Why Trust Is Hard – Challenges in e-Mediated Services

Christer Rindebäck and Rune Gustavsson

School of Engineering, Blekinge Institute of Technology,
S-372 25 Ronneby, Sweden
{christer.rindeback, rune.gustavsson}@bth.se

Abstract. Design and maintenance of trustworthy electronically mediated services is a major challenge in supporting trust of future information systems supporting e-commerce as well as safety critical systems in our society. We propose a framework supporting a principled life cycle of e-services. Our application domain is distributed health care systems. We also include comparisons with other relevant approaches from trust in e-commerce and trust in agents.

1 Background

Trust has been identified to be a key issue when it comes to the design of user-accepted behavior of complex computer systems [1]. Examples of such systems include Multi agent systems (MAS) and emergent systems such as Network Enabled Capabilities (NEC) in defense and efforts related to European EC Programmes in Ambient Intelligent Systems (AmI). Furthermore R&D efforts in GRID computing and web services have a clear focus on issues related to design and maintenance of trustworthy information systems. Although trust and trustworthiness are common denominators in those efforts the approaches are quite different illustrating the complexities of the subject matter as well as the different backgrounds. Reputation and brand naming are examples of trust creating signs in the real world. The purpose of our contribution is to combine different approaches toward aspects of trust and trustworthiness into a framework that allows us to have a principled approach toward engineering of trustworthy behavior of computer mediated services (e-services). For instance, these systems need to be designed in a way that allows the involved entities to exchange information securely and in a trusted way, and that tasks can be delegated to parties that can be trusted to perform the task as expected by the delegating party.

E-services is advocated by large industry consortia as well as by international research communities as a promising future paradigm of the on-line environment providing electronically delivered service based on assembling and coordination of other services. A particular important societal application area is the organization of future health care utilizing information technology. We have had several projects focusing on future home health care based on emergent technologies. In the home health care area we are investigating support systems for health care personnel, home care personnel and patients to establish a trusted support for all parties involved in home health care. The application area is rich and challenging with respect to different trust models. Our suggested framework is based on our current understanding of trust aspects related to assessments of our prototypes and projects in the area.

R. Falcone et al. (Eds.): Trusting Agents, LNAI 3577, pp. 180–199, 2005.

Trust is largely a subjective issue [2]. Actors may trust, for example, a low security system, among other possible reasons, because they do not know better or because they think that security is irrelevant for the particular system. Trust has also a contextual relation to risk assessment. Obviously this assessment is fundamentally different in nature if, e.g., your life, reputation, or (part of) your economy is at stake. But trust is a concept with many dimensions directed toward different objects between multiple actors, e.g. agents.

This contribution focuses on principal challenges regarding understanding, designing, implementing and monitoring *trustworthy* information systems. Most aspects of trustworthy information systems, agent mediated or not, have been addressing trust (risk assessments) related to economic risks (e-commerce) or reputation (privacy concerns related to e-commerce). We are addressing areas where your life might be at stake, i.e., health care in home environments (e-health) or e-services used in emergency situations in our society.

To set the scene; we regard in our setting trust as a relation between a subject and an object regarding the behavior of the object in a given situation (context). The trust evaluation is a subjective assessment of the object behavior (actual or expected) based on the subject's relevant criteria. In the case that the object is an artifact, the subjective assessment can be supported or refuted by the perceived trustworthiness of the system. Since systems are engineered we are looking for design and maintenance criteria that supports (enforces) trustworthiness in our framework.

In the following Section 2 *Trust and Agents - a Background* we investigate the relationship between a number of identified dimensions and corresponding objects of trust and specifically trust in relation to MAS. Thereafter, in section 3 *Why Trust is Harder than Trustworthiness*, we identify the main issues of the paper as well a research agenda toward that end. The following Section 4 *A Framework Enabling Assessing Trustworthiness*, describes our approach in more details. We illustrate our approach with an example in designing trustworthy systems in a following section 5, *Trust in e-Services in Home Health Care*. We conclude the paper with two sections of comparisons with other approaches, *Models of Trust - Other Approaches*, and, *Trust in Agents - Other Approaches*. The final section, *Conclusions and Further Research*, includes self assessments and pointers to further investigations on the important issue of trust in electronically mediated services.

2 Trust and Agents – a Background

During the last decade two complementary views on agents and trust have emerged with roots either in agent technologies or in models of trust. In short, we have witnessed research agendas on aspects of trust, from a *user* point, in behavior of agent systems on the one side or research agendas focusing on models of trust *between agents* in agent societies on the other side. Sometimes it is not entirely clear what the focus is in papers on agents and trust. In this paper we claim that the first view is a sound one where as the latter view is more troublesome given present state-of-the-art in agent technologies and models of trust. To support our claim we first give a short overview of relevant models of trust followed by an (also short) overview of state-of-the-art of agent technologies.

2.1 Models of Trust

Below we give a short overview of contemporary models of trust along two dimensions; (1) subjects/objects of trust, more precisely between humans, human entities and organizations, social/natural order, and, artifacts, (2) dimensions of trust, that is, ethical/moral behavior, professional competence, and specifically concerning artifacts, functionality and reliability. It should be noted that artifacts in this overview corresponds to physical artifacts or embedded software control systems (e.g., a sledge hammer or a VCR). We will return to agent-based artifacts later. The following table (Table 1) captures the relevant relationships marked with references to relevant work.

Table 1. A matrix of trust models and their relation to objects and dimensions of trust

Subject / Object	**Ethical/moral behavior**	**Professional competence**	**Action fulfillment**	**Reliability**	**Functionality**
Artifacts	N/A	N/A	N/A	Muir[3]	Muir[3]
Humans Deutsch[4], Rempel et. al[5].	Barber[6], Baier[7]	Barber[6]	Gambetta[2]	N/A	N/A
Communities Giddens[8]	Barber[6]	Barber[6]	Gambetta[2]	N/A	N/A
Trust in Social Natural order & Confidence - Barber[6], Luhmann[9]					

Trust is complex not just in the sense that we may speak about what to trust by whom, with regard to who, or what, but also with respect to the dimension of the behavior of the object the subject have trust in. The table above depicts a number of subjects/objects of trust as well as dimensions of trust. For example; a human might trust that another human has professional competence in a specific context, or trust that a VCR has the intended functionality and reliability. However, we do not even think of ethical behavior from a VCR but this has emerged as a major concern regarding downloaded software (spyware and malware). The subjects/objects in a trust relation are actors (phenomena) involved. In the model of trust above four categories of subjects/objects of trust are presented. We can, for instance, investigate the trust of an individual in the behavior of a society or the other way around. Depending on what the subject-object roles are we can have quite different models and outcomes of assessments of the relevant trust example; an illustrative example is the different views and concerns related to privacy in our societies. The following subjects/objects of trust are part of our model:

- *Trust in social/natural order and confidence* - Our society rests on basic assumptions about what will and will not happen in most situations. For instance we have trust in the natural order, that the heaven won't fall down or that the natural laws will cease not to be true. There is also a general trust related to the social order in most of our societies, that is that the governmental representatives will do the best for the citizens and countries they represent and follow laws and norms as well as follow established practices accordingly. This mutual trust isn't something that actors in general reflect consciously about. The non-reflective trust serves as a basic trust/confidence level for our daily actions where in general there isn't any alternatives to the anticipated risks. The notion *confidence* [9] is sometimes used in situations where actors in reality have no choice. It isn't a viable option to stay in bed all day due to concerns about the social or natural order.

- *Trust in communities* - Humans are often part of a larger community. In the society we have for instance companies, non-profit organizations, governmental institutions and other groups of humans, which often act according to policies, and interests of the community. In many cases the trust may be attributed primarily (or at least in part) in the behavior of a community e.g. a hospital. On the other hand, a hospital may be perceived as more trustworthy than another due to better reputation regarding the perceived treatment and quality of their staff. Depending on the context, trust by a subject may be placed on the object being a community, an individual representing the community, or both.
- *Trust in humans* - In many situations we attribute trust toward other humans, we may trust a particular person about his capabilities or trust his intentions about a particular action. When buying a used car for instance we may trust a car salesman to a certain degree or trust a neighbor being an honest person. Trust between humans has been studied among others by [4, 2, 5].
- *Trust in artifacts* - Trust in human made objects such as cars, computers, VCR:s are in some cases discussed in a manner which implies that these objects can be seen as objects in which trust is placed. For instance 'I trust my car' or 'they trusted the bus to arrive on time'. This means that our expectations regarding the objects with respect to reliability are in some sense confused with or attributed for trust in humans enabling the intended behavior (the design and implementation team of a company, the driver of the bus employed by a public transport company). Since it is unusual or questionable to discuss classical (non-software) artifacts as trustworthy entities with bad will (or good) toward others or as in possess of emotions the use of the notion trust in artifacts is not classically applicable in those settings.

The trust by a subject defines toward what object the trust is attributed and along which dimensions. The following classifications of trust dimensions has been identified in literature on trust models:

- *Functionality* - The functionality of an artifact is an important and natural quality of trust, e.g., the tools are expected to function as they should. An implicit trust condition is that an artifact or tool is not behaving in an unexpected or undesired way by its design [3]. As we have indicated earlier, this situation is quite different when it comes to computer (software) based artifacts, that is, e-services. Firstly, the available functionalities, or affordances, are more complex (flexible). Secondly, and more important from a trust perspective the software can be designed by purpose or by affording vulnerabilities to create dysfunctional behavior that can be very harmful to the user or her system. The explicitly available functions and their appearance and accessibility shape the e-service from the perspective of its users. The user has to trust that these services meet her trust criteria in a trustworthy way without unwanted results. In our framework we indicate how we can meet these requirements from a designers point of view.

- *Reliability* - The reliability of an artifact is another important criteria of trust in classical artifacts. The tools should be resistant to tear and wear in a reasonable way and the VCR should function flawless for some years. Reliability thus means that an artifact can be expected to function according to the presented functionality and is working when needed. In some contexts reliability can be interpreted as safety, for instance, a safe electric equipment has protection (fuses) against short-circuits that could be harmful. Again, when it comes to software mediated services the trust dimensions of reliability and safety need to be assessed from different aspects.
- *Trust in Action Fulfillment* - In cooperation a specific trust dimension surfaces in most contexts. That is, can a subject trust that an object will indeed fulfill a promise or obligation to do a specified action? In a subcontractor scenario or in a health care situation where a doctor has prescribed a treatment to be carried out concerns may arise about whether the treatment will be carried out or not. Similar concerns can be identified in e-services regarding whether e.g an ordered product will be delivered or not.
- *Trust in Professional Competence* - When a decision to delegate a task to another actor is taken this is often based on a perception of that actors professional competence. This refers to expectations about the professional abilities [6] of e.g. a doctor or banker and suggests further refinements of trust expectations. We can trust somebody to have the right competence for carrying out actions associated with their profession. We trust doctors' judgments about medical needs and we trust them in their ability to adjust treatments in accordance with new findings within their area of expertise. In many situations humans can't gain complete insight in all qualities, aspects and problems characterizing professionalism in certain domains where we need help or assistance. Instead the decision weather or not to engage in a relationship with, i.e., a doctor or act according to the recommendations by a professional is based on trust in the professional competence of that actor.
- *Trust in Ethical/moral Behavior* - Trust isn't only related to professionalism in dealing with tasks as such, it is also suggested to be linked to values and less tangible nuances such as ethical and moral premises. If a trusted professional acts in a manner that is perceived as being against common ethical and moral norms we can choose to distrust this person in a given context despite his professional skills. Examples include certain types of medical experiments or other acts that can be regarded as unethical or even criminal if detected. Trust in moral or ethical behavior is, of course, very context dependent. Moral trust or as it is put forward in [6] as trust in fiduciary obligations means that some others in our social relationships have moral obligations and responsibility to demonstrate a special concern for other's interests above their own. The lack of control will give the trustee the possibility the possibility to exploit or harm the truster [7]. The ethical/moral trust dimension is based on a scenario when there is a risk for betrayal based on ethical and moral reasons. For instance in an e-service the information handled by the involved organization about individual clients can be misused in unethical manners in a way that is perceived as unmoral and would harm the truster. This is also connected to willingness from the trustee to put the truster's interest before his or her own. For instance, an e-service designed for health monitoring is expected to be mainly ben-

eficial for it's customers. The data collected could be used by the service provider as statistical data that could be passed on to the highest bidding parties.

2.2 Models of Agent Capabilities

We note the obvious fact that most contemporary trust models are related to trust and trust dimensions in human-human relations, see previous section. In the same way, trust models related to artifacts are complementary in the sense that human capabilities and expectations in the form of ethical/moral behavior, professional competence and action fulfillment is replaced by the technical requirements of functionality and reliability. For instance a human can be trusted to act in a moral ethical manner in a certain context whilst it makes no sense to claim that an artifact is acting by itself in this manner [10]. Having said that, there are many open issues related to trust in e-services and software agent-mediated services. For instance, can an (agent-mediated) artifact be instructed or designed in a way that measures up to or comprises ethical/moral behavior? What is the meaning of demanding accountability or liability on agent behavior? Our position is that these kinds of responsibilities are only meaningful and enforceable on the owner of the (agent)services.The purpose of this section is to revisit the (classical) discussion on trust as summarized in Table 1 into the situations where we have either a subject assessing her trust in agent mediated services or in situations where it is justified to model trust within the (software) agent societies. This is done in our proposed extension table 2. The Subject/Object heading indicates the two different interpretations of trust. The first row of the table is the situation where a user can assess her trust dimensions regarding the behavior of the agent-mediated services offered. In short, a user can judge (direct or indirect) whether or not the system behavior is either of ethical/moral, competent, fulfilling, functional, or reliable. The second row is the second reading of trust; between software agents themselves. The bottom line is that we regard the first user assessed trust dimensions to be the only viable stance given state-of-the-art agent technologies today and in a foreseeable future. That is, we claim that state-of-the-art agent system can not be trusted to have ethical/moral behavior even if the system can be design to have a formal rational behavior, e.g., regarding to problem solving in a technical domain. The reason for this stance is that we regard ethical/moral behavior to reflect on the societal consequences of, e.g., rational behavior. On the other hand we believe it is possible to implement self assessment models in an agent system to allow for decisions on (formal) competence, action fulfillment, functionality, or, certain aspects of reliability. Our framework supporting design and maintenance of trustworthy systems is based on that assumption. The rationale for the statements of the matrix of table 2

Table 2. A matrix of trust models related to agent mediated services and within agent societies

Subject / Object	**Ethical/moral behavior**	**Professional competence**	**Action fulfillment**	**Functionality**	Reliability
Agent system behavior	Yes	Knowledge Based Systems	Yes	Yes	Possible
Within agent systems	No	Possible	Possible	Possible	Possible

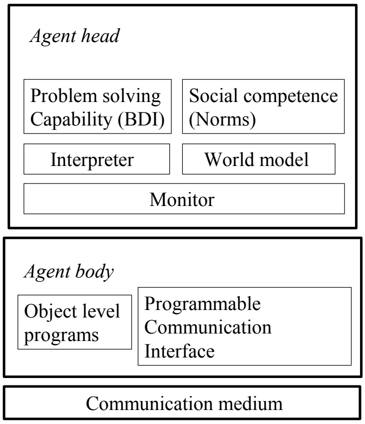
Agent head
Problem solving Capability (BDI)
Social competence (Norms)
Interpreter
World model
Monitor
Agent body
Object level programs
Programmable Communication Interface
Communication medium

Fig. 1. A reference architecture for agents in a multi-agent system (Belief-Desire-Intentions) architecture that also implements a local rational behavior accordingly. Technologies supporting MAS are focusing on Agent Communication Languages (ACL) and coordination patterns as well as community (institution, society) models [12, 14]. The latter are at present based on natural and social systems (normative behaviors)

are as follows [11, 12, 13, 14], where [12] includes a state-of-the-art overview of agent technologies and a road map up to 2009 and onwards. The different definitions of an agent from the agent research communities are emphasizing an agent as an autonomous computation software entity with a rational behavior. For multi-agent systems the focus is on interactions and co-ordinations of individual agents to achieve a common task or behavior. The capabilities of individual agents and a MAS are determined by which architectures that can be implemented [14, 15]. The capabilities of individual agents have hitherto been focusing on problem solving capabilities manifested by the well-known BDI (Beliefs-Desires-Intentions) architecture. A minor upgrade of the traditional BDI agent architecture given in [15]. Figure 1 summarizes the current state-of-the-art of agent (head-body) architectures (i.e., capabilities that can be implemented by individual agents or in a MAS). We argue that state-of-the-art agent technologies allow us to have trust in professional competence, and action fulfillment in the behavior of MAS as indicated in Table 2. Examples include knowledge-based systems with explanation capabilities. Furthermore, we can implement a MAS is such a way that the system will indeed have the desired functionality and reliability [12, 13]. However, it should be noted that this goal is not yet achieved concerning reliability but rather a stated goal of the road map of [12]. In that road map, reliability is especially addressing security concerns of MAS. As a matter of fact it limits trust concerns and hence building and maintaining trustworthy systems to issues related to reputation mechanisms, reliability testing, security and verifiability, and electronic contracts. Our framework thus includes and extends issues related to trustworthy systems as expressed in the road map of [12]. Precisely, for that reason we argue that indeed it is possible to have a grounded belief of trustworthiness by the user in agent based behavior such as e-services. That is, an example of ethical/moral behavior of agent systems in table 2.

On the other hand we argue that trust within agent systems in the sense of trust equal to the phenomena of human trust is beyond state-of-the-art in a foreseeable future mainly due to the fact that we do not have a corresponding complementary component (e.g., consciousness) complementing the architectural components Problem solving capability and Social competence of Figure 1. Regarding the other qualities within an agent society such as professional competence those qualities requires implemented mechanisms supporting self-adjustments, negotiations, learning, and semantic control. Those and similar mechanisms can be available within the next 5-10 years [12]. We will return to some of those topics in Section 7 on Trust in agents - other approaches.

3 Why Trust Is Harder Than Trustworthiness

We model trust in e-services as an individual assessment of trustworthiness of that service taking into account the given context. Our approach toward enabling trust by users and societies in e-services is consequently to focus on designing and building trustworthy systems based on a principled approach of handling trust concerns of system actors, e.g. users, and transforming those concerns into design principles and signs to be assessed by the users evaluating the trustworthiness. Our framework identifies a conceptual structure and some important processes toward a methodology to that end.

E-services is not a unambiguously defined concept [16] but a common definition is: ”Interactive software-based information systems received via the Internet” [17]. The information system typically involves many system components, e.g. software and artifacts. Typically e-services are composed of other services provided by third parties. For instance in order to distribute sensible health care data a suitable certificate may be used to create the necessary trust in the service.

When buying anti-virus software we are rather buying a service than a product. The software is bought with an initial subscription. When new viruses are discovered information about the viruses is added to a database that supports downloading upgrades to subscribers of the service. This service oriented approach also leads to a continuous relationship between the service provider and it’s users. In health care we are seeing similar tendencies where patients are treated over longer time spans compared to earlier than they just visited hospitals when they were ill and left upon recovery. Recovery and care will to a larger extent take place in the home of those needing care assisted by health care personnel.

The structure of the framework, i.e., the basic concepts and their relations are described in Section 4. In our model we take, as have earlier been said, into account relevant trust concerns, aspects, mechanisms and signs supporting user’s trust assessment. Further details on that strand are given in the next section. Needless to say, much research and experiments remains to be done to assess and refine our approach to meet the goals expressed above.

The inherent difficulty with qualities such as trust, and other related qualities such as security, privacy, and usefulness, is its *systemic* nature. That is, these qualities can only be assessed at the system level. From an engineering point of view these systemic qualities are sometimes called *non-functional* because:

- The quality cannot be decomposed into qualities of components.
- Two components can have the quality but not their composition. This aspect is particularly important in the area of composition of e-services. Reasons behind this non-compositional nature of systemic qualities include loss of quality due to uncontrolled (unforeseen) interactions between components or due to incomprehensible complexity of the conjunction of services perceived by the user.

An example of a systemic quality is traffic security. We have learned that by engineering vehicles taking into account traffic security *concerns* (expressing aspects that are manifested in *mechanisms* such as reliable brakes, air bags, belts, and crash zones) the risk assessment by users are simplified to assessing *signs* (mediated through brand names, accident statistics, or reputation) *associated* to the vehicle. The society has on its side developed an infrastructure supported by another set of traffic security concerns (road systems minimizing collisions, vehicle control authorities, education, monitoring authorities, legal frameworks) aiming at a higher traffic security in the society at hand. We all know that accidents still cannot be avoided but still we all trust the traffic system enough to use it on a daily base at our own decisions. Of course, the efforts of creating a trustworthy traffic system are ongoing processes with the explicit and *measurable systemic goal* to decrease the numbers and the severities of accidents. In effect, our societies have identified, since the last century, a set of traffic security concerns and aspects that have been translated into mechanisms implemented in different subsystems (components), i.e., more trustworthy vehicles, safer roads, and better monitoring measures. Our societies thus have furthermore developed a strategy and means toward attaining trustworthy traffic systems that each user can decide to trust (or not) at their will to use in an appropriate way. Of course, nobody believes that building and testing trustworthy components in itself will replace continuous traffic security assessments at the system level. The aim of our contribution is to propose and illustrate a similar comprehensive approach, as in the traffic example, toward supporting trust in e-services. In short, from an engineering perspective, we can only aim at designing, implementing, and maintaining trustworthy systems and components. Our success in gaining acceptance and trust by users of the systems will depend upon how well we have succeeded in translating trust concerns into aspects and mechanisms that can be implemented in a trustworthy manner by providing appropriate signs. At this point in time we, however, do not have an appropriate metric on the systemic level (compared to statistics and assessments of traffic accidents) to enable us to claim that we have a good strategy for supporting the users to gain more confidence in their trust assessment of electronically supported services by, e.g., providing more appropriate signs. Thus, making trustworthy information systems is hard but supporting users trust in them is at the moment very much harder. Our contribution is to outline a framework and processes to enable the first concern and to narrowing the latter divide.

Trust and trustworthiness are two notions we need to use wisely in order to emphasize the differences between the two. Trustworthiness is what designers of systems can implement [18] as mechanisms into the system manifested by an appropriate set of signs. An actor performing risk assessment related to trust then assesses if the system is trustworthy by inspecting the signs. The judgment whether the system is trusted or not is thus taken by the observer or user of the e-service. It is not possible to directly code trust into the system, see our discussion in Section 2. We use the term *actor* to denote stakeholders in the system. This is because of the fact that different actors may have different considerations related to trust affecting the design considerations needing attention [19]. Trust is obviously very context dependent [6]. We can, for instance, have trust in one actor providing an e-service and then not trust the very same actor in another service. We may also lose trust in some services if something disruptive has happened such as introduction of new technologies or unexpected breakdowns. Identification of and maintaining trust aspects that should be *sustainable* during change are in many applications crucial. There are several approaches aiming at modeling non-functional requirements such as trust by introduction of some measurable quality. A problem with this approach is that the qualities identified and measured have turned out to be quite arbitrary [20]. For instance, user satisfaction, being a systemic quality, has been approximated by a set of measurable qualities by different models. To infer user satisfaction relying naively on numeric calculations of numbers related to those measurable qualities could be misleading at least or totally wrong at worst. Furthermore, it has turned out to be difficult to compare different numerical models or to make predictions due to changes in the system. In the context of trust research there is no consensus about what to quantify, measure or investigate in order to reach a conclusion on whether a system is to be considered as trustworthy or not. This state of affairs imposes challenges on system designers in design, development, and use of tools enabling evaluation of computer systems trustworthiness. Our proposed framework and associated processes are steps in that direction. We compare our approach with contemporary approaches toward trust in electronically mediated services or agents in Sections 6 and 7.

4 A Framework Enabling Assessing Trustworthiness

The following Figure 2 captures the main ingredients of our framework supporting design, implementation and monitoring of trustworthy e-services. The components of the framework are:

- The *context*, including: actors, e-services, artifacts, location, and, time. The context also includes other components and factors such as contracts, ownerships, responsibilities, legal frameworks, work practices, and, organizational aspects.
- *Trust concerns* addressed: e.g., loss of life, threat of privacy, loss of money, loss of reputation, responsibilities, or time and duration of engagement.
- *Trust aspects* that can be derived from trust concerns: legal aspects, responsibilities at breakdowns, information integrity, security, privacy aspects, or explanations of functionality.
- *Mechanisms* that implement trust aspects, e.g., explanations, in a trustworthy way.
- *Signs* ensuring correct implementation of trust mechanisms that can be inspected by the observer of the e-service [21].

The relationships depicted by arrows in the figure are typically many - too - many. That is, a trust concern can be broken down into one or many aspects or vice versa. The same argument holds between aspects and mechanisms as well as between mechanisms and signs. An example from the earlier mentioned traffic domain is traffic signs (a mechanism and a sign) that implement trusted traffic information by alternating between

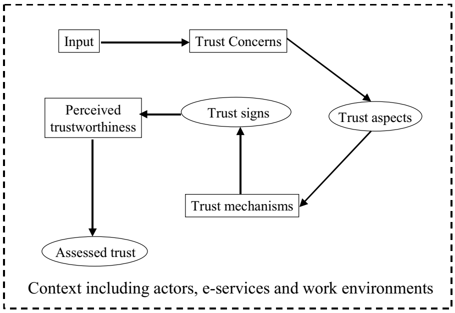
Input
Trust Concerns
Trust aspects
Trust signs
Perceived trustworthiness
Trust mechanisms
Assessed trust
Context including actors, e-services and work environments

Fig. 2. Main components of our framework supporting design and maintenance of trustworthy electronically mediated services

sending out red, yellow, or green light. The trustworthiness and trust rely on that all agents involved trust that (almost) all other agents knows the intended reading of the signals and follow suit. As another sign all legal drivers can, on request, provide a valid driver's license.

Another example is security that can be and often is a trust concern. Typically aspects related to security are information security, network security, computer security, or physical security. A typical mechanism related to the aspects of confidentiality, access control and information integrity is encryption. As yet, however, there is no well-accepted sign that encryption has been trustworthy implemented. Certificates - issued by 'trusted third parties' are in many senses to weak today to serve as a trust sign in our setting.

The degree of trust in services possessed by individuals is by no means static. As reported by [5] trust between individuals tends to grow stronger in close relationships, the familiarity factor. Familiarity with services is a strong support of trust. Trust does not only increase it can decline and hence has a dynamic nature [22]. The following factors exemplifies what contributes to the dynamics of trust:

- The actors' experiences through interactions with the e-service and involved actors, experience-based trust [23].
- Changes in society [8]. The climate and attitude toward providers or components of an e-service may change in society in general. If Internet Banks would be claimed to take to high fees in general the trust in Internet Banks in general may decrease.
- Changes in the composition of services, objects and artifacts. If the composition of an e-service changes, i.e. new technology or a new actor is introduced the trust concerns raised by actors may change.

We have a plentiful of potential changes among the concerns and aspects of trust. In a dynamic society these reasons for changes will prevail. The design of trustworthy e-services therefore is an effort that needs attention not just during the design phase, but also during the whole life cycle of the e-service. The dynamic nature of trust suggest that we continually must re-evaluate and eventually redesign mechanisms and signs of our framework to support efficient and reliable risk assessment concerning trust. Our framework supports this process and is part of our *trust management* process. The following semi-formal notations and definitions clarify the different dependencies of Figure 2 and provide a backbone for our methodology of trust management of e-services. The intended reading is that users express trust concerns in a given context. This input can be translated into a set of trust aspects. One category of trust concerns often mentioned individuals are related to misuse of owned or generated information that might lead to loss of life, loss of freedom, loss of money, loss of reputation, or receiving unwanted commercial offers. The generic term of those concerns is privacy. However, the given context will qualify the aspects (types) of privacy that are relevant for the expressed trust concerns indicated above. The identified aspects can then be translated into relevant mechanisms, e.g., secure end-to-end information exchange between mutually identified and trusted end users, and validated by appropriated signs. The components of an e-service include providers - the provider of the service to the user, third party actors - enabling the creation and distribution of the service, content of the service - the information and products distributed, computer-based artifacts used to provide the service to the user, access points to the service, and implemented trust mechanisms that are coupled to the relevant trust aspects formally defined as:

e-service = < Provider, Third_Party, Content, Computer_based_artifacts,
Access_points, Trust_mechanisms >

Definitions of the concepts Situation, Trustworthiness, and, Trust:

Situation = < Time-interval, Location >
Trustworthiness = < e-service, Situation, Context >
Trust = < User, Trustworthiness, Signs >

A situation is a binary relation between a time-interval and a location (where the service is delivered and used). Trustworthiness of an e-service connects the service to a situation and a context. The context is specified in the design phase of a particular e-service, c.f., our case scenario of Figure 3. Finally, the perceived trust by the user is a three-valued relation connecting the user, and signs that manifests the trustworthiness of the e-service. The value of trust can be of any type that supports reasoning and modeling in the framework. Examples include Boolean values (Yes, No), numerical values modeling strength of Belief in the trust, c.f., [24], or measuring fuzziness. In more elaborated modeling where partial ordering might be useful we can use lattices as the value domain of Trust. Given those definitions in a formal language we can define and reason about properties and invariance of properties of and between components of our framework in Figure 2 by introducing a suitable logical framework and notations. Given that logical framework we can for instance state precisely what we mean by ”Trustworthiness of an e-service independent of a set of situations”, ”Trustworthiness of an e-service independent of a set of mechanisms” or other invariants by introducing restrictions of formulas over sets.

We will return to the methodological processes related to the structure of framework in Figure 2 later. That is support for design and trust management. Design and trust management is modeled after Boehm’s risk driven spiral model[25]. Eventually we hope to supplement the framework with guidelines on how to design and implement mechanisms and signs supporting trustworthiness.

5 Trust in e-Services in Home Health Care

Distributed health care (e-health) utilizing Information and Communication Technology (ICT) is a vibrant area of research and development worldwide. First and foremost there is an international societal-economical need to assess current models of health care. Not the least in health care for people with special needs.

The underlying idea behind e-health in homes is that given the proper support a patient (e.g., elderly person) can stay longer in his/her home and thus have a higher quality of life than otherwise. At the same time the society gains is expected to be lower total costs and fewer burdens on hospitals and other health institutions. E-health systems are typically very complex socio-techno systems and a shift toward future e-health systems requires and understanding of the socio-economical aspects as well as of systemic possibilities and considerations. Systemic *invariants* such as, e.g., ”good care” and trust, have to be sustained during introduction of ICT in order for e-health to get acceptance by involved parties We have developed our framework to support design, implementation and maintenance of this change of institutional centric health care into a distributed patient-centric health care while preserving trust in the necessary services by all agents involved, not least by the patients.

The following figure, Figure 3, captures our ’patient-centric view’ of e-health. We have investigated this scenario in several national and international projects [1] in distributed health care. A result of those investigations, based on lessons learned and insights, is the framework presented in this paper.

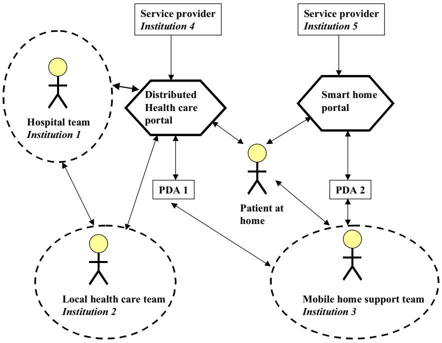

Fig. 3. Teams and institutions involved in distributed home health care

The scenario above involves five *institutions*, three concerning health care and home support services and two service providers. Furthermore we have three types of teams, hospital *teams*, local health care teams, and mobile home support teams. Two portals - e-service systems, including stationary and mobile access points) support the activities of the scenario. The Health care portal provides sensors supporting monitoring of the

[1] EC Alfebiite - http://alfebiite.ee.ic.aac.uk

health of the patient. The related information is transformed into suitable formats for assessments of the teams and the patient in a role-based manner. The smart home portal has sensors and actuators supporting the patient in his daily life at home.

A successful transition from today's health care organizations and practices to the situation depicted in Figure 3 would typically mean that large parts of the treatments of the patient now conducted in hospitals or similar institutions will take place in the homes of the patients. This also implies that the work situation for Institution 3 (Home support team) will be more qualified although their basic education will mainly be as now. The home support team thus needs good support from the artifacts delivered by Institutions 4 and 5 and a mutually trusted case-based delegation of tasks from the hospitals in order to do their job in a satisfactory way. To summarize the new situation: The health care authorities that are responsible for the health care have to have a grounded trust in that the new organization will deliver high-quality health care services in a cost-efficient way. The involved organizations must also have similar beliefs. The persons involved, not the least the patient, must trust that the new work situation will provide sufficient support for the new work flows. Last but not least the overall systemic goal of ”good care” has to be maintained during the ICT enabled transformation. We have investigated several different partial scenarios related to the scenario given above. One set of investigations was related to equipment for measuring the health status of patients (related to the Health Care portal of Figure 3). Another set of investigations is related to improving the learning and knowledge sharing in teams utilizing Peer-to-Peer technologies [2] (Institution 2 in Figure 3). A third set of investigations was focusing on issues of shared awareness and work flow management where we have actors from more than one institution, institutions 1 and 2 of the scenario. Two applications [3] in this setting are SHINE - Sustaining health and interaction in networked environments - and DICE - Delegation and interaction in care environments. In the DICE application we had doctors and nurses from either of Institutions 1 or 2. Furthermore, we have nursing assistants belonging to either of institutions 1, 2, or 3.

The workflows will typically be supported by digital information management systems with different types of access possibilities. That is, the primary asset is information and a primary concern is trustworthy management of the information. One important aspect of trustworthiness is thus related to dependability (e.g., security, integrity of persons and data, and accessibility). Field personnel using new digital artifacts (DICE) have frequently raised the following trust concerns during our evaluation tests:

- How do we know that we do the right kind of tasks or actions in the right way?
- What happens if something goes wrong?
- Can our employer spy on us or misuse the information the system provides about our work?
- We have a very dynamic environment. Can we have a flexible system taking care of our mentioned concerns?

The system requirements can be formulated as: Trusted role and context based access control to services in e-health. Intuitively we can presuppose that enforcing normative

[2] E.g. WoundDoc - an information sharing tool for health care personnel.

[3] For more information visit http://www.soclab.bth.se

behavior could be a way to support trustworthiness but here we have to strike a balance between being too restrictive and in that way hampering a needed flexibility in the work flow processes (e.g., as in the DICE system). Our approach to achieve flexibility is to identify and implement context dependent normative behavior.

Background material on different aspects of trust was developed in the EC project Alfebiite. In effect: three supporting frameworks of trust: A logical Framework for Norm-Governed Behavior, A Conceptual Framework on Operational Model of Normative Behavior, and, Communicative Acts and Interactions Patterns developed in the project have largely influenced our approach. The following different concepts of trust have been proposed in those deliverables:

- A mere mental attitude (prediction and evaluation toward an other agent;
- A decision to rely upon the other, i.e., an intention to delegate and trust, which makes the truster 'vulnerable';
- A behavior, i.e., the intentional act of trusting, and the consequent relation between the truster and the trustee.

In our case we focus on the latter two concepts, since they open up for a methodological approach toward creating trust. In the conceptual framework we have models connecting trust and delegation (weak or strong). The models presuppose human agents but some models could also be used in the situation of trust in artifacts (which is of our main concern in our investigations). For instance, we model the trust in artifacts as strong delegation. In the same deliverable we also find the notions of internal and external trust useful for our investigations. The concept of a three party relationship based trust model is also very appropriate in our approach.

Another interesting concept for us is Adjustable Social Autonomy [26] modeling time dependent levels of delegation. Especially, we share the beliefs that "A very good solution (of adjustable social autonomy) is maintaining a high degree of interactivity during the collaboration, providing both the man/delegator/client and the machine/delegee/contractor the possibility of having initiative in interaction and help (mixed initiative) and of adjusting the kind/level of delegation and help, and the degree of autonomy run time. This means that channels and protocols - on the delegator's side - for monitoring (reporting, observing, and inspecting), and for discretion and practical innovation: for both client and contractor channels and protocols are needed for communication and re-negotiation during the role-playing and the task execution". As a matter of fact, our implementation of the DICE system is designed to meet such requirements concerning run-time observations and adjustments of systems.

6 Models of Trust – Other Approaches

One driver behind the interest in trust and e-services is that higher trust in e-service providers are likely to affect the willingness to engage in relationships and utilize the provided services. As a fact, trust has been defined as a willingness to depend or rely on other actors [27]. From the truster's perspective trust is a mechanism used to reduce complexity [9, 8] under situations of risk where we can choose our path of action based on expectations. One model proposed to deal with trust in risky environment such as e-commerce is the model of trust in electronic commerce (MoTech). It aims to explain the factors that affect a person's judgment of an e-commerce site's trustworthiness [23]. MoTech contains of a number of dimensions intended to reflect the stages visitors goes through when exploring an e-commerce website. The dimensions pre-interactional filter, interface properties, informational content and relationship management will be described below. Each of these components addresses factors that have been observed to affect consumers' judgment of an on-line vendor's trustworthiness.

Pre-interactional filters refer to factors that can affect people's perceptions before an e-commerce system has been accessed for the first time. The factors presented are related to user psychology or pre-purchase knowledge. The first group refers to factors such as propensity to trust and trust toward IT in general and the Internet. Pre-purchase knowledge is related to Reputation of the industry, company and Transference (off-line and on-line). The second dimension of MoTech is concerned with interface properties that affect the perception of a website. Here the components are branding and usability. Factors in the branding component are appeal and professionalism. The usability component factors are organization of content, navigation, relevance and reliability. The next dimension, informational content contains components related to competence of the company and the products and services offered and issues regarding security and privacy. The fourth and last dimension reflects the facilitating effect of relevant and personalized vendor-buyer relationship. The components Pre-purchase Interactions and Post-purchase interactions are related to factors such as responsiveness, quality of help and fulfillment.

The model structures e-commerce designers work and give directions toward important trust considerations during the discussed dimensions. In the light of our framework we would interpret the four dimensions or stages as four situations. For instance the pre-interactional stage is the situation before any interaction has taken place with the e-commerce web site. The factors are related to concerns, aspects, and mechanisms in our framework. The MoTech components privacy and security are trust aspects and the factors proposed are mechanisms such as policy, encryption and contractual terms in our framework. To summarize: we can model the MoTech approach in our framework whereby we also get a more principled approach for evaluation and maintenance. MoTech is developed for e-commerce applications but has also been tested in other contexts such as on-line gambling.

7 Trust in Agents – Other Approaches

Current state of the art tries to capture and reason about norms in agent societies. These so called normative agents trends are the one lying closest to human behavior as of today. Instead of acting based on reactive stimuli or a message related to problem solving a norm based agent can act based on social norms in order to achieve some kind of goal in isolation or in a team. However the state of the art within the area of MAS-architectures and agent models today merely reaches a desired level of a mixture of normative behavior and reflective behavior in key applications, Section 2 and Section 5. Another approach is to view human and computational agents differently. This is especially obvious when relying on some of the more common definitions of the term agent e.g. [28] who defines an agent to be *"a computer system that is situated in some environment that is capable of autonomous action in its environment in order to meet its design objectives"*. This definition excludes human beings from the agent metaphor at least in a computational setting. This explicit differentiation between human and computational agents opens for our approach of trust in agents, i.e., trust in human-agent interaction. Where the human is the subject and the agent(system) is the object, see Section 2. Models of trust in agent behavior have been an active research area for more than a decade. Different aspects of trust models have been proposed and sometimes implemented. Concerns of trust in agent behavior goes back to mid seventies where the corresponding systems under investigation was expert systems or in a later terminology knowledge based systems. The tasks performed by the systems were knowledge intensive problem solving in areas such as diagnosis, planning, scheduling, and monitoring. The problem solving capability was captured and engineered to mimic human expertise in selected areas. By necessity the knowledge systems had to handle inherent weaknesses such as brittleness and sometimes assessment conflicts between experts. In short, there were concerns by the users how to trust decisions suggested by the systems. The following trust aspect was identified to remedy these concerns. The users requested explanations of the support for the conclusions drawn by the system. Two explanation facilities, or mechanisms, were identified, i.e., answers to the questions "Why?" and "What if?" c.f., our framework in Figure 2. Different strategies of reasoning and implementations of those mechanisms have been evaluated since that time. A good exposition of trust concerns related to explanations and corresponding mechanisms are described in [29]. Below we assess contemporary efforts in designing and building trustworthy agent systems and e-services. In our discussion we frequently refer to concepts from our framework, Figure 2. Furthermore we base our assessments on our discussion in Section 2, Figure 1, and, Table 2.

MAS (Multi Agent Systems) designers and programmers investigate trust due to it's importance in human interpersonal relationships where trust seems to affect how we make decisions about what to delegate and to whom or whey we choose to act in a way rather than another. For instance why do we trust A to do a task for us instead of B? Here reputation has been identified as a major factor to be aware of. Thus by implementing mechanisms into MAS the intention is to create agent-to-agent trust in trade and interaction between agents in the systems ultimately enabling them to act independently of the agent owner and make deals and commit to tasks on behalf of its owner. These kind of trust supports are mechanism-oriented and it is often hard to assess in what ways those mechanisms are related to trust concerns as expressed in our framework, Figure 2.

The Foundation for Intelligent Physical Agents (FIPA) has proposed MAS Security Models. A good overview of relevant material "Specifying Standard Security Mechanisms in Multi-agent Systems" is provided in [30]. From the point of our framework the focus is on mechanisms. The FIPA requirements are collected from a set of scenarios, related to e-commerce, from which a set of security issues is derived. The corresponding architectural elements, or trust aspect in our terminology, are found to be authentication, authorization, integrity, and privacy. Some generic safeguards are the proposed. There is no attempt by FIPA to address the concerns that might lead to the mentioned set of trust aspects. Neither mechanisms nor signs are explicitly addressed. In practice, it might be difficult to assess how well the FIPA efforts supports trust in agent-mediated services in the selected domain e-commerce. Security aspects of trust concerns of agent behavior are addressed by several researchers beside the FIPA efforts. The application area is typically e-commerce [31]. Again most efforts is devoted to discuss similar trust aspects as in the FIPA case but sometimes introducing other mechanisms and sometimes signs (certificates).

A research agenda addressing challenges for trust, fraud, and deception research in multi-agent systems has recently been proposed [32]. The areas identified are: Trust model discrimination, Building reputation without interaction, benchmarking trust modeling algorithms. Trust as measuring a reputation based quantity is also the basic mechanism in studies of objective-trust based agents [33]. The underlying assumptions with these approaches are that, in our terminology, the chain of trust concerns - trust aspects - mechanisms can be compiled into a metric (sign) calculated by an algorithm. These approaches are of cause possible to model in our framework and of relevance in specific circumstances. On the other hand, addressing trust in life threatening situations such as distributed e-health require a more elaborate approach.

Recent advancements in semantic web technologies as well as in web services and semantic Grid computing make introduces the concept of ”smart” or ”intelligent” services. In our view those kinds of services can and perhaps should be modeled as agent mediated services. This approach allows a fruitful interaction between the high-level agent approach and the bottom-up approach provided by the web service and Grid computing communities. Both communities have their preferred approach toward trustworthy systems with specific advantages and disadvantages. Our framework is aiming at a common ground for designing and maintaining trustworthy intelligent e-services.

8 Conclusions and Further Research

Creating and maintaining information systems that users can decide to trust is a hard challenge. In effect we ask the user to trust, economically and in some cases even with their life, the behavior of electronically mediated services, e-services. To that end we propose a framework and a methodological approach aiming at designing, developing and maintaining trustworthy systems. The framework is based on the idea that trust is a subjective assessment that is highly context dependent. To capture the anatomy of those assessments we introduce the following concepts in our methodology; trust concerns, trust aspects, trust mechanisms, and, signs. Typically, users articulate trust concerns and they look for signs that will assist them in their assessments. Trust aspects are design tools allowing designers to decide proper mechanisms to be implemented and to provide signs that verify that those mechanisms have been properly implemented. Trust concerns thus give insight into hypothetical or validated concerns related to trust among providers, third party actors and users of e-services, hence trust aspects are operationalizations of the different concerns. Trust mechanisms are implemented trust aspects, e.g., explanations, encryption algorithms etc. The signs, trade marks, documentation, certificates, and so on, provide actors and end users with credentials belonging to an e-service enabling the actors to form their judgment of whether or not trust the service.

We have also included a case study to illustrate and validate our framework related to trustworthy e-services. Our applications are primary distributed e-services supporting a patient and associated health care and home care teams from a home-centric point of view. We have chosen this application area for two reasons. Firstly, the application area is of high societal importance worldwide; secondly, the application amply illustrates the different aspects and challenges of trust in artifact-mediated services. In fact your life might depend on some of those services. We have developed our models in different application projects and based our approach on contemporary R&D on trust and trustworthiness.

In our definition of e-services we have taken into account different aspects of their context, i.e., other actors than the user, other e-services, artifacts and contextual qualities such as contracts, ownerships, responsibilities, legal frameworks, work practices, organizational aspects, and, time. Furthermore, we make some comparison of our approach with contemporary approaches toward trust in system behavior from the e-commerce area and the Multi Agent System domain. The approach and models are to a high degree work in progress and will be refined in other upcoming projects where we have to trust artifact-mediated services where life might be at stake.

References

1. Gefen, D., Straub, D.W.: Managing user trust in b2c e-services. e-Service Journal **2** (2003) 7–24
2. Gambetta, D.: Can we trust trust? In Gambetta, D., ed.: Trust : Making and Breaking Cooperative Relations. B. Blackwell, New York (1988)
3. Muir, B.M.: Trust in automation .1. theoretical issues in the study of trust and human intervention in automated systems. Ergonomics **37** (1994) 1905–1922
4. Deutsch, M.: The resolution of conflict; constructive and destructive processes. Yale University Press, New Haven, (1973)
5. Rempel, J., Holmes, J., Zanna, M.: Trust in close relationships. Journal of Personality and Social Psychology **49** (1985) 95–112
6. Barber, B.: The logic and limits of trust. Rutgers University Press, New Brunswick, N.J. (1983)
7. Baier, A.: Trust and antitrust. Ethics **96** (1986) 231–260
8. Giddens, A.: The consequences of modernity. Polity Press ;, Cambridge (1990)
9. Luhmann, N.: Familiarity, confidence, trust: Problems and alternatives. In Gambetta, D., ed.: Trust : Making and Breaking Cooperative Relations. Basil Blackwell, New York, NY (1988) 94–110
10. Friedman, B., Kahn, P.H., Howe, D.C.: Trust online. Communications of the Acm **43** (2000) 34–40
11. Luck, M.: Challenges for agent-based computing. Special Issue of Autonomous Agents and Multi-agent Systems **9** (2004) 203–252
12. Luck, M., McBurney, P., Priest, C.: A manifesto for agent technology: Towards next generation computing. Special Issue of Autonomous Agents and Multi-agent Systems **9** (2004) 253–283
13. Zambonelli, F., Omicini, A.: Challenges and research directions in agent-oriented software engineering. Special Issue of Autonomous Agents and Multi-agent Systems **9** (2004)
14. Sierra, C.: Agent-mediated electronic commerce. Special Issue of Autonomous Agents and Multi-agent Systems **9** (2004) 285–301

15. Hgg, S., Ygge, F.: Agent-oriented programming in power distribution automation : an architecture, a language, and their applicability. Dept. of Computer Science Lund University, Lund (1995)
16. Stafford, T.F.: E-services. Communications of the Acm **46** (2003) 27–28
17. Featherman, M.S., Pavlou, P.A.: Predicting e-services adoption: a perceived risk facets perspective. International Journal of Human-Computer Studies **59** (2003) 451–474
18. Sisson, D.: e-commerce: Trust & trustworthiness (2000)
19. Shankar, V., Urban, G.L., Sultan, F.: Online trust: a stakeholder perspective, concepts, implications, and future directions. Journal of Strategic Information Systems **11** (2002) 325–344
20. Kotonya, G., Sommerville, I.: Requirements engineering : processes and techniques. Worldwide series in computer science. J. Wiley, Chichester ; New York (1998)
21. Bacharach, M., Gambetta, D.: Trust as type detection. In Castelfranchi, C., Tan, Y.H., eds.: Trust and deception in virtual societies. Kluwer Academic Publishers, North Holland (2001)
22. mcKnight, H.D., Cummings, L.L., Chervany, N.L.: Initial trust formation in new organizational relationships. Academy of Management Review **23** (1998) 473–490
23. Egger, F.N.: From Interactions to Transactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce. PhD thesis, Technische Universiteit Eindhoven (2003)
24. Singh, M.P.: Trustworthy service composition: Challenges and research questions. Trust, Reputation, and Security: Theories and Practice **2631** (2003) 39–52
25. Boehm, B.W.: Software risk management. IEEE Computer Society Press, Washington, D.C. (1989)
26. Falcone, R., Castelfranchi, C.: The human in the loop of a delegated agent: The theory of adjustable social autonomy. Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans **31** (2001) 406–418
27. Dobing, B.R.: Building trust in user-analyst relationships. Ph. d., University of Minnesota (1993)
28. Wollridge, M.: An introduction to Multi-Agent Systems. Wiley, Chichester, England (2002)
29. Shapiro, S.C.: Encyclopedia of artificial intelligence. 2nd edn. Wiley, New York (1992)
30. Poslad, S., Charlton, P., Calisti, M.: Specifying standard security mechanisms in multi-agent systems. Trust, Reputation, and Security: Theories and Practice **2631** (2003) 163–176
31. Tan, J.J., Titkov, L., Poslad, S.: Securing agent-based e-banking services. Trust, Reputation, and Security: Theories and Practice **2631** (2003) 148–162
32. Barber, K.S., Fullam, K., Kim, J.: Challenges for trust, fraud and deception research in multi-agent systems. Trust, Reputation, and Security: Theories and Practice **2631** (2003) 8–14
33. Witkowski, M., Aritikis, A., Pitt, J.: Trust and cooperation in a trading society of objective-trust based agents. In Falcone, R., Singh, M., Tan, Y.H., eds.: Worskhop on Deception, Fraud, and Trust in Agent Societies, Barcelona, National Reserach Council, Institute of Psychology, Rome Italy (2000) 127 – 136