Gutowska A, Sloane A, Buckley K A. On desideratum for B2C E-commerce reputation systems. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 24(5): 820–832 Sept. 2009

On Desideratum for B2C E-Commerce Reputation Systems

Anna Gutowska, Andrew Sloane, and Kevan A. Buckley

School of Computing and Information Technology, University of Wolverhampton, Wolverhampton, WV1 1SB, U.K.

E-mail: ag37@rice.edu; {A.Sloane, K.A.Buckley}@wlv.ac.uk

Revised June 1, 2009.

Abstract This paper reviews existing approaches to reputation systems, their constraints as well as available solutions. Furthermore, it presents and evaluates a novel and comprehensive reputation model devoted to the distributed reputation system for Business-to-Consumer (B2C) E-commerce applications that overcomes the discussed drawbacks. The algorithm offers a comprehensive approach as it considers a number of issues that have a bearing on trust and reputation such as age of ratings, transaction value, credibility of referees, number of malicious incidents, collusion and unfair ratings. Moreover, it also extends the existing frameworks based on information about past behaviour, with other aspects affecting online trading decisions which relate to the characteristic of the providers, such as existence of trustmark seals, payment intermediaries, privacy statements, security/privacy strategies, purchase protection/insurance, alternative dispute resolutions as well as the existence of first party information.

Keywords decision support, distributed systems, electronic commerce, online information services

1 Introduction

The process of globalization creates new challenges and opportunities for companies by offering an access to new markets that were previously closed due to cost, regulations, etc. The adoption of the Internet, in particular Internet-enabled B2C E-business solutions, allows many Small and Medium Enterprises (SMEs) to respond to these challenges and opportunities by extending the geographic reach of their operations. Very often, however, websites created for sales purposes are simple in design and functionality and therefore, do not arouse trust at first glance. Furthermore, in contrast to "big brands" which have already established their reputation in the online marketplaces, SMEs are unknown to many E-commerce customers.

In the E-commerce environment, which does not require the physical presence of the participants, there is a high level of "uncertainty" regarding the reliability of the services, products, or providers. Although many technologies exist to make the transaction more secure, there is still the risk that the unknown provider will not comply with the protocol used. Thus, the decision of who to trust and with whom to engage in a transaction becomes more difficult and falls on the shoulders of the individuals. In such an environment, reputation systems come in place to assist consumers in decision making. One of the important aspects in such decisions is a third party's reputation based on the various parameters of past behaviour.

There are a number of existing consumer-to-consumer (C2C) on-line reputation systems such as those used by eBay[1] or Amazon[2]. However, unlike C2C E-commerce marketplaces, most B2C sites do not provide users with feedback information. There are some centralized services/websites though, which do offer store ratings and reviews to their users, such as BizRate[3] or Resellerratings[4]. All of them, however, rely only on simple algorithms calculating the average rating based on the given feedback.

Nevertheless, much academic work on reputation systems has been devoted to the C2C part of E-commerce (Peer-to-Peer networks) which can be found in [5–8]. Unlike the existing centralized approaches (e.g., eBay, Amazon) which are single-factor based, many authors proposed distributed reputation systems which still tend to be "one issue-centric"[9–12] (addressing only one of many problems existing in the reputation systems[6–8]). Even in studies attempting to provide more complex reputation methods, for example work on Histos/Sporas[13], some issues are still not taken into consideration, such as the transaction value, age of rating, or the credibility of referees.

Many of the problems addressed in C2C reputation models also apply to the B2C E-commerce environment. Not many authors, however, concentrate on the latter model of the marketplace. The only work known to the authors addressing it is [14] and [15].

Regular Paper

Nevertheless, to the best of the authors' knowledge, there are no studies which focus on deriving reputation ratings in B2C E-commerce environment taking into account the characteristics of the providers.

At first, this paper describes a selection of reputation systems that constitute a good representation of current research. It reviews existing approaches, constraints of current systems and available solutions, and discusses some aspects of trust and reputation systems that require more attention. Based on the above, the reputation metric which is suitable for B2C E-commerce is presented together with its evaluation and results. The proposed reputation model offers a comprehensive approach by including age of rating, transaction value, credibility of referees, and number of malicious incidents, as well as preventing collusion and inclusion of unfair ratings. Furthermore, in addition to the information about past behaviour it also incorporates other aspects affecting online trust which are based on providers' characteristics. Past behaviour is not the only information source affecting trust/reputation rating of an online vendor. According to previous research[16−17], there are many issues influencing online trust-based decisions such as the existence of trustmark seals, payment intermediaries, privacy statements, security/privacy strategies, purchase protection/insurance, alternative dispute resolutions as well as the existence of first party information. The extended approach presented in this paper yields a promising improved distributed B2C reputation mechanism.

2 Reputation Systems Classification

A reputation system collects, distributes, and aggregates feedback about participants, services or products which can assist other users in the decision making in the future. They aim to encourage trustworthy behaviour, and help prevent participation by those who are dishonest[18].

Online reputation systems can be classified based on the network architecture, E-commerce model and source of reputation information.

Reputation network architecture[6] determines how ratings are gathered and stored in reputation systems. The two main types are centralised and distributed architectures. Initial efforts at trust management in electronic communities were based on *centralised* trust databases. In these systems information about the performance of participants is collected as ratings and stored in a central authority (reputation centre). All reputation scores are publicly available so participants can use them when deciding which party to transact with. Many existing reputation systems are centralized, e.g., eBay, Amazon, BizzRate, and Resellerratings. Also many have been studied in the context of online communities and marketplaces[13,19]. *Distributed* reputation systems have no central location for submitting feedback or obtaining reputation ratings from. Instead, each participant is responsible for obtaining and collecting ratings from other participants. There can be distributed stores where ratings can be submitted, or each participant records their opinion about transactions with other parties. Therefore, to learn about the reputation of a potential transaction partner, a participant needs to find the distributed stores or try to obtain ratings from other participants who had direct contact with the target party[6]. The user calculates the reputation ratings based on the received scores.

Depending on the E-commerce model, reputation mechanisms can be broadly classified into *bidirectional* and *unidirectional*[14]. The first type exists in the context of C2C E-commerce model, such as online auctions and peer-to-peer services (e.g., eBay, Amazon) where users can act as both buyers and sellers thus, they can rate and be rated at the same time. In this case, the user reputation can be extracted from explicit ratings assigned to them. The latter type relates to the B2C model where sellers, products and services are rated by the users (buyers) or selected evaluators (e.g., BizzRate, Resellerratings).

On the basis of the source of the reputation information, the reputation systems can be classified into *explicit* and *implicit* ones[14]. The former group of reputation systems uses explicit feedback information, i.e., users' reputation as evaluated by other users as in bidirectional systems. The *implicit* reputation mechanism uses implicit reputation information, e.g., derived from analyzing the position of each user within the social network[20].

3 Constraints of Reputation Systems

There are several problems that exist in all commercial and academic reputation systems mainly caused by opportunistic behaviour of market participants. Some of these constraints have already been addressed while others still require more attention from the research community. They are identified and presented below.

3.1 Unfair Ratings

Avoiding or trying to reduce the influence of unfair ratings, either unfairly positive or negative, constitutes the fundamental problem in reputation systems as they rely on feedback given by others[6]. This problem often relates to collusion attacks in which a group of market participants try to manipulate their own reputation or the reputation of others. On one hand, a colluding group can give unfairly high ratings to a single buyer in order to improve his reputation, which is called ballot stuffing[19]. On the other hand unfairly negative ratings can be given to a seller in order to drive him out of the market, i.e., bad-mouthing[19].

The common way to deal with the problem of unfair ratings is to filter out the incorrect inputs[21]. Some researchers go even a step further, e.g., by proposing a punishment mechanism[22]. In this scenario, if feedback messages from two peers involved in the transaction disagree it means that at least one of the sides is lying, thus, both sides are punished as it is difficult to figure out which one was untrustworthy. There are other systems[23] that attempt to improve the accuracy of the transaction feedback by requiring proof of interaction, i.e., a transaction certificate that has to be signed by two parties when rating the transaction. While this may not prevent the participants from lying about the outcome, it does prevent from submitting fraudulent feedback about sellers they have not interacted with.

There are still some systems, however, which assume that buyers always provide unbiased feedback and therefore the dilemma of unfair ratings is not addressed[12].

3.2 Credibility of Referees

This issue relates to the situation when a trusting participant needs to gather reputation ratings about a potential transaction partner from others (indirect ratings, witness information). There is a possibility that some peers could be dishonest in order to obtain some benefit from lying. Some of the reputation models do not deal with the possibility that referees (users providing feedback) may lie about their ratings of another agent and assume that the majority of users are honest and well-behaved[24].

There are researchers, however, who work on the issue of lying witnesses. The general assumption while dealing with this problem is that participants in the system are not fully trusted. One of the solutions is to evaluate credibility of reputation providers by judging them on their reputation. In the bidirectional rating mechanism, which can be found in the C2C E-commerce models (e.g., online auctions and other P2P services like eBay), the credibility of the raters/referees can be easily extracted from explicit ratings (explicit feedback information) as the users act both as raters and ratees (rated users)[14]. This solution can be found in [9–10] where the credibility of reputation providers is evaluated by judging them on their reputation. There is an assumption that raters with low reputation are likely to give unfair ratings. In the model presented in [25] recommendations from a recommender with a high reputation has the same importance as from a direct interaction while the ones coming from an agent with bad reputation are not taken into account. In that study, after the transaction, the agent also compares the recommendation with the real behaviour of the recommended agent and, based on that, the reputation of the recommender is increased or decreased. Huynh, Jennings and Shadbolt[11] also count the difference between the actual performance of the transaction partner and its rating received from the referee. Using this information, however, they calculate the credibility of a referee instead of relying on his reputation. Similar approach is also presented in [9, 26] where provider's reputation is judged on the perceived accuracy of its past opinions.

3.3 Changing Identities

Reputation systems are based on the assumption that identities are long lived and the reputation ratings about a particular party from the past are related to the same party in the future[6]. Changing identity is closely related to the problem of a new entity entering the system. Some models may encourage changing identity, e.g., when a new entrant receives some positive reputation at the start — an initial credit. Users with bad reputation can easily use this opportunity to drop their pseudonyms to clear their past low-performance record and get a fresh start. In [27], the problem is referred to as the dilemma of cheap pseudonyms and market participants changing identities referred to as whitewashers.

The extensive work on the problem of strangers' policies and whitewashing has been done by Feldman *et al.*[28] who propose a "stranger adaptive" strategy which uses information on all first-time interactions to estimate the probability of being cheated by another stranger.

Different models deal differently with reputation for the new entrants. Papaioannou and Stamoulis[22] assign a low initial reputation value in order to limit the incentive for name changes. Bamasak and Zhang[10] set initial trust and reliability values to zero which are greater than those of a malicious agent which means that a new entrant will not be treated unfairly. On the other hand, this scheme gives an incentive for a cheating party to change its identity and to start from "zero". In the system proposed by Huynh, Jennings and Shadbolt[11], at first, each agent receives the default credibility value as it cannot provide any references about its previous behaviour. An end user can decide how to refer to the new participants: to discredit them till they prove to be credible or to consider them to be accurate and honest until proven otherwise.

Garg *et al.*[29] propose a novel solution of lending reputation. To enable a start for the new entity an existing member who knows the new one can choose to recommend it by "lending" part of its reputation, which from that moment is at risk. If the new entity behaves well the lender is getting the lent reputation back plus some reward. On the other hand if it behaves badly the lender loses the portion of the reputation which was at risk. This mechanism does not encourage participants to change their identities as to enter the system again they would have to find someone who would lend them a part of their reputation.

3.4 Reputation Life Time

The behaviour and performance of market participants change over time therefore trustworthiness does not remain the same value either. There are two main ways of dealing with this issue. One of them is to apply only the time window from which the transactions are taken into consideration[10]. The problem of determining the appropriate time threshold for a specific reputation calculation, however, is not well understood[21]. The other method is to use a decay function, that is assigning more weight to recent ratings than to the older ones. This solution is applied by Fan *et al.*[12] where the impact of history is controlled by an exponentially smoothed function of the previous reputation scores. Similarly, in the work presented by Huynh *et al.*[30] the age relevance of a reputation rating is calculated by an exponential decay function based on its recency. Also, the authors of REGRET[31] give much more relevance to the last referrals over the previous ones, using a normalized weight based on a time-dependant function. The combination of the two methods mentioned above can be found in [9], where each broker applies a different time threshold to decide whether or not the reputation rating should be taken at the full value. If the recommendation was reported within the time threshold, the time differential factor is taken into account.

Many of the systems, however, assume that the behaviour of agents does not change over time and therefore do not take the time factor into account[26].

3.5 Transaction Value

Transaction value constitutes an issue, which has not received enough attention from the academic community. While counting reputation ratings, the value of the transactions should also be taken into account as it would prevent the dishonest provider from building a high reputation by cooperating in many small transactions and then cheating in a very large transaction.

Bamasak and Zhang[10] make an attempt to address this issue. However, the transaction value in this model is not taken into account while calculating the reputation ratings. There is only a reputation threshold value applied which is proportional to the value of the transaction. The higher the transaction value, the higher the threshold. Fan *et al.*[12] propose a scheme where the information of the transaction value is incorporated in the reputation score algorithm but only in case of a possible malicious behaviour of value snipping, i.e., a transaction where the item value is much higher than the rest of the transactions. In this algorithm, an additional penalty value is applied which is proportional to the difference between the new high transaction value and 1 (i.e., in this scheme the item value for the rest of the transactions is set to 1). The study concentrates more on the mechanism of exponential smoothing, in particular how the value of the smoothing factor can affect the amount of cheating. This is also tested for the case where the value of the sale item follows a long normal distribution, which means it does not deal with high variation of the sale item value. Therefore, incorporating transaction values in the reputation rating requires further studies.

4 Desiderata for B2C Reputation Systems

The problems discussed in the preceding section relate mainly to C2C marketplaces and have been mainly studied in that context. Many of them, however, do apply to B2C environments and therefore, the solutions to them are considered as the requirements for a successful B2C online reputation system. The reputation model proposed in this study covers these desirable characteristics plus other B2C specific issues. In the authors' opinion, they constitute the desiderata for B2C E-commerce reputation systems which are presented below.

1) The feedback is a vector, not a scalar value (like in eBay), reflecting other users' evaluation of different aspects of the provided service quality and consists of the following components: *transaction outcome*, i.e., if the product/service was received, *fulfilling provider's signals*[21], e.g., if the delivery time, the product were as promised, and *customer service/support*.

2) The behaviour and performance of providers change over time. Thus, in order to model the dynamic nature of reputation, the reputation value in this study is decayed, as a *function of time*. In this way the more recent ratings are considered more important and are valued higher comparing to the older ones (see Subsection 6.2). Furthermore, as in [13], the memory of the reputation system is considered which disregards very old ratings.

3) In counting reputation ratings the *value of the transactions* is also taken into account based on the

exponential function (see Subsection 6.2). Also, the transaction value range depends on the context to which the reputation system will be applied, i.e., the maximum price of sold goods/services in the marketplace.

4) Whilst choosing the group of users to require the data from to calculate indirect reputation, it is important to take their *credibility as referees* into account (see Subsection 5.1). The reason for that is threefold. Firstly, it is often too costly or impossible to collect ratings results from all interactions with the provider in question[6]. Secondly, to avoid the inclusion of dishonest feedback into reputation calculation from users demonstrating colluding behaviour or leaving unfair ratings. Thirdly, to choose the right subset of users with "similar opinions". Namely, different people have different standards and they tend to trust the opinions of people who have the same standards with themselves[13].

5) Some participants may rate higher/lower in general than others. The reputation metric in this study applies the weight based on the *rating tendency* concept (see Subsection 6.2) inspired by [14]. This mechanism decreases the rating from the rater who has a tendency to rate higher than others, and vice versa.

6) Malicious behaviour should be prevented in the long term. As in [10], in the proposed model the reputation value is reduced to the minimum when a party reaches a certain threshold of malicious incidents (see Subsection 6.2). Up to that threshold the appropriate weight is applied based on the exponential function.

7) Characteristics of a provider which affect trust and the decision process in online transactions[7], i.e., the existence of: trustmark seals, payment intermediaries, first party information, privacy statements, security/privacy strategies, purchase protection/insurance, and of alternative dispute resolution, are part of the final reputation value (see optional parameters in Subsection 5.3).

8) *New users* have an initial reputation which is calculated based on their characteristics as providers (see Subsection 5.3).

5 Proposed Approach

The reputation mechanism presented in this paper is designed for the distributed B2C E-commerce model and it addresses the desiderata from the preceding section. It is based on the unidirectional rating mechanism where two main roles are considered: *buyer agent*, i.e., agent representing a user and *provider* (a Web service).

The overall model of the proposed reputation system rating is divided into 3 main stages (Fig.1): the implicit users' reputation/credibility extraction, the "*n*-best/most suitable raters" group generation and the provider's reputation calculation.

5.1 Implicit Users' Reputation/Credibility Extraction (Stage 1)

In the first stage, the credibility of users/referees is calculated. In the C2C E-commerce models (as discussed in Subsection 3.2.), the credibility of the raters/referees can be easily extracted from explicit ratings. In the B2C model, however, with anonymous users (feedback providers) who do not have reputation assigned to them, which would be based on explicit feedback on their performance given by other users, the above mechanism is inadequate. Therefore, the solution is to extract users' reputation automatically and implicitly from their past transaction rating data and

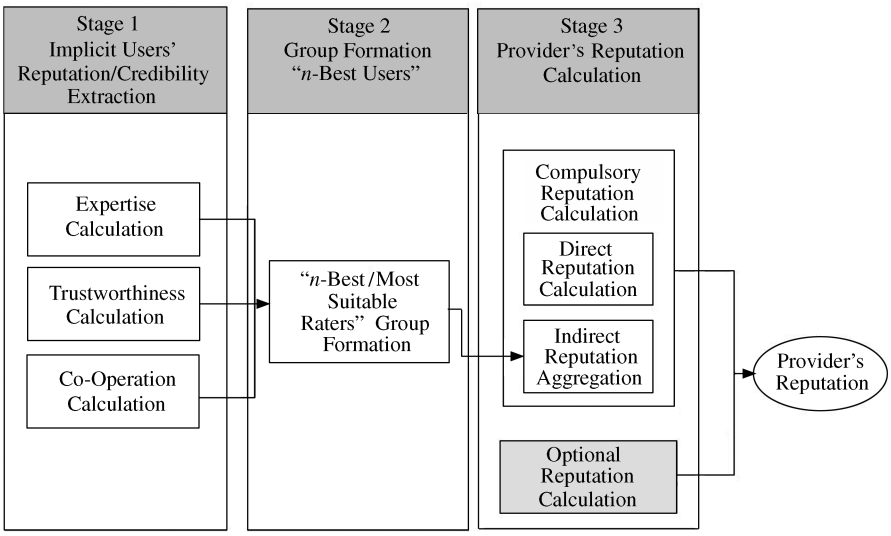

Fig.1. Process of the proposed reputation system model.

use it to choose "*n best/most suitable raters*". The method presented here is inspired by [14] and uses raters' ratings to estimate the raters' underlying credibility (i.e., implicit users reputation). It is based on the source credibility theory[32] which employs a few schemes of collaborative filtering methods (using similarities between a target rater and the rest of the users). The mechanism applied in this study applies the measured values of the source credibility factors, i.e., expertise, trustworthiness, and co-orientation (see Subsection 6.1). The theory was shown to support rating mechanisms both in the B2B[15] and B2C[14] E-commerce.

5.2 Raters Group Formation (Stage 2)

In the second stage the three above credibility factors are used to form a group of the "n-best/most suitable raters". Among several possibilities of combining these three elements into user's reputation (e.g., by arithmetic mean, harmonic mean, multiplication), the proposed model employs the filtering mechanism (expertise with the threshold of trustworthiness and co-orientation). This choice has been made based on the experimental results presented in [14] which show that the best outcomes (giving almost 34% performance gain over a model with randomly selected n users) are obtained with the application of the filtering mechanism of all three source credibility factors.

5.3 Provider's Reputation Calculation (Stage 3)

In the proposed model the reputation of the provider consists of two main parts: the compulsory reputation and the optional reputation. The compulsory reputation is calculated based on the information from the direct interactions (i.e., of the users requiring the reputation calculation) and indirect interactions (i.e., of the "n-best/most suitable raters" chosen in Stage 2). The calculations include the transaction ratings, time and value of the transactions as well as rating tendency of the raters.

In addition to the compulsory reputation, a user may choose to include some or all of the *optional parameters* into calculations, which will influence the rating value of a provider. They constitute the chosen characteristics of the providers that affect trust in the online trading decision making process which are taken from the Trust Taxonomy based on the results from the conducted survey[16]. The optional parameters are the existence of: trustmark seals, payment intermediaries, first party information, privacy statements, security/privacy strategies, purchase protection/insurance, and alternative dispute resolution, and are further described in [7, 16]. Optional parameters form the basis of the initial reputation for newcomers as at that point there is no information of past behaviour available.

A detailed description of how the reputation is calculated is presented in the next section.

6 Establishing the Reputation Metric

From the desiderata formed above a general reputation metric formula based on the weighted average has been established. Let p be a provider whose reputation value is calculated in any instance of time t. Similarly, let a be an agent representing a user/buyer that belongs to the buyers' community (i.e., users of the reputation system). To calculate the reputation of the provider in question, agent a uses his own feedback from the previous transactions with provider p (direct interactions) as well as feedback provided by other agents representing the users from the community (indirect interactions). The subset of the agents/users required for feedback is established based on the implicit user reputation (Subsection 6.1).

The reputation value of provider p is calculated as the arithmetic mean of the compulsory reputation (Subsection 6.3) and the optional reputation (Subsection 6.4). In addition the weight $wm(p)$ based on the number of malicious incidents is applied (Subsection 6.2).

If the optional reputation metric is not chosen to be applied then the reputation metric takes the value of the compulsory reputation metric multiplied by $wm(p)$. Further, the full rating scale of trust is [0, 1].

6.1 Implicit User Reputation

The implicit user reputation is generated based on the measured values of the source credibility factors (as discussed in Subsection 5.1) which are combined by the filtering mechanism (as discussed in Subsection 5.2).

Expertise Measurement. The expertise factor is defined as the degree of a user's competency to provide an accurate prediction[14]. The expertise of user u is

$$IR_{\mathrm{E}} = aw(u)\Big(1 - \frac{\sum_{p\in P}\sum_{a\in A(p)} |\bar{g}_{u,p} - \bar{g}_{a,p}|}{N_P}\Big). \quad (1)$$

where: $g_{u,p}$, $g_{a,p}$ are the average transaction ratings for provider p given accordingly by user u and user a; $A(p)$ the group of users who assigned rating for provider p; N_P is the cardinality of P; $aw(u)$ is the activity weighting and is defined as $1 - 1/m$ (m: the number ratings provided by user u) in order to obtain a higher value of expertise with more rating activities.

Trustworthiness Measurement. The trustworthiness factor is defined as the degree to which a user is perceived as providing information that reflects his actual feelings or opinions[14]. It is measured by the similarity between his rating and the mean of the ratings of the other users where the Pearson's correlation coefficient is employed.

$$IR_{\mathrm{T}} = sw(u) \cdot \frac{\sum_{p\in P}(g_{u,p} - \bar{g}_u)(g_{A,p} - \bar{g}_A)}{\sqrt{\sum_{p\in P}(g_{u,p} - \bar{g}_u)^2 \sum_{p\in P}(g_{A,p} - \bar{g}_A)^2}} \tag{2}$$

where $g_{u,p}$ is the average transaction rating of user u (rater) for provider p; $g_{A,p}$ is the average transaction rating of all the users who rated provider p; $\bar{g}_u$, $\bar{g}_A$ are the sample/set means; $sw(u)$ is the significance weighting which is 1 if the number of ratings of user u is over 50, otherwise $n/50$; a higher trustworthiness value is obtained on a user who has provided many ratings.

Co-Orientation Measurement. The co-orientation factor is defined as the degree to which a user is similar to the other users in the community that he belongs to [14].

$$IR_{\mathrm{C}} = \Bigg(\sum_{a\in A} sw(u) \cdot \frac{\sum_{p\in P(u)}(g_{u,p} - \bar{g}_u)(g_{a,p} - \bar{g}_a)}{\sqrt{\sum_{p\in P(u)}(g_{u,p} - \bar{g}_u)^2 \sum_{p\in P}(g_{a,p} - \bar{g}_a)^2}}\Bigg) \Big/ N_A \tag{3}$$

where $g_{u,p}$, $g_{a,p}$ is the average transaction rating for provider p given accordingly by user u (rater) and user a; g_u, g_a are the sample/set means; N_A is the cardinality of the A; $sw(u)$ is the significance weighting which is 1 if the number co-rated providers between user u and user a is over 50; otherwise it is $n/50$; this also assigns a higher co-orientation value to a user with many co-ratings with the other users.

6.2 Weights

There are four weights used in the proposed reputation metric which are associated with the following compulsory parameters: number of malicious incidents, reputation lifetime, transaction value and source of feedback. The application of the weights in calculating the reputation reflects the stated desiderata 2, 3, 5 and 6 (Section 4). The weights' equations are presented below. They can take values 0 to 1. The first three are based on the exponential function.

The weight for the *malicious incident* component is $wm(p)$ and is calculated as follows:

$$\begin{cases} wm(p) = \alpha^{-m}, & \text{if } 0 \leqslant m < M; \\ wm(p) = 0, & \text{if } m \geqslant M; \end{cases} \tag{4}$$

where

$$\alpha = \sqrt[x]{1/M}, \quad x \to 0 \tag{5}$$

where m is a number of malicious incidents of provider p that occurred within the transactions taken into calculation; M is the set threshold of the number of malicious incidents above which the reputation value is reduced to minimum; α is used to scale $wm(p)$ and $\alpha > 1$.

The weight associated with the *reputation life time* is defined as:

$$wt_x = \beta^{-\Delta t(x)} \tag{6}$$

where $\Delta t(x)$ is the time difference between the current time (i.e., time of request) and the time when the transaction x took place. β is used to scale $\Delta t(x)$ and $\beta > 1$. The time weight is applied to the reputation metric in a recursive algorithm (Subsection 6.3.1).

The other weight wv_x is associated with the *transaction value* and is calculated using the formula below:

$$wv_x = 1 - \gamma^{-v(x)} \tag{7}$$

where

$$\gamma = \sqrt[x]{1/vMax}, \quad x \to 0 \tag{8}$$

where $v(x)$ is the value of transaction x and $vMax$ is the transaction range, i.e., the maximum value of the goods/services in the marketplace (based on the context to which the reputation system is applied). γ is used to scale $v(x)$ and $\gamma > 1$.

The weight $ws(u)$ associated with *the source of feedback* parameter is based on the "rating tendency" concept inspired by [14].

$$ws(u) = 1 - (\bar{g}_u - \bar{g}_{A(u)}) \tag{9}$$

where g_u is the average transaction ratings from a rater u; $g_{A(u)}$ is the average ratings of the other users from the subset of the "best/most suitable users" (for the providers that the rater u rated).

6.3 Compulsory Reputation Metric

Compulsory reputation is defined as the arithmetic mean of *aggregated direct and indirect ratings* (see below). The rating scale for compulsory reputation metric is [0, 1].

6.3.1 Computing Aggregated Ratings

The aggregated ratings are calculated with the application of the recursive algorithm applied to the list of the transaction data records sorted according to the time value.

The *aggregated direct rating* value is calculated based on the data stored in the requesting agent a database, i.e., regarding its direct interactions:

$$\begin{aligned} AGRD_{a,x}(p) = {} & UR_{a,x}(p) \cdot [wt_x/(wt_x + wt_{x-1})] + \\ & AGRD_{a,(x-1)} \cdot [wt_{x-1}/(wt_x + wt_{x-1})]. \end{aligned} \tag{10}$$

For the case where $x = 0$ the aggregated direct rating is equal to $UR_{a,j}(p)$ — the updated rating for that transaction (see Subsection 6.3.2), where x is the index of the last transaction on the list $(n - 1)$.

The *aggregated indirect rating* values are calculated in the same manner as above but are based on the list of the transaction data from the group of the "n-best/most suitable users". In addition, the weight ws is applied for each user providing information.

6.3.2 Computing Updated Ratings

Updated reputation rating $UR_{a,x}(p)$ is calculated by agent a for transaction x in which a was involved with provider p. In general, each provider is reputed by an agent after each transaction by providing a transaction rating g. This is the average of two components: fulfilling the provider's signals and customer service, where both can take values [0, 1]. In addition, appropriate weight wv based on the transaction value is applied.

6.4 Optional Reputation Metric

The optional reputation is based on the set of *optional parameters* (providers' characteristics) (Subsection 5.3) which take values [0, 1] and is presented by the average of the above parameters which have been chosen to be included into calculation. The rating scale for optional reputation metric is [0, 1].

7 Evaluation and Results

The discussed reputation metric was evaluated by simulation with the use of a slightly modified version of Repast[33] which is a free and open source agent-based modeling toolkit written in Java. The strength of the metric was measured by how truly it reflects the agents (providers) behaviour and in particular by its resistance against different hostile agents.

7.1 Simulation Overview

The simulation framework models different types of behaviours in the B2C marketplace. The simulation is based on discrete time ticks. At each tick buyer agents are supposed to initiate a transaction with a provider and rate him afterwards. After the agents finished their actions the data is collected and represented graphically. The effectiveness of a reputation system and its metric depends on its resistance against malicious behaviours. The success of non-honest agents is its measurement for the quality of the metric[34]. The detailed description of the framework can be found in [35].

7.2 Modeling the Buyers

The buyers in the simulation framework differ in types. The buyer agent type is a combination of its trust disposition and its expectations.

Disposition to trust and the same risk attitude refer to the fact that people have a baseline attitude when they approach any trust situation. Some of them have a higher baseline level of trust than others thus, some individuals may find it easier/more difficult to trust. The disposition to trust affects the decision of either the buyer agent wants to engage in a transaction with the provider or not (see the acceptance function in Subsection 7.4.). Based on the above there are different types of the buyer agents in the simulation:

Risk Taking. This type of buyers is willing to take risks easily which means they accept the high value transactions even with the provider with low reputation.

Cautious. This type of buyers is risk averse and they are very careful with their decisions. They accept the transactions only if the provider has high reputation.

Conservative. Buyers representing this type come between the two above extremes.

In the presented framework the buyer agents have also different expectations towards the outcome of the transaction which affects the way they rate the transaction (see the rating function in Subsection 7.5.). As in [36], there are three types of the buyers agents in this study: optimists, realists, and pessimists. Optimists will be expecting a very positive outcome, pessimists on the other hand a rather bad outcome, and realists will come somewhere between the two extremes.

Combining the two attributes discussed above the following types of buyers agents were implemented in the simulation framework: Risk Taking Optimists, Risk Taking Realists, Risk Taking Pessimists, Cautious Optimists, Cautious Realists, Cautious Pessimists, Conservative Optimists, Conservative Realists, and Conservative Pessimists.

7.3 Modeling the Providers

There are different types of providers implemented in the framework which are called Trustworthy, Shady, Player, and Fly-By-Night of which the last three are malicious. They differ in their behaviour while transacting which is correlated with their characteristics, i.e., the cheating probability (*ChP*) and the range of the transaction outcomes they produce in terms of customer service and fulfilling providers' signals (in other words the quality of services they provide). The remaining attributes constitute the optional parameters in the reputation metric (see Subsection 5.3) which take values between 0 and 1 where 0 means no existence of the attribute. In this way, each type of the provider has the optional reputation (*OP*) value based on the above which constitutes the initial reputation value for any new provider in the system. In the reputation system, the OP values would be provided by a selected evaluator or a devoted agent that would gather this information from the providers' websites. The properties of different providers are as follows.

Trustworthy. This type of providers do not cheat in the transactions ($ChP = 0$) and provides high service quality. All the parameters mentioned above have high values ($OP = 0.92$).

Shady. This agent does not have a particular pattern in its behaviour ($ChP = 50$). It provides false statements on its website which results in high values of the optional parameters apart from Trustmark Seals and Payment Intermediaries ($OP = 0.63$). The quality of the services it provides is low.

Player. This type of provider tries to build high reputation by not cheating ($ChP = 0$). When it achieves its goal, however, it starts behaving in a malicious way ($ChP := 100$). When its reputation falls down below the threshold it starts being honest again ($ChP := 0$). A Player agent has got high values for First Party Information, Privacy Statements and Security Strategies ($OP = 0.43$). When it does not cheat the services provided are of a high quality.

Fly-By-Night. This agent's goal is to cheat ($ChP = 100$). It provides false information about the services it offers. The way of payment is direct to the bank account ($OP = 0.51$). The quality of the services it provides is low.

7.4 Transaction Acceptance Function

In the presented simulation the buyer agents have a trust disposition which allows them to make different decisions when it comes to engaging in a transaction with a provider.

In this work the assumption is that no buyer agent will transact voluntarily with a non-trustworthy provider, i.e., the provider with low reputation. The other factor taken into consideration while making the decision is the value of the transaction. The acceptance function, therefore, is a correlation between the provider's reputation and the value of the transaction. The higher the value of the transaction, the higher the reputation for the buyer to engage in this transaction. As different people have different disposition to trust, in the presented framework, different types of buyer agents have different acceptance functions. In this way different types of agents accept the transaction of a specific value at different reputation levels.

Users' willingness to trust, however, can be changed by experience[37]. In the proposed framework all buyer agents representing a specific type start with the same acceptance function which is affected/changed later on by the outcome of the transaction (experience) and in particular by the providers' malicious incidents. The calculation of the acceptance threshold for a specific transaction value with a specific provider is based on Lagrange Interpolation[38].

7.5 Rating Function

In the proposed framework each buyer agent rates each transaction he has been involved in and collects these ratings (Subsection 5.3) in his database.

In a real marketplace, different people will rate a transaction differently based on their experience and their expectations towards the transaction outcome. In the discussed simulation framework, three cases are considered: optimists, realists, and pessimists. When it comes to the transaction, optimists will be expecting a very positive outcome, pessimists on the other hand a rather bad outcome, and realists will come somewhere between the two extremes. The simulation framework addresses the above scenario in a way that the optimist agent will hope for the best outcome (in terms of customer service and provider's signals) he has had so far with the provider in question, the pessimist agent will anticipate the worst one, and the realist agent will expect the average result based on his experience. If the expected outcome (*expOut*) is higher than the actual one (*realOut*), the buyer agent applies the punishment value (p) to the transaction rating (*rating*) which is a difference between the expected and the real outcome value. If the expected outcome value is equal to or lower than the actual one, the ratings reflect the outcome. The above rules are presented below:

```
p:= expOut - realOut
if p > 0 then
   if  p <= realOut then
    rating:= realOut - Random(0,p)
   else
    rating:= realOut - Random(0,realOut)
else
   rating:= realOut
```

Apart from the transaction rating, the final reputation value includes also the other component which is Optional Reputation discussed in Subsection 6.4.

7.6 Evaluation Criteria

The strength of the metric is measured by how truly it reflects the agents (providers) behaviour and in particular by its resistance against different hostile agents. In the simulation the average requested reputation, the market honesty, the acceptance rate, the average number of transactions and the average number of malicious incidents are calculated separately for each type of the provider agents.

Average requested reputation is the mean value of all reputation ratings of providers from a specific type as if calculated/received by a buyer when requesting reputation rating. This is based on the rating information stored in the buyers' databases.

Market honesty is the mean value of the actual outcomes from the transactions produced by the provider agents (not ratings). These are stored in providers' databases.

Acceptance rate is the proportion of accepted /completed transactions with all initiated transactions with providers of a specific type.

Average number of transactions is the average number of transaction that a provider of a specific type was involved in (accepted transactions).

Average number of malicious incidents is the average number of malicious incidents for a specific type of a provider.

7.7 Simulation Results

There is no work known to the authors that introduces a reputation metric for B2C E-commerce reputation systems taking into account provider's characteristics. This paper, therefore, presents pioneering results and it is not possible to compare them with the efficiency of any other reputation metric.

The simulation results are depicted in Figs. 2∼7. The horizontal axis in the figures represents the time. In Figs. 2 and 3 the vertical axis corresponds to the computed reputation, in Figs. 4, 5 and 7 it represents the number of transactions and in Fig.6 the acceptance rate.

Market honesty (Fig.2) and average requested reputation (Fig.3) show that the reputation metric correctly

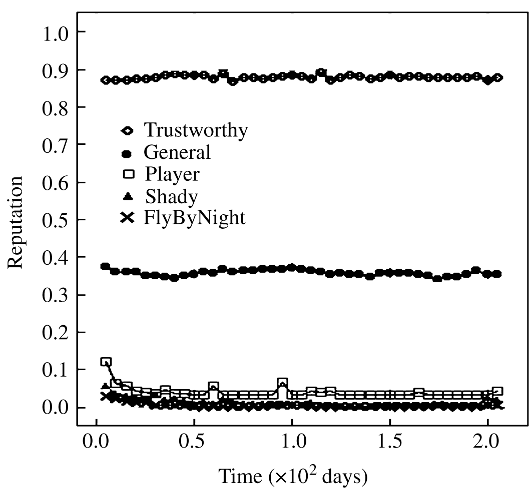

Fig.2. Market honesty.

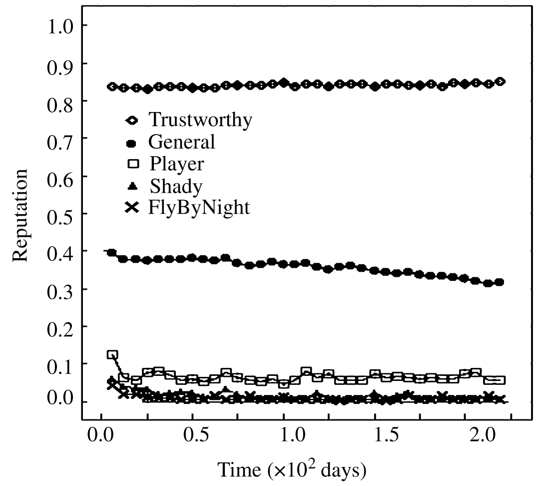

Fig.3. Average requested reputation.

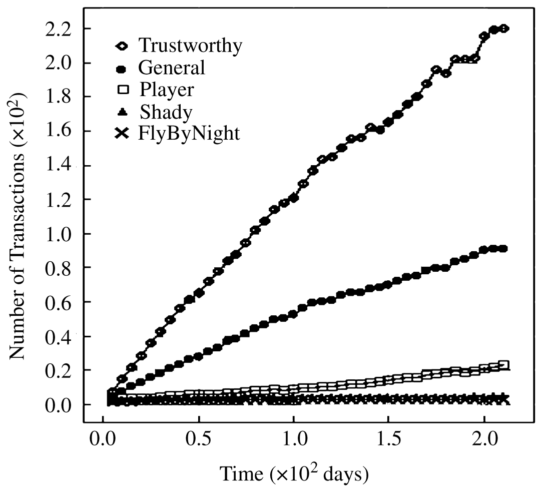

Fig.4. Average Number of transactions.

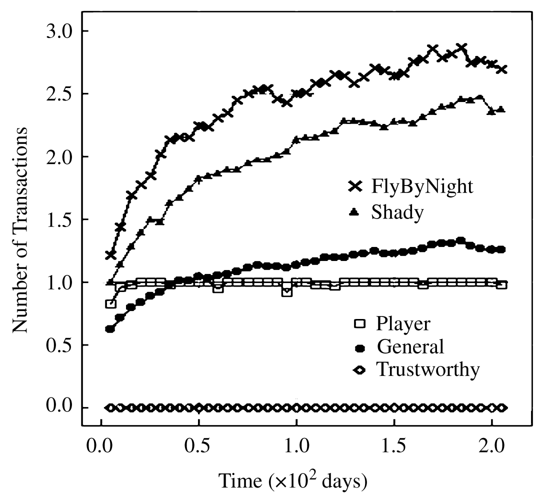

Fig.5. Average number of malicious incidents.

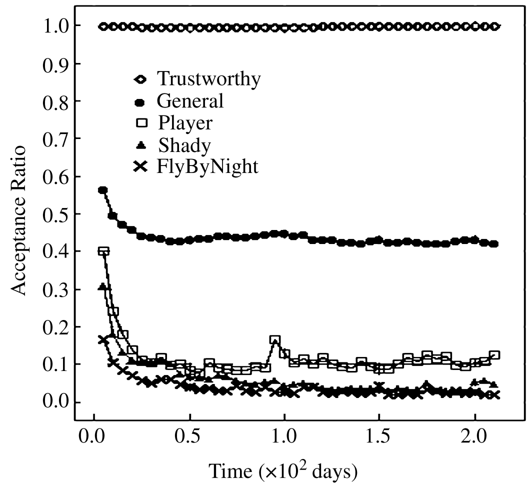

Fig.6. Acceptance rate.

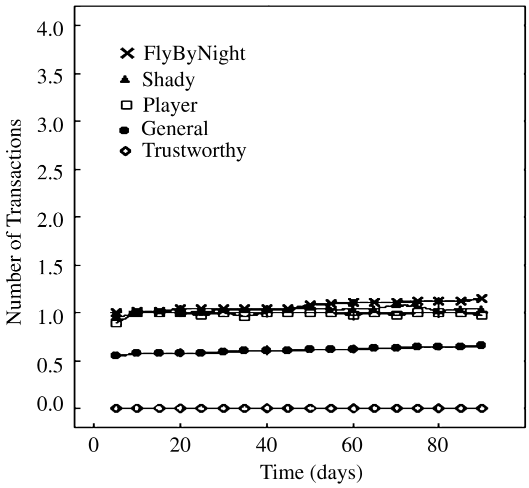

Fig.7. Average number of malicious incidents, $M = 1$, ceteris paribus.

reflects the behaviour of different types of providers, i.e., Trustworthy agents keep their high reputation scores throughout the experiment and the different types of malicious agents have low reputation due to their transaction history. It is noticeable that initially the reputation of the malicious agents is a bit higher and it decreases with time. This is caused by the fact that the initial reputation for new providers with no transaction records is their optional reputation which in many cases is based on the false information provided by them on their websites. When the transaction information comes into the equation, however, the reputation algorithm appropriately deals with the scenario and decreases the reputation value.

The slight difference in values between market honesty and average requested reputation reflects the fact that different types of buyer agents rate the transactions differently which does not always match real outcomes. The dissimilarity, however, is not significant which strongly suggests that the reputation metric closely mirrors the behaviours in the marketplace.

The results shown in Figs. 4∼6 indicate that malicious agents are not involved in many transactions (Fig.4) due to their low reputation. The acceptance rate (Fig.6) decreases as the buyer agents do not accept transactions with providers with low reputation. The average number of malicious incidents (Fig.5) is kept stable which is controlled by the maximum number of malicious incidents simulation parameter. If the parameter is set as $M = 1$ then the reputation metric will decrease the reputation of this provider to 0 which means it will not be accepted as a transaction partner anymore and will not get a chance to gain profit by cheating. This scenario is illustrated in Fig.7, ceteris paribus (i.e., while other parameters stay unchanged).

Overall, the results showed that the proposed reputation metric closely reflects different types of behaviour in the marketplace and the method is particularly resistant to malicious behaviour.

8 Conclusions

This paper reviewed the existing approaches to current reputation systems, their constraints as well as solutions available to them. Following these, the list of characteristics of a successful B2C reputation system was presented. Based on it, a novel comprehensive reputation model was proposed which meets the above requirements and also extends the existing reputation frameworks based only on information on past behaviour with other aspects affecting online trust, i.e., the providers' attributes.

The discussed metric was evaluated by simulation and the results show that it closely reflects different types of behaviours in the marketplace and the method is particularly resistant to malicious behaviour.

One of the assumptions of the proposed system, i.e., that there are no external parties included in the framework can be easily amended in the future by including the information coming from other systems or reputation authorities.

References

[1] eBay. Accessed on July 01, 2008, http://www.ebay.co.uk/.

[2] Amazon. Accessed on July 01, 2008, http://www.amazon.co.uk/.

[3] Bizrate. Accessed on July 01, 2008, http:/www.bizrate.com/.

[4] Resellerratings. Accessed on July 01, 2008, http://www.resellerratings.com/.

[5] Sabater J, Sierra C. Review on computational trust and reputation models. *Artificial Intelligence Review*, 2005, 24(1): 33–60.

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

[7] Gutowska A, Bechkoum K. A distributed agent-based reputation framework enhancing trust in e-commerce. In *Proc. IASTED International Conference on Artificial Intelligence and Soft Computing*, Palma de Mallorca, Spain, August 29–31, 2007, pp.92–101.

[8] Marti S, Garcia-Molina H. Taxonomy of trust: Categorizing P2P reputation systems. *Computer Networks*, 2006, 50(4): 472–484.

[9] Lin K J, Lu H, Yu T, Tai C E. A reputation and trust management broker framework for Web applications. In *Proc. IEEE International Conference on e-Technology, e-Commerce and e-Service*, Hong Kong, China, March 29–April 1, 2005, pp.262–269.

[10] Bamasak O, Zhang A N. A distributed reputation management scheme for mobile agent-based E-commerce applications. In *Proc. IEEE International Conference on e-Technology, e-Commerce and e-Service*, Hong Kong, China, March 29–April 12, 2005, pp.270–275.

[11] Huynh T D, Jennings N R, Shadbolt N R. Certified reputation: How an agent can trust a stranger. In *Proc. International Conference on Autonomous Agents and Multi-Agent Systems*, Hakodate, Japan, May 8–12, 2006, pp.1217–1224.

[12] Fan M, Tan Y, Whinston A B. Evaluation and design of online cooperative feedback mechanism for reputation management. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(2): 244–254.

[13] Zacharia G, Moukas A, Maes P. Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, 2000, 29(4): 371–388.

[14] Cho J, Kwon K, Park Y. Q-rater: A collaborative reputation system based on source credibility theory. *Expert Systems with Applications*, 2009, 36(2P2): 3751–3760.

[15] Ekstrom M, Bjornsson H. A rating system for AEC e-bidding that accounts for rater credibility. In *Proc. the CIB W65 Symposium*, 2002, pp.753–766.

[16] Gutowska A. Research in online trust: Trust taxonomy as a multi-dimensional model. Rechnical Report, School of Computing and Information Technology, University of Wolverhampton, 2007. http://www.scit.wlv.ac.uk/∼in0345/Research.htm.

[17] Gutowska A, Bechkoum K. The issue of online trust and its impact on international curriculum design. In *Proc. the Third China-Europe International Symposium on Software Industry-Oriented Education*, Dublin, Ireland, February 6–7, 2007, pp.134–140.

[18] Resnick P, Kuwabara K, Zeckhauser R, Friedman E. Reputation systems. *Communications of the ACM*, 2000, 43(12): 45–48.

[19] Dellarocas C. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proc. the 2nd ACM Conference on Electronic Commerce*, Minneapolis, USA, October 17–20, 2000, pp.150–157.

[20] Pujol J M, Sangüesa R, Delgado J. Extracting reputation in multi agent systems by means of social network topology. In *Proc. the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, Bologna, Italy, July 15–19, 2002, pp.467–474.

[21] Rein G L. Reputation information systems: A reference model. In *Proc. Annual Hawaii International Conference on System Sciences*, Big Island, Hawaii, USA, January 1–6, 2005, p.26a.

[22] Papaioannou T G, Stamoulis G D. An incentives' mechanism promoting truthful feedback in Peer-to-Peer systems. In *Proc. International Symposium on Cluster Computing and the Grid*, Cardiff, UK, May 9–12, 2005, pp.275–283.

[23] Singh A, Liu L. TrustMe: Anonymous management of trust relationships in decentralized P2P systems. In *Proc. the 3rd IEEE International Conference on Peer-to-Peer Computing*, Lingoping, Sweden, September 1–3, 2003, pp.142–149.

[24] Yu B, Singh M P. An evidential model of distributed reputation management. In *Proc. the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, Bologna, Italy, July 15–19, 2002, pp.294–301.

[25] Carbo J, Molina J, Davila J. Comparing predictions of SPORAS vs. a fuzzy reputation agent system. In *Proc. the 3rd WSES International Conference on Fuzzy Sets and Fuzzy Systems*, Interlaken, Switzerland, February 11–15, 2002, pp.147–153.

[26] Teacy W T L, Patel J, Jennings N R, Luck M. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proc. the 4th International Conference on Autonomous Agents and Multiagent Systems*, Ultrecht, Netherlands, July 25–29, 2005, pp.997–1004.

[27] Friedman E J, Resnick P. The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, 2001, 10(2): 173–199.

[28] Feldman M, Papadimitriou C, Chuang J, Stoica I. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 2006, 24(5): 1010–1019.

[29] Garg A, Montresor A, Battiti R. Reputation lending for virtual communities. In *Proc. the 22nd International Conference on Data Engineering Workshops*, Atlanta, USA, April 3–8, 2006, p.22.

[30] Huynh T D, Jennings N R, Shadbolt N R. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems*, 2006, 13(2): 119–154.

[31] Sabater J, Sierra C. Reputation and social network analysis in multi-agent systems. In *Proc. the First International Joint Conference on Autonomous Agents and Multiagent Systems*, Bologna, Italy, July 15–19, 2002, pp.475–482.

[32] Best R, Corney K, Hawkins D. Consumer Behavior: Building Marketing Strategy. McGraw-Hill/Irwin, 2003.

[33] Schlosser A, Schlosser A. Technische Universitat Darmstadt: Reputation System Simulator Documentation, Accessed on July 01, 2008, http://www.inferenzsysteme.informatik.tu-darmstadt.de/users/schlosser/.

[34] Schlosser A, Voss M, Brückner L. On the simulation of global reputation systems. *Journal of Artificial Societies and Social Simulation*, 2005, 9(1): 4.

[35] Gutowska A, Sloane A. Evaluation of reputation metric for the B2C E-commerce reputation system. In *Proc. the 5th International Conference on Web Information Systems and Technologies*, Lisbon, Portugal, March 23–26, 2009, pp.489–498.

[36] Michalakopoulos M, Fasli M., On different trust attitudes and their effects on the electronic marketplace. In *Proc. IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Compiegne, France, September 19–22, 2005, pp.102–108.

[37] Shneiderman B. Designing trust into online experiences. *Communication of the ACM*, 2000, 43(12): 57–59.

[38] Cheney W, Kincaid D. Numerical Mathematics and Computing. International Thomson Publishing, 1998.

Anna Gutowska received her first class honours M.Sc. degree in management and marketing specialised in computer applications for economy from the Technical University of Bialystok, Poland in 2003, and a first class honours B.Sc. degree in computer science from the same university in 2004. She is currently a lecturer in the School of Computing and Information Technology at the University of Wolverhampton, UK, where her main teaching lies in the areas of information systems, databases and websites. At the same time she conducts a part-time Ph.D. study. She received the Best Student Paper Award at the 5th International Conference on Information Systems and Technologies (WEBIST 2009). Her current research focuses on developing an agent-based reputation system for E-commerce that improves aspects of trust.

Andrew Sloane is a professor of telematics in the School of Computing and IT, University of Wolverhampton, UK. His main subject interest is computer communications. He is the author of three books on communications, computer communications (1993 and 2nd Ed. 1999), multimedia communication (1996) and Internet multimedia (2005). Current projects are concerned with interpersonal communication in the home, and communication in business and education.

Kevan A. Buckley obtained his B.Sc. degree in computing science from Teesside Polytechnic before moving to University of Wales, Cardiff to undertake postgraduate study and research. Whilst there he received an M.Sc. degree in systems engineering, a Ph.D. degree in artificial intelligence and robotics and subsequently occupied a post-doctoral research post in the Cardiff Manufacturing Engineering Centre. He is currently a principal lecturer in the School of Computing and Information Technology at University of Wolverhampton, where his main teaching is focused on computer programming and artificial Intelligence. His research interests include the transfer of AI technologies into new application areas and he is currently working on cyber emotions: http://www.cyberemotions.eu.