Evaluating the STORE Reputation System in Multi-Agent Simulations

Jonas Andrulis[1], Jochen Haller[1], Christof Weinhardt[2], and Yuecel Karabulut[3]

[1] SAP Research, Vincenz-Priessnitz-Str. 1, 76131 Karlsruhe, Germany,
{firstname.lastname}@sap.com
[2] Information Systems and Management, University of Karlsruhe, Englerstr. 14,
76131 Karlsruhe, Germany, {firstname.lastname}@iw.uni-karlsruhe.de
[3] SAP Labs, USA, {firstname.lastname}@sap.com

Abstract. In recent global business environments, collaborations among organisations raise an increased demand for swift establishment. Such collaborations are formed between organisations entering Virtual Organizations (VOs), crossing geographic borders and frequently without prior experience of the other partner's previous performance. In VOs, every participant risks engaging with partners who may exhibit unexpected fraudulent or otherwise untrusted behaviour. In order to cope with this risk, the STochastic REputation system (STORE) was designed to provide swift, automated decision support for selecting partner organisations in the early stages of the VO's formation. The contribution of this paper first consists of a multi-agent simulation framework design and implementation to evaluate the STORE reputation system. This framework is able to simulate dynamic agent behaviour, agents hereby representing organisations, and to capture the business context of different VO application scenarios. A configuration of agent classes is a powerful tool to obtain not only well or badly performing agents for simulation scenarios, but also agents which are specialized in particular VO application domains or even malicious agents, attacking the VO community. The second contribution comprises of STORE's evaluation in two simulation scenarios, set in the VO application domains of Collaborative Engineering and Ad-hoc Service provisioning. Besides the ability to clearly distinguish between agents of different classes according to their reputation, the results prove STORE's ability to take an agent's dynamic behaviour into account. The simulation results show, that STORE solves the difficult task of selecting the most trustworthy partner for a particular VO application domain from a set of honest agents that are specialized in a wide spread of VO application domains.

1 Introduction

Virtual Organizations (VOs) have emerged as a business model in application domains with a high demand for cross-domain collaborative business processing. Increased collaboration among business partners and focusing on an organization's core competencies requires such novel models to address business opportunities. A VO is defined as a set of

sovereign, geographically dispersed organizations that temporarily pool their resources to jointly address a business opportunity one organization alone is not able to master [1]. A VO follows a phased life cycle where speed is an essential requirement. Especially the initial identification and formation phases, dealing with potential VO partner identification, selection and integration have to be swiftly conducted to stay ahead of competitors. Afterwards, the VO enters the operational phase, executing Business Processes, while the evolution phase is entered whenever exceptions occur or compensation is necessary. The dissolution phase concludes the VO life cycle, dispersing the VO assets among the participants. VOs form for instance in response to a government issued tender in the Collaborative Engineering (CE) application domain, e.g. to upgrade a plane or in domains such as Ad-Hoc Service Provisioning (AH), e.g. offering electronic services to business travellers. The VO manager role (the trustor) is responsible for the decision making which potential members (the trustees) are invited and subsequently selected to join the VO, a crucial decision with respect to the entire VO's success [2]. The possibility of a VO partner performing badly during the VO's operational phase or announcing bankruptcy endangers the investment taken in integrating their processes and infrastructure for the purpose of the VO. A reputation system can provide additional decision support besides the a priori knowledge from quotations and bidding to avoid events such as VO partner replacement by minimizing the risk [3] of choosing unreliable partners in the first place. The STochastic REputation (STORE) system for VOs [4] is designed to offer exactly this kind of decision support. It aggregates observable data about a trustee's behavior, following principles of probability theory. To achieve this, reputation, an objective trust measure, is aggregated from multiple independent trust sources, so-called Trust Indicators (TIs), inherently characterizing an organization's reliability. Such TIs provide measurable trust values about an organization from heterogeneous domains, e.g. capturing operational aspects such as timely delivery of a service, organizational stability measuring employee turnover or environmental risks due to factory locations. Trust aspects captured by the TIs are, in contrast to reputation, highly subjective properties conveying the probability of an organization's expected reliable behavior [5]. To allow for the desired predictions of an organization's future performance, a stochastic modeling approach is chosen [6]. More concretely, this approach integrates a stochastic trust management approach in a reputation system.

This publication provides a simulation based evaluation of the STORE reputation system. The contribution is three-fold:

1. Design and implementation of a Multi-Agent based Simulation (MAS) framework to evaluate the STORE reputation system.
2. An interface extension of the STORE reputation system offering a trustor the capability to express application scenario specific trust preferences about a trustee.
3. Evaluation of reputation based VO member selection in two different VO application scenarios from Collaborative Engineering (CE) and Ad-Hoc (AH) service provisioning.

In the following section, related work from Trust and Reputation Management as well as MAS is discussed. Section three starts off with a brief introduction of STORE's model and architecture, followed by the detailed design and implementation of our simulation framework. Section four then comprises of the evaluation results from the two VO application scenarios. Section five concludes and offers an outlook on future work.

2 Related Work

This section provides a brief enumeration of related work from the research field of Trust and Reputation Management, followed by a more detailed assessment of related work in the area of (Agent-Based) Simulation.

Trust is a complex sociological phenomenon. In the context of a VO, we define trust as the subjective probability by which the trustor expects the trustee to perform actions captured in a role specification within the context of a VO. This definition is based on work from Gambetta [5] and Jøsang [7]. In the area of information technology, the term trust management was introduced by [8], who defined the term "trust management (problem)" as the collective study of security policies, security credentials and trust relationships. This technical perspective resulted in a system providing access control for distributed environments. Following this ground breaking publication, a multitude of trust management approaches were developed and published in parallel. The most recent and successful ones are surveyed in [7]. On the higher level of business to consumer e-commerce, Egger at al.[9] present a related trust model, directly observable factors that characterizing online vendors. While the general approach to root trust in observable indicators is comparable to our work, the application domain and hence the relevant indicators are different.

Reputation is a known concept in many disciplines, equally broad as and closely related to trust [10]. Reputation is frequently interpreted as the general opinion of a group towards an individual, another group of people or an organization. Broken down to the field of trust management in VOs, we define reputation as the perception a VO has about the intentions and norms of another organization. This perception develops through past actions and through objective indicators. A general reputation can thereby be mapped to an individual binary (directed) trust relationship. Related reputation systems based on stochastic principles are Jøsang's Beta Reputation System [11], TRAVOS [12] and Regan's Bayesian modeling approach [13]. While the Beta and TRAVOS reputation systems are based on feedback from personal experience, the latter, in the same manner, relies on modeling with one single distribution. In contrast, STORE takes a different approach with a richer trust model, rooting trust in a set of trust sources instead of one. Each TI is modeled with a TI specific probability distribution function instead of the same. This approach then demands a more sophisticated aggregtion of the trust sources. While for instance the Beta distribution itself aggregates

the homogeneous trust sources to a reputation value, STORE's heterogeneous trust sources require an aggregation layer based on a tree shaped Bayes Network. The Beta Reputation System was recently extended to multi-valued feedback with contributions from one of the authors [14]. Reputation systems, mostly focusing on systems using transaction feedback as only input parameter, are frequently evaluated using MAS. Two reputation systems, following a probabilistic approach comparable to the STORE system, are the TRAVOS model by Teacy et al [12] and the Beta reputation system by A. Jøsang et al. [11]. To evaluate these systems, a MAS marketplace was created. We also adopt a similar agent based framework design, where different kinds of agents interact. This approach is able to capture a high degree of complexity, as encountered in VO environments. In many MAS models, "bad behavior" is a choice an agent takes out of self-interest. This is not only the case for the two examples mentioned above, but also for the large class of reputation systems that are evaluated by playing a prisoner's dilemma game. Lik Mui et al. [15] summarize the recent work in this area and illustrate the resulting strategies. In these models the real quality of the agent is determined by his choice of strategy in a game, that allows for exactly two different strategies and the reputation is an indicator for the kind of strategy a certain agent is playing. Our approach differs in that our agents do not "choose" to perform badly, but cooperate with their best possible quality, while not every agent is capable of offering the same high quality. We believe, this approach delivers a better model for business environments, where the outcome for all participants heavily depends on the quality of the individual inputs.

All related reputation systems define one function to compute a reputation value. By offering, in addition to a consolidated reputation value, reputation values for four trust classes, STORE offers several functions that allow the setting of application specific trust preferences per class. Furthermore the change of an agent's behavior is swiftly captured as an adjusted reputation value.

3 Simulation Framework

This section introduces the MAS framework, designed to evaluate the STORE reputation system. First, the requirements derived from both VO scenarios are analyzed, accompanied by a brief scenario description. Second, the agent model and framework design is presented. Detailed information about the framework's setting for the different scenarios conclude this section.

3.1 Requirements from Application Scenarios

The goal of the STORE reputation system design, as published in [4], is to provide reputation based decision support for selecting trustworthy partners in VOs. Such support is intended to be delivered for all classes of VOs and not a specific class in particular. The EU funded IST project TrustCom published, as one of its deliverables, a VO classification along

with class specific security and trust requirements [16]. The outcome of this classification exercise was a nearly continuous set of VO classes that could be distinguished, among others, according to properties such as expected lifetime and expected volume of business transactions. In this publication, we will evaluate STORE in the two contrasting VO scenarios from this classification.

This is, first, a CE VO from the aerospace industry, addressing a government issued tender for a plane upgrade. A VO manager such as British Aerospace Systems, enacting the manager role, seeks to attract the most trustworthy design specialists, several storage providers and high performance computing providers. In summary, the the integrator seeks around five trustworthy organisations to start a VO. The operation then starts with the integrator submitting design upgrade tasks to the design specialists. The data is maintained by a storage provider and retrieved for design activities from this location by the design specialists. Updated design data is uploaded back to the storage provider. A similar work pattern then continues with the design validation where the design specialist simulates the design data on the high performance computing provider's infrastructure and e.g. reports faulty aerodynamic behaviour, triggering another design step. This VO requires weeks to be set, may last up to several decades and single transactions conducted among the members reach a high financial volume, up to the M€.

In an AH VO, the members, consisting of mobile operators, local and global electronic service and infrastructure providers as well as restaurants and hotels, provide electronic services to business travellers. Assuming the VO manager maintains for instance the local WLAN network infrastructure, at least two organisation - the traveller's home operator and a content provider - are needed for a VO. Such a traveller, residing at a foreign location e.g. in a hotel (enacting the VO manager role) seeks access to electronic services providing data and other digital content. Such services range from local tourist information, weather forecast or entertainment to the traveller's home operator's portal with e-mail or stock information. A telco scenario, consisting of the hotel as integrator, local and home operator as well as the service providers is set up in seconds. It may last for hours up to days (billing in the end taking most of the time) and transactions with little volume are conducted. For the remainder of this paper, these scenarios will be abbreviated as the aerospace and telco scenarios.

To judge the claim of STORE being able to provide reputation based decision support for VO scenarios in general, the evaluation methodology must allow a comparison of the reputation's decision support on the partner selection across different scenarios. Due to STORE's Bayesian approach on reputation computation, as described in the following sections, and the intrinsic uncertainty of an organisation's trustworthy behaviour, the reputation mechanism can not be analysed analytically in closed form. We adopted a MAS methodology because it is, to our knowledge, the only choice that integrates the stated VO properties as simulation settings and meets all other requirements and conditions.

3.2 The STORE reputation system

The left box in Figure 1 provides an overview of STORE's model and architecture, as far as required to follow the coming sections in this paper. The system's details are published in [4]. Organizations are represented by their TIs, characterising their observable trustworthy behaviour. "Delivery Delay" is a sample TI, characterising one operational aspect of the organization's trustworthy behavior. TIs are classified into $m = 4$ Trust Classes (TCs) of environmental, financial, operational and organizational TIs. This taxonomy of TCs is extensible, further classes can be defined. The TCs are derived from extensive research in domains related to reputation management such as Operations Research, Risk Management[4]. Each TI is modelled with a Probability Distribution Function (PDF), which is updated with each new data observation, according to TI specific update intervals. TIs are aggregated towards the reputation vector using a Bayes Network[4]. TIs are modeled in the leaf nodes. Their heterogeneous PDFs are normalized according to a common ordinal scale by the set of preference or π-nodes in the network. With the help of the preference nodes, the heterogeneous TIs become comparable. All TIs are aggregated by the generalized reputation node R, while the TC nodes only aggregate the TIs belonging to their class. The STORE reputation system finally delivers a reputation vector consisting of the TCs and R to a reputation requester. Each vector component's value consists of the expectation value of its node's PDF and is defined in the interval $[0, 1]$. This reputation vector extends the originally in [4] published interface which only returned the generalized reputation value R. The new TC components allow a VO manager to express trust preferences according to his VO application context, e.g. by emphazising operational trust aspects. The form, trust preferences are expressed in, is described in Subsection 3.3.

3.3 Agent model and MAS Framework

The MAS framework simulates a full VO lifecycle, one in each round. Agents, representing organisations, are matched to form a VO based on their reputation in each round. An agent enacts exactly one of the two VO business roles, manager or member. Its behavior is determined by a set of modelled TIs which characterize this agent's observable, trustworthy behaviour. In this paper, we focus on the six TIs, listed in Table 1, covering all four TCs. TIs are observed in regular time intervals. From these TI data observations, an agent specific reputation vector is computed by STORE and the MAS framework derives a QoS value (defining the agent's production input) in the interval $[0, 1]$ from a scenario specific QoS rule set. A value of one is interpreted as the maximally achievable performance. The framework also maintains a virtual "bank account" for each agent. The right hand box in Figure 1 illustrates these described components and their relations.

---

[4] A Bayes network is graphically represented as a directed, acyclic graph; nodes maintain random variables, edges denote conditional probability relationships between the connected nodes.
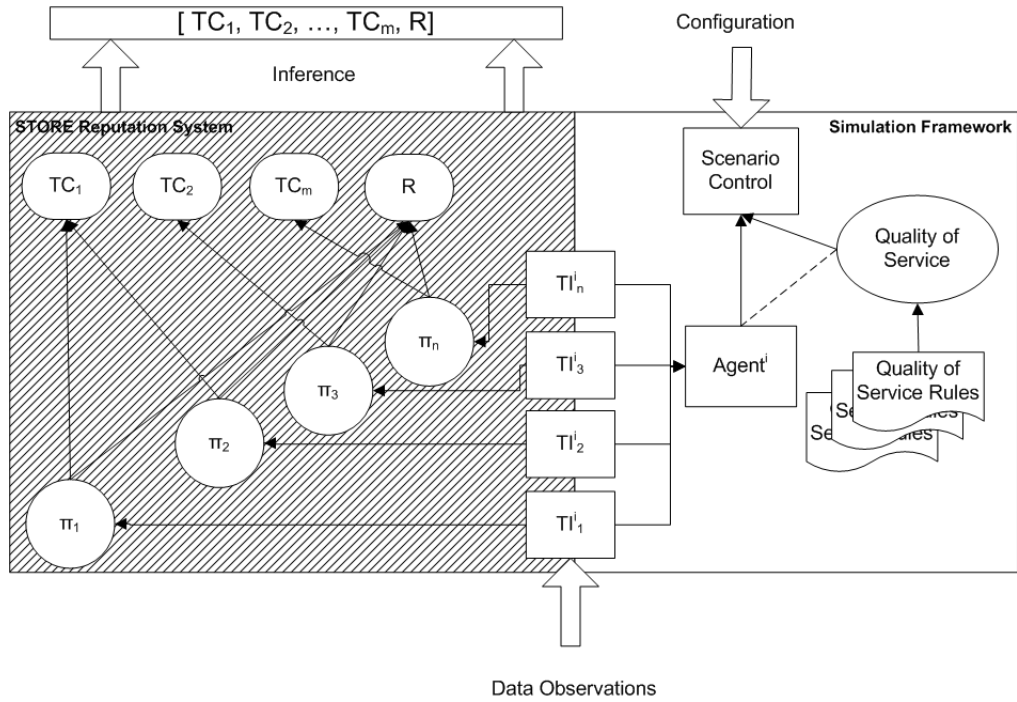
Fig. 1. STORE and MAS Framework architecture

Agent classes are assigned to agents, configuring their behaviour for a planned simulation. For this paper, four agent classes are defined. Class 1 defines a uniformly, in all TCs, well performing agent, Class 2 defines an agent who is specialized in telco scenarios and performs well in the operational and organizational TCs. Similarly, Class 3 defines an aerospace scenario specialized agent which performs especially well in the financial and environmental TCs, while Class 4 defines a uniformly badly performing agent. Table 1 lists the different agent classes and as well as their performance with respect to TIs and TCs..

TI	Trust Class	Class 1	Class 2	Class 3	Class 4
Country Bond Spread	Environmental	+	-	+	-
Cash Flow Margin	Financial	+	-	+	-
Complaint Rate	Operational	+	+	-	-
Delivery Delay	Operational	+	+	-	-
System Downtime	Operational	+	+	-	-
Employee Fluctuations	Organizational	+	+	+	-

Table 1. TI settings for different agent classes (+ high, - low)

The framework generates TI data for each agent by drawing from each individual TI's inverse PDF. Technically, assigning a class to an agent sets the inverse PDF's parameters, hereby determining the agent's later behaviour during the simulation. For example the Delivery Delay TI follows an Exponential distribution with one parameter λ and its mean $\frac{1}{\lambda}$. Assigning Class 4 to an agent sets λ to a smaller value, hereby shifting the exponential distribution's entire support and its mean towards higher values of delay. For instance setting $\lambda = \frac{1}{2}$ results in an expected delay of 2 (days). Assigning Class 1 with e.g. $\lambda = 1$ would result in the opposite, a shift towards smaller delay times of one (day). Adopting this statistical approach allows for a realistic model of an organization's trustworthy behaviour. Business partners typically follow a certain strategy, e.g. honest or malcicious, which is captured by the agent classes and the inverse PDF's mean. Furthermore, organisations show small deviations around this strategy in their behaviour. Even honest ones may exhibit e.g. temporary lapses in trustworthiness due to events out of their control such as accidents or infrastructure malfunction. This volatility is captured by setting the inverse PDF's variance.

The simulation differentiates agents into managers and members where an agent of a certain type can only interact with his counterpart. Agents are matched with the Gale and Shapley [17] algorithm. Each agent computes a preference relation over every agent of the opposite VO role that is only based on the agent's reputation vector. Such a matching leaves no incentive for any two partners to enforce a different matching and is therefore called "stable". Adding now the VO application domain context to the matching, the reputation vector $\vec{R} \in \mathbb{R}^5$ for agent i as provided by STORE consists of a value for each TC ("environmental", "financial", "operational" and "organizational"), as well as the generalized reputation value R. ψ defines the metric for an agent-specific preference relation, $\psi : \mathbb{R}^5 \rightarrow \mathbb{R}$, as $\psi(\vec{\omega}) = \vec{R} \cdot \vec{\omega}$ where $\vec{\omega}$ captures the agent specific "weighting" that sets the trust preferences for ω's components. For example $\vec{\omega}_0^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}$ represents a preference vector that only incorporates the generalized reputation value and ignores the trust classes. A scenario specific preference vector for the separate TCs captures their relevance in different VO scenarios.

The MAS framework entails a scenario control component, controlling the progress of a running simulation scenario. Simulations are executed in rounds, in each round performing the following actions:

- Generate TI data for every agent.
- Calculate a "Quality of Service" (QoS) value for every agent from the TI data according to the QoS rule set.
- Request the updated reputation vectors from STORE for every agent.
- Compute a preference relation for each agent over all potential partners of the opposite VO role.
- Perform matching according to the preference relations. The matched agents form a new VO.
- Calculate production for newly formed VO and resulting pay-off for every agent.
- STORE receives the TI data as new data observations, resulting in a Bayes Network update.

The production or productivity y of a VO is bilaterally calculated using a Cobb-Douglas productivity function $y = x_{Manager} \cdot x_{Member}$, where in our model the input factors x_i are the QoS values. They are dimensionless and represent a productivity rate compared to the possible optimum. The productivity between a manager and member is equally divided between both parties, resulting in a pay-off rate for every agent. A higher QoS of the transaction partners leads to an increased pay-off for both of them. This way, we create an incentive for each agent to transact with a partner of higher QoS and a method to assess how well STORE compute's an agent's reputation by comparing it with the cash rates of agent representatives from the four agent classes.

The MAS framework is tailored for STORE's evaluation. This becomes apparent when examining the TIs characterising an organisation's trustworthy behaviour. The MAS takes the aggregation of multiple, heterogeneous trust values towards a reputation value into account. The framework can very well be used for the evaluation of other reputation systems, e.g. aggregating homogeneous feedback values, that provide a less challenging evaluation task. The QoS rules serve well as a measure to compare different reputation systems with each other in the same application scenario.

3.4 Technical scenario settings

The MAS framework is implemented in the Java programming language and offers a set of configuration options to define VO application domain specific settings. First, the agent set up can be freely defined including the number of manager and member agents, as well as their class assignments. The agent ratio, how many members are required to form a VO with one manager is separately defined. Regarding the general simulation properties, the number of rounds is part of the setting, along with a number of so-called blind rounds. The latter do not contribute to the generated result sets, but are defined to bridge the system's tune in phase. This allows for equal, well defined cold start conditions of simulations which are supposed to compared afterwards. To avoid large unjustified financial increases of the agent's accounts, transaction costs per round can also be set. The length of one round in real time and the number of total rounds played, vary in both selected scenarios. The length of a round is an important factor for the reaction speed of the reputation value and should be comparable to the expected lifetime of the VO (when simulating a VO with an expected lifetime of 10 years with rounds that last for one week real time does not make sense).

4 Evaluation

In this section, the results of our evaluation are presented. First, the setting for the two VO scenarios is described. Second, we show how STORE is able to separate agents of different classes by reputation. Third, as an additional reputation independent measure, an agent's cash rate is introduced and used to compare simulation results across different VO application scenarios. A reputation system independent random matching for

the potential partners is used as a benchmark against the two scenario specific STORE matchings. The same simulation results are used to discuss a VO's productivity increase when relying on STORE's decision support. The Section concludes with a sensitivity analysis. All presented data is average over 50 runs of the identical scenario configuration. A transaction cost per round and agent of 0.15 currency units is set.

4.1 Simulation Setting

The aerospace environment from 3.1 consists of few organisations with few high-turnover transactions, due to the high initial investment necessary for joining. The aerospace market in our simulation consists of one Class 1 manager seeking five member agents for the formation of an aerospace VO. There are in total 9 potential members: 2 Class 1, 2 Class 2, 2 Class 3 and 3 Class 4 agents. One round lasts for about half a year real time and 20 rounds are played in one scenario. The STORE system is configured to react slow to change. The CE weighting vector $\vec{\omega}_{CE}^T = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 1 \end{pmatrix}$ includes, with weight 0.5, the TCs "environmental" and "financial" along with the generalized reputation value. The two trust classes are emphasised, since physical goods are transported in such VOs relying on proper infrastructure and since high volume transactions take place. This preference vector takes, by preferring the less frequently observed "environmental" TC, also the long term orientation of an aerospace scenario into account. These TCs tend to capture a sustainable development with the rarely observed TIs "Country Bond Spread", capturing a region's environmental risk, and "Cash Flow Quote".

In the telco scenario from 3.1, many transactions are executed during one day and the typical transaction volume is low. This lowers the entry barrier for new and small companies. In the telco simulation environment, 3 managers (2 Class 1 and 1 Class 4) are matched with 2 members each. On the market there are 1 Class 1, 2 Class 2, 2 Class 3 and 2 Class 4 potential members available. The STORE reputation system reacts quickly to change. 50 rounds are played where one round represents about five days real time. The telco trust preference vector $\vec{\omega}_{AH}^T = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \end{pmatrix}$ only weights the TC "operational", as this class focusses on the current quality of service ignoring long-term trends, and again the generalized reputation value. In such a short-lived environment, fast and timely service ranging from low-level network speed to higher-level content delivery of stock data is essential. For example the TIs "System Downtime" and "Delivery Delay" belong to this TC and are both frequently observed. In this setting, the sums of both trust preference vector components are equal, simplifying cross-VO scenario comparisons.

4.2 Separation of agents from different classes

To illustrate how STORE can separate agents by reputation according to their class assignments, we take a closer look at the development of the reputation values compared to the reputation independent QoS measure. This first simulation evaluates this very basic function in a static setting without dynamic changes in the agent's behavior.

Figure 2 depicts the reputation of the four classes of agents over the simulation rounds. The generalized reputation value of the two specialized agents (Class 2 and 3) is clearly, as desired, in between the reputation of agents from classes Class 1 and 4. This good result becomes even more appreciated when compared to the QoS measurements in Figure 3. While the agents representing the four classes are clearly and continually separated by reputation, their actual, real performance is not. For instance the measured QoS for the agents of Classes 1, 2 and 3 even intersects in round three. The agent classes are on purpose defined that the resulting inverse PDF parametrisations constitute a worst case where agent classes only marginally differ in their real performance measured by QoS. This first functional test scenario was conducted in a telco scenario setting without introducing a VO context in form of a trust preference vector yet.
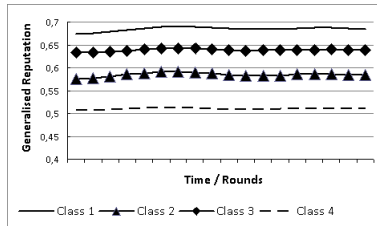
Fig. 2. Generalized Reputation of the four agent classes changing over time in an telco scenario
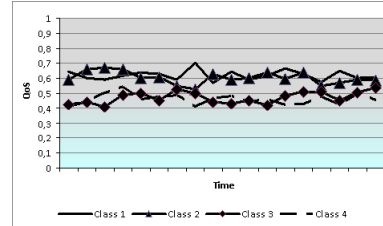
Fig. 3. QoS of the four agent classes changing over time in an telco scenario

This neglect of VO context and TCs explains, why the Class 3 agent carries a higher generalized reputation value in a telco scenario setting. The following simulation scenario now addresses this lack of VO context by joining results from both, telco and aerospace scenario simulation runs, compared with the random matching benchmark scenario. Each scenario is executed with its specific setting and the manager agents use the above defined trust preference vectors $\vec{\omega}_{CE}$ and $\vec{\omega}_{AH}$. If STORE delivers the desired, VO specific reputation based decision support, the most trustworthy agents (with the best QoS for their VO scenario) should interact with each other and thus receive better pay-off rates. Their overall cash rate should be higher than in the random matching benchmark. To cross-evaluate the newly extended reputation interface with the trust preference vectors, we additionally evaluate each scenario with the correct scenario specific trust preference vector against the same setting with the vector from the other scenario (an aerospace scenario with weighting vector $\vec{\omega}_{AH}$ and vice versa). We expect, that with the correct scenario specific trust preference vectors, agents specialized in the same VO domain are preferably selected along with Class 1 agents, resulting in higher cash rates. The random matching benchmark represents a VO environment with no reputation based decision support.

The resulting cash rates for these three configurations are illustrated in the following Figures, first clearly indicating the superior performance of the reputation based matching compared to the random matching. In the random matching benchmark (Figure 4), every agent performs badly since being randomly matched even to Class 4 agents. In all scenarios and with both trust preference vectors, the Class 1 and 4 agents are continually clearly separated by their cash rates and, as designed and expected, exhibit superior and inferior performance respectively. Figure 5 shows the results of an aerospace scenario with AH trust preferences, as e.g. expressed by an uninformed VO manager. As shown in Section 4.2, the wrong trust preferences still manage to separate Class 1 and 4 agents, but struggle to identify scenario specialized agents. The specialized Class 2 and 3 agents perform better as in the random matching, but are not separated as well as with the correct CE trust preferences depicted in Figure 6. This scenario, assuming an informed VO manager, delivers the best cash rates, by preferring Class 1 and the CE specialized Class 3 agents at all times (overlayed lines of both classes in the diagram). The Class 2 agent still receives a pay-off comparable to the random matching benchmark and the Class 4 agent is only selected if absolutely necessary.
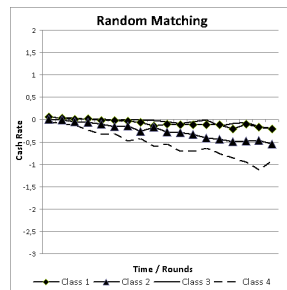

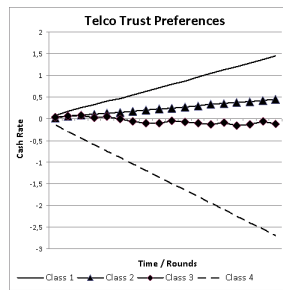**Fig. 4.** Cash rates for the four agent classes in a random matching aerospace scenario.

Fig. 5. Cash rates for the four agent classes in an aerospace scenario with wrong trust preferences.
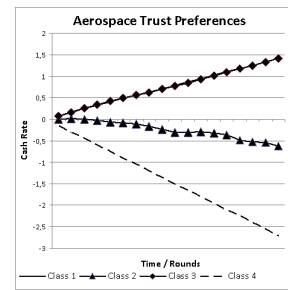
Fig. 6. Cash rates for the four agent classes in an aerospace scenario with correct trust preferences.

This clearly shows that STORE's reputation based decision support for selecting members of VOs in general meets the expectations, but at the same time emphasizes the importance of VO scenario specific trust preferences expressed by an informed VO manager. The reputation based matching executed with an improper trust preference vector generates significantly worse results especially for the specialized agents. The trust preference vector is an exogenous input in this simulation and has to be set by a business expert, the VO manager.

The results in an telco scenario are similar to the ones presented above: The reputation based member selection, using a wrong trust preference vector (Figure 8) manages to separate Class 1 from 4 agents, but only

a correct trust preference vector (Figure 9) is capable of differentiating specialized agents.
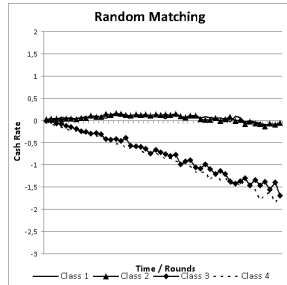

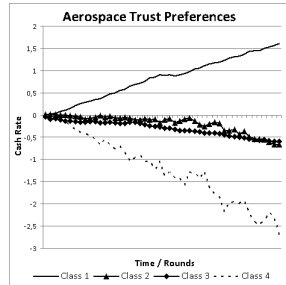**Fig. 7.** Cash rates for the four agent classes in a random matching telco scenario scenario.

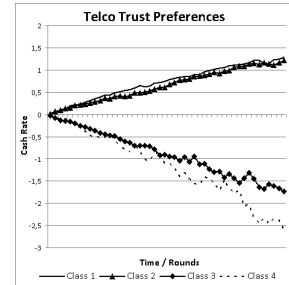**Fig. 8.** Cash rates for the four agent classes in a telco scenario with wrong trust preferences.

Fig. 9. Cash rates for the four agent classes in a telco scenario with correct trust preferences.

4.3   Sensitivity analysis of dynamic agent behavior

Many reputation systems struggle when observing the effects of a change in an agent's behavior. In our MAS framework, an agent's behaviour can be changed in each round with well defined configuration settings. With this powerful tool, research with more than the four previously defined, static agent classes becomes possible. We can simulate VO environments with agents exhibiting decaying or increasing performance, even modelling attacker agents with, for instance, oscillating performance, becomes possible. This allows to evaluate the STORE reputation system with different configurations, for instance reacting faster or slower to changes in an agent's behaviour. We begin with a scenario of four equally defined member agents. We configure one manager seeking three members to form a VO, so that every round, one of the potential members is not selected for the VO. Due to STORE's stochastic model, the observed generalized reputation value for the four agents is initially at a similar level with random diffusion. Which one of potential members is selected by the manager is random in these first rounds. After some rounds we change the behavior of all agents: two of the potential members (0 and 1) suffer a large decrease in their performance, while the other two (2 and 3) only decrease slightly. In Figure 10, left hand chart, we observe a significant change in the generalized reputation value for the first and a less severe for the second group. This affects the selection behavior of the manager which now prefers the members 2 and 3 for the rounds five to 17, suffering from decreased performance, in the right hand chart. When all agents recover to their initial performance, the selection becomes random over the entire group of potential members again.
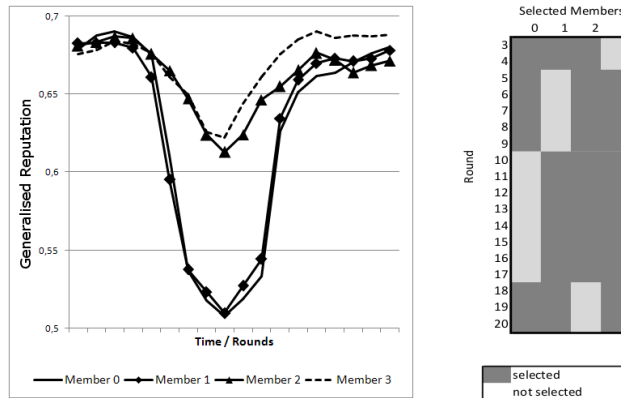
Fig. 10. The generalized reputation values for four potential members changing their quality over time and a table indicating which agents are selected for VO membership.

This illustrates that STORE reacts well to gradual change in an observed agent's behavior and that it enables every manager to react to an altered situation, by supporting his decision to select only the most trustworthy partners at all times.

The following scenario evaluates STORE's reaction speed to changes in an agent's behavior. This speed heavily depends on a STORE internal configuration, how far back in time observed TI data is considered for the reputation computation. A larger time window results in slower reaction which is suitable for a long-lived aerospace scenario, while a smaller time window and quicker reaction is more appropriate for an agile telco scenario. An exemplary comparison of the reaction speed for both configurations is shown for the generalized reputation value change in Figure 11 and for the "environmental" TC change in in Figure 12. This same change of behavior, occuring in round four, as in the previous setting is employed.
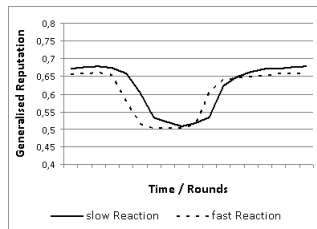

Fig. 11. Reputation reaction to quality change with STORE configured for slow and fast reaction.
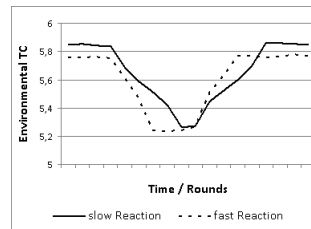

Fig. 12. TC reaction to quality change with STORE configured for slow and fast reaction.

We observe that STORE can be adapted to react slower or faster to behavioral changes, as demanded by a particular VO scenario. In fast paced telco scenarios, a change in the computed reputation can already be observed within the same round and fully adapts to the change within two more. A slower configuration, better suited for long lived aerospace scenarios, requires two more rounds which is still acceptable since one round amounts for appr. half a year, while aerospace scenarios may last for decades. However, configuring STORE's reaction speed should be carefully planned, since extreme settings in both directions, e.g. allows attacks on the reputation mechanism. While reacting too fast, older TI data is disregarded and an attacker may aim at boosting his reputation with well performed, low volume transactions, while reaping the benefits when cheating in singular high volume ones.

5 Conclusion and future work.

In this publication, the STORE reputation system has been evaluated adopting a MAS methodology. Due to STORE's trust model based on the TI taxonomy and the rich TI model itself, a dedicated MAS framework has been designed, in which the agents form VOs, selected according to their reputation vector computed by STORE. Member agents transact with a manager, receiving a pay- off based on their reputation independent productivity measure. Four classes of agents with varying adequacy for the two selected aerospace and telco scenarios have been introduced. Unfortunately, for space reasons, the detailed mapping of the agent classes to technical TI attributes, e.g. the PDF parameters, could not be provided in this paper. The simulation results show that STORE's reputation measure can not only separate polarized agents, only performing exceptionally well or badly. It also provides decision support in difficult cases, where agents are specialized in particular VO application domains and perform on similar levels, but have to be selected for application specific roles. The results show that STORE's one configurable architecture caters for a wide range of VO scenarios. With the reputation interface, that has been extended to a vector carrying besides the generalized reputation value, also the four TCs, an informed VO manager can obtain even better decision support. He expresses his trust preferences, basically his subjective opinion about a member's trustworthy behaviour, broken down as a vector with weights for the TCs. Well expressed trust preferences, capturing the VO context, lead to a better support for the partner selection and increased productivity during the VOs operation. Finally, it has been shown that STORE very well captures gradual and temporal changes in an agent's behavior.
So far, STORE was evaluated in an aerospace and telco scenario. We plan to derive settings for simulating critical events in a VO's lifetime, e.g. replacement of a misperforming member, and expect thereby further refinement of the existing settings. We also plan to simulate further VO scenarios with properties in between the presented ones, such as an eLearning VO. Further simulation results analysing first STORE's resilience against attacks on the system itself and the reputation mechanism are available. Having started with an attack model, leading to

an attack classification, we identified and simulated one novel type of freeriding attack. In this scenario, STORE's ability to recognize misbehaving agents is assessed. Second, we compared STORE to Jøsang's Beta reputation system. Both results will be published separately. Since STORE explores an approach based on a richer trust model, it takes some effort to present such a comparison. Due to STORE's flexibility, a VO scenario specific configuration can be derived that e.g. reacts faster to an organisation's changes in trustworthy behavior than the Beta system, but introduces a tradeoff to unwanted sensitivity in short term behavior. In short, STORE with a reasonable scenario specific configuration performed consistently better in the sensitivity analysis than the Beta system, but not by a large margin.

References

1. Strader, T., Lin, F., Shaw, M.: Information structure for electronic virtual organization management. Decision Support Systems (1998) 75–94
2. Robinson, P., Karabulut, Y., Haller, J.: Dynamic virtual organization management for service oriented enterprise applications. In: The First International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2005). (2005)
3. Cruz, M.G.: Modeling, measuring and hedging operational risk. Wiley, Chichester [u.a.] (2003)
4. Winkler, T.J., Haller, J., Gimpel, H., Weinhardt, C.: Trust indicator modeling for a reputation service in virtual organizations. (2006)
5. Gambetta, D. In: Can We Trust Trust? Basil Blackwell (1988) 213–237 Reprinted in electronic edition from Department of Sociology, University of Oxford, chapter 13, pp. 213-237.
6. Haller, J.: A stochastic approach for trust management. In: International Workshop on Security and Trust in Decentralized/Distributed Data Structures (STD3S). (2006)
7. Josang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems (August 2004)
8. Blaze, M., Feigenbaum, J., Lacy, J.: Decentralized trust management. In: SP '96: Proceedings of the 1996 IEEE Symposium on Security and Privacy, Washington, DC, USA, IEEE Computer Society (1996) 164
9. Egger, F.N. In: From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce. (2003)
10. Mui, L., Mohtashemi, M., Halberstadt, A.: A computational model of trust and reputation. 35th Annual Hawaii International Conference on System Sciences (HICSS-35) **7** (2002) 188
11. Ismail, R., Josang, A.: The beta reputation system. In: Proceedings of the 15th Bled Conference on Electronic Commerce. (2002)
12. Teacy, W.T.L., Patel, J., Jennings, N.R., Luck, M.: Travos: Trust and reputation in the context of inaccurate information sources. (2006)

13. Regan, K., Cohen, R., Poupart, P.: Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In: Twenty-First National Conference on Artificial Intelligence (AAAI). (2006)
14. Josang, A., Haller, J.: Dirichlet reputation systems. In: The 2nd International Conference on Availability, Reliability and Security (ARES). (2007)
15. Mui, L., Mohtashemi, M., Halberstadt, A.: Notions of reputation in multi-agents systems: a review. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems, ACM Press (2002) 280–287
16. (editor), P.K.: Case study scenarios- WP11 problem definition. TrustCoM (IST Framework 6 integrated project, grant 001945) public deliverable (2004)
17. Gale, D., Shapley, L.S.: College admissions and the stability of marriage. American Mathematical Monthly **69** (1962) 9–14