TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
ISSN 2029-4913 / eISSN 2029-4921

2017 Volume 23(1): 81–110
doi:10.3846/20294913.2015.1022810

QADE: A NOVEL TRUST AND REPUTATION MODEL FOR HANDLING FALSE TRUST VALUES IN E–COMMERCE ENVIRONMENTS WITH SUBJECTIVITY CONSIDERATION

Eva ZUPANCIC, Denis TRCEK

Faculty of Computer and Information Science, University of Ljubljana, Trzaska 25, SI–1000 Ljubljana, Slovenia

Received 29 October 2012; accepted 02 January 2014

Abstract. Trust is essential to economic efficiency. Trading partners choose each other and make decisions based on how much they trust one another. The way to assess trust in e-commerce is different from those in brick and mortar businesses, as there are limited indicators available in online environments. One way is to deploy trust and reputation management systems that are based on collecting feedbacks about partners' transactions. One of the problems within such systems is the presence of unfair ratings. In this paper, an innovative QADE trust model is presented, which assumes the existence of unfairly reported trust assessments. Subjective nature of trust is considered, where differently reported trust values do not necessarily mean false trust values but can also imply differences in dispositions to trust. The method to identify and filter out the presumably false values is defined. In our method, a trust evaluator finds other agents in society that are similar to him, taking into account pairwise similarity of trust values and similarity of agents' general mindsets. In order to reduce the effect of unfair ratings, the values reported by the non-similar agents are excluded from the trust computation. Simulations have been used to compare the effectiveness of algorithms to decrease the effect of unfair ratings. The simulations have been carried out in environments with varying number of attackers and targeted agents, as well as with different kinds of attackers. The results showed significant improvements of our proposed method. On average 6% to 13% more unfair trust ratings have been detected by our method. Unfair rating effects on trust assessment were reduced with average improvements from 26% to 57% compared to the other most effective filtering methods by Whitby and Teacy.

Keywords: e-commerce, trust and reputation management systems, false trust values, subjectivity.

JEL Classification: C63, O31.

Introduction

Internet has become an important business medium and there are growing number of participants engaging in electronic commerce. Consumers are reluctant to conduct business over Internet due to concerns about trust and trustworthiness of participating entities (Com-

Corresponding author Eva Zupancic
E-mail: *eva.zupancic@fri.uni-lj.si*

mission of the European Communities 2009). In order to protect entities against malicious ones, security and trust mechanisms should be deployed. Protection with security services, such as authentication, is referred to as *hard security* (Rasmusson, Sverker Jansson 1996) and is unable to detect entities that will act deceitfully or provide misleading information after (legally) entering the e-commerce system. Therefore, additional control mechanisms should be deployed to provide protection against such type of threats. Such mechanisms are referred to as *soft security* mechanisms (Rasmusson, Sverker Jansson 1996), of which trust and reputation management systems are among the most important.

In recent years, many authors have presented computational models of trust (Pinyol, Sabater-Mir 2013; Josang *et al.* 2007), in order to develop trust and reputation management techniques. Trust and reputation management systems represent a method to promote trust between unknown entities in online environments. The main purpose is to raise the number of good interactions between agents, avoid bad interactions and mitigate risk involved in transactions. Additionally, studies show that sellers with better reputations were more likely to sell their items and at a higher price (Resnick 2002; Lucking-Reiley *et al.* 2007).

There are various open problems concerning trust and reputation management systems. One of the most important ones is the existence of unfair ratings in e-commerce systems (Hoffman *et al.* 2009; Yang *et al.* 2009). The problem is fundamental because an entity in an e-commerce system computes trust based on the ratings from other entities, of which it cannot control the sincerity. The assessed trust value is misleading if false ratings from other entities are taken into the trust computation. This could result in wrong decisions with all the related consequences. It is clearly seen that the problem of unfair ratings should be resolved and is of great importance for e-commerce systems.

Although the problem of unfair ratings has been recognized, the current solutions do not consider underlying psycho-sociological factors of trust phenomenon. Namely, according to one of the most authoritative definition of trust from Gambetta (2000), "trust is the subjective probability by which an individual A expects that another individual B performs a given action on which its welfare depends". As such, agents in e-commerce system may provide contradictory trust assessment towards a certain agent for two reasons. Firstly, they might be dishonest and *intentionally* provide unfairly high or unfairly low trust values about trading partners they interacted with, irrespective of the real experiences. The basic motivation for providing misleading trust values is to deform the reputation of other agents in the system. For example to degrade the reputation of rivals or to upgrade the reputation of associates. Secondly, they might be honest but they perceive trust differently. They provide true trust values based on their experience, which might be *unintentionally* misleading for other agents due to their different perception of trust (Fang *et al.* 2012).

In this paper, an innovative trust model called QADE (Qualitative Assessment Dynamics – Extended) is presented, which provides a genuine solution for resolving unfair trust ratings while considering agents' different trust tendencies. The proposed QADE trust model is an extension of the Qualitative Assessment Dynamics (QAD trust model) (Trcek 2009). Our solution, the QADE trust model, implies that trust is a *personal* and *subjective* phenomenon based on different factors, which have not been appropriately considered in existing methods for handling false trust values. Further, an original approach for handling false trust ratings

in e-commerce environments is proposed, which considers agents' various trust evaluation characteristics. Our proposed model and filtering technique deal exclusively with trust as one of the criteria in e-commerce purchasing process. For the subsequent process of evaluation and selection any decision-making method can be used (Zavadskas, Turskis 2011; Kim *et al.* 2008; Zavadskas *et al.* 2004; Kersuliene, Turskis 2011). The contributions of this paper are:

- Definition of *QADE trust model*. The QADE trust model presents mathematical formalization of trust relationships in e-commerce environments. It includes innovative components, such as "private trust vector", "general mindset", "similarity function", etc. The QADE trust model formalizes the existence of unfair trust ratings and defines trust considering its subjective nature.
- Definition of *QADE filtering algorithm* for trust value calculation. The proposed algorithm detects unfair trust ratings and excludes them from the trust value assessments.

Extensive simulations of trust relations between users (agents) in an e-commerce system have been made in order to evaluate the QADE trust model. The simulations have included fraudulent agents who report false trust values. Our approach has been compared with other most important algorithms, including the method proposed by Whitby *et al*. (2004) and the filter proposed in TRAVOS trust model (Teacy *et al.* 2006). The statistical comparison of results showed the convincing improvements of our proposed algorithm as it performed 20% to 67% better compared to the other most representative existing methods.

The rest of the paper is organized as follows. The following section presents Qualitative Assessment Dynamic (QAD) formalization of trust. Section "QADE Trust Model" describes the extensions to QAD that include mechanisms related to unfairly reported trust values. A method for handling with unfair trust values is proposed in Section "Unfair ratings", followed by the presentation and analysis of simulation results. Section "Related work" contains a review of related work and Section "Conclusions" completes the article.

1. Brief overview of Qualitative Assessment Dynamics

A considerable amount of research has focused on the development of trust and reputation models in e-commerce environments. The simplest trust and reputation models are online reputation models that are used in e-commerce sites such as eBay[1], Amazon[2] and OnSale[3]. They lack in reliability measures, consideration of false information or cheating and temporal issues (Pinyol, Sabater-Mir 2013). These issues are considered in more advanced trust models that base frequently on the following methods for trust computation: probability calculus and Bayesian networks (Wang, Vassileva 2005; Ismail, Josang 2002; Teacy *et al.* 2006), theory of evidence (Josang 2001; Yu, Singh 2002), game theory (Schillo *et al.* 2000), fuzzy logic (Victor *et al.* 2009; Liu *et al.* 2013; Sabater-Mir, Paolucci 2007; Manchala 2000), and discrete value approaches (Abdul-Rahman, Hailes 2000; Trcek 2009; Cahill *et al.* 2003). The extensive overview of status of trust in computing environment

[1] www.eBay.com.

[2] www.Amazon.com.

[3] www.onSale.com.

could be found in (Pinyol, Sabater-Mir 2013; Keung, Griffiths 2010; Grandison, Sloman 2000; Sabater, Sierra 2005; Josang *et al.* 2007).

The Qualitative Assessment Dynamics (QAD) trust model (Trcek 2009) takes into account agents' different trust perception and trust evaluation that stem from their subjective properties. According to the definition from the Oxford Dictionaries (2013), which defines trust as "firm belief in the reliability, truth, or ability of someone or something", trust results from belief or from faith. Further, personality psychologists conceptualize trust as a belief, expectancy or feeling, which are rooted in the personality (Grabner-Kräuter, Kaluscha 2003). Many other definitions (Hussain, Chang 2007) define trust with terms, such as belief, faith, willingness, hope or intention. These factors are based on, or influenced by, person's subjectivity, which is also emphasized in the Gambetta's definition of trust (Gambetta 2000). Based on this, in QAD trust model diversities in trust evaluation processes are assumed. It is supposed that the trust assessment is different for each agent and is affected by unobservable actions, which originate from agents' various personalities and their subjective concerns.

The basis for QAD trust model is the research done in the area of developmental psychology done by Piaget (2002) that provides a perspective on trust as a kind of reasoning and judgment process. The QAD model identifies and formalizes factors that drive trust based on Piaget's work. The QAD trust model complements existing trust management methodologies and is presented in the rest of the section.

In this paper, e-commerce environment is defined with multi-agents system approach. An agent represents an autonomous software or human entity that can act in the system, perceive events and reason (Sterling, Taveter 2009). In QAD formalization of trust, e-commerce system consists of communicating agents that represent buyers and sellers, and certain trust relations between them. The values of relationships between agents depend on each agent's *character* and can change over time.

Definition 1. Trust is a relationship between agents a_i and a_j, which means agent's a_i trust assessment towards agent a_j. The trust assessment value is taken from assessment set Ω and it is denoted by $\omega_{i,j}$.

In general, trust value between agents is context dependent. The agent's a_i trust attitude towards agent a_j may differ in another context. As only one context will be considered in the rest of the paper, the context is omitted from Definition 1. Further, a trust relation is generally not reflexive, not symmetric and non-transitive.

Chang *et al.* (2006) define a *trusting* agent as an entity who has faith or belief in another entity in a given context and at a given time slot. Further, they define a *trusted* agent as an entity in whom faith or belief has been placed by another entity in a given context and at a given time slot. In terms of QAD, trusted agent is an agent that trusting agent has defined, or known, trust relationship towards him, even if it is *untrusted*. In Definition 1, an agent a_i represents a trusting agent and an agent a_j represents a trusted agent. This terminology is used consistently through the paper.

Definition 2. The assessment set consists of five values, $\Omega = \{2,1,0,-1,-2\}$, which describe trusted, partially trusted, undecided, partially distrusted and distrusted trust relationship

between agents, respectively. If trust relationship between two agents is either not defined or unknown, it is denoted by "–".

Definition 3. In a given context, trust assessment values between agents are given by a trust matrix $\mathcal{M}$, where elements $\omega_{i,j}$ denote agent's a_i trust attitude towards agent a_j.

A general form of a trust matrix of a certain society with *n* agents is as follows:

$$\mathcal{M} = \begin{bmatrix} \omega_{1,1} & \cdots & \omega_{1,n} \\ \vdots & \ddots & \vdots \\ \omega_{n,1} & \cdots & \omega_{n,n} \end{bmatrix}. \tag{1}$$

Agents assess trust and spread trust assessment values through social interactions. Note that a trust matrix changes through time – namely, after agents had interactions that changed trust relationship between them (refer to Def. 6).

Definition 4. In a trust matrix, row k represents agent's k public trust vector. It gives agent's a_k trust assessments towards other agents and is denoted as $\mathcal{M}_{k,n} = \left(\omega_{k,1}, \omega_{k,2}, \ldots, \omega_{k,n}\right)$. Further, $\underline{\mathcal{M}_{k,n1}} = \left(\omega_{k,1}, \omega_{k,2}, \ldots, \omega_{k,n1}\right)$ is agent's k public trust sub-vector where " – " values are omitted.

Definition 5. In a trust matrix $\mathcal{M}$, column k represents society trust vector. It holds assessments (given by society) about particular agent a_k and is denoted as $\mathcal{M}_{n,k} = \left(\omega_{1,k}, \omega_{2,k}, \ldots, \omega_{n,k}\right)$. Further, $\underline{\mathcal{M}_{n1,k}} = \left(\omega_{1,k}, \omega_{2,k}, \ldots, \omega_{n1,k}\right)$ is society trust sub-vector with omitted " – " values.

The society trust vector is input for trust computation. For example, to compute trust of an agent a_i in an agent a_k, the agent a_i aggregates values from the society trust sub-vector $\underline{\mathcal{M}_{n1,k}}$. The society trust sub-vector $\underline{\mathcal{M}_{n1,k}}$ contains all trust assessments provided by agents that already have had interactions with the agent a_k. The aggregation of trust assessments from the society trust sub-vector is defined in the Definition 6. The QAD trust model considers agents' different trust propensities and the assessment of trust relationships between agents depends on "characters" of the agents. Our trust model defines QAD operators that are assigned to the agents and model their trust evaluation.

Definition 6. QAD operators are elements of set $\Psi = \left\{\Uparrow, \Downarrow, \rightsquigarrow, \leftrightarrow, \uparrow, \downarrow\right\}$, where the symbols denote extreme-optimistic assessment, extreme-pessimistic assessment, centralistic consensus-seeker assessment, non-centralistic consensus-seeker assessment, moderate-optimistic assessment and moderate-pessimistic assessment. The operators are functions $op_i \in \Psi$, such that $op_i : \underline{\mathcal{M}_{n1,j}} = \left(\omega^-_{1,j}, \omega^-_{2,j}, \ldots, \omega^-_{n1,j}\right) \rightarrow \omega^+_{i,j}$, where superscript " – " denotes pre-operation value and superscript " + " denotes post-operation value. Mappings for particular operators are defined as follows:

$\omega^-_{i,j} \neq$ "–":

$$\Uparrow_i : \max\left(\underline{\mathcal{M}_{n1,j}}\right) \rightarrow \omega^+_{i,j}; \tag{2}$$

$$\Downarrow_i : \min\left(\underline{\mathcal{M}_{n1,j}}\right) \rightarrow \omega^+_{i,j}; \tag{3}$$

$$\rightsquigarrow_i : \begin{cases} \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \to \omega^+_{i,j} & \text{if } \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- < 0 \\ \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \to \omega^+_{i,j} & \text{ohterwise} \end{cases}; \tag{4}$$

$$\leftrightarrow_i : \begin{cases} \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \to \omega^+_{i,j} & \text{if } \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- > 0 \\ \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \to \omega^+_{i,j} & \text{ohterwise} \end{cases}; \tag{5}$$

$$\uparrow_i : \begin{cases} \omega^-_{i,j} \to \omega^+_{i,j} & \text{if } \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \le \omega^-_{i,j} \\ \omega^-_{i,j} + 1 \to \omega^+_{i,j} & \text{ohterwise} \end{cases}; \tag{6}$$

$$\downarrow_i : \begin{cases} \omega^-_{i,j} \to \omega^+_{i,j} & \text{if } \frac{1}{\left|\underline{\mathcal{M}_{n1,j}}\right|} \sum_{\omega^- \in \underline{\mathcal{M}_{n1,j}}} \omega^- \ge \omega^-_{i,j} \\ \omega^-_{i,j} - 1 \to \omega^+_{i,j} & \text{ohterwise} \end{cases}; \tag{7}$$

$$\omega^-_{i,j} = \text{“}-\text{”}: \ \text{“}-\text{”} \to \omega^+_{i,j}. \tag{8}$$

The properties of operators can be informally stated as follows. If trust value is undefined, it remains undefined also after trust evaluation (refer to Eq. 8); otherwise, it is changed according to operator. Extreme-optimistic assessment operator ($\Uparrow$) filters out the most positive assessment value among existing assessments given by society about particular agent (refer to Eq. 2). Just oppositely, extreme-pessimistic assessment operator ($\Downarrow$) filters out the most negative assessment value (refer to Eq. 3). Centralistic consensus seeker assessment operator ($\rightsquigarrow$) computes the average value and rounds this value towards zero (refer to Eq. 4). Non-centralistic consensus seeker assessment operator ($\leftrightarrow$) results in a value, which is (contrary to the previous operator) “average rounded away from the 0 value” (refer to Eq. 5). Moderate optimistic assessment operator ($\uparrow$) results in next higher qualitative level if the average assessment of the rest of community is more optimistic than the agent's trust assessment is; otherwise it results in the same value (refer to Eq. 6). Moderate pessimistic assessment operator ($\downarrow$) is affected with more negative values only. It means the expressed assessment is “weakened” to the next lower qualitative level, narrowing the gap towards the aggregated assessment of the rest of community if this is more pessimistic than the agent's trust is (the value changes one level downwards, refer to Eq. 7).

Further explanation of QAD operators and examples of an agent's behavior modeled with each operator can be found in (Zupancic, Trcek 2011).

2. QADE trust model

The QAD trust model assumes that opinions of other agents stored in the trust matrix $\mathcal{M}$ are accurate and an agent evaluates its trust value towards others based on values from $\mathcal{M}$. In the reality, this is not always true and the agent should not assume the other actors in the e-commerce system always report their opinions truthfully. As pointed by Dellarocas (2000), some trading partners may provide unfairly high feedback to increase others' reputations, or they may provide unfairly low feedback to decrease others' reputations.

To address the problem of unfair ratings, in this paper the QADE trust model is proposed. It is an extension of the QAD trust model. In the QADE trust model 10 extensions are proposed that introduce support for modeling unfair ratings. Additionally, trust is considered subjectively when modeling unfair trust ratings. The extensions over the original trust model include agent's private trust vectors, historical trust matrices, historical private trust vectors, multisets of public/private trust values between two agents, new QADE operators definition, attacker agent definition, agent's general mindset definition, similarity function definition, similarity matrix, and QADE filtering algorithm. Firstly, an agent's *private* trust vector is introduced.

Extension #1: Definition 7. An agent's a_i truthful trust values towards other agents in the society are given by agent's a_i private trust vector $Z_i = \left(\zeta_{i,1}, \zeta_{i,2}, \ldots, \zeta_{i,n}\right)$, where $\zeta_{i,j} \in \Omega$ denotes the agent's a_i truthful trust relationship towards agent a_j. Further, $\underline{Z_i}$ is agent's a_i private trust vector with omitted " – " values.

In contrast to values from public trust matrix $\mathcal{M}$, the values from private trust vector are not shared with other agents.

The values in the trust matrix $\mathcal{M}$ represent the newest trust values. The extended (QADE) model takes into account the historical values as well.

Extension #2: Definition 8. History of trust values between agents is given by historical trust matrices $\mathcal{M}^{-l}$, where $\left\{l : l \in \mathbb{Z}, l \geq 0\right\}$. Element $\omega_{i,j}^{-l}$ denotes agent's a_i l-th latest trust value towards agent a_j.

The latest trust values between agents, i.e. the trust values assessed after the last interaction, are stored in the trust matrix $\mathcal{M}^0$. The second latest trust values, i.e. trust values assessed after the previous interaction, are stored in the trust matrix $\mathcal{M}^{-1}$. The trust values assessed after the pre–previous interaction, are stored in the trust matrix $\mathcal{M}^{-2}$, etc.

Additionally, the QADE trust model considers the history of private trust values.

Extension #3: Definition 9. An agent's a_i history of truthful trust values towards other agents is given by historical private trust vectors Z^{-l}, where $\left\{l : l \in \mathbb{Z}, l \geq 0\right\}$. Element $\zeta_{i,j}^{-l}$ denotes agent's a_i l-th latest private trust value towards agent a_j.

Extension #4: Definition 10. Let t represents the total number of interactions between agents a_i and a_j. A multiset of public trust values between agents a_i and a_j (including the current and previous trust values) is given by $\omega_{i,j}^{0:-t} = \left\{\omega_{i,j}^{0}, \omega_{i,j}^{-1}, \ldots, \omega_{i,j}^{-t}\right\}$ and a multiset of private trust values between agents a_i and a_j (including the current and previous trust values) is given by $\zeta_{i,j}^{0:-t} = \left\{\zeta_{i,j}^{0}, \zeta_{i,j}^{-1}, \ldots, \zeta_{i,j}^{-t}\right\}$.

Further, the following notation is introduced. Let $\mathcal{M}_{k,n}^{0:-t} = \left\{\omega_{k,1}^{0:-\mathrm{t}}, \omega_{k,2}^{0:-\mathrm{t}}, \ldots, \omega_{k,n}^{0:-t}\right\}$ and $\mathcal{M}_{n,k}^{0:-t} = \left\{\omega_{1,k}^{0:-\mathrm{t}}, \omega_{2,k}^{0:-\mathrm{t}}, \ldots, \omega_{n,k}^{0:-t}\right\}$ denote a multiset of agent's a_k historical trust assessments

toward other agents and multiset of historical trust assessments about agent a_k, respectively. Similarly, let $\underline{\mathcal{M}_{k,n1}^{0:-t}} = \left\{\omega_{k,1}^{0:-t}, \omega_{k,2}^{0:-t}, \ldots, \omega_{k,n1}^{0:-t}\right\}$ and $\underline{\mathcal{M}_{n1,k}^{0:-t}} = \left\{\omega_{1,k}^{0:-t}, \omega_{2,k}^{0:-t}, \ldots, \omega_{n1,k}^{0:-t}\right\}$ denote a multiset of agent's a_k historical trust assessments toward other agents and multiset of historical trust assessments about agent a_k with omitted " – " values, respectively.

In the rest of the paper, $\mathcal{M}$, $\omega_{i,j}$, Z_i, $\zeta_{i,j}$, $\mathcal{M}_{k,n}$, $\mathcal{M}_{n,k}$, $\underline{\mathcal{M}_{k,n1}}$ and $\underline{\mathcal{M}_{n1,k}}$ denote $\mathcal{M}^0$, $\omega_{i,j}^0$, Z_i^0, $\zeta_{i,j}^0$, $\mathcal{M}_{k,n}^0$, $\mathcal{M}_{n,k}^0$, $\underline{\mathcal{M}_{k,n1}^0}$ and $\underline{\mathcal{M}_{n1,k}^0}$, respectively, unless otherwise stated.

Extension #5: The agents assess trust value toward other agents based on their operators, as explained in previous section. In QADE trust model, the trust value between agents a_i and a_j is calculated based on agent's a_i current and historical *private* trust assessments and influenced by *public* (also referred to as *reported*) trust values of other agents, which also includes historical values. The QADE operators are functions $op_i \in \emptyset$ (as described in Def. 6), such that $op_i : \widetilde{\underline{\mathcal{M}_{n1,j}}} = \left(\left(\underline{\mathcal{M}_{n1,j}^{0:-t}} \setminus \underline{\mathcal{M}_{i,j}^{0:-t}}\right) \cup \zeta_{i,j}^{0:-t}\right) \to \zeta_{i,j}^{+}$ (in contrast to $\underline{\mathcal{M}_{n1,j}} = \left(\omega_{1,j}^{-}, \omega_{2,j}^{-}, \ldots, \omega_{n1,j}^{-}\right)$ domain in Def. 6). Hence, the mappings for operators are redefined as follows:

$\zeta_{i,j}^{-} \neq$ "–":

$$\Uparrow_i : \max\left(\widetilde{\underline{\mathcal{M}_{n1,j}}}\right) \to \zeta_{i,j}^{+}; \tag{9}$$

$$\Downarrow_i : \min\left(\widetilde{\underline{\mathcal{M}_{n1,j}}}\right) \to \zeta_{i,j}^{+}; \tag{10}$$

$$\rightsquigarrow_i : \begin{cases} \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \to \zeta_{i,j}^{+} & \text{if } \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- < 0 \\ \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \to \zeta_{i,j}^{+} & \text{ohterwise} \end{cases}; \tag{11}$$

$$\leftrightarrow_i : \begin{cases} \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \to \zeta_{i,j}^{+} & \text{if } \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- > 0 \\ \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \to \zeta_{i,j}^{+} & \text{ohterwise} \end{cases}; \tag{12}$$

$$\uparrow_i : \begin{cases} \zeta_{i,j}^{-} \to \zeta_{i,j}^{+} & \text{if } \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \leq \zeta_{i,j}^{-} \\ \zeta_{i,j}^{-} + 1 \to \zeta_{i,j}^{+} & \text{ohterwise} \end{cases}; \tag{13}$$

$$\downarrow_i : \begin{cases} \zeta_{i,j}^{-} \to \zeta_{i,j}^{+} & \text{if } \dfrac{1}{\left|\widetilde{\underline{\mathcal{M}_{n1,j}}}\right|} \sum_{\omega^- \in \widetilde{\underline{\mathcal{M}_{n1,j}}}} \omega^- \geq \zeta_{i,j}^{-} \\ \zeta_{i,j}^{-} - 1 \to \zeta_{i,j}^{+} & \text{ohterwise} \end{cases}; \tag{14}$$

$$\zeta_{i,j}^{-} = \text{“}-\text{”}: \text{“}-\text{”} \rightarrow \zeta_{i,j}^{+}. \tag{15}$$

The agent's private vector contains its real trust opinions about other agents. The values from private vector can be reported to other agents (i.e. stored into the public trust matrix $\mathcal{M}$) in a modified form. An agent that reports false trust values about other agents will be referred to as *attacker* agent.

Extension #6: Definition 11. An agent a_i is an attacker if $Z_i \neq \mathcal{M}_{i,n}$, i.e. the agent's a_i private trust vector and public trust vector differ.

The QADE model includes formalization of an attacker, which is an agent who intentionally reports false trust assessments. Other types of attacks, which are typical for trust and reputation management systems (e.g. Sybil attack, short-term abuse of the system, denial of service attack, etc.), are out of scope of this research work. Therefore, they are not considered in the QADE trust model.

The operator is a trust assessment function that takes into account the possibility of agents' different trust assessments originating from their subjective concerns. The operator function may produce different outputs from the same input trust values, as the agents perceive the environment in different ways. If two agents a_i and a_j report different trust values about an agent a_k, it does not necessarily mean that one of them is attacker spreading false information. It may be that a_i and a_j assess the agent a_k with different assessment function, i.e. with different QADE operator. However, the agent a_i may treat trust opinions from other agents distinctly. In the real world example, people treat opinions of those who are complaining about everything differently from those who are enthusiastic about everything.

In QADE trust model, the agents define how similar they are with other agents in the society. The agent a_i perceives another agent a_j as similar if they assess trust in a same way, i.e. the agents a_i and a_j use the same operator for trust evaluation. Using the same operator, the agent's a_i opinions about other agents are similar to the agent's a_j opinions about these agents and they have similar general trust attitude towards society.

The similarity computation between agents a_i and a_j is twofold: it considers *pairwise similarity* of trust opinions about other agents and *similarity of general mindset* of agents a_i and a_j that reflex their trust assessment operators.

Extension #7: Definition 12. The general mindset of agent a_i is given by a normalized histogram of the distribution of trust values in the private trust vector Z_i:

$$htg(Z_i) = \left(htg_i^{[-2]}, htg_i^{[-1]}, htg_i^{[0]}, htg_i^{[1]}, htg_i^{[2]} \right), \tag{16}$$

where $htg_i^{[-2]}, htg_i^{[-1]}, htg_i^{[0]}, htg_i^{[1]}$ and $htg_i^{[2]}$, respectively, denote relative frequency of distrusted, partially distrusted, undecided, partially trusted and trusted relationship values towards other agents in the society, respectively.

Extension #8: Definition 13. The similarity function $sim : \underline{Z_i} \times \underline{\mathcal{M}_{j,n1}} \to [0,1]$ is defined as follows:

$$s_{i,j} = 1 - \sqrt{\sum_{\substack{\zeta_{i,k} \in \underline{Z_i} \\ \omega_{j,k} \in \underline{\mathcal{M}_{j,n1}}} \frac{\left(\left(\zeta_{i,k} - \omega_{j,k} \right) \cdot \left(\left(1 - htg_i^{\left[\omega_{j,k} \right]} \right) + \frac{\left(\left| \overline{\zeta_i} - \overline{\omega_j} \right| \right)}{maxDst} \right) / 2 \right)^2}{\min\left(\left| \underline{Z_i} \right|, \left| \underline{\mathcal{M}_{j,n1}} \right| \right) \cdot maxDst^2}}, \tag{17}$$

where $\overline{\zeta_i} = \frac{1}{\left| \underline{Z_i} \right|} \sum_{\zeta_{i,k} \in \underline{Z_i}} \zeta_{i,k}$, $\overline{\omega_j} = \frac{1}{\left| \underline{\mathcal{M}_{j,n1}} \right|} \sum_{\omega_{j,k} \in \underline{\mathcal{M}_{j,n1}}} \omega_{j,k}$ and $maxDst = \max(\Omega) - min(\Omega)$.

The similarity between agents a_i and a_j is computed as weighted Euclidean distance between agent's a_i private trust vector and agent's a_j public trust vector, where weights are defined such that they fit agent's a_i general mindset. Euclidean distance in mathematics represents the most fundamental distance metrics. It was chosen as a basis for our similarity function definition. In the proposed similarity function, the Euclidean distance between trust vectors represents *pairwise similarity*, as it computes the pairwise distance between pairs $\left(\zeta_{i,k}, \omega_{j,k} \right)$, where $\zeta_{i,k}$ is an element from the agent's a_i private trust vector and $\omega_{j,k}$ is an element from the agent's a_j public trust vector. Further, weights have been introduced in order to regulate effects of certain pairwise comparisons on the overall distance. The similarity function sim reduces the distance between compared trust values if the value $\omega_{j,k}$ fits the agent's a_i general mindset. If this value does not fit its general mindset, the distance between the compared trust values is not reduced. The weight consists of two parts: $\left(1 - htg_i^{\left[\omega_{j,k} \right]} \right)$ and $\frac{\left(\left| \overline{\zeta_i} - \overline{\omega_j} \right| \right)}{maxDst}$. The first part of the weight is proportional to the relative frequency of value $\omega_{j,k}$ in the general mindset of the agent a_i. If this value does not exist in the agent's a_i general mindset, it means it does not correspond with its general mindset. In that case, the value of the first part of the weight is the highest possible, i.e. 1. On the other hand, if the relative frequency of the value $\omega_{j,k}$ in the agent's a_i general mindset is high, i.e. meaning that such value corresponds with its general mindset, then the value of the first part of the weight is less then 1. The closer the value is to zero, the smaller the distance between the compared trust vectors. The second part of the weight reduces the distance accordingly to the tendency of the agents' general mindsets towards positive or negative values. It compares the average trust rating in the agent's a_i private trust vector with the average trust rating in the agent's a_j public trust vector. The smaller distance between average values means the smaller weight value. As such, the weights imply the measurement of agents *general mindset similarity*. Finally, the distance value is normalized in order to achieve value between 0 and 1. Smaller distance between agents' vectors means greater similarity, as the function sim computes similarity as the opposite

value of the distance value, to which value 1 is added in order to achieve value on the scale $\left[0,1\right]$. The similarity value $s_{i,j}$ closer to 1 means greater similarity between agents a_i and a_j, while the value closer to 0 means lower similarity between them.

For example, let us compare agent *A* with agents *B* and *C*. The agents have properties as described in Table 1.[4] The histograms representing general mindset of the agents are given in Figure 1. First, let us compute normalized Euclidean distance (*NED*) between trust vectors. The results $1-NED\left(A,B\right)=0.79$ and $1-NED\left(A,C\right)=0.80$ imply that the agent *A* perceives the agent C more similar to itself than agent *B* considering only pairwise comparisons of elements from trust vectors. Further, let us consider general mindsets of agents *A*, *B* nd *C*, too. The agents A and B assess trust with moderate-pessimistic trust operator ($\downarrow$) and they are inclined to more negative trust assessment, i.e. assigning (partially) distrusted values. The agent C assesses trust with centralistic consensus seeker trust operator ($\rightsquigarrow$), meaning that it leans to trust assessments around undecided trust value (0). Thus, the pairwise comparisons of trust values $\zeta_{A,j}=1$ with $\omega_{B,j}=0$ and $\omega_{C,j}=2$ (bold valus) indicate semantic difference, although the both values are one level apart from $\zeta_{A,j}$. The value $\omega_{C,j}=2$ implies that the agent's *C* trust value distribution includes trusted trust relationships and therefore does not fit into the agent's *A* general mindset. The proposed similarity function *sim* reduces the distance between elements if the element of the compared agent's trust vector corresponds to comparing agent's general mindset. Computing the similarity among agents *A*, *B* and *C* with the proposed similarity function *sim* results in the following similarity values: $s_{A,B}=0.85$ and $s_{A,C}=0.77$. Now, the agent *A* perceives the agent *B* more similar to itself than the agent *C*.

Table 1. Properties of agents A, B and C

Agent	Operator	Trust values
A	$\downarrow$	$Z_A=\left(0,-1,-2,-2,0,-2,0,\mathbf{1},0,1,0,-1,-1,0,-2,0,-2,-2\right)$
B	$\downarrow$	$\mathcal{M}_{B,n}=\left(1,-2,-2,-1,1,-1,1,\mathbf{0},1,0,1,-2,-2,0,-2,0,-2,-1\right)$
C	$\rightsquigarrow$	$\mathcal{M}_{C,n}=\left(1,-1,-1,-1,0,-1,1,\mathbf{2},1,1,0,-1,-1,0,-1,1,-1,-1\right)$

Extension #9: Definition 14. The similarity factors among agents are given by similarity matrix $\mathcal{S}$, where elements s_{ij} define the similarity between agents a_i and a_j, measured by agent a_i.

A general form of similarity matrix of a certain society with *n* agents is as follows:

$$\mathcal{S}=\begin{bmatrix} s_{1,1} & \cdots & s_{1,n} \\ \vdots & \ddots & \vdots \\ s_{n,1} & \cdots & s_{n,n} \end{bmatrix}. \tag{18}$$

Note that values $s_{i,j}$ and $s_{j,i}$ may not be equal, since $s_{i,j}$ defines similarity between agents a_i and a_j based on the agent's a_i private vector Z_i; and $s_{j,i}$ defines similarity between agents

[4] The data are taken from simulations, which are described in section "Simulation results".

a_i and a_j based on the agent's a_j private vector Z_j. If the agents in the society apply the same operator and if there are not any fraudulent agents in the society, then the similarity matrix $\mathcal{S}$ is symmetric. In other case, it may not be.

The QADE trust model is a mathematical formalization of trust evaluation in e-commerce environment. It defines trust as "a relationship between two agents" and it does not specify concrete factors that define the relationship value. For example, one might evaluate the trust of a seller in an online marketplace based on different factors, such as product price, quality and/or delivery time. As another example, one might evaluate the trust of peers in file trading P2P network based on percentage of valid files received and/or their availability. Trust evaluations may differ according to agents' subjective considerations, such as their benevolence, honesty, willingness or faith, related risk, etc. The QADE trust model presents a generalized theoretical framework and includes an *overall* trust definition that is applicable to various contexts. The QADE trust model can be integrated with existing trust formalizations where agents evaluate trust between each other irrespectively of *how* they evaluate it under the condition that the factors determining trust must be aggregated into a single value. For example, if trust between a trusting agent (user) and a trusted agent (mobile application) is defined based on the number of usages, elapsed usage time, usage frequency and experienced features with respect to certain context (Yan *et al.* 2013), these factors must be aggregated into an aggregate value in order to apply the QADE trust model. If such context-specific trust formalization does not propose its own aggregation method, a weighted sum or any other aggregation method can be used. However, in the QADE trust model the method for trust factors aggregation is *not* prescribed.

In the QADE model, a five-level scale is proposed to assess trust with the following values: trusted (2), partially trusted (1), undecided (0), partially distrusted (–1) or distrusted (–2) relationship value. Representation of trust assessment with discrete values is less accurate and less expressive than representation with real numbers or with probabilistic distributions.

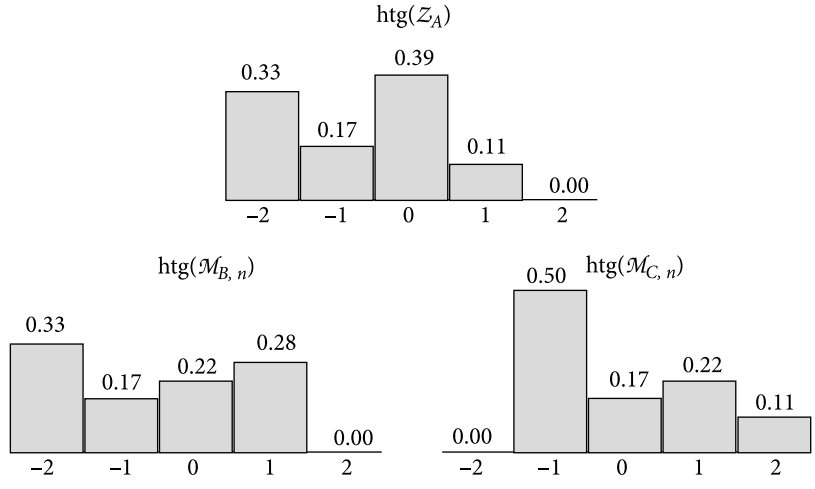

Fig. 1. Example: histograms of trust values distributions for agents A, B and C

However, a study (Trcek 2009) shows that humans prefer expressing trust values on a qualitative scale. The model was designed on the assumption that humans (or, software agents managed by human users) will use it, therefore the five-level scale has been proposed.

Further, the QADE model defines *general mindset* and *similarity function* based on the proposed five-level scale. However, the proposed scale may be extended or reduced with a simple modification of general mindset and similarity function definitions. Other definitions are independent from the proposed five-level scale. Or, vice versa, if another trust formalization has different trust values representation than QADE trust model, the trust values from another formalization could be translated. For example, if another (context-specific) trust formalization describes trust value (denoted with x) with continuous value on scale $[0,1]$, a simple value conversion would be as follows: $x \in [0,0.2)$ translates to -2, $x \in [0.2,0.4)$ translates to -1, $x \in [0.4,0.6)$ translates to 0, $x \in [0.6,0.8)$ translates to 1 and $x \in [0.8,1]$ translates to 2. An analytical study of trust value conversion, including estimation of uncertainty involved in the conversion, can be found in (Pinyol *et al.* 2007). Note that the QADE trust model does *not* propose a value conversion method.

The proposed QADE trust formalization serves as a basis for implementation of trust and reputation management systems (TRMS). The TRMSs are not standalone systems, but they extend existing e-commerce systems. The trust and reputation management systems could be implemented as distributed or centralized systems (Josang *et al.* 2007; Hoffman *et al.* 2009). However, the system model of a trust and reputation management system is not part of the presented research work.

3. Unfair ratings

The problem of unfair rating is considered when an agent in e-commerce system reports trust assessment value in the public trust matrix $\mathcal{M}$, which does not reflect its real experience. Typical attacks based on reporting of false trust values, which were studied in various experiments (Dellarocas 2000; Yang *et al.* 2009; Yu, Singh 2003; Kerschbaum *et al.* 2006), are as follows: "ballot-stuffing attack" and "bad-mouthing attack". Ballot-stuffing attack means that the users report unfairly *high* trust values about trading partners they interacted with, irrespective of the real experiences. This will inflate trusted partners' reputation. Bad-mouthing attack means that the users provide unfairly *low* ratings to partners agents they had interactions with, which will lower their reputation.

The attacks can be further classified into two categories: individual user attacks and collaborative user attacks (Yang *et al.* 2009). In the first type of attacks, an independent user reports unfairly high or low opinions about others. In the second type of attacks, a group of users provides unfair trust values to downgrade or to boost the reputation of the targeted group of other agents in an e-commerce system.

Defense technique

An agent evaluates trust towards other agents, such that it aggregates reported trust assessments from trust matrices $\mathcal{M}^{-l}$ and values from private trust vectors Z^{-l}. Relative frequencies

of different trust values represent the agent's *representative* trust attitude as shown in Figure 2.[5] Hence, the unfairly reported trust values could deform its trust assessments and *skew* the representation of its trust attitude, as seen in Figure 2. The agent must be able to detect falsely reported opinions and ignore or adjust them in order to reduce their effect on trust aggregation and thus *correct* the representations of its trust attitude, as shown in Figure 2.

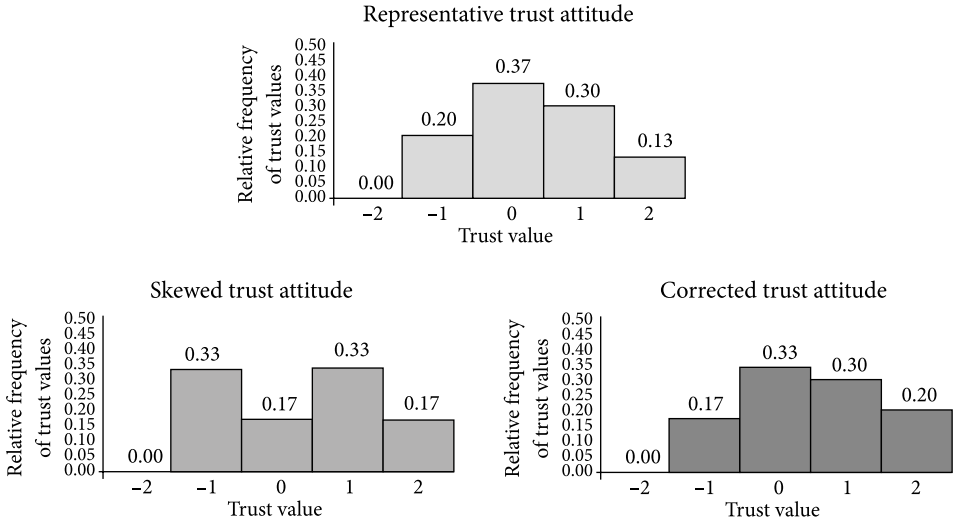

Fig. 2. An agent's representative, skewed and corrected trust attitude

In general, there are two basic approaches for handling with opinions judged to be false: *endogenous* and *exogenous* methods (Josang *et al.* 2007). The former exclude presumed unfair ratings based on statistical comparison of reported values alone, while the latter use the externally determined reputation of the rater to determine if the values are fairly or unfairly reported.

Our solution is both endogenous and exogenous. It is exogenous because it finds similar agents and consider them as reputable source of information. Further, the agents compare their trust values with trust values of similar agents – in that is our proposed method endogenous. A trust evaluation method filtering out presumed unfair ratings is defined as described in Algorithm 1 (**Extension #10**). Algorithm 1 is described in Table 2. It contains four main functions: *deriveGeneralMindset(agent), computeSimilarity(agent, agent, generalMindset), computeTrustWithQADEOp(agent, agent, operator, assessmentSet)* and *reportToPublicTrustMatrix(trustValue, agentFraudulence).* Algorithm 1 describes the procedure for calculating trust value between a trusting agent a_i and a trusted agent a_j. Firstly, the trusting agent a_i derives its general mindset from the private trust values. The function *deriveGeneralMindset(agent)* implements the derivation of an agent's general mindset as defined in Definition 12. In the next step, the agent a_i computes its similarity with other agents that participate in e-commerce system. To compute similarity, the agent a_i computes the distance between its

[5] The histograms are taken from simulation results, which are presented in section "Simulation results".

private trust values with public trust values from other agents in e-commerce system and weights it accordingly to its general mindset. The function *computeSimilarity(agent, agent, generalMindset)* implements similarity function *sim* that is defined in Def. 13. Next, the trusting agent a_i assesses trust value towards a_j regarding its QADE operator, whereby only trust assessments of similar agents are included in the trust computation. Trust value computation is implemented with *computeTrustWithQADEOp(agent, agent, operator, assessmentSet)* function. Based on its operator, the agent a_i computes trust towards a_j such that the trust value fits the agent's representative trust attitude. The agent a_i stores the real trust assessment $\zeta_{i,j}$ in its private vector. At least, it reports trust assessment $\omega_{i,j}$ in the trust matrix $\mathcal{M}$. Note that the real trust assessments computed with different QADE operators may differ. However, the agent a_i could also report false value intentionally. The agent reports fairly or unfairly, depending on its fraudulence. The function *reportToPublicTrustMatrix(trustValue, agentFraudulence)* implements the possible alternation of computed trust value and storage of (alternated) trust value in the trust matrix.

Table 2. QADE filtering algorithm (Extension #10)

Algorithm 1 Assess trust value excluding unfair ratings (QADE filter)

Input: trusting agent a_i , trusted agent a_j
$htg\left(Z_i\right) \leftarrow deriveGeneralMindset(a_i)$ //as defined in Def. 12
for $a_k \in \mathcal{A}, a_k \neq a_i$ **do**
$s_{i,k} \leftarrow computeSimilarity\left(a_k, a_i, htg\left(Z_i\right)\right)$ //as defined in Def. 13
end for
for all $s_{i,k} \geq thSim$ **do**
add $\underline{\omega_{k,j}^{0:-t}}$ to assessments set *AS*
end for
add $\zeta_{i,j}^{0:-t}$ to assessments set *AS*
$\zeta_{i,j}^{+} \leftarrow computeTrustWithQADEOp\left(a_i, a_j, op_i, AS\right)$ //as defined in Extension #5
$\omega_{i,j}^{+} \leftarrow reportToPublicTrustMatrix\left(\zeta_{i,j}^{+}, \overline{a}_i\right)$

4. Simulation results

In order to evaluate the proposed QADE model and the filtering algorithm, a simulation tool was implemented based on event-driven web services (Juric 2010; Potocnik, Juric 2013) and cloud infrastructure (Dukaric, Juric 2012). The performance of our filtering algorithm was compared with an exogenous filtering technique proposed in TRAVOS model (Teacy *et al.* 2006), referred to as TRAVOS filter in the rest of the paper; and with an endogenous filtering method proposed by Whitby *et al.* (2004), referred to as Whitby filter. The Whitby filter handles unfair ratings such that it filters out those ratings that are not in the majority amongst other ones. Trust values of the agents whose trust values are significantly different from average are rejected from trust computation. TRAVOS proposes a method to filter out inaccurate values, where an agent that provides trust value is judged on the perceived accuracy of its previously provided trust values. Based on this judgment, trust values providing agent's influence on a

trusting agent's assessment of a trusted agent is reduced. The comparison of the filtering techniques effectiveness was performed, i.e. how they reduce the effect of inaccurate trust values. The trust value was computed according to QADE trust model.

The basic simulation environment is as follows. The simulated society consists of n participating agents. The agents represent buyers and sellers in an e-commerce system. They have assigned operators and trust relationships between them. At the beginning, the trust relationship among agents are undefined, i.e. $\omega_{i,j} = \text{“–”}, \forall i, j$. Some agents in the system are good and the remaining are bad (the difference between these two types of agents is explained later in this paragraph). Further, each agent could act in both ways – as a seller (service provider) or a buyer (service requester). There are two agents involved in a single transaction, where one of them requests a service and another one provides a service. After the transaction, the service requester evaluates the service provider. The simulation runs in discrete time steps and in each step an agent a_i (as service requester/trusting agent) and an agent a_j (as service provider/trusted agent) are randomly selected. The selected agents conduct a transaction that defines the new trust value between them. If a_j is a good agent that provides a high quality service, then the agent's a_i new trust value towards a_j is 0, 1 or 2, which represent undecided, partially trusted or trusted relationship value. Otherwise, if a_j is a bad agent and provides low quality service, then the agent's a_i new trust value for a_j is –2, –1 or 0, which represent distrusted, partially distrusted or undecided relationship value. The trust values are selected randomly in each transaction in order to simulate the dynamics of the service provider, i.e. variations in quality of services provided over time and depending on a service requester. The trust value, which is obtained as an outcome of transaction, is then aggregated with other trust assessments about the agent a_j. The aggregation of the trust values is based on agent's a_i QADE operator. After computations, the agent a_i stores the trust value in private and public trust matrix. Next, some agents in the society are attackers (*att*) and report false trust values about other agents, which are referred to as targeted agents (*trg*). An attacker agent reports different trust assessment about a targeted agent in public trust matrix as it stores in private trust vector. To avoid taking false opinions into trust computation, the agent a_i considers only opinions from similar agents, which means that similarity factor $s_{i,\star}$ is above similarity threshold (*thSim*).

Environment with agents who perceive and evaluate trust in different ways have been simulated. The agents have assigned different kind of QADE operators. For example, agents that apply $\uparrow$ (moderate-optimistic assessment operator; refer to Eq. 13) assess trust in other agents with higher values than agents that apply $\downarrow$ (moderate-pessimistic assessment operator; refer to Eq. 14), despite the same observations, i.e. trust ratings reported by other agents. With application of different operators, the subjective nature of trust is reflected. The operators are assigned to agents, such that there are subsequent amount of different operators: 16.67% of $\Uparrow$, 16.67% of $\Downarrow$, 16.67% of $\uparrow$, 16.67% of $\downarrow$, 16.67% of $\rightsquigarrow$ and 16.67% of $\leftrightarrow$. The simulated society consists of 50% of good agents and 50% of bad agents. The similarity threshold level is set to *thSim* = 0.7. The simulations have been run with different number of attackers and targeted agents, as described in Table 3. Each simulation configuration has been executed with 300 agents and for 10000 time steps.

Table 3. Percentage of attackers and targeted agents in conducted simulations

Configuration	Att	Tar	Trust computation
Representative	0%	0%	QADE
Bad-mouthing attack (individual)	5%–50%	20%	QADE, QADE filter, TRAVOS filter, Whitby filter
	20%	5%–50%	
Bad-mouthing attack (collaborative)	5%–50%	20%	QADE, QADE filter, TRAVOS filter, Whitby filter
	20%	5%–50%	
Ballot-stuffing attack (individual)	5%–50%	20%	QADE, QADE filter, TRAVOS filter, Whitby filter
	20%	5%–50%	
Ballot-stuffing attack (collaborative)	5%–50%	20%	QADE, QADE filter, TRAVOS filter, Whitby filter
	20%	5%–50%	

The purpose of simulations has been to evaluate our proposed filtering method. "Representative" configuration simulation run was executed in order to capture agents' behavior with different QADE operators assigned. After 10 000 steps, the statistical properties of the distribution of trust values was observed, such as relative frequencies of –2, –1, 0, 1 and 2 trust values, mean, median, maximum and minimum values, standard deviation, variance, skewness and kurtosis. There were no unfairly reported opinions. The resulting behavior is considered as *agents' representative (real) trust attitude* and it reflects their general mindsets. Next, attackers, i.e. agents that report false values, were included in simulations. The configurations with varying numbers of attackers and targeted agents were performed, in which trust was calculated in two ways, firstly without filtering out alleged false values; and then with filtering. Statistical properties of trust values distribution for trust computation without filtering were computed, referred to as *skewed trust attitude* and for trust computation with different filtering methods, referred to as *corrected trust attitude*. Statistical parameters between representative, skewed and corrected trust values distribution were compared in order to compare effectiveness of filtering methods. Next, the percentage of detected unfair trust ratings for each filtering method was computed. Based on that, the performance comparison of Whitby filter, TRAVOS filter and QADE filter was performed.

Sections "Bad-mouthing attack" and "Ballot-stuffing attack" describe further simulation parameters specific to each set of experiment and present the results.

4.1. Bad-mouthing attack

"Bad-mouthing attack (individual)" and "Bad-mouthing attack (collaborative)" configurations with different number of attackers and targeted agents were executed. The number of attackers and targeted agents ranged between 5% and 50%. Altogether 300 different agents' trust assessments distributions were collected. More specifically, 50 trust assessment distributions for each type of QADE trust operators were collected, as there are six different QADE trust operator, which are distributed equally among the agents. The difference between representative trust attitude and skewed trust attitude, and the difference between representative trust attitude and corrected trust attitude for every agent was computed. In the second part, the

percentages of unfair trust ratings that were recognized and excluded from trust computation for different filtering methods were compared.

Figure 3 shows the mean trust value in the society with varying numbers of attackers after 10 000 steps. As the simulations are performed in the society with uniform distribution of QADE operators and there are 50% of good agents and 50% of bad agents, the mean value in the scenario without attack is around 0. In our simulations, mean value equals 0.01. Introducing the attackers into the society that report unfairly low trust values results in lower mean value. It ranges from –0.11 in the society with 5% of attackers to –0.45 in the society with 50% of attackers in the (individual) bad-mouthing attack scenario and from –0.11 to –0.24 in the collaborative bad-mouthing attack scenario. The difference between representative trust values distribution and skewed trust values distribution is smaller in the collaborative attack scenario, as there is limited number of values that could be unfair – only trust assessments of 20% of targeted agent could be unfair. In the configuration with bad-mouthing attackers, the distance between representative trust attitude and corrected trust attitude is the smallest in the scenario with QADE filter and the biggest when Whitby filter is used, as seen in Figure 3.

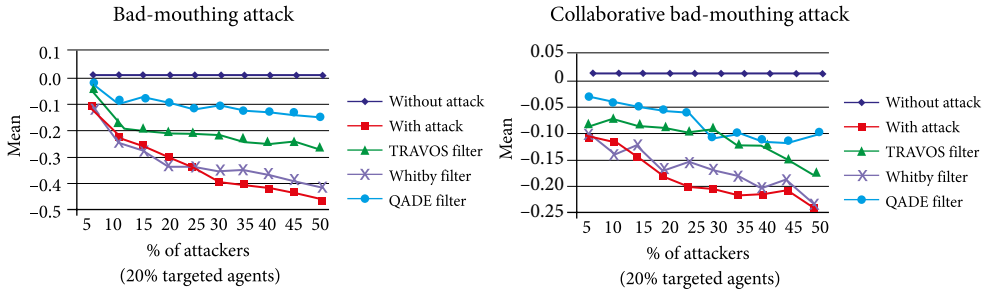

Fig. 3. The mean trust value in the "Bad-mouthing attack" configurations with different number of attackers

Figure 4 shows the mean trust value in the society with constant number (20%) of attackers and varying number of targeted agents after 10 000 steps. Bad-mouthing attackers decrease the mean of trust values distribution, such that it ranges from –0.11/–0.03 in the society with 5% targeted agents to –0.49/–0.50 in the society with 50% targeted agents in the bad-mouthing/collaborative bad-mouthing attack scenario. As seen in Figure 4, the difference between corrected trust values distribution and representative trust distribution is the smallest when using our QADE filter. For example, in the society with 50% of targeted agents and 20% of collaborative bad-mouthing attackers, the mean of trust values distribution equals –0.15 when using QADE filtering method, –0.24 when using TRAVOS filter and –0.40 with filtering technique proposed by Whitby, which is 17% and 49% improvement of QADE filter in comparison with TRAVOS and Whitby filter, respectively.

The distance in the mean difference between representative, skewed and corrected trust attitude on the [0, 1] scale was measured. The distance between the representative trust attitude (in scenarios without attackers) and skewed trust attitude (in scenarios with attackers) equals $dist(repr, skew) = 1$, for every configuration with varying percentage of attackers

and targeted agents, and represents "total effect" of unfair ratings. The purpose of filtering out unfair ratings is to reduce the effect of such ratings, i.e. to "correct" the trust attitude to resemble the representative trust attitude. It follows, the smaller $dist(repr, corrFilter)$ value means the better filtering and 0 value means "no effect" of unfair ratings. On average, distance between representative and corrected trust attitude (using different filtering methods) equals $dist(repr, corrQADE) = 0.35$, $dist(repr, corrTRAVOS) = 0.64$ and $dist(repr, corrWhitby) = 0.95$ in the "Bad-mouthing attack (individual)" configurations, which is 29% and 61% improvement, respectively. In the "Bad-mouthing attack (collaborative)" configurations, it equals $dist(repr, corrQADE) = 0.42$, $dist(repr, corrTRAVOS) = 0.61$ and $dist(repr, corrWhitby) = 0.87$. This represents 20% and 45% improvement of our filtering approach in comparison with TRAVOS and Whitby filter and clearly indicates the effectiveness of QADE filtering algorithm. The QADE filtering method significantly outperforms other two in the scenario with individual and collaborative bad-mouthing attackers.

Further, the amounts of unfair trust ratings recognized with different filtering algorithms were compared. Figure 5 shows the percentage of detected unfair trust ratings in the society with constant number (20%) of targeted agents and varying number of attackers, while Figure 6 shows the percentage of detected unfair trust ratings in the society with constant number (20%) of attacker agents and varying number of targeted agents. The results in Figure 5 and Figure 6 show the average percentage of recognized unfair trust ratings for 10 000 trust evaluations, i.e. one trust evaluation between two different agents in each time step.

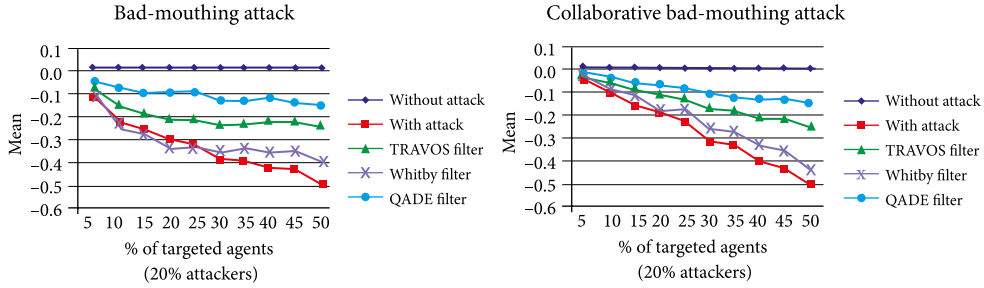

Fig. 4. The mean trust value in the "Bad-mouthing attack" configurations with different number of targeted agents

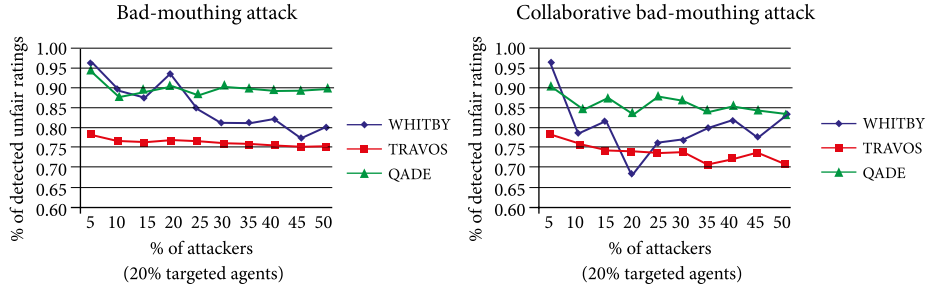

Fig. 5. The percentage of detected unfair ratings in the "Bad-mouthing attack" configurations with different number of attackers

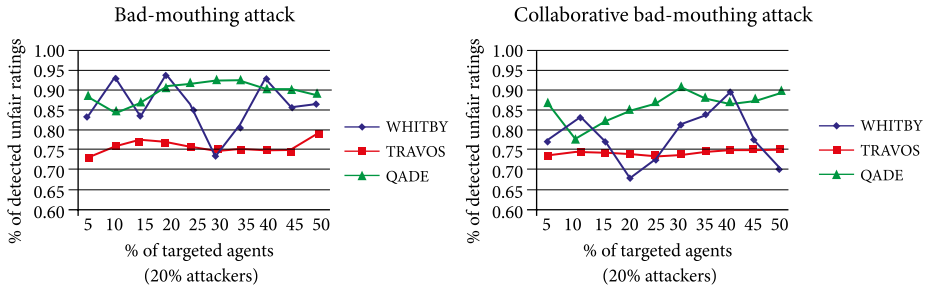

Fig. 6. The percentage of detected unfair ratings in the "Bad-mouthing attack" configurations with different number of targeted agents

As seen in Figure 5 and Figure 6, TRAVOS filter identified 73% to 79% and 70% to 78% unfairly reported trust assessments in the scenarios with individual bad-mouthing attackers and collaborative bad-mouthing attackers, respectively. QADE filter recognized 85% to 95% unfair trust ratings in configurations with (individual) bad-mouthing attackers and 78% to 91% unfair trust assessments in configurations with collaborative bad-mouthing attackers. Whitby filter detected 74% to 96% unfair ratings in "Bad-mouthing attack (individual)" and 68% to 97% unfair trust ratings in "Bad-mouthing attack (collaborative)" configurations. The performance of Whitby filter varied strongly between the configurations. The Whitby filtering method excludes trust assessments that are not in majority amongst provided trust ratings. For example, if there are many transactions between agents that tend to assess trust with low values (i.e. agents with extreme-pessimistic and moderate-pessimistic assessment operator), the unfairly low trust values are "hidden" in majority of other (fairly low) trust assessments. Therefore, Whitby filter lacks stability, which can be seen in Figure 5 and Figure 6.

4.2. Ballot-stuffing attack

"Ballot-stuffing attack (individual)" and "Ballot-stuffing attack (collaborative)" configurations with different number of attackers and targeted agents (varying from 5% to 50%) were executed. As previously described, the differences between representative, skewed and corrected behavior, and differences in percentage of unfair trust ratings that were correctly detected for each configuration were compared.

The means of trust values in the society with varying number of attackers and constant number of targeted agents after 10000 time steps are shown in Figure 7. The mean value in the configuration without attack is the same as in the previous set of simulations, i.e. it equals 0.01. The mean value moves towards more positive trust values in the societies with ballot-stuffing attackers, as they unfairly report too high trust values. It ranges from 0.10 if there are 5% attackers to 0.46 if there are 50% attackers in the society in the case of ballot-stuffing attack. In the collaborative ballot-stuffing attack scenario, it ranges from 0.10 for the society with 5% attackers to 0.24 in the society with 50% attackers. As seen in the Figure 7, the best corrected trust values distribution was achieved with QADE filtering technique, followed by TRAVOS and Whitby filtering method, in both ballot-stuffing and collaborative ballot-stuffing attack scenarios.

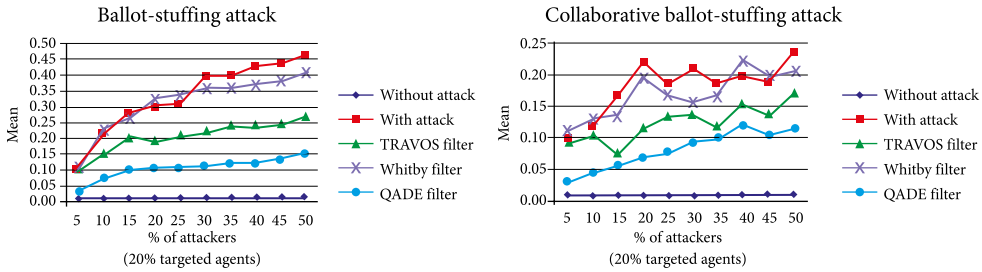

Fig. 7. The mean trust value in the "Ballot-stuffing attack" configurations with different number of attackers

Figure 8 shows the mean of trust value distributions in the society with constant number (20%) of attackers and varying number of targeted agents after 10 000 time steps. The mean increases due to unfairly high reported trust values, such that it ranges from 0.09/0.05 in the society with 5% targeted agents to 0.50/0.47 in the society with 50% targeted agents in the ballot-stuffing/collaborative ballot-stuffing attack scenario. The difference between representative distribution of trust values and distribution of corrected trust values is the best in the scenario with QADE filtering method, as seen in Figure 8.

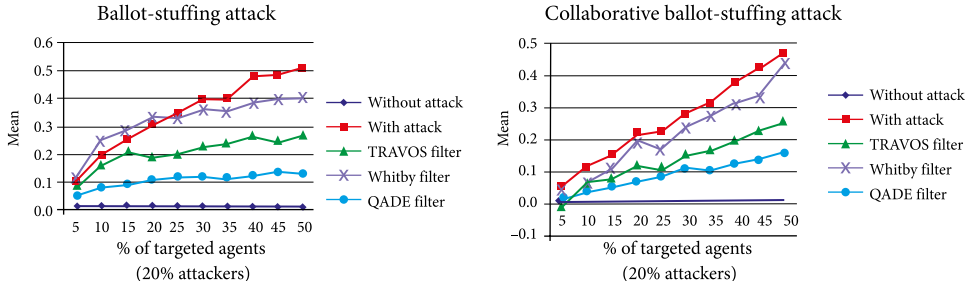

Fig. 8. The mean trust value in the "Ballot-stuffing attack" configurations with different number of targeted agents

On average, distance between representative and corrected trust attitude equals $dist(repr, corrQADE) = 0.30$, $dist(repr, corrTRAVOS) = 0.64$ and $dist(repr, corrWhitby) = 0.97$ in the "Ballot-stuffing (individual)" configurations, which is 34% and 67% improvement, respectively. It equals $dist(repr, corrQADE) = 0.34$, $dist(repr, corrTRAVOS) = 0.55$ and $dist(repr, corrWhitby) = 0.88$ in the "Ballot-stuffing (collaborative)" configurations, representing 21% and 54% improvement of our proposed filter. QADE filtering technique outperforms TRAVOS method and Whitby's filtering method in both scenarios with individual and collaborative ballot-stuffing attackers.

Figure 9 shows the percentage of detected unfair trust ratings in the society with constant number (20%) of targeted agents and varying number of attackers. Figure 10 shows the percentage of identified unfair trust ratings in the society with constant number (20%) of

attacker agents and varying number of targeted agents. Whitby filter identified and excluded 73% to 96% and 76% to 96% of all unfair trust ratings in configuration with individual and collaborative ballot-stuffing attackers, respectively. As explained in previous section, Whitby filter lacks stability, which holds true also for ballot-stuffing and collaborative ballot-stuffing attack scenarios. TRAVOS filter detected 74% to 78% unfair trust ratings in "Ballot-stuffing (individual)" configurations and 69% to 76% unfair trust values in "Ballot-stuffing (collaborative)" configurations. Our QADE filter identified 86% to 90% unfairly reported trust assessments in configurations with (individual) ballot-stuffing attackers and 81% to 88% unfair trust ratings in configurations with collaborative ballot-stuffing attackers. Based on above results, it follows that QADE filter algorithm outperforms the other two filtering algorithms in identification of unfair trust ratings.

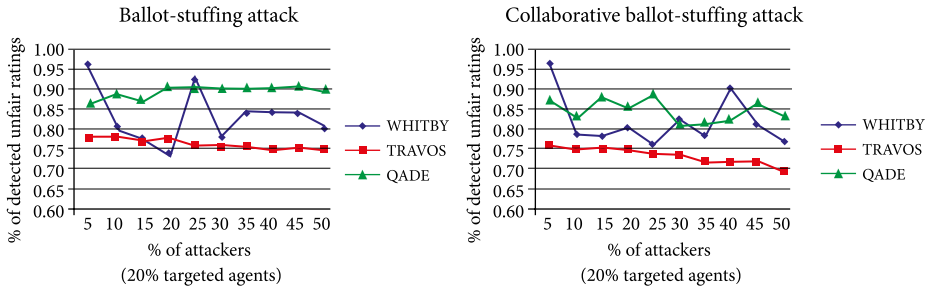

Fig. 9. The percentage of detected unfair ratings in the "Ballot-stuffing attack" configurations with different number of attackers

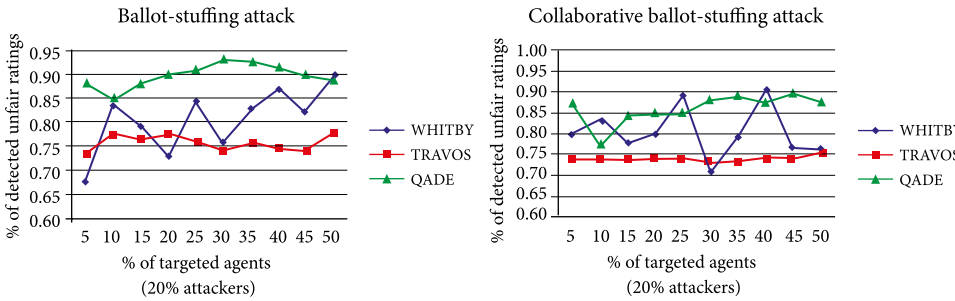

Fig. 10. The percentage of detected unfair ratings in the "Ballot-stuffing attack" configurations with different number of targeted agents

The simulation results showed that the proposed QADE filter algorithm corrects agents' trust attitude significantly more successful than the most representative endogenous filtering technique proposed by Whitby *et al.* and the most representative exogenous filtering method proposed in TRAVOS model, with average improvement of 57% over Whitby and 26% over TRAVOS filter. It holds true for the society with individual and collaborative ballot-stuffing agents as well as for the society with individual and collaborative bad-mouthing attackers. Additionally, QADE filtering method outperforms Whitby's and TRAVOS filtering technique in both cases: with varying number of attackers and with varying number of targeted agents.

5. Related work

Formalizations of trust in computing environment, along with techniques for managing unfair trust values, have recently attracted the attention of the information science research community. They presented various trust models and unfair ratings management methods that are discussed in this section.

Common approach to build trust models and to handle unfair trust ratings is to employ statistical methods. Whitby *et al.* (2004) propose trust model based on Bayesian Reputation Systems (BRS) with a statistical filtering technique for excluding unfair ratings. They extend BRS to filter out those ratings that are outside the q% quantile and (1 – q)% quantile of the majority opinion, whereby the feedback provided by others is represented by a beta distribution. They do not assume that more positively or more negatively rating could be a result of raters' different trust attitude. In contrast, our approach does not have this limitation.

Teacy *et al.* (2006) propose the TRAVOS model, which is based on the beta probability density function. Their method for handling unfair opinions is two-part. In the first stage, they estimate the probability of accuracy of reported opinions, based on comparison of trusting agent's and reporter agent's beta distribution functions that are constructed using outcomes of all previous interactions with trusted agent. In the second stage, based on this value, they modify opinions given by reporter agent according to its accuracy. In contrast to them the QADE trust model is defined on the assumption that different trust values can also be a result of different socio-cognitive processes that trust is driven by and are therefore handled in such way.

Noorian *et al.* (2011) also adopt probability theory to build trust model and introduce similarity metric to handle unfair trust values. They propose a two-layered filtering algorithm that combines cognitive and probabilistic view of trust. As such, authors allow human dispositions (optimism, pessimism and realism) to be incorporated into trust evaluations. In the first layer, they filter out dishonest trust values according to similarity between agents. In the second layer, they further discount trust values according to behavioral tendencies of agents. However, the proposed two-layered filtering algorithm does not assume (in the first layer) that contradictory trust values towards a certain agent might origin from their different general mindsets. Therefore, such ratings might be falsely identified as dishonest ratings and filtered out from further trust computation. In contrast, our approach does not have these limitations as it considers agents' general mindsets when comparing trust values provided by them.

The similarity concept has been used for collaborative filtering and its applications for recommender systems (Su, Khoshgoftaar 2009; Terveen, McDonald 2005). The general idea of recommender systems is to find other people who have similar preferences or interests and use this information to predict the likeability for further items. Our approach differs in that it aims to find agents who have similar general mindsets/trust attitudes, which means that they perceive and assess trust in similar ways, by which their preferences or interests may differ.

The elements from collaborative filtering have been used in approach proposed by Dellarocas (2000). His model uses collaborative filtering technique to find the nearest neighbors of a trusting agent based on their preference similarity with the trusting agent on commonly

rated trusted agents. Unfairly high ratings are then filtered out using cluster filtering approach, which divides neighbors' ratings into the lower and the higher rating cluster. Ratings in the higher rating cluster are considered as unfairly high ratings provided by ballot-stuffing attackers, and therefore are excluded. In Dellarocas' model, similarity between agents is based on commonly rated agents. In our approach, similarity is based on pairwise similarity of trust values towards commonly rated agents and comparison of general trust attitudes of compared agents.

Chen and Singh (2001) also use the elements from collaborative filtering. Their method computes reputation of raters in three steps. First, they compute quality and confidence values of rater's trust values for each object in the category, referred to as local match and local confidence. Then they compute the cumulated quality and confidence values of all trust values for each category of objects, referred to as global match and global confidence. Finally, they compute the rater's reputation based on the rater's global match and global confidence for each category. Opinions from less reputed raters are considered with less weight. Contrary to our approach, their method is highly objective and they do not consider subjectivity of trust phenomenon.

Tavakolifard *et al.* (2009) presented trust management based on similarity. In Tavakolifard *et al.* (2009), similarity is defined based on an agent network, in terms of traditional friend-of-a-friend (FOAF) network. Their addition to traditional FOAF network is in that two agents are similar either if they trust the same agents or if they are trusted by the same agents. Although using similarity, this approach is different to our QADE trust model. In our approach, pairwise similarity is computed based on trust assessments about an agent in which trust is computed, and based on similarity of trusting agents' general mindset. In our similarity computation, trust values that are reported by the agent, for which trust is computed, are not considered.

Liu *et al.* (2013) propose trust model based on fuzzy logic. Their model formalizes trust as weighted average of all provided trust assessments towards trusting agent. The weights are calculated by taking into consideration forgetting factor, subjective differences between agents and confidence of agents that provided trust assessments. They measure the subjective differences using collaborative filtering approach: as differences in agents' preferences and interests. On the contrary, our trust model defines subjective difference between agents as difference in their general mindsets, irrespective from their interests.

Common drawback of the typical methods to mitigate effect of unfair ratings is that they do not assume agents' different trust evaluations that reflect their subjective facets. Namely, they do not assume that an agent might provide trust rating that (unintentionally) differs from other trust ratings due to agents' differences in mindsets. However, several authors presented mathematical formalizations of trust (Castelfranchi, Falcone 2001; Mezzetti 2004; Marsh 1994; Yan *et al.* 2013) that are based on research in psychology, sociology and (human) user surveys.

Castelfranchi and Falcone (2001) presented a cognitive model of trust. They defined trust as a mental state, compound of other more elementary mental attitudes, such as beliefs and goals, and delegation as an action and the resulting relation between two agents, whereby trust is the mental background of delegation. However, in their socio-cognitive model, the

authors do not refer to the possibility of having dishonest agents with respect to their spreading of (false) trust values.

Mezzetti (2004) proposes a socially inspired reputation model that introduces a jurisdiction sub-context, implying that an agent having authority over a particular context or situation, can be trusted for providing reliable recommendations about other agents within that context. In contrast to the QADE model, they do not consider subjective nature of trust and they do not handle the possibility of having dishonest agents and ways to deal with them.

Marsh (1994) formalized users' dispositions of trusting behavior: optimism, pessimism and realism. The paper describes the differences in trust assessments that originate from different dispositions. The author also discussed the effects of these differences. However, the paper does not consider the potential existence of unfair trust assessments.

Yan *et al.* (2013) propose a trust-behavior-based reputation and recommender system for mobile applications based on data collected via large-scale user survey. They formalized trust behavior considering "usage behavior" that is reflected by the number of application usages, elapsed usage time, usage frequency and experienced features; "reflection behavior" that concerns the usage behaviors after confronting application problems or errors; and "correlation behavior" that concerns the usage behaviors correlated to a number of similar functioned mobile applications. The proposed trust behavior formalization is context-specific and specially designed for a user's trust evaluation of a mobile application. Yan *et al.* (2012) also presented a concrete solution for trust and reputation formalization in pervasive social chatting environment using similar approach. The proposed formalizations present an alternative to our work, which formalizes trust broadly and assesses it on the basis of assessments made by other agents. As such, our proposed model is not limited to a particular context use case.

Our work differs in a number of ways. It presents a solution for identification of unfairly reported values with consideration of underlying subjective nature of trust. By that, the QADE trust model narrows the gap between trust formalizations that are based on complex psychological theories and robust methods for unfair trust value mitigation. Our proposed trust model is appropriate for e-commerce systems and includes an effective method for handling unfairly reported trust assessments.

Conclusions

In this paper, the QADE trust model has been presented, which is an extension of the Qualitative Assessment Dynamics (Trcek 2009). The model introduced 10 extensions related to the agents' private trust vectors, historical trust matrices, historical private trust vectors, multisets of public/private trust values between two agents, new QADE operators definition, attacker agent definition, agents' general mindset definition, similarity function definition, similarity matrix, and QADE filtering algorithm. The extensions address one of the most important aspects of trust and reputation management in e-commerce systems – handling of unfair ratings. Unfair ratings are common in e-commerce environments and have to be considered by participants (agents), otherwise it is impossible to calculate trustworthy ratings. A trusting agent's trust in another agent depends on trust assessments about that agent reported by other agents in a society and on the trusting agent's private trust values obtained

via direct interactions. Based on these values and its QADE operator, the agent evaluates the trust towards the other agents from the society, referred to as representative trust attitude. If these values were falsely reported, the trusting agent would compute trust values that were not accurate. Hence, its trust attitude would be skewed.

The presented QADE filtering algorithm provides a solution for identification of unfairly reported values, which are excluded from the trust computation. With filtering out false values, the trusting agent's trust attitude can be corrected. The proposed QADE algorithm provides effective filtering of unfair ratings in order to correct the agent's trust attitude. It considers subjective nature of trust and provides means to deal with similarities among agents. The similarity between agents is twofold: it considers pairwise similarity of trust opinions between agents and similarity of their general mindsets, which emphasizes human factors included in our model.

Proposed QADE method was compared with two most representative existing methods, the endogenous filtering technique proposed by Whitby *et al.* (2004) and with the exogenous filtering method defined in TRAVOS model (Teacy *et al.* 2006). Simulations were carried out in order to evaluate our novel approach. The distances in the mean difference between the representative, skewed and corrected trust attitudes on the scale of [0, 1] were measured. Values closer to 0 mean better filtering. On average, the distance between the representative and the corrected trust attitude equals $dist\left(repr,corrQADE\right)=0.39$, $dist\left(repr,corrTRAVOS\right)=0.62$ and $dist\left(repr,corrWhitby\right)=0.91$ in the bad-mouthing attack scenarios, considering individual and collaborative attacks. Further, it equals $dist\left(repr,corrQADE\right)=0.32$, $dist\left(repr,corrTRAVOS\right)=0.60$ and $dist\left(repr,corrWhitby\right)=0.93$ for the ballot-stuffing attack configurations. The above results show that the effectiveness of our QADE method is on average 26% to 57% better than the other two most representative endogenous and exogenous filtering techniques. Moreover, the percentages of unfair ratings that were detected with each filtering technique were measured. The best performance was achieved with the QADE filter that detected 88% unfair trust ratings, followed by the Whitby filter and the TRAVOS filter that identified on average 82% and 75% unfair trust assessments, respectively.

In the proposed QADE trust model, an agent computes the trust in a trusted agent based on the previous experiences with that agent and based on the similar agents' opinions. As such, the efficiency of the QADE filter algorithm is relatively low for agents with small number of previous interactions. The efficiency increases by increasing the number of interactions among the agents. However, TRAVOS and Whitby filters have the same characteristics, and the QADE filter still outperforms them. In order to eliminate this weakness, our future work will include the introduction of transitivity. *A similarity network* will be proposed, so that agents will (to some extent) consider also the trust assessments reported by the "friends of friends", where a friend is considered as an agent with similar general mindset, until the agents collect enough experiences by themselves.

Further extensions to the QADE trust model will include time-related issues. Namely, the current model does not differentiate between older and newer trust assessments. *An aging factor* of trust assessments will be introduced, as well as the assessments weighting according to their age. Additionally, our future work will include the assumption that the agents' behavior may change over time. The model will be enhanced to achieve resistance to attacks

based on time strategies, such as "Betrayal attack" (meaning that an agent suddenly turns into a malicious one after it maintains good reputation for some time) or "On-off attack" (meaning that a malicious agent repeatedly changes its behavior from honest to dishonest), which will be empirically evaluated.

References

Abdul-Rahman, A.; Hailes, S. 2000. Supporting trust in virtual communities, in *Proc. of the 33rd Hawaii International Conference on System Sciences (HICSS '00)*, 4–7 January 2000, Maui, Hawaii. IEEE Computer Society, Volume 6. 9 p. http://dx.doi.org/10.1109/HICSS.2000.926814

Cahill, V.; Gray, E.; Seigneur, J.-M.; Jensen, C. D.; Chen, Y.; Shand, B.; Dimmock, N.; Twigg, A.; Bacon, J.; English, C.; Wagealla, W.; Terzis, S.; Nixon, P.; Di Marzo Serugendo, G.; Bryce, C.; Carbone, M.; Krukow, K.; Nielson, M. 2003. Using trust for secure collaboration in uncertain environments, *IEEE Pervasive Computing* 2(3): 52–61. http://dx.doi.org/10.1109/MPRV.2003.1228527

Castelfranchi, C.; Falcone, R. 2001. Social trust: a cognitive approach, in C. Castelfranchi, Y. H. Tan (Eds.). *Trust and deception in virtual societies*. Springer, 55–90.

Chang, E.; Dillon, T.; Hussain, F. K. 2006. *Trust and reputation for service-oriented environments: technologies for building business intelligence and consumer confidence*. John Willey & Sons. 374 p. http://dx.doi.org/10.1002/9780470028261

Chen, M.; Singh, J. P. 2001. Computing and using reputations for internet ratings, in *Proc. of the 3rd ACM conference on Electronic Commerce (EC '01)*, 14–17 October, Tampa, Florida, USA, 154–162. http://dx.doi.org/10.1145/501158.501175

Commission of the European Communities. 2009. *Report on cross-border e-commerce in the EU*, Brussels, 2009.

Dellarocas, C. 2000. Immunizing online reputation reporting systems against unfair ratings and discriminatory behaviour, in *Proc. of the 2nd ACM conference on Electronic commerce (EC '00)*, 17–20 October 2000, Minneapolis, Minnesota, 150–157. http://dx.doi.org/10.1145/352871.352889

Dukaric, R.; Juric, M. B. 2012. Towards a unified taxonomy and architecture of cloud frameworks, *Future Generation Computer Systems* 29(5): 1196–1210. http://dx.doi.org/10.1016/j.future.2012.09.006

Fang, H.; Zhang, J.; Şensoy, M.; Thalmann, N. M. 2012. SARC: subjectivity alignment for reputation computation, in *Proc. of the 11th International Conference on Autonomous Agents and Multiagent Systems,* 4–8 June, 2012, Valencia, Spain, 3: 1365–1366.

Gambetta, D. 2000. Can we trust trust, Chapter 13 in D. Gambetta (Ed.) *Trust: making and breaking cooperative relations.* Blackwell Pub, 213–237.

Grabner-Kräuter, S.; Kaluscha, E. A. 2003. Empirical research in on-line trust: a review and critical assessment, *International Journal of Human-Computer Studies* 58(6): 783–812. http://dx.doi.org/10.1016/S1071-5819(03)00043-0

Grandison, T.; Sloman, M. 2000. A survey of trust in internet applications, *IEEE Communications Surveys & Tutorials* 3(4): 2–16. http://dx.doi.org/10.1109/COMST.2000.5340804

Hoffman, K.; Zage, D.; Nita-Rotaru, C. 2009. A survey of attack and defense techniques for reputation systems, *ACM Computing Surveys (CSUR)* 42(1): 1–19. http://dx.doi.org/10.1145/1592451.1592452

Hussain, F. K.; Chang, E. 2007. An overview of the interpretations of trust and reputation, in *Proc. of Third Advanced International Conference on Telecommunications (AICT 2007)*, 13–19 May 2007, Morne. 30 p. http://dx.doi.org/10.1109/AICT.2007.11

Ismail, R.; Josang, A. 2002. The beta reputation system, in *Proc. of the 15th Bled Conference on Electronic Commerce*, 17–19 June 2002, Bled, Slovenia, 1–14.

Josang, A. 2001. A logic for uncertain probabilities, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 9(3): 279–311. http://dx.doi.org/10.1142/S0218488501000831

Josang, A.; R. Ismail, R.; Boyd, C. 2007. A survey of trust and reputation systems for online service provision, *Decision Support Systems* 43(2): 618–644. http://dx.doi.org/10.1016/j.dss.2005.05.019

Juric, M. B. 2010. WSDL and BPEL extensions for event driven architecture, *Information and Software Technology* 52(10): 1023–1043. http://dx.doi.org/10.1016/j.infsof.2010.04.005

Kerschbaum, F.; Haller, J.; Karabulut, Y.; Robinson, P. 2006. Pathtrust: a trust-based reputation service for virtual organization formation, in *Trust Management, Proceedings 4th International Conference iTrust 2006,* 16–19 May 2006, Pisa, Italy. Springer, 193–205.

Kersuliene, V.; Turskis, Z. 2011. Integrated fuzzy multiple criteria decision making model for architect selection, *Technological and Economic Development of Economy* 17(4): 645–666. http://dx.doi.org/10.3846/20294913.2011.635718

Keung, S. N. L. C.; Griffiths, N. 2010. Trust and reputation, in N. Griffiths, K. M. Chao (Eds.). *Agent-based service-oriented computing*. Advanced Information and Knowledge Processing. Springer, 189–224.

Kim, D. J.; Ferrin, D. L.; Rao, H. R. 2008. A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents, *Decision Support Systems* 44(2): 544–564. http://dx.doi.org/10.1016/j.dss.2007.07.001

Liu, S.; Yu, H.; Miao, C.; Kot, A. C. 2013. A fuzzy logic based reputation model against unfair ratings, in *Proc. of the 12th International Conference on Autonomous Agents and Multi-agent Systems*, 6–10 May 2013, Saint Paul, Minnesota, USA, 821–828.

Lucking-Reiley, D.; Bryan, D.; Prasad, N.; Reeves, D. 2007. Pennies from eBay: the determinants of price in online auctions, *The Journal of Industrial Economics* 55(2): 223–233. http://dx.doi.org/10.1111/j.1467-6451.2007.00309.x

Manchala, D. W. 2000. E-commerce trust metrics and models, *IEEE Internet Computing* 4(2): 36–44. http://dx.doi.org/10.1109/4236.832944

Marsh, S. 1994. Optimism and pessimism in trust, in *Proc. of the Ibero-American Conference on Artificial Intelligence (IBERAMIA '94)*, October 1994, Caracas, Venezuela. McGraw-Hill Publishing, 1–12.

Mezzetti, N. 2004. A socially inspired reputation model, in S. Katsikas, *et al.* (Eds.). *Public key infrastructure*. Lecture Notes in Computer Science 3093. Springer, 191–204. http://dx.doi.org/10.1007/978-3-540-25980-0_16

Noorian, Z.; Marsh, S.; Fleming, M. 2011. Multi-layer cognitive filtering by behavioral modeling, in *Proc. of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '11),* 2–6 May 2011, Taipei, Taiwan, 2: 871–878.

Oxford Dictionaries. 2013. *Trust* [online], [cited 01 December 2013]. Available from Internet: http://www.oxforddictionaries.com/definition/english/trust

Piaget, J. 2002. *Judgment and reasoning in the child*. Reprint. Taylor & Francis. 268 p.

Pinyol, I.; Sabater-Mir, J. 2013. Computational trust and reputation models for open multi-agent systems: a review, *Artificial Intelligence Review* 40(1): 1–25. http://dx.doi.org/10.1007/s10462-011-9277-z

Pinyol, I.; Sabater-Mir, J.; Cuni, G. 2007. How to talk about reputation using a common ontology: from definition to implementation, in *Proc. of the Ninth Workshop on Trust in Agent Societies*, 2007, Hawaii, USA, 90–101.

Potocnik, M.; Juric, M. B. 2013. Towards complex event aware services as part of SOA, *IEEE Transactions on Services Computing*, 7(3): 486–500. http://dx.doi.org/10.1109/TSC.2013.7

Rasmusson, L.; Sverker Jansson, S. 1996. Simulated social control for secure internet commerce, in *Proc. of the 1996 Workshop on New Security Paradigms (NSPW '96)*, 17–20 September 1996, Lake Arrowhead, CA, USA, 18–25. http://dx.doi.org/10.1145/304851.304857

Resnick, P. 2002. Trust among strangers in internet transactions: empirical analysis of ebay's reputation system, *Advances in Applied Microeconomics* 11: 127–157. http://dx.doi.org/10.1016/S0278-0984(02)11030–3

Sabater-Mir, J.; Paolucci, M. 2007. On representation and aggregation of social evaluations in computational trust and reputation models, *International Journal of Approximate Reasoning* 46(3): 458–483. http://dx.doi.org/10.1016/j.ijar.2006.12.013

Sabater, J.; Sierra, C. 2005. Review on computational trust and reputation models, *Artificial Intelligence Review* 24(1): 33–60. http://dx.doi.org/10.1007/s10462-004-0041-5

Schillo, M.; Funk, P.; Rovatsos, M. 2000. Using trust for detecting deceitful agents in artificial societies, *Applied Artificial Intelligence: An International Journal* 14(8): 825–848. http://dx.doi.org/10.1080/08839510050127579

Sterling, L. S.; Taveter, K. 2009. *The art of agent-oriented modeling*. MIT Press. 392 p.

Su, X.; Khoshgoftaar, T. M. 2009. A survey of collaborative filtering techniques, *Advances in Artificial Intelligence* 4. 9 p. http://dx.doi.org/10.1155/2009/421425

Tavakolifard, M.; Herrmann, P.; Knapskog, S. J. 2009. Inferring trust based on similarity with TILLIT, in *Trust Management III*. IFIP Advances in Information and Communication Technology, Volume 300. Springer, 133–148. http://dx.doi.org/10.1007/978-3-642-02056-8_9

Teacy, W.; Patel, J.; Jennings, N.; Luck, M. 2006. Travos: trust and reputation in the context of inaccurate information sources, *Autonomous Agents and Multi-Agent Systems* 12(2): 183–198. http://dx.doi.org/10.1007/s10458-006-5952-x

Terveen, L.; McDonald, D. W. 2005. Social matching: a framework and research agenda, *ACM Transactions on Computer-Human Interaction (TOCHI)* 12(3): 401–434. http://dx.doi.org/10.1145/1096737.1096740

Trcek, D. 2009. A formal apparatus for modeling trust in computing environments, *Mathematical and Computer Modelling* 49(1–2): 226–233. http://dx.doi.org/10.1016/j.mcm.2008.05.005

Victor, P.; Cornelis, C.; De Cock, M.; Herrera-Viedma, E. 2009. Practical aggregation operators for gradual trust and distrust, *Fuzzy Sets and Systems* 184(1): 126–147. http://dx.doi.org/10.1016/j.fss.2010.10.015

Wang, Y.; Vassileva, J. 2005. Bayesian network trust model in peer-to-peer networks, in *Agents and Peer-to-Peer Computing.* Lecture Notes in Computer Science 2872. Springer, 23–34. http://dx.doi.org/10.1007/978-3-540-25840-7_3

Whitby, A.; Josang, A.; Indulska, J. 2004. Filtering out unfair ratings in bayesian reputation systems, in *Proc. of the 7th Intl. Workshop on Trust in Agent Societies*, 19–23 July 2004, New York, 106–117.

Yan, Z.; Chen, Y.; Shen, Y. 2013. A practical reputation system for pervasive social chatting, *Journal of Computer and System Sciences* 79(5): 556–572. http://dx.doi.org/10.1016/j.jcss.2012.11.003

Yan, Z.; Zhang, P.; Deng, R. H. 2012. TruBeRepec: a trust-behavior-based reputation and recommender system for mobile applications, *Personal and Ubiquitous Computing* 16(5): 485–506. http://dx.doi.org/10.1007/s00779-011-0420-2

Yang, Y. F.; Feng, Q. Y.; Sun, Y.; Dai, Y. F. 2009. Dishonest behaviors in online rating systems: cyber competition, attack models, and attack generator, *Journal of Computer Science and Technology* 24(5): 855–867. http://dx.doi.org/10.1007/s11390-009-9277-5

Yu, B.; Singh, M. P. 2002. Distributed reputation management for electronic commerce, *Computational Intelligence* 18(4): 535–549. http://dx.doi.org/10.1111/1467-8640.00202

Yu, B.; Singh, M. P. 2003. Detecting deception in reputation management, in *Proc. of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '03)*, 14–18 July 2003, Melbourne, VIC, Australia, 73–80. http://dx.doi.org/10.1145/860575.860588

Zavadskas, E. K.; Kaklauskas, A.; Kaklauskiene, J.; Trinkunas, V.; Krutinis, M. 2004. Web-based multiple criteria analysis of ethical and trust problems, *Journal of Foundations of Civil Engineering and Environmental Engineering* 6: 1–57.

Zavadskas, E. K.; Turskis, Z. 2011. Multiple criteria decision making (MCDM) methods in economics: an overview, *Technological and Economic Development of Economy* 17(2): 397–427. http://dx.doi.org/10.3846/20294913.2011.593291

Zupancic, E.; Trcek, D. 2011. The evaluation of Qualitative Assessment Dynamics (QAD) methodology for managing trust in pervasive computing environments, in *Proc. of 6th International Conference on Pervasive Computing and Applications (ICPCA)*, 26–28 October 2011, Port Elizabeth, South Africa, 67–73. http://dx.doi.org/10.1109/ICPCA.2011.6106480

Eva ZUPANCIC. She holds a BSc in computer and mathematical science from University of Ljubljana, Slovenia and a PhD in computer and information science from University of Ljubljana, Slovenia. Research interests: trust, reputation, trust and reputation modelling, trust and reputation management systems, false ratings, subjectivity, human factors, personalization, e-environment.

Denis TRCEK is with the Faculty of Computer and Information Science, University of Ljubljana, where he heads the Laboratory of e-media. He has been involved in the field of computer networks and IS security and privacy for over twenty years. He has taken part in various EU and national projects in government, banking and insurance sectors (projects under his supervision total approx. one million EUR). His bibliography includes over one hundred titles, including a monograph published by renowned publisher Springer. He has served (and still serves) as a member of various international bodies and boards (MB of the European Network and Information Security Agency, etc.). His interests include security, trust management, privacy and human factor modeling.