A Classification of Trust Systems

Sebastian Ries*, Jussi Kangasharju, and Max Mühlhäuser

Department of Computer Science
Darmstadt University of Technology
Hochschulstrasse 10
64289 Darmstadt, Germany
{ries, jussi, max}@tk.informatik.tu-darmstadt.de

Abstract. Trust is a promising research topic for social networks, since it is a basic component of our real-world social life. Yet, the transfer of the multi-facetted concept of trust to virtual social networks is an open challenge. In this paper we provide a survey and classification of established and upcoming trust systems, focusing on trust models. We introduce a set of criteria as basis of our analysis and show strengths and short-comings of the different approaches.

1 Introduction

Trust is a well-known concept in everyday life, which simplifies many complex processes. Some processes are just enabled by trust, since they would not be operable otherwise. On the one hand, trust in our social environment allows us to delegate tasks and decisions to an appropriate person. On the other hand, trust facilitates efficient rating of information presented by a trusted party. Computer scientists from many areas, e.g., security, ubiquitous computing, semantic web, and electronic commerce, are still working on the transfer of this concept, to their domain. In Sect. 2 we will introduce, the main properties of social trust, in Sect. 3 we provide our own set of criteria and the analysis of a selected set of trust systems from different areas, and in Sect. 4 we give a short summary and derive ideas for our future work.

2 Properties of Trust

There is much work on trust by sociologists, social psychologists, economists, and since a few years also by computer scientists. In general trust can be said to be based on personal experience with the interaction partner in the context of concern, on his reputation, or on recommendations. Furthermore, trust is connected to the presence of a notion of uncertainty, and trust depends on the expected risk associated with an interaction. [1, 2, 3, 4, 5, 16]

* The author's work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the PhD program "Enabling Technologies for Electronic Commerce" at Darmstadt University of Technology.

R. Meersman, Z. Tari, P. Herrero et al. (Eds.): OTM Workshops 2006, LNCS 4277, pp. 894–903, 2006.

The following properties are regularly assigned to trust, and are relevant when transferring the concept to computer sciences. Trust is subjective and therefore asymmetric. It is context dependent, and it is dynamic, meaning it can increase with positive experience and decrease with negative experience or over time without any experience. This makes also clear that trust is non-monotonic and that there are several levels of trust including distrust. A sensitive aspect is the transitivity of trust. Assuming Alice trusts Bob and Bob trusts Charlie, what can be said about Alice trust in Charlie? In [2], Marsh points out that trust is not transitive. At least it is not transitive over arbitrary long chains, since this will end in conflicts regarding distrust. Yet recommendation and reputation are important factors for trust establishment.

McKnight and Chervany state in [1] that there are three principle categories of trust: personal / interpersonal trust, impersonal / structural trust, and dispositional trust. Interpersonal trust describes trust between people or groups. It is closely related to the experiences, which people had with each other. Structural trust is not bound to a person but raises from social or organizational situation. Dispositional trust can be explained as a person's general attitude towards the world. As shown in [6] much work is done on transferring interpersonal trust to computer sciences, whereas there is little work supporting the other categories.

Although, trust is a well-known concept and despite there is a set of properties on which most researchers agree, it is hard to define trust. A couple of definitions are provided from several scientific areas with different focuses and goals (cf. [2, 6]). A definition which is shared or at least adopted by some researchers [3, 7, 8, 9], is the definition provided by the sociologist Diego Gambetta:

> "... trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [we] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [our] own action." [16]

3 Classification Criteria and Analysis

Having introduced the general aspects of trust, we will now give a survey how the concept of trust is realized in different areas of computer science. We derive our coarse-gained classification from the work provided in [2,4,5,10,11]. As main categories we see *trust modeling*, *trust management* and *decision making* [12]. In this classification, trust modeling deals with the representational and computational aspects of trust values. Trust management focuses on the collection of evidence and risk evaluation. Although decision making is actually a part of trust management, we treat it separately, since it is such an important aspect.

Due to the limitations of this paper and our own research interests we focus for a more fine-grained classification only on trust modeling, especially on the aspects of *domain*, *dimension*, and *semantics* of trust values.

Trust values are usually expressed as numbers or labels, thus their domain can be binary, discrete, or continuous. A binary representation of trust allows

only to express the two states of "trusted" and "untrusted". This is actually near to certificate- or credential-based access control approaches, where access is granted, if and only if the user presents the necessary credentials. But since most researchers agree that trust has several levels, binary models are considered as not sufficient. Trust can also be represented using more than two discrete values. This can be done either by using labels or by using a set of natural numbers. The advantage of this approach is, that trust values can be easily assigned and understood by human users [3, 13]. Continuous trust values are supported by well-known mathematical theories depending on the semantics of the trust values.

The dimension of trust values can be either one- or multi-dimensional. In one-dimensional approaches this value usually describes the degree of trust an agent assigns to another one, possibly bound to a specific context. Multi-dimensional approaches allow to introduce a notion of uncertainty of the trust value.

The semantics of trust values can be in the following set: rating, ranking, probability, belief, fuzzy value. As rating we interpret values which are directly linked with a trust related semantics, e.g., on a scale of natural numbers in the interval $[1, 4]$, 1 can be linked to "very untrusted",..., and 4 to "very trusted". Whereas, the trust values which are computed in ranking based models, e.g., [14], are not directly associated with a meaningful semantics, but only in a relative way, i.e. a higher value means higher trustworthiness. Therefore, it is only possible to assign an absolute meaning to a value, if this value can be compared to large enough set of trust values of other users. Furthermore, trust can be modeled as probability. In this case, the trust value expresses the probability that an agent will behave expected. The details of belief and fuzzy semantics are explained together with 'Subjective Logic' and ReGreT (see below). A summary of our classification is presented in Table 1.

Table 1. Classification of trust models

	Domain	Dimension	Sem.	Trust management	Decision making
Marsh	cont. in [-1,1)	1 (situational trust)	rating	– (but risk evaluation)	threshold-based
TidalTrust	disc. in $[1, 10]$	1 (rating)	rating	global policy (no risk evaluation)	–
Abdul-Rahman & Hailes	disc. labels	1 (trust value)	rating	–	–
SECURE Project (exemplary)	disc. in $[0, \infty]$	2 (evid.-based)	prob.	local policies (incl. risk evaluation)	threshold-based
	cont. in $[0, 1]$	3 (bel., disbel., uncert.)	belief		
Subjective Logic	disc. in $[0, \infty]$	2 (evid.-based)	prob.	not directly part of SL	not directly part of SL
	cont. in $[0, 1]$	3 (b, d, u)	belief		
ReGreT	disc. fuzzy values	2 (trust, confidence)	fuzzy values	local policies (fuzzy rules)	–

3.1 Model Proposed by Marsh

The work of Marsh [2] is said to be the seminal work on trust in computer science. Marsh concentrates on modeling trust between only two agents. He introduces knowledge, utility, importance, risk, and perceived competence as important aspects related to trust. The trust model should be able to answer the questions: With whom should an agent cooperate, when, and to which extend? The trust model uses real numbers in $[-1; 1)$ as trust values. He defined three types of trust for his model. Dispositional trust $\mathcal{T}_x$ is trust of an agent x independent from the possible cooperation partner and the situation. The general trust $\mathcal{T}_x(y)$ describes the trust of x in y, but is not situation specific. At last, there is the situational trust $\mathcal{T}_x(y, a)$, which describes the trust of agent x in agent y in situation a. The situational trust is computed by the following linear equation:

$$\mathcal{T}_x(y, a) = \mathcal{U}_x(a) \times \mathcal{I}_x(a) \times \widehat{\mathcal{T}_x(y)} \ , \tag{1}$$

where $\mathcal{U}_x(a)$ represents the utility and $\mathcal{I}_x(a)$ the importance, which x assigns to the trust decision in situation a. Furthermore, $\widehat{\mathcal{T}_x(y)}$ represents the estimated general trust of x in y.

The trust management provided by Marsh does not treat the collection of recommendations provided by other agents, he only models direct trust between two agents. The aspect of risk is dealt with explicitly based on costs and benefits of the considered engagement.

The decision making is threshold based. Among other parameters the cooperation threshold depends on the perceived risk and competence of the possible interaction partner. If the situational trust is above the value calculated for the cooperation threshold, cooperation will take place otherwise not. Furthermore, the decision making can be extended by the concept of "reciprocity", i.e. if one does another one a favor, it is expected to compensate at some time.

3.2 TidalTrust

In [13] Golbeck provides a trust model which is based on 10 discrete trust values in the interval $[1, 10]$. Golbeck claims that humans are better in rating on a discrete scale than on a continuous one, e.g., in the real numbers of $[0, 1]$. The 10 discrete trust values should be enough to approximate continuous trust values. The trust model is evaluated in a social network called FilmTrust [15] with about 400 users. In this network the users have to rate movies. Furthermore, one can rate friends in the sense of "[...] if the person were to have rented a movie to watch, how likely it is that you would want to see that film" [13].

Recursive trust or rating propagation allows to infer the rating of movies by the ratings provided by friends. For a source s in a set of nodes S the rating r_{sm} inferred by s for the movie m is defined as

$$r_{sm} = \frac{\sum_{i \in S} t_{si} \cdot r_{im}}{\sum_{i \in S} t_{si}} \ , \tag{2}$$

where intermediate nodes are described by i, t_{si} describes the trust of s in i, and r_{im} is the rating of movie m assigned by i. To prevent arbitrary long recommendation chains, the maximal chain length or recursion depth can be limited. Based on the assumption that the opinion of the most trusted friends are the most similar to opinion of the source, it is also possible to restrict the set of considered ratings, to those provided by the most trusted friends.

Although the recommendation propagation is simple, the evaluation in [13] shows that it produces a relatively high accuracy, i.e. the ratings based on recommendation are close to the real ratings of the user. Since this approach does not deal with uncertainty, the calculated trust values can not benefit in case that their are multiple paths with the similar ratings. The trust value is calculated as a weighted sum. For the same reason, the path length does not influence the trust value. The values for trust in other agents on the path are used for multiplication and division in each step. Since each node aggregates its collected ratings and passes only a single value to its ancestor in the recursion, the source cannot evaluate which nodes provided their rating. The approach does not deal with any form of risk or decision making.

3.3 Model Proposed by Abdul-Rahman and Hailes

The trust model presented by Abdul-Rahman and Hailes [7] is developed for use in virtual communities with respect to electronic commerce and artificial autonomous agents. It deals with a human notion of trust as it is common in real world societies. The formal definition of trust is based on Gambetta [16].

The model deals with direct trust and recommender trust. Direct trust is the trust of an agent in another one based on direct experience, whereas recommender trust is the trust of an agent in the ability of another agent to provide good recommendations. The representation of the trust values is done by discrete labeled trust levels, namely "Very Trustworthy", "Trustworthy", "Untrustworthy" and, "Very Untrustworthy" for direct trust, and "Very good", "good", "bad" and, "very bad" for recommender trust.

A main aspect of this trust model is to overcome the problem that different agents may use the same label with a different subjective semantics. For example, if agent a labels an agent c to be "Trustworthy" based on personal experience, and a knows that agent b labels the same agent c to be "Very Trustworthy". The difference between these two labels can be computed as "semantic distance". This "semantic distance" can be used to adjust further recommendations of b.

Furthermore, the model deals with uncertainty. Uncertainty is introduced if an agent is not able to determine the direct trust in an agent uniquely, i.e. if an agent has e.g., as much "good" as "very good" experiences with another agent. But it seems unclear how to take benefit from this introduction of uncertainty in the further trust computation process. The combination of recommendations is done as weighted summation. The weights depend on the recommender trust and are assigned in an ad-hoc manner.

Although the model drops recommendations of unknown agents for the calculation of the recommended trust value, those agents get known by providing

recommendations, and their future recommendations will be used as part of the calculation.

It is important to mention that the direct trust values are only used to calculate the semantic distance to other agents, but are not used as evidence which could be combined with the recommendations.

Trust management aspects are not considered. The collection of evidence is only stated for recommendations of agents which have direct experience with the target agent. It is not explicitly described how to introduce recommendations of recommendations. Furthermore, the system does not deal with risk. Decision making seems to be threshold based, but is not explicitly treated.

3.4 SECURE Project Trust Model

The trust model and trust management in the SECURE project [5, 17] aims to transfer a human notion of trust to ubiquitous computing.

A main aspect of the trust model is to distinguish between situations in which a principal b is "unknown" to a principal a, and situations in which a principal b is "untrusted" or "distrusted". The principal b is unknown to a, if a cannot collect any information about b. Whereas b is "untrusted" if a has information, based on direct interaction or recommendations, stating that b is an "untrustworthy" principal.

This leads to define two orderings on a set of trust values $\mathcal{T}$ denoted as $\preceq$ and $\sqsubseteq$. The first ordering $(\mathcal{T}, \preceq)$ is a complete lattice. For $X, Y \in \mathcal{T}$ the relation $X \preceq Y$ can be interpreted as Y is more trustworthy than X. The second ordering $(\mathcal{T}, \sqsubseteq)$ is a complete partial order with a bottom element. The relation $X \sqsubseteq Y$ can be interpreted as the trust value Y is based on more information than X.

The set of trust values can be chosen from different domains as long as the orderings have the properties described above. It is possible to use intervals over the real numbers in $[0, 1]$ [17]. This allows for an interval $[d_0, d_1]$ to introduce the semantics of belief theory by defining d_0 as belief and $1 - d_1$ as disbelief. Uncertainty can be defined as $d_1 - d_0$. Another possibility would be to define the trust values as pair of non-negative integers (m, n). In this case m represents the number of non-negative outcomes of an interaction and n the number of negative ones. These approaches seem to be similar to the trust model provided by Jøsang, but they do not provide a mapping between these two representations. It is also possible to define other trust values e.g., discrete labels.

The trust propagation is based on policies. This allows users to explicitly express whose recommendations are considered in a trust decision. Let $\mathcal{P}$ be the set of principals, the policy of a principal $a \in \mathcal{P}$ is π_a. The local policy allows to assign trust values to other agents directly, to delegate the assignment to another agent, or a combination of both. Since it is possible to delegate the calculation of trust values, the policies can be mutually recursive. The collection of all local policies π can be seen as global trust function m. This function m can be calculated as the least fixpoint of Π, where Π is $\Pi : \lambda p : \mathcal{P}.\pi_p$.

The trust management also deals with the evaluation of risk. Risk is modeled based on general cost probability density functions, which can be parameterized

by the estimated trustworthiness of the possible interaction partner. The evaluation of risk can based on different risk policies, which e.g., describe if the risk is independent from the costs associated to an interaction or if it increases with increasing costs.

The decision making is threshold based. For the application in an electronic purse [5] two thresholds are defined by the parameters x, y ($x \leq y$). If the situation specific risk value (parameterized by the trust value corresponding to the interaction partner) is below x, the interaction will be performed (money will be payed), if it is above y the interaction will be declined. In case the risk value is between x and y the decision will be passed to the user.

3.5 Subjective Logic

The trust model presented by Jøsang [10], named "subjective logic", combines elements of Bayesian probability theory with belief theory. The Bayesian approach is based on beta probability density function (pdf), which allows to calculate posteriori probability estimates of binary events based on a priori collected evidence. For simplification we do not explain the concept of atomicity, which is introduced by Jøsang to use his model also for non-binary events.

The beta probability density function f of a probability variable p can be described using the two parameters α, β as:

$$f(p \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad \text{where } 0 \leq p \leq 1,\ \alpha > 0,\ \beta > 0\ . \tag{3}$$

By defining $\alpha = r+1$ and $\beta = s+1$, it is possible to relate the pdf directly to the priori collected evidence, where r and s represent the number of positive and negative evidence, respectively. In this model trust is represented by opinions which can be used to express the subjective probability that an agent will behave as expected in the next encounter. It is possible to express opinions about other agents and about the truth of arbitrary propositions. The advantage of this model is that opinions can be easily be derived from the collected evidence.

An approach to deal with uncertainty is called belief theory, which tempts to model a human notion of belief. In belief theory as introduced in [10] an opinion can be expressed as a triple (b, d, u), where b represents the belief, d the disbelief, and u the uncertainty about a certain statement. The three parameters are interrelated by the equation $b + d + u = 1$. Jøsang provides a mapping between the Bayesian approach and the belief approach by defining the following equations:

$$b = \frac{r}{r+s+2}, \quad d = \frac{s}{r+s+2}, \quad u = \frac{2}{r+s+2} \qquad \text{where } u \neq 0\ . \tag{4}$$

Furthermore, he defines operators for combining (consensus) and recommending (discounting) opinions. In contrast to the belief model presented in [18] the consensus operator is not based on Dempster's rule. Moreover, the model supports also operators for propositional conjunction, disjunction and negation.

In [19] it is shown how "subjective logic" can be used to model trust in the binding between keys and their owners in public key infrastructures. Other papers introduce how to use "subjective logic" for trust-based decision making in electronic commerce [20] and how the approach can be integrated in policy based trust management [21].

Another approach modeling trust based on Bayesian probability theory is presented by Mui et al. in [8], an approach based on belief theory is presented by Yu and Singh in [18].

3.6 ReGreT

ReGreT tries to model trust for small and mid-size environments in electronic commerce [22]. The system is described in detail in [23, 24]. A main aspect of ReGreT is to include information which is available from social relations between the interacting parties and their environments. In the considered environment the relation between agents can be described as competitive (*comp*), cooperative (*coop*), or trading (*trd*).

The model deals with three dimensions of trust or reputation. The individual dimension is based on self-made experiences of an agent. The trust values are called direct trust or outcome reputation. The social dimension is based on third party information (witness reputation), the social relationships between agents (neighborhood reputation), and the social role of the agents (system reputation). The ontological dimension helps to transfer trust information between related contexts. For all trust values a measurement of reliability is introduced, which depends on the number of past experience and expected experience (*intimate* level of interaction), and the variability of the ratings.

The trust model uses trust or reputation values in the range of real numbers in $[-1; 1]$. Overlapping subintervals are mapped by membership functions to fuzzy set values, like "very good", which implicitly introduce semantics to the trust values. In contrast to the probabilistic models and belief models, trust is formally not treated as subjective probability that an agent will behave as expected in the next encounter, but the interpretation of a fuzzy value like "very good" is up to the user or agent.

Since the fuzzy values are allowed to overlap, this introduces also a notion of uncertainty, because an agent can be e.g., "good" and "very good" at the same time to a certain degree.

The inference of trustworthiness is based on intuitively interpretable fuzzy rules. The trustworthiness assigned by agent a to agent c with respect to providing information about agent b, e.g., can depend on the relation between the agents b and c, as shown in the following example. In the example the social trust of a in information of b about c is "very bad" if the cooperation between b and c is high.

IF $coop(b; c)$ is $high$

THEN $socialTrust(a; b; c)$ is $very\ bad$.

Further information concerning risk evaluation and decision making is not given.

4 Conclusion

In this paper we have provided a short survey of trust systems based on different approaches. Furthermore, we provided a set of criteria to analyze systems dealing with trust, on a top level by distinguishing between trust model, trust management and decision making, and for the main aspects of trust modeling in detail. As we can see from our survey, it is possible to reason about trust models without especially addressing aspects of trust management, and the other way around. The comparison of trust models is yet difficult, since they are often developed for different purposes and use different semantics for modeling trust. Furthermore, most authors define their own way of trust management to evaluate their trust models. The trust propagation chosen by Golbeck seems to be a simple and yet an accurate way to evaluate recommendations in social networks.

By analyzing the trust models, we came to the conclusion that the models need to be able to represent a notion of uncertainty or confidence, since it is a main aspect of trust. The approach taken in ReGreT allows to define a subjective component for confidence, but the approach seems to be done in an ad hoc manner. The approach taken by belief models binds uncertainty to belief and disbelief. In conjunction with the Bayesian approach uncertainty depends directly of the number of collected evidence, but it is not related to a subjective and context-dependent measurement. For our future work we favor the Bayesian approach, since it allows to easily integrate the collected evidence. We will try to find a new way to derive uncertainty from the relation between the amount of collected evidence and an amount of expected evidence based on this approach. By giving the user the opportunity to define an expected amount of evidence, uncertainty gets a subjective and most notably a context-dependent notion.

References

1. McKnight, D.H., Chervany, N.L.: The meanings of trust. Technical report, Management Information Systems Research Center, University of Minnesota (1996)
2. Marsh, S.: Formalising Trust as a Computational Concept. PhD thesis, University of Stirling (1994)
3. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. In: Decision Support Systems. (2005)
4. Grandison, T., Sloman, M.: A survey of trust in internet applications. IEEE Communications Surveys and Tutorials **3**(4) (2000)
5. Cahill, V., et al.: Using trust for secure collaboration in uncertain environments. IEEE Pervasive Computing **2/3** (2003) 52–61
6. Abdul-Rahman, A.: A Framework for Decentralised Trust Reasoning. PhD thesis, University College London (2004)
7. Abdul-Rahman, A., Hailes, S.: Supporting trust in virtual communities. In: Proc. of Hawaii International Conference on System Sciences. (2000)
8. Mui, L., Mohtashemi, M., Halberstadt, A.: A computational model of trust and reputation for e-businesses. In: Proc. of the 35th Annual HICSS - Volume 7, Washington, DC, USA, IEEE Computer Society (2002)

9. Teacy, W.T., et al.: Travos: Trust and reputation in the context of inaccurate information sources. Autonomous Agents and Multi-Agent Systems **12**(2) (2006)
10. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **9**(3) (2001) 279–212
11. Grandison, T., Sloman, M.: Specifying and analysing trust for internet applications. In: I3E '02: Proc. of the IFIP Conference on Towards The Knowledge Society, Deventer, The Netherlands, Kluwer, B.V. (2002) 145–157
12. Ries, S.: Engineering Trust in Ubiquitous Computing. In: Proc. of Workshop on Software Engineering Challenges for Ubiquitous Computing, Lancaster, UK (2006)
13. Golbeck, J.: Computing and Applying Trust in Web-Based Social Networks. PhD thesis, University of Maryland, College Park (2005)
14. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: Proc. of the 12th international conference on World Wide Web, New York, USA, ACM Press (2003) 640–651
15. Golbeck, J., Hendler, J.: Filmtrust: Movie recommendations using trust in web-based social networks. In: Proc. of the Consumer Communications and Networking Conference. (2006)
16. Gambetta, D.: Can we trust trust? In Gambetta, D., ed.: Trust: Making and Breaking Cooperative Relations. Basil Blackwell, New York (1990) 213–237
17. Carbone, M., Nielsen, M., Sassone, V.: A formal model for trust in dynamic networks. In: Proc. of IEEE International Conference on Software Engineering and Formal Methods, Brisbane, Australia, IEEE Computer Society (2003)
18. Yu, B., Singh, M.P.: An evidential model of distributed reputation management. In: Proc. of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems, New York, NY, USA, ACM Press (2002) 294–301
19. Jøsang, A.: An algebra for assessing trust in certification chains. In: Proc. of the Network and Distributed System Security Symposium, San Diego, USA, (1999)
20. Jøsang, A.: Trust-based decision making for electronic transactions. In: Proc. of the 4th Nordic Workshop on Secure IT Systems, Stockholm, Sweden (1999)
21. Jøsang, A., Gollmann, D., Au, R.: A method for access authorisation through delegation networks. In: 4th Australasian Information Security Workshop (Network Security) (AISW 2006). Volume 54 of CRPIT., Hobart, Australia, ACS (2006)
22. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review **24**(1) (2005) 33–60
23. Sabater, J., Sierra, C.: Reputation and social network analysis in multi-agent systems. In: Proc. of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems, New York, NY, USA, ACM Press (2002) 475–482
24. Sabater, J.: Trust and reputation for agent societies. PhD thesis, Institut d'Investigacion en Intelligencia Artificial, Spain (2003)