinds of agents in the electronic market, i.e., buyer agents, seller agents, and search agents. Each buyer agent maintains a whitelist (which is used to store those buyers who are trustworthy), a blacklist (which is used to store those buyers who are untrustworthy) and a list of historical transaction records (which is used to store those sellers who have been traded with it). The function of search agents is to help buyers select some sellers who can satisfy their purchase demand, and simultaneously list each seller's reviewers (see the central box of Fig. 1).

Once getting the recommended sellers returned by search agent, the honest buyers who defending various attacks will select a suitable seller as trading partner based on their trustworthiness. So, rational defending buyers should evaluate each seller's trustworthiness according to its own

experience and reviewers' advises. As some reviewers are selfish and dishonest by giving unfair ratings, therefore, the defending buyer should evaluate each reviewer's trustworthiness before referring their ratings. For each reviewer, if the honest buyer and the reviewer once traded with same sellers, the buyer can evaluate the reviewers' trustworthiness according to its own experience directly. Otherwise, the honest buyer will seek the reviewers it trusts for help. These trusted reviewers can be found by the social network that are constructed based on whitelist and blacklist of the honest buyer. Once getting reviewers' trustworthiness, the honest buyer will generate a customized optimal advisors list from reviewers for evaluating each recommended seller. After selecting trading seller and getting the goods or enjoying the service, the honest buyer will rate the trading seller according to its subjective feel. Simultaneously, the defending buyer's whitelist and blacklist should be updated according to this experience of trading. The symbols and their meanings of this framework are summarized in Table 1.

In order to explicitly define the research scope and background of this paper, we assume that the defense agents

Table 1 Symbols used in the framework and their meanings

Symbols	Meaning of symbols	Symbols	Meaning of symbols
$B = \{b_i i = 1, \dots, m\}$	The set of buyers	r_{b_i, s_j}	The rating of b_i to s_j
$S = \{s_j j = 1, \dots, n\}$	The set of sellers	$B_{b_i}^T$	The set of those buyers who are trusted by b_i
$B_{s_j}^H$	The set of those buyers who have interacted with s_j	$B_{b_i}^D$	The set of those buyers who are distrusted by b_i
$S_{b_i}^H$	The set of those sellers who have interacted with b_i	$R_{b_i, T}(b_j)$	The trust value of b_i to b_j
$S^{candidate}$	The set of those sellers who are recommended by search agent	$R_{b_i, D}(b_j)$	The distrust value of b_i to b_j
TN_{b_i}	The social network of b_i	$STD_{b_i}(b_j)$	The synthetic value of b_i to b_j
WL_{b_i}	The whitelist of b_i	$A_{b_i}^{s_j}$	The best advisor list of b_i to b_j
BL_{b_i}	The blacklist of b_i		

in the electronic commerce platform with whitelist and blacklist mechanism follows the following assumptions.

Assumption 1 Buyer agents pay more attention to the ratings that are given by the advisors who have similar opinions to their own. Some inexperience buyer agents often need to consider other buyer agents' advices. Similar to the experience of humans, buyer agents are more incline to refer to the advisors who have similar opinions and believe that agents who often give positive/negative ratings belong to the same category.

Assumption 2 Buyer agents' acceptance toward other reviewers' ratings decrease over time. The more recent the rating time, the more accurate the rating can reflect current trustworthiness of the seller and the more significance it is to the seller's reputation prediction.

Assumption 3 We assume that the buyer agents are not competitive in general and are willing to share their whitelist and blacklist with others. This is quite common in major e-marketplaces and travel agent portals.

Assumption 4 We rule out the influence of various prices in the selection of trading seller in this study as we concentrate on the effect of trustworthiness computation under the environment that the prices of the provided products or services are similar.

4 The whitelist and blacklist co-evolutionary strategy

Every time after search agent returning the recommended sellers $S^{candidate}$ and each seller's recent reviews $B_{S^{candidate}}^H$ to the honest buyer b_i , it will adopt the whitelist and blacklist co-evolutionary defending strategy (abbr. WBCEA) and implement following steps.

- (1) Based on whitelist and blacklist, the honest buyer b_i constructs its social network TN_{b_i} . And then, through the propagation of trust and distrust in the social network TN_{b_i} , buyer b_i tries to find the reviewers (i.e. buyers) $B_{b_i}^T$ that can be trusted by it (see Algorithm 1).
- (2) For each reviewer who rates the candidate seller $s_j^{candidate}$ in the $B_{S^{candidate}}^H$, if it also traded with some sellers who traded with buyer b_i , b_i can evaluate the reviewer's trustworthiness according to its own experience directly. Otherwise, b_i will seek trustworthy buyers in $B_{b_i}^T$ (that is gained from algorithm 1) for reference. Based on the synthesized trustworthiness of each reviewer, the defending buyer b_i further

generates an optimal advisors list $A_{b_i}^{s_j^{candidate}}$ for evaluating each recommended seller $s_j^{candidate} \in S^{candidate}$ (see Algorithm 2).

- (3) The honest buyer b_i evaluates each candidate seller $s_j^{candidate}$'s trustworthiness according to its own experience and optimal advisors' advices based on Algorithm SRCA.
- (4) The honest buyer b_i rates the selected seller after the transaction, and updates its' whitelist and blacklist according to this experience of trading (see Algorithm 3).

The following sub-sections illustrate these steps in detail.

4.1 The trust network construction algorithm

An agent's trust network is constructed based on its social network. In this subsection, we first explain the concepts of social network, distance, and layer. And then, we illustrate the principle of finding trustable buyers through an example intuitively. Finally, the trust network construction algorithm is given.

Definition 1 A social network of a buyer are networks that are constructed based on its whitelist and blacklist. It is composed of many vertexes and directed edges. Each vertex represents a buyer agent and each directed edge represents a trust or distrust relationship between the two connected agents.

The trust relationship is represented by solid arrow labeled with "trust" (see Fig. 2a), and the distrust relationship is represented by dotted arrow labeled with "distrust" (see Fig. 2b). The agent that connects arrowhead is named "trustee" (who receives trust statement), and the one who connects tail is called "trustor" (who specifies trust statement). In constructing a social network for an honest buyer (e.g., b_i), all the members in its whitelist and blacklist are added to its social network by solid and dotted arrows respectively. We assume in this paper that the buyer agents are willing to share their whitelists and blacklists with other buyer agents (Assumption 3). We note that the members in b_i 's whitelist and blacklist also have their respective whitelists and blacklists. Hence, we can say that an agent trusts another agent if the latter is in the whitelist of the

Fig. 2 Representation of trust and distrust relationship between agents

former, or an agent distrusts another agent if the latter is in the blacklist of the former. Meanwhile, the six degrees of separation theory asserts that “any two persons in the world can be connected by at most six persons”. Based on this theory, we construct a social network for an agent b_i by including all other agents that have at most 6 trust or distrust relation with b_i . From an honest buyer’s view, the trustworthiness and untrustworthiness of a buyer in its social network is determined by the distance that the latter buyer to the honest buyer in the trust/distrust chains. Example 1 explains these concepts of distance and layer.

Example 1 Supposing that there is a sample social network with trust and distrust labels (see Fig. 3a). In this figure, b_i is an honest buyer. b_i ’s whitelist WL_{b_i} is $\{b_a, b_b, b_c\}$, b_a ’s whitelist WL_{b_a} is $\{b_d, b_e\}$, b_f is in b_b ’s whitelist and in b_e ’s blacklist simultaneously, b_g is in b_d ’s whitelist and in b_e ’s blacklist simultaneously, and b_h is in b_f ’s whitelist and b_c ’s blacklist simultaneously. The whitelist or blacklist of each buyer that is not listed above is empty. Obviously, we can find a chain $\langle b_i, b_a, b_e, b_g \rangle$ from Fig. 3a. According to the basics of graph theory, the distance from b_i to itself is zero, and the distance from b_i to b_a is 1. Therefore, the distance from b_i to b_g in chain $\langle b_i, b_a, b_e, b_g \rangle$ is 3. In the social network of b_i , the agents who have equal distance from b_i are located in the same layer. In Fig. 3a, if b_i is located in the first layer, b_a, b_b and b_c are located in the second layer. Moreover, an agent may belong to multiple chains with similar or different distance (or layer). For another example, b_f is located in the third layer of chain $\langle b_i, b_b, b_f \rangle$ and in the fourth layer of chain $\langle b_i, b_a, b_e, b_f \rangle$ simultaneously. Similarly, b_g is located in the fourth layer of chains $\langle b_i, b_a, b_d, b_g \rangle$ and $\langle b_i, b_a, b_e, b_g \rangle$ respectively.

Propagation of trust and distrust along social network chains is similar to the “word of mouth” propagation of information for humans [1]. Josang et al. [8] concluded

that “trust will be weakening in the chain of propagation” and “in the case where an agent receives conflicting recommended trust, e.g. both trust and distrust, it needs some methods for combining these conflicting recommendations”. Based on these conclusions, in this paper, we define the following rules to determine the trustworthiness or untrustworthiness of an agent in social network. These rules consider the fact that “trust will be weakening in the chain of propagation”, therefore, it is believed the shorter the propagation chain, the less the weakening effect. Therefore, “layer” (or distance to the honest agent) of agents is considered in these rules. Example 2 explains these two rules intuitively.

Rule 1: If an agent is trusted and distrusted by (i.e., in the whitelist and the blacklist of) different buyers who are located in the same layer simultaneously, this agent’s trustworthiness is considered to be uncertain.

Rule 2: If an agent is trusted and distrusted by (i.e., in the whitelist and the blacklist of) different agents who are located in different layer, the agent’s trustworthiness can be judged based on the trust/distrust of its upper layer agents. If the upper agent whose layer is smallest trust this agent, this agent is considered as trustworthy, otherwise, it is believed to be not trustworthy.

Example 2 In Fig. 3a, b_g is trusted by b_d and distrusted by b_e simultaneously, b_d and b_e are located in the same layer (i.e., 3). Therefore, b_g is located in the same layer of chains $\langle b_i, b_a, b_d, b_g \rangle$ and $\langle b_i, b_a, b_e, b_g \rangle$. Moreover, b_g is trusted by b_d and distrusted by b_e simultaneously. Hence, according to Rule 1, b_i cannot give certain judge whether b_g is trustworthy or not. For another example, in Fig. 3a, b_f is located in the third layer and the forth layer of chains $\langle b_i, b_b, b_f \rangle$ and $\langle b_i, b_a, b_e, b_f \rangle$ respectively. Moreover, b_b trust b_f , while b_e distrust b_f . As b_f ’s upper layer agents,

Fig. 3 An example of social network and trust network

b_b and b_e 's layer is 2 and 3 respectively. As b_b 's layer is the smallest one of b_f 's upper layer agents, therefore, according to Rule 2, b_f is considered as trustworthy based on b_b 's trust.

Algorithm 1 implements above ideas about how to construct a trust network by using two auxiliary queues, i.e., Q_t and Q_d . Q_t is used to temporarily store agents who are in whitelist, and Q_d is used to temporarily store agents who are in blacklist. Moreover, we define a variable named

$depthLimit$ to represent the upper bound of chain length (i.e. 6) in the resulting trust network. The main steps of this algorithm are as follows. First, the queues Q_t and Q_d (see step (1) in Algorithm 1) are initialized. Secondly, we find the trustworthy buyers $B_{b_i}^T$ for b_i (see step (3-5) in Algorithm 1). Thirdly, we find the buyers that are not trustworthy (denoted as $B_{b_i}^D$) for buyer b_i (see step (6-8) in Algorithm 1). Finally, the queues Q_t and Q_d (see step (9-11) in Algorithm 1) updated.

Algorithm 1 Trust network construction algorithm

Input: (1) The set of buyers $B = \{b_i | i = 1, \dots, m\}$; (2) The whitelist and blacklist of buyer b_i ; (3) Three assistant empty queues Q_t, Q_d, T_{b_i} ; (4) The upper bound of chain length, i.e., $depthLimit$;

Output: The set of buyers who are trusted by b_i , denoted as $B_{b_i}^T$

Process:

- (1) Add the agents who are in b_i 's whitelist into Q_t . Add the agents who are in b_i 's blacklist into Q_d ; $T_{b_i} = \emptyset$;
- (2) while($Q_t \neq \text{NULL} \ \&\& \ \text{Depth} < depthLimit$) do
- (3) for each b_k in Q_t
- (4) if($b_k \notin Q_d \ \&\& \ Q_d \notin B_{b_i}^T \ \&\& \ b_k \notin B_{b_i}^D$)
- (5) add b_k into $B_{b_i}^T$, add b_k into T_{b_i} ;
- (6) for each b_k in Q_d
- (7) if($b_k \notin Q_t \ \&\& \ Q_d \notin B_{b_i}^T \ \&\& \ b_k \notin B_{b_i}^D$)
- (8) add b_k into $B_{b_i}^D$;
- (9) Clear Q_t and Q_d ;
- (10) Add the agents who in the whitelist of those agent who are located in T_{b_i} into Q_t ;
- (11) Add the agents who in the blacklist of those agent who are located in T_{b_i} into Q_d ;
- (12) clear T_{b_i} ;
- (13) Depth++
- (14) Return $B_{b_i}^T$

If we execute algorithm 1 over agent b_i 's social network shown in Fig. 3a, we can get the resulting trust network shown in Fig. 3b. From this figure, we can see that b_a, b_b, b_c, b_d, b_e , and b_f are the agents who are trusted by b_i . In comparison, according to the trust network construction method given in previous work [5, 6, 20, 21] that do not consider “distrust” labels, all the agents $b_a, b_b, b_c, b_d, b_e, b_f, b_g$ and b_h will be added into the trust network. Therefore, algorithm 1 further purifies the trust network using “distrust” information.

4.2 The optimal advisor lists generation algorithm

As there may be many sellers who satisfy the buyer agents' requirement, only one fixed advisor list may lack pertinence in evaluating all of the recommended sellers' trustworthiness. That is because some members in a fixed advisor list may be experienced in trading with one seller, while inexperienced in trading with others, therefore, their advices about

the latter sellers will be inaccuracy. Similar to the idea of finding a most suitable advisor list (called *optimal advisor list*) for evaluating a recommended seller we have proposed in literature [5], this paper also tries to find an *optimal advisor list* for each seller. Algorithm 2 shows the main idea of the optimal advisor lists generation algorithm based on whitelist and blacklist, which is composed of four steps. First, buyer b_i calculates pair-wised similarities between the itself and any buyer that rated $S^{candidate}$ (e.g. b_k) according to equation (1) or (2) (see step (3)-(6) in Algorithm 2). Psychology research [10] showed that people have both the trustworthy aspect and the untrustworthy aspect. Based on their results, we randomly set two initial values (denoted as $R_{b_i,T}(b_k) \in [0, 1]$ and $R_{b_i,D}(b_k) \in [0, 1]$ respectively) to each buyer agent for denoting these two aspects. Moreover, we specify that buyer b_i updates b_k 's trustworthy aspect and untrustworthy aspect according to equation (3) (see step (7) in Algorithm 2). Thirdly, considering the trustworthy aspect and the untrustworthy aspect simultaneously, the

synthesized trustworthiness of b_k is calculated according to equation (4) (see step (8) in Algorithm 2). Finally, buyer b_i

updates and generates the optimal advisor lists for evaluating the recommended seller (see step (10) in Algorithm 2).

Algorithm 2 optimal advisor lists generation algorithm

Input: (1) The honest buyer b_i ; (2) The set of sellers who are recommended by search agent $S^{candidate}$; (3) reviewers who rated each recommended seller $B_{S^{candidate}}^H$;

Output: Each recommended seller's optimal advisor lists $A_{b_i}^{S^{candidate}}$;

Process:

- (1) for each $s_j^{candidate}$ in $S^{candidate}$
- (2) for each buyer b_k in $B_{S^{candidate}}^H$
- (3) if (b_i and b_k have common transactions)
- (4) Calculate the similarity between b_i and b_k according to formula (1);
- (5) else
- (6) Calculate the similarity between b_i and $B_{b_i}^T$ according to formula (2);
- (7) Update $R_{b_i,T}(b_k)$ and $R_{b_i,D}(b_k)$ according to formula (3) and (4);
- (8) Calculate the synthetic value $STR_{b_i}(b_k)$ according to formula (5);
- (9) end for
- (10) Add the buyers who have high trustworthiness (bigger than 0.6) into $A_{b_i}^{S^{candidate}}$;
- (11) end for
- (12) Return Each recommended seller's optimal advisor lists $A_{b_i}^{S^{candidate}}$.

The following definitions illustrate the formulas used in this algorithm. It should be noted that we do not consider b_k 's layer (or distance) in social network in definitions 2-4. That is caused by two reasons. One is that, for any buyer b_k in $B_{S^{candidate}}^H$ while not in $B_{b_i}^T$, there is not any relationship between b_k and b_i , let alone layers. That is to say, layers do not exist in all the similarity calculation cases. The other reason is that, similarity of trustor and trustee's viewpoints affects their trust, while it is not necessarily true inversely. For example, the degree that a trustor trusts a trustee is affected by factors such as the familiarity between the trustor and the trustee, their interaction frequency, the consistency/similarity of their viewpoints, the number of common friends, and so on. However, the similarity of two persons' behaviors or viewpoints are not necessarily affected by their trust. In real life, a person's behavior or viewpoints may be similar to the ones of distant strangers or even his/her competitors and enemies.

Definition 2 If the defending agent b_i and any reviewer agent b_k once traded with same sellers, the similarity between agent b_i and agent b_k is determined by the ratings they rated these sellers, the average values of their ratings. Equation (1) defined the calculation method.

$$sim_1(b_i, b_k) = \frac{\sum_{s_j \in S_{b_i, b_k}} (r_{b_i, s_j} - \bar{r}_{b_i})(r_{b_k, s_j} - \bar{r}_{b_k})}{\sqrt{\sum_{s_j \in S_{b_i, b_k}} (r_{b_i, s_j} - \bar{r}_{b_i})^2} \sqrt{\sum_{s_j \in S_{b_i, b_k}} (r_{b_k, s_j} - \bar{r}_{b_k})^2}} \quad (1)$$

where, S_{b_i, b_k} represents the set of sellers who once traded with b_i as well as b_k , r_{b_i, s_j} is the rating that b_i rated s_j ($s_j \in S_{b_i, b_k}$), $\bar{r}_{b_i}$ represents the average of ratings that b_i rated its trading partners, r_{b_k, s_j} represents the rating that agent b_k rated s_j ($s_j \in S_{b_i, b_k}$), $\bar{r}_{b_k}$ represents the average of ratings that b_k scored its trading partners.

Definition 3 If defending agent b_i and any reviewer agent b_k did not trade with any same seller in their histories, the similarity between them is determined by characteristics of b_k 's rating and the ratings given by the buyers who are trusted by b_i . Equation (2) defines the calculation method.

$$sim_2(b_i, b_k) = \frac{\sum_{s_j \in S_{b_k}^H} (r_{b_k, s_j} - \bar{r}_{b_k})(\bar{r}_{s_j} - \bar{r})}{\sqrt{\sum_{s_j \in S_{b_k}^H} (r_{b_k, s_j} - \bar{r}_{b_k})^2} \sqrt{\sum_{s_j \in S_{b_k}^H} (\bar{r}_{s_j} - \bar{r})^2}} \quad (2)$$

where, $S_{b_k}^H$ is the sellers that historically traded with b_k , r_{b_k, s_j} is the rating that b_k rated s_j ($s_j \in S_{b_k}^H$), $\bar{r}_{b_k}$ is the average of ratings that agent b_k rated its trading partners, $\bar{r}_{s_j}$ is the average of ratings that $B_{b_i}^T$ (i.e., the set of agents who are trusted by agent b_i) scored s_j , $\bar{r}$ is the average of ratings that all the members in $B_{b_i}^T$ scored the sellers that they traded with.

Definition 4 From defending agent b_i 's viewpoint, the trustworthy aspect and untrustworthy aspect of agent b_k can be updated based on their similarity. Equations (3) and (4) defines the updating method respectively.

$$\begin{cases} R_{b_i,T}(b_k) = R_{b_i,T}(b_k) + R_{b_i,T}(b_k) \times (sim(b_i, b_k) - \omega) \times |1 - \beta_1^{sim(b_i, b_k) - \omega}| \\ R_{b_i,D}(b_k) = R_{b_i,D}(b_k) - R_{b_i,D}(b_k) \times (sim(b_i, b_k) - \omega) \times |1 - \beta_2^{sim(b_i, b_k) - \omega}| \end{cases} \quad (3)$$

$$sim(b_i, b_k) = \begin{cases} (sim_1(b_i, b_k) + 1)/2 & \text{if } b_i \text{ and } b_k \text{ have traded with same sellers} \\ (sim_2(b_i, b_k) + 1)/2 & \text{otherwise} \end{cases} \quad (4)$$

where, $R_{b_i,T}(b_k) \in [0, 1]$ represents the trustworthy aspect of b_k from b_i 's viewpoint, $R_{b_i,D}(b_k) \in [0, 1]$ represents the untrustworthy aspect of b_k from b_i 's viewpoint; $sim(b_i, b_k) \in [0, 1]$ is calculated according to equation (4) (i.e., the similarity between b_i and b_k is the normalized value of $sim_1(b_i, b_k)$ if b_i and b_k have traded with same sellers, otherwise, their similarity is a normalized value of $sim_2(b_i, b_k)$); $\omega \in (0, 1)$ is a classification factor to divide the growth of trustworthy aspect and untrustworthy aspect into positive, negative and zero; β_1 and β_2 are factors to control the increment speed of trust and distrust respectively.

In experiment, $\omega \in (0, 1)$ is set to 0.5. This setting ensures that the trustworthy aspect and untrustworthy aspect of b_k keep unchanged if the similarity between b_i and b_k is 0.5 (which means their similarity is not obviously very large or very small). If the similarity between b_i and b_k is larger than 0.5 (which means that their similarity is obviously very large), the trustworthy/untrustworthy aspect of b_k increases/decreases by a certain amount; otherwise, the trustworthy/untrustworthy aspect of b_k decreases/increases by a certain amount. It is also important to note that factors β_1, β_2 should satisfy the condition $0 < \beta_2 < \beta_1 < 1$. The constraints $\beta_2 < \beta_1$ is set to ensure that the speed of increase in the trustworthy aspect is less than that of the untrustworthy aspect, and the speed of decrease in the trustworthy aspect is greater than that of the untrustworthy aspect. This constraint is consistent with the research result that "people devote more attention to negative information than to positive information" [17].

The synthesized trustworthiness of an agent can be gained by synthesizing its trustworthy aspect and untrustworthy aspect (see Definition 5). In constructing formula of $STD_{b_i}(b_k)$, two thresholds $\theta_1 (0 < \theta_1 < 1)$ and $\theta_2 (0 < \theta_2 < 1)$ are introduced based on perception of humans that: a person is trustable when his/her trustworthy aspect is greatly larger than his/her untrustworthy

facet; a person is not trustable when his/her trustworthy facet is slightly smaller than his/her untrustworthy facet [10]. Therefore, θ_1 should be larger than θ_2 . In experiment, θ_1 and θ_2 are assigned with 0.8 and 0.2 respectively. That is to say, if a person's trustworthy aspect is larger than its untrustworthy aspect by 0.8 (i.e., $R_{b_i,T}(b_k) - R_{b_i,D}(b_k) > \theta_1$, $\theta_1 = 0.8$), his/her synthesized trustworthiness is believed to be 1. If a person's trustworthy aspect is smaller than its untrustworthy aspect by 0.2 (i.e., $R_{b_i,T}(b_k) - R_{b_i,D}(b_k) < -\theta_2$, $\theta_2 = 0.2$), his/her synthesized trustworthiness is believed to be 0. Otherwise, the synthesized trustworthiness of this agent is calculated according to formula $\frac{1}{\theta_1 + \theta_2}(R_{b_i,T}(b_k) - R_{b_i,D}(b_k) + \theta_2)$ (see third case of formula (5)). This formula is constructed by linear connection of the two points $(\theta_1, 1)$ and $(-\theta_2, 0)$. Figure 4 illustrates the construction principle visually. Its horizontal axis is $R_{b_i,T}(b_k) - R_{b_i,D}(b_k)$, and vertical axis is the synthesized trustworthiness of a person.

Definition 5 To facilitate comparison, the trustworthy aspect and untrustworthy aspect of b_k can be synthesized to one value by b_i . Equation (4) defines the synthetization method.

$$STD_{b_i}(b_k) = \begin{cases} 1 & \text{if } R_{b_i,T}(b_k) - R_{b_i,D}(b_k) > \theta_1 \\ 0 & \text{if } R_{b_i,T}(b_k) - R_{b_i,D}(b_k) < -\theta_2 \\ \frac{1}{\theta_1 + \theta_2}(R_{b_i,T}(b_k) - R_{b_i,D}(b_k) + \theta_2) & \text{otherwise} \end{cases} \quad (5)$$

where, $R_{b_i,T}(b_k)$ represents b_k 's trustworthy aspect from b_i 's view, $R_{b_i,D}(b_k)$ represents b_k 's untrustworthiness from b_i 's aspect; $\theta_1 (0 < \theta_1 < 1)$ and $\theta_2 (0 < \theta_2 < 1)$ are two thresholds.

4.3 Seller's reputation evaluation algorithm

To decrease purchase risk, honest buyers often evaluate sellers' reputation first. Therefore, buyer b_i must be endowed

Fig. 4 Construction principle of synthesized trustworthiness

with the ability of evaluating sellers. For accurately evaluating each seller's reputation, an honest buyer comprehensively considers its private trust (which is gained by their own experience) to the seller and the public reputation of the seller (which is calculated according to their optimal advisors' comment). Algorithm 3 illustrates the new seller's reputation calculation algorithm [5]. The main idea of this algorithm is as follows: (1) Buyer b_i first calculates each seller's private trustworthiness according to its own trading experience with this seller (see step (2)-(5) in Algorithm 3). (2) Buyer b_i calculates each seller's public reputation according to the optimal advisor lists obtained from Algorithm 3 (see step (6) in Algorithm 3). (3) The private trustworthiness and the public reputation are combined to get the perceived reputation of each seller (see step (7) in Algorithm 3). The formulas used in this algorithm are defined in the following definitions, which are similar to the ones given by Zhang and Cohen [22] and Ji et al. [5].

Definition 6 b_i 's private trust to seller $s_j^{candidate}$ is calculated according to b_i 's rating to this seller. Formula (6) defines the calculation method [22].

$$R_{b_i, pri}^{s_j^{candidate}} = \frac{\sum_{t=1}^n N_{b_i, pos}^{s_j^{candidate}} \lambda^{t-1} + 1}{\sum_{t=1}^n (N_{b_i, pos}^{s_j^{candidate}} + N_{b_i, neg}^{s_j^{candidate}}) \lambda^{t-1} + 2} \quad (6)$$

where, $N_{b_i, pos}^{s_j^{candidate}}$ represents the number of positive rating that b_i rated to $s_j^{candidate}$, $N_{b_i, neg}^{s_j^{candidate}}$ represents the number of negative rating that b_i rated to $s_j^{candidate}$, λ is a discount factor, t ($t = 1, 2, \dots, n$) is the time windows of rating.

Definition 7 b_i 's public reputation to seller $s_j^{candidate}$ is calculated according to the ratings of $A_{b_i}^{s_j^{candidate}}$. Formula (7) defines the calculation method.

$$R_{b_i, pub}^{s_j^{candidate}} = \frac{\left[\sum_{k=1}^m \sum_{t=1}^n P_{a_k, pos}^{s_j^{candidate}} \lambda^{t-1} \right] + 1}{\left[\sum_{k=1}^m \sum_{t=1}^n (P_{a_k, pos}^{s_j^{candidate}} + P_{a_k, neg}^{s_j^{candidate}}) \lambda^{t-1} \right] + 2} \quad (7)$$

where, $P_{a_k, pos}^{s_j^{candidate}}$ represents the trust weighted probability of positive rating that advisor $a_k \in A_{b_i}^{s_j^{candidate}}$ (which is evolved

in Algorithm 3) rated $s_j^{candidate}$, $P_{a_k, neg}^{s_j^{candidate}}$ represents the trust weighted probability of negative rating that advisor a_k rated $s_j^{candidate}$, and λ is a discount factor, t ($t = 1, 2, \dots, n$) is the time windows of rating. The calculation principle of $P_{a_k, pos}^{s_j^{candidate}}$ and $P_{a_k, neg}^{s_j^{candidate}}$ are presented in formula (8), which are adapted from Zhang and Cohen [22] and the formulas given by Jøsang and Ismail [9] and Yu and Singh [20] based on Dempster-Shafer theory.

$$\begin{cases} P_{b_i, pos}^{a_k} = \frac{2STD_{b_i}(a_k)N_{b_i, pos}^{a_k}}{(1-STD_{b_i}(a_k))(N_{b_i, pos}^{a_k} + N_{b_i, neg}^{a_k}) + 2} \\ P_{b_i, neg}^{a_k} = \frac{2STD_{b_i}(a_k)N_{b_i, neg}^{a_k}}{(1-STD_{b_i}(a_k))(N_{b_i, pos}^{a_k} + N_{b_i, neg}^{a_k}) + 2} \end{cases} \quad (8)$$

where $STD_{b_i}(a_k)$ is a_k 's synthetic trustworthiness estimated by buyer b_i , $N_{b_i, pos}^{a_k}$ represents the number of positive rating that b_i rated to a_k , and $N_{b_i, neg}^{a_k}$ represents the number of negative rating that b_i rated to a_k .

Definition 8 The perceived reputation by buyer b_i of a given seller $s_j^{candidate}$ is the weighted combination of b_i 's private trustworthiness and public reputation given by $A_{b_i}^{s_j^{candidate}}$, as defined by formula (9) [22].

$$R_{b_i}^{s_j^{candidate}} = wR_{b_i, pri}^{s_j^{candidate}} + (1-w)R_{b_i, pub}^{s_j^{candidate}} \quad (9)$$

where, w is calculated according to formulas (10) and (11).

$$w = \begin{cases} \frac{N_{all}^{b_i}}{N_{min}} & \text{if } N_{all}^{b_i} < N_{min} \\ 1 & \text{otherwise} \end{cases} \quad (10)$$

$$N_{min} = -\frac{1}{2\varepsilon^2} \ln \frac{1-\eta}{2} \quad (11)$$

where, $N_{all}^{b_i}$ is the total number of ratings provided by b_i for the seller $s_j^{candidate}$, and N_{min} is a threshold calculated according to formula (11), which is similar to that defined in literature [22]. If $N_{all}^{b_i} \geq N_{min}$, buyer b_i will be confident about the private trustworthiness estimated based on its own ratings, therefore, the weight of private trustworthiness is simply assigned as 1. Otherwise, b_i will also consider public reputation estimated based on advisors ratings. ε represents

the maximal level of error that can be accepted, η represents confidence.

Algorithm 3 SRCA (Seller's reputation calculation algorithm)

Input: (1) The honest buyer b_i ; (2) The set of the sellers who are recommended by search agent $S^{candidate}$; (3) Every Recommended seller's optimal advisor lists $A_{b_i}^{s^{candidate}}$;

Output: Recommended sellers' reputation

Process:

- (1) for each $s_j^{candidate}$ in $S^{candidate}$
 - (2) if ($s_j^{candidate} \in S_b^H$)
 - (3) Based on b_i 's rating to $s_j^{candidate}$, calculate the private reputation $R_{b_i, pri}^{s_j^{candidate}}$ according to formula (6);
 - (4) else
 - (5) $R_{b_i, pri}^{s_j^{candidate}} = 0$;
 - (6) Based on the ratings given by $A_{b_i}^{s_j^{candidate}}$ calculate $s_j^{candidate}$'s public reputation $R_{b_i, pub}^{s_j^{candidate}}$ according to formulas (7);
 - (7) Fuse $R_{b_i, pri}^{s_j^{candidate}}$ and $R_{b_i, pub}^{s_j^{candidate}}$ to get the total reputation $R_{b_i}^{s_j^{candidate}}$ of $s_j^{candidate}$ according to formulas (9);
 - (8) Return Recommended sellers' reputation
-

4.4 The whitelist and blacklist updating algorithm

After the transaction, the defending buyer b_i can rate the selected seller and then update its' own whitelist and blacklist according to its experience. This updating makes the whitelist and the blacklist always a timely record of the buyers, the honest buyer trust, and distrust. The main idea of this updating process (see Algorithm 4) is as follows: (1)

considering whether each member in $B_{s_j^{candidate}}^H$ (which is denoted as b_k) should be exchanged into b_i 's whitelist and blacklist (see step (1) in algorithm 4); (2) if b_k 's syntactic trust is larger than the one of the most untrustworthy buyer in WL_{b_i} (denoted as b_{mu}), b_k will replace b_{mu} (see step (2-4) in Algorithm 4); if b_k 's synthetic trust is smaller than the one of the most trustworthy buyer in BL_{b_i} (denoted as b_{mt}), b_k will replace b_{mt} (see step (5-7) in Algorithm 4).

Algorithm 4 The whitelist and blacklist updating algorithm

Input: (1) The set of buyers that have interacted with the selected seller $s_j^{candidate}$, denoted as $B_{s_j^{candidate}}^H$; (2) WL_{b_i} and BL_{b_i} of buyer b_i ;

Output: WL_{b_i} and BL_{b_i}

Process:

- (1) for each b_k in $B_{s_j^{candidate}}^H$
 - (2) Select the buyer b_{mu} who have the smallest synthetic value from WL_{b_i} ;
 - (3) if ($STD_{b_i}(b_k) > STD_{b_i}(b_{mu})$)
 - (4) Remove b_{mu} from WL_{b_i} , add b_k into WL_{b_i} ;
 - (5) Select the buyer b_{mt} who have the greatest synthetic value from BL_{b_i} ;
 - (6) if ($STD_{b_i}(b_k) < STD_{b_i}(b_{mt})$)
 - (7) Remove b_{mt} from BL_{b_i} , add b_k into BL_{b_i} ;
 - (8) end for
-

5 Experimental results

To verify the performance of WBCEA given in Section 4, we design a set of experiments. Similar to experiments designed in previous studies [5, 6], 6 typical kinds of attacks

including *AlwaysUnfair*, *Camouflage*, *Sybil*, *Whitewashing*, *Sybil-Camouflage*, and *Sybil-Whitewashing* are selected to attack the reputation system. *AlwaysUnfair* attackers always give high reputation to dishonest sellers while rate low to the honest seller. *Camouflage* attackers intermittently tell

the truth or call white black (i.e., give unfairly high scores to dishonest sellers and unfairly low scores to honest sellers). In experiments of this paper, each *Camouflage* attacker will rate honestly in the first 20 days, while give unfair ratings to both the duopoly dishonest and honest seller in following days. Similar to *AlwaysUnfair* attackers, buyers who adopt *Sybil* attack always called white black. Different from *AlwaysUnfair* attack, the number of dishonest buyers in *Sybil* is greatly larger than the one in *AlwaysUnfair* attack. The buyers who use the *Whitewashing* attack strategy always whiten their low reputation by recreating a new account. In experiments, the *Whitewashing* attackers provide an unfair rating each day and then recreate a new account in later day to whitewash their sham action.

Naive strategy and the oracle strategy are designed as baselines. Naive strategy means that the buyers who adopt this strategy believe all raters are good and their ratings are true. Oracle strategy assumes that the buyers are omniscient. Therefore, they always know each seller's real reputation. Moreover, strategies such as *iClub*, *Personalized*, *MET*, *PEALGA*, *GBR* and *Multi-faceted* are selected to compare with our strategy (*WBCEA*). *iClub* [11, 12], *Personalized* [22], *MET* [6] are typical strategies that only consider trust factor in sellers' and buyers' evaluation process. These strategies belong to the filtering, discounting and evolutionary category respectively. *PEALGA* [5] is another evolutionary strategy that only considers trust. But it constructs different customized optimal advisor list for different candidate sellers. *GBR* [13] and *Multi-faceted* [4] are compared because they consider both trust and distrust information in the evaluation of sellers and buyers. In particular, this paper constructs an algorithm named *WBCEA.S*. The only difference between *WBCEA* and *WBCEA.S* is that the latter algorithm only evolves one advisor list to evaluate all the candidate sellers. This algorithm is constructed for verifying whether customized optimal advisor list can outperform one advisor list in evaluating candidate sellers when both trust/distrust are considered and whitelist/blacklist are maintained.

5.1 Experimental settings

To facilitate the analysis of various defense strategies, Jiang et al. [6] simulated a market with two duopoly sellers, 99 dishonest common sellers, and 99 honest ordinary sellers. The ordinary sellers are selected as noisy when honest buyers' selecting duopoly sellers. For comparing with the strategy given by Jiang et al. [6], we choose a similar market setting and relist the electronic market parameters and the values in trust attack/defense experiment (see Table 2). According to Table 2, under all attacks except for *Sybil*, there are 12 dishonest buyers and 28 honest buyers in the

Table 2 Parameters in simulation

Key parameters	Values
Number of dishonest duopoly sellers	1
Number of honest duopoly sellers	1
Number of dishonest common sellers	99
Number of honest common sellers	99
Number of dishonest buyers($ D_B $)	12/28*
Number of honest buyers($ H_B $)	28/12*
Simulation days(Days)	100
Dominance Ratio(Ratio)	0.5

Non-Sybil-based Attack/Sybil-based Attack

market. Under *Sybil* attack, the dishonest buyers and honest buyers are 28 and 12 respectively. Besides, 100 days of transactions are simulated in total. It should be noted that the initial trustworthy aspect (i.e., $R_{b_i,T}(b_k)$) and untrustworthy aspect (i.e., $R_{b_i,D}(b_k)$) of reviewers is randomly assigned with a value range from 0 to 1 respectively. In each day, every buyer makes one transaction with a partner. The ratings they score sellers are ranged from 0 to 1. It also should be noted that, recommendation algorithm design is not the topic of this paper, the recommendation list about sellers is randomly generated in experiments. The settings of parameters in *WBCEA* are listed in Table 3.

5.2 Evaluative criteria

To compare the experimental results, we choose similar criteria given in Jiang et al. [6] to evaluate the performance of each strategy. One criterion is robustness, which is used to evaluate the feasibility (i.e., anti-attack ability) of each defense strategy from macroscopic scale. Formula (12)

Table 3 Setting and meaning of variables or parameters used in *WBCEA*

Parameters	Meanings	Value
m	Length of whitelist	4
n	Length of blacklist	4
r	The number of recent reviewers who rate each sellers	20
Depth	The depth of Network	6
β_1	Discounted factor for trust value	0.4
β_2	Discounted factor for distrust value	0.3
θ_1	Threshold	0.8
θ_2	Threshold	0.2
Ratio	ratio for selecting duopoly sellers	0.5
λ	Discounted factor for historical rating	0.9
ε	Maximal level of error that can be accepted	0.25

defines the function of robustness. According to this formula, the value of robustness ranges from -1 to 1. The more transactions a defense agent trades with the honest duopoly seller, the higher its correct selection rate is, the larger the value of robustness is, and therefore the better the defending ability is.

Definition 9 Robustness of a defense strategy (abbr., *Def*) against an attack model (abbr., *Atk*) is the average transaction difference between the two kinds of duopoly sellers and the honest buyers in the simulation days. Its formula (12) defines the calculation principle [6].

$$R(Def, Atk) = \frac{|Tran(s^H)| - |Trans(s^D)|}{|B^H| \times Days \times Ratio} \quad (12)$$

where $Tran(s^H)$ is the transaction volume of duopoly honest seller, $Trans(s^D)$ is the transaction volume of duopoly dishonest seller, B^H is the number of honest buyers, *Days* is the total transaction days, and *Ratio* is the selection probability of duopoly sellers.

The mean absolute error (abbr., MAE) of seller's reputation is used to measure the accuracy of trust models in modeling seller's reputation. Formula (13) defines the calculation function of MAE. The smaller the MAE, the accuracy the defense strategy's prediction is, and therefore the better the defense strategy is.

Definition 10 The mean absolute error (MAE) of a seller s_j 's reputation is defined according to its real reputation and estimated reputation when buyer adopt a given defense strategy. Formula (13) [6] defines the calculation function.

$$MAE(s_j) = \frac{\sum_t \sum_{b_i} |R^t(s_j) - \tilde{R}_{b_i}^t(s_j)|}{|B^H| \times Days} \quad (13)$$

where B^H is the number of honest buyers, *Days* is the total transaction days, $R^t(s_j)$ is the actual reputation of seller s_j in day t ($t \in [0, Days]$), and $\tilde{R}_{b_i}^t(s_j)$ is the estimated reputation of seller s_j in day t , which is calculated according to the ratings of advisors of $b_i \in B^H$.

5.3 Results and analysis

5.3.1 Robustness analysis

The experimental results about the robustness of compared strategies are listed in Table 4. Each row represents the robustness of one defending strategy. Each column represents the robustness of comparison strategies when defending one attack. The number before “±” is the mean value of robustness. The number after “±” is the mean square deviation. The larger the mean value or the smaller the mean square deviation, the more robust the corresponding strategy will be. The following paragraphs explain these results in detail.

The naïve strategy assumes that all raters are good and their ratings are true. If majority of reviewers are attackers, following these reviewers' advices, naïve strategy may make the defending agent falsely judge the trustworthiness of sellers and falsely choose dishonest seller to trade with. The *oracle* strategy assumes agents always know the real trustworthiness of each reviewer and can always choose the honest seller according to honest reviewers' advices. Therefore, *oracle* can always reach the highest robustness. *Oracle* and *naïve* values can be regarded as baselines. The nearer a robustness to *oracle*, the more robust the corresponding strategy is.

Comparing all the strategies, we can find that *PEALGA* and *WBECA* achieve the best result or nearly the best result upon defending various pure and combined attacks. In particular, *PEALGA* and *WBCEA* achieve the best performance when defending against *Sybil* and *Sybil & whitewashing* attacks. Other compared strategies score poorly on one

Table 4 Robustness of compared strategies

Classifications	Strategies	alwaysUnfair	Camouflage	Whitewashing	Sybil	Sybil&Camouflage	Sybil&Whitewashing
Baselines	<i>naïve</i>	0.89 ± 0.03	0.93 ± 0.02	0.88 ± 0.19	−0.98 ± 0.07	−0.50 ± 0.07	−1.00 ± 0.07
	<i>oracle</i>	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Trust	<i>iClub</i>	0.99 ± 0.03	0.98 ± 0.02	0.76 ± 0.15	0.23 ± 0.32	0.92 ± 0.10	0.26 ± 0.30
	<i>Personalized</i>	0.98 ± 0.02	0.99 ± 0.02	0.98 ± 0.03	0.78 ± 0.30	0.94 ± 0.09	−0.97 ± 0.21
	<i>MET</i>	0.99 ± 0.03	0.99 ± 0.03	0.99 ± 0.03	0.79 ± 0.26	0.92 ± 0.07	0.87 ± 0.15
	<i>PEALGA</i>	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.03	0.98 ± 0.07	0.98 ± 0.06	0.97 ± 0.05
	<i>GBR</i>	0.96 ± 0.03	0.96 ± 0.03	0.97 ± 0.03	0.84 ± 0.43	0.91 ± 0.12	0.75 ± 0.08
Trust & distrust	<i>Multi-faceted</i>	0.98 ± 0.03	0.93 ± 0.04	0.76 ± 0.09	0.91 ± 0.10	0.80 ± 0.10	0.40 ± 0.21
	<i>WBCEA-S</i>	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.03	0.88 ± 0.09	0.97 ± 0.06	0.74 ± 0.45
	<i>WBCEA</i>	0.98 ± 0.03	0.98 ± 0.03	0.98 ± 0.03	0.96 ± 0.08	0.98 ± 0.06	0.97 ± 0.08

Table 5 MAE and variance of dishonest duopoly seller's reputation

Classifications	Strategies	alwaysUnfair	Camouflage	Whitewashing	Sybil	Sybil&Camouflage	Sybil&Whitewashing
Baselines	<i>naive</i>	0.76 ± 0.01	0.65 ± 0.02	0.76 ± 0.05	0.54 ± 0.01	0.49 ± 0.02	0.55 ± 0.02
	<i>oracle</i>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Trust	<i>iClub</i>	0.85 ± 0.20	0.73 ± 0.11	0.80 ± 0.08	0.06 ± 0.01	0.70 ± 0.11	0.06 ± 0.01
	<i>personalized</i>	0.45 ± 0.06	0.46 ± 0.06	0.82 ± 0.04	0.23 ± 0.07	0.59 ± 0.08	0.25 ± 0.04
	<i>MET</i>	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.10 ± 0.06	0.12 ± 0.03	0.36 ± 0.15
	<i>PEALGA</i>	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.04 ± 0.01	0.04 ± 0.02
Trust & distrust	<i>GBR</i>	0.13 ± 0.07	0.10 ± 0.05	0.06 ± 0.02	0.52 ± 0.10	0.46 ± 0.06	0.19 ± 0.04
	<i>Multi-faceted</i>	0.06 ± 0.11	0.07 ± 0.05	0.14 ± 0.05	0.04 ± 0.03	0.15 ± 0.04	0.32 ± 0.10
	<i>WBCEA_S</i>	0.01 ± 0.00	0.01 ± 0.00	0.08 ± 0.04	0.03 ± 0.00	0.05 ± 0.01	0.27 ± 0.06
	<i>WBCEA</i>	0.01 ± 0.01	0.02 ± 0.00	0.02 ± 0.01	0.02 ± 0.00	0.06 ± 0.02	0.03 ± 0.04

or more scenarios. For example, though *GBR* and *Multi-faceted* strategies consider trust and distrust information simultaneously, they cannot defend the attacks including *Sybil* very well. *WBCEA_S* is greatly inferior to *PEALGA* and *WBCEA* when defending against *Sybil* (0.88 ± 0.09) and *Sybil&Whitewashing* (0.74 ± 0.45) attacks. The high performances of *PEALGA* (only considers trust in the pre-evolution of the optimal customized advisor list) and *WBCEA* (which considers trust and distrust factors in the co-evolution of the whitelist and blacklist) are caused by the fact that both of them emphasize the idea of evolving an optimal customized advisor list to evaluate each candidate trading seller, which enable defenders to accurately predict duopoly sellers' reputations and accurately choose the honest duopoly seller as transaction partner. Therefore, we can conclude that the idea of generating an optimal customized advisor list is more influential than the idea of simultaneously consideration of trust and distrust and maintaining of whitelist and blacklist in accurately evaluating sellers' trustworthiness.

5.3.2 Accuracy analysis

Though Table 4 can reveal the difference of robustness in defending against various attacks, it cannot differentiate these defending strategies' estimation accuracy about duopoly sellers' reputation. Tables 5 and 6 list the MAE and variances of dishonest and honest duopoly sellers' reputation respectively. As *oracle* strategy assumes that agent knows each seller's real reputation, agents who are oracles can always accurately predict seller's reputation (i.e., MAE equals to zero). Therefore, *oracle* strategy can be taken as the lowest (best) baseline. The closer a strategy's MAE to *oracle* strategy, the more accurate it is. In contrast, as *naive* strategy assumes that agents trust all the buyers' ratings are true, therefore, difference between real reputation and predicted reputation gained by *naive* strategy is large (i.e., MAE is very large). Especially, in predicting the reputation of dishonest duopoly seller, the MAE is as large as 0.76 in some cases. Therefore, *naive* strategy can be taken as the highest (worst) baseline. The closer a strategy's MAE

Table 6 MAE and variances of honest duopoly seller's reputation

Classifications	Strategies	alwaysUnfair	Camouflage	Whitewashing	Sybil	Sybil&Camouflage	Sybil&Whitewashing
Baselines	<i>naive</i>	0.19 ± 0.01	0.10 ± 0.01	0.20 ± 0.08	0.98 ± 0.01	0.47 ± 0.02	0.97 ± 0.01
	<i>oracle</i>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Trust	<i>iClub</i>	0.01 ± 0.00	0.01 ± 0.00	0.12 ± 0.07	0.35 ± 0.16	0.01 ± 0.01	0.34 ± 0.15
	<i>personalized</i>	0.02 ± 0.00	0.01 ± 0.00	0.06 ± 0.00	0.18 ± 0.13	0.05 ± 0.00	0.96 ± 0.07
	<i>MET</i>	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.13 ± 0.13	0.10 ± 0.03	0.14 ± 0.10
	<i>PEALGA</i>	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.04 ± 0.01	0.02 ± 0.01
Trust & distrust	<i>GBR</i>	0.07 ± 0.01	0.05 ± 0.01	0.04 ± 0.01	0.33 ± 0.15	0.23 ± 0.01	0.19 ± 0.03
	<i>Multi-faceted</i>	0.02 ± 0.00	0.04 ± 0.02	0.12 ± 0.04	0.05 ± 0.03	0.12 ± 0.04	0.30 ± 0.09
	<i>WBCEA_S</i>	0.01 ± 0.00	0.01 ± 0.00	0.07 ± 0.03	0.03 ± 0.00	0.05 ± 0.01	0.28 ± 0.16
	<i>WBCEA</i>	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.05 ± 0.01	0.02 ± 0.02

to the one of *naïve* strategy, the more inaccurate the strategy is. From Tables 5 and 6, we can see that the MAEs of *PEALGA* and *WBCEA* can approach the ones of *oracle* strategy very well under almost all the attacks. *WBCEA_S* also perform well when predicting duopoly seller's reputation under attacks except for *Sybil&Whitewashing*. *GBR* and *Multi-faceted* strategies can quite accurately predict honest duopoly seller's reputation, especial when attacks without *Sybil*. However, their prediction about dishonest duopoly seller is not accurate (i.e., the MAE is as large as 0.52 ± 0.10 , 0.46 ± 0.06 , 0.32 ± 0.10).

Except for Tables 5 and 6, Figs. 5 and 6 illustrate variation trends of the averaged estimated reputation of duopoly sellers when defenders (i.e., honest buyers) adopt various defending strategies under multifarious attacks in detail. In these figures, the horizontal axis represents the 100 days of trading, while the vertical axis is the averaged estimated reputation of duopoly sellers. From these figures, we can see that *oracle* can always accurately estimate honest and dishonest sellers' reputation (the real reputation of dishonest seller is 0, and the real reputation of honest seller is 1). Therefore, the closer a curve of given strategy

Fig. 5 Reputations of dishonest duopoly sellers when buyer agents adopt various strategies defending attacks

to the *oracle* curve, the more accurate this strategy's estimation is. Inversely, as *naïve* strategy assumes the defender trust all others' ratings, it can be regarded as the worst baseline. The following paragraph analyzes and compares these strategies' ability in defending against various attacks in detail.

In general, *PEALGA* and *WBCEA* outperform all the trust strategies and all the trust and distrust strategies, under almost all attacks, especially under attacks such as *Sybil*, *Sybil&whitewashing*, and *Sybil&camouflage*. Moreover, *PEALGA* and *WBCEA* perform similarly and achieve the best performance under all attacks as there is

Fig. 6 Reputations of honest duopoly sellers when buyer agents adopt various strategies defending attacks

very little difference between their reputation MAE (see the 7th row and the 11th row in Tables 5 and 6). That can also be further explained according to Figs. 5 and 6, in which *PEALGA* and *WBCEA* always approach the *oracle* baseline best under all attacks. What's more, *WBCEA* outperform other strategies including *PEALGA* in defending *Sybil* as it is the quickest strategies to reach their real reputation (see Figs. 5d and 6d). These results are achieved because simultaneous consideration of trust and distrust can improve defending strategy's prediction accuracy slightly, while the adaptation of customized optimal advisor list for evaluating each candidate seller can improve its prediction accuracy greatly.

6 Conclusions

In this paper, based on psychology research result of Lewicki et al. [10], we assume that each buyer agent has both the trustworthy aspect and the untrustworthy aspect, and propose to assign two scores to each buyer agent for denoting its trustworthy aspect and untrustworthy aspect respectively. The synthetization of these two aspects can be used as criteria for evaluating a buyer's synthetized trustworthiness. Besides, each buyer maintains a whitelist and a blacklist, which is evolved according to the new algorithm called *WBCEA* for defending multifarious attacks. The *WBCEA* algorithm is composed of several sub-algorithms, such as the trust network construction algorithm, the optimal advisor lists generation algorithm, the seller's reputation calculation algorithm and the whitelist and blacklist updating algorithm. According to the whitelist and blacklist, a buyer can construct its own trust network (see algorithm 1). By doing so, the buyer can select trustworthy advisors for evaluating each candidate seller and chose the most trustworthy seller as trading partner.

A set of experiments is designed and implemented to compare the performance of our strategy with recent typical defending strategies and baseline ones. Experimental results show that the *WBCEA* strategy and the *PEALGA* strategy have similar performance in defending against all attacks. Moreover, they can outperform existing related trust strategies and all the trust and distrust strategies in robustness and MAE of seller's reputation when defending various attacks. In particular, *WBCEA* slightly outperform *PEALGA* when defending against *Sybil* attack.

The strategies compared in this paper are suitable for B2B electronic market, where sellers' behaviors are not changing frequently. Whether the strategy proposed in this paper is suitable for the C2C market in which the sellers' behaviors often change is a problem that needs to be explored. Moreover, it is significant for us to study the stability of this strategy when the configuration (number of sellers, the ratio of dishonest buyers) changes. Besides,

using real data to verify the accuracy, robustness and stability of this strategy is our future target.

Acknowledgments This paper is supported in part by the Natural Science Foundation of China (Nos. 71403151, 61572035, 61402011, 61433012, 61502281), the Natural Science Foundation of Shandong Province (Nos. ZR2013FM023, ZR2013FQ030, ZR2014FP011, ZR2011FL002), China's Post-doctoral Science Fund (No. 2014M561948), Postdoctoral innovation project special funds of Shandong Province (No. 201403007), Applied research project for Qingdao postdoctoral researcher, the Qingdao Science and Technology Development Project (KJZD-13-29-JCH), Project of Shandong Province Higher Educational Science and Technology Program (J14LN33), the Leading talent development program of Shandong University of Science and Technology and Special Project fund of Taishan scholar of Shandong Province.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Abdul-Rahman A, Hailes S (2000) Supporting trust in virtual communities. In: Hawaii international conference on system science. IEEE Computer Society, vol 6, p 6007
2. McKnight DH, Chervany NL (2001) Trust and distrust definitions: one bite at a time. *Lect Notes Comput Sci* 107(12):27–54
3. McKnight DH, Choudhury V (2006) Distrust and trust in B2C e-commerce: do they differ? *International Conference on Electronic Commerce: the New E-Commerce - Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet*, 2006, Fredericton, New Brunswick, Canada, pp 482–491
4. Fang H, Guo G, Zhang J (2015) Multi-faceted trust and distrust prediction for recommender systems. *Decis Support Syst* 71(C):37–47
5. Ji S, Ma H, Zhang S, Leung HF, Chiu D, Zhang CJ et al (2016) A pre-evolutionary advisor list generation strategy for robust defending reputation attacks. *Knowl-Based Syst* 103(C):1–18
6. Jiang S, Zhang J, Ong Y (2013) An evolutionary model for constructing robust trust networks. In: *Proceedings of the 12th international conference on autonomous agents and multiagent systems (AAMAS)*, pp 813–820
7. Jøsang A (2012) Robustness of trust and reputation systems: does it matter? *Ifip Advances in Information & Communication Technology*, pp 253–262
8. Jøsang A, Gray E, Kinatader M (2003) Analysing topologies of transitive trust. In: *Proceedings of the workshop of formal aspects of security & trust*
9. Jøsang A, Ismail R (2002) The beta reputation system. *Bled Conference on Electronic Commerce*. pp. 324–337.
10. Lewicki RJ, Mcallister DJ, Bies RJ (1998) Trust and distrust: new relationships and realities. *Acad Manag Rev* 23(3):438–458
11. Liu S, Zhang J, Miao C, Theng Y, Kot AC (2011) ICLUB: an integrated clustering-based approach to improve the robustness of reputation systems. *Inter Conf Auton Agents Multiagent Syst* 3:1151–1152
12. Liu S, Zhang J, Miao C, Theng Y, Kot AC (2014) An integrated clustering-based approach to filtering unfair multi-nominal testimonies. *Comput Intell* 30(2):316–341

13. Liu X, Wang Y, Zhu S, Lin H (2013) Combating Web spam through trust–distrust propagation with confidence. *Pattern Recogn Lett* 34(13):1462–1469
14. Victor P, Cornelis C, Cock MD et al (2011) Trust-and Distrust-based recommendations for controversial reviews. *IEEE Intell Syst* 26(1):48–55
15. Victor P, Verbiest N, Cornelis C, Cock MD (2013) Enhancing the trust-based recommendation process with explicit distrust. *ACM Trans Web (TWEB)* 7(2):42–59
16. Singh S, Bawa S (2007) Privacy, trust and policy based authorization framework for services in distributed environment. *Int J Comput Sci* 2(2):85–92
17. Smith NK, Larsen JT, Chartrand TL, Cacioppo JT, Katafiasz HA, Moran KE (2006) Being bad isn't always good: affective context moderates the attention bias toward negative information. *J Pers Soc Psychol* 90(2):210–20
18. Teacy WTL, Patel J (2006) TRAVOS: trust and reputation in the context of inaccurate information sources. *Auton Agent Multi-Agent Syst* 12(2):183–198
19. Whitby A, Jøsang A, Indulska J (2004) Filtering out unfair ratings in Bayesian reputation systems. *Int Joint Conf Auton Agent Syst* 37(3):106–117
20. Yu B, Singh MP (2002) Detecting deception in reputation management Second international joint conference on autonomous agents & multiagent systems, vol 2, pp 73–80
21. Yu B, Singh MP (2002). An evidential model of distributed reputation management. In: *Proceedings of the first international joint conference on autonomous agents and multiagent systems: part*. ACM pp 294–301
22. Zhang J, Cohen R (2011) A framework for trust modeling in multiagent electronic marketplaces with buying advisors to consider varying seller behavior and the limiting of seller bids. *ACM Trans Intell Syst Technol* 4(2):55–73
23. Zhang L, Jiang S, Zhang J, Ng WK (2012). Robustness of trust models and combinations for handling unfair ratings. In: *Ifip advances in information & communication technology*, pp 36–51

Yongquan Liang is a professor and dean of the College of Computer Science and Engineering in Shandong University of Science and Technology. He received his M.Sc. degree in Computer software and Theory from Beijing University of Aeronautics and Astronautics and Ph.D degree in Computer software and Theory from Institute of Computing Technology, Chinese Academy of Sciences. Now he is a member of Chinese Association for Artificial Intelligence, and Senior member of CCF.

HoFung Leung received his BSc and MPhil degrees in Computer Science from The Chinese University of Hong Kong, and his PhD degree and DIC (Diploma of Imperial College) in Computing from Imperial College London. Now, he is a Professor and the Chairman of the Department of Computer Science and Engineering at The Chinese University of Hong Kong. He is a Chartered Fellow of the BCS, a Fellow of the HKIE, a Senior Member of

both the ACM and the IEEE, and a full member the HKCS. He is a Chartered Engineer registered by the Engineering Council.

Shujuan Ji is an Associate Professor of the College of Computer Science and Engineering in Shandong University of Science and Technology. She received her B.Sc., M.Sc. and Ph.D degrees in Computer software and Theory from Shandong University of Science and Technology, Qingdao, P.R.China.

Haiyan Ma is a Master student of Computer software and Theory in Shandong University of Science and Technology, Qingdao, P.R.China.

Chunjin Zhang is an Engineer of the Center of Modern Education in Shandong Shandong University of Science and Technology. He received his B.Sc. degree in Computer Science and Technology from College of Information Science and Engineering at Shandong University of Science and Technology, and received his M.Sc. degree in Control Theory and Control Engineering from College of Information and Electrical Engineering at Shandong University of Science and Technology, Qingdao, P.R.China.