

Trust in Ubiquitous Computing

Vom Fachbereich Informatik
der Technischen Universität Darmstadt
genehmigte

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

von

Dipl.-Inform. Sebastian Ries

geboren in Miltenberg



Referenten

Prof. Dr. Max Mühlhäuser (TU Darmstadt)
Prof. Dr. Audun Jøsang (University of Oslo)

Tag der Einreichung: 06.05.2009
Tag der mündlichen Prüfung: 02.07.2009

Darmstadt 2009
Hochschulkenziffer D17

Acknowledgements

This work would not have been possible without the continuous support and encouragement of my colleagues and friends over the last years, which I would like to acknowledge here.

First and foremost, I would like to thank my advisor, Max Mühlhäuser, for his faith in my work and for giving me excellent advice on many issues concerning this work and beyond. Next, I am grateful to Audun Jøsang, for the fruitful discussions concerning this thesis and for acting as a second referee.

I am grateful to all at Telecooperation and RBG for providing me with a friendly and supportive place to work. Special thanks go to Jussi Kangasharju for supervising my work in its beginnings, and to Andreas Heinemann for acting as a supervisor afterwards.

It was a pleasure for me to co-author scientific publications with Erwin Aitenbichler, Dirk Bradler, Stephan Borgert, Michael Hartle, Andreas Heinemann, Gina Häußge, Jussi Kangasharju, Max Mühlhäuser, Daniel Schreiber, Julian Schröder-Bernhardi, Georg Turban, and Stefan Georg Weber.

Many people improved this text with their reviews and comments. Thanks to Alexander Behring, Melanie Hartmann, Andreas Heinemann, Audun Jøsang, Max Mühlhäuser, Guido Rößling, Daniel Schreiber, Jürgen Steimle, and Stefan Georg Weber. Thanks to Lara Schwarz for proof-reading the final version.

Finally, thanks to my parents, brother, and friends, especially Steffi Freigang, for their mental support and patience during the course of this work.

Abstract

In the vision of ubiquitous computing, the activities of daily life are supported by a multitude of heterogeneous, loosely coupled computing devices. The support of seamless collaboration between users, as well as between their devices, can be seen as one of the key challenges for this vision to come true.

This thesis provides a trust based approach to supporting the selection of trustworthy interaction partners. The goal of this approach is to estimate an entity's trustworthiness as accurately as possible in order to improve the average quality of the entity's interactions.

In this thesis, the trustworthiness of an entity is derived from evidence gained during past interactions. To this end, current Bayesian trust models are extended and improved regarding the following aspects: (i) better integration of the characteristics of the application context, (ii) more intuitive access to the trust model, and (iii) better integration of recommendations by third parties. The last aspect is important as there are numerous situations in which direct evidence between entities is rare. The proposed approach provides means for the robust integration of recommendations provided by third parties, especially considering attacks by entities intentionally providing misleading recommendations.

Scientific Contribution: The scientific contribution of this thesis is summarized as follows:

- The trust model that is provided in this thesis extends Bayesian trust models in order to improve the integration of context-dependent parameters, such as *dispositional trust* and *aging* of evidence. Furthermore, a parameter called *maximum number of evidence units* allows the user to define the number of evidence that is expected to be sufficient for being representative for an entity's behavior within a certain application context. In the proposed model, the dispositional trust can be assessed according to the preference of the user; alternatively, a new approach for deriving the dispositional trust from the behavior of previously encountered entities is provided.
- The proposed interrelation between the aging and the maximum number of expected evidence units allows the limitations of current Bayesian trust models to be overcome. The thesis shows that in those models,

aging either does not have an impact on the expectation value in the absence of evidence, or it narrows the range of the expectation value.

- A second representation of trust - called the Human Trust Interface (HTI) - is proposed providing for an easier access to the model by human users. This representation is based on a simple set of parameters. These parameters are also the basis for a graphical representation allowing users to interpret and adjust the trust values of other entities intuitively.
- As the model supports two different representations a mapping between both representations is required in order to switch between both representations. The provided mapping allows users and developers of trust models to benefit from the advantages of both representations.
- The distributed computational model that is proposed for the aggregation of direct evidence and recommendations has been designed to be especially robust to so-called Sybil attacks, which occur when a single party tries to multiply the influence of its recommendations by creating a high number of seemingly independent entities. This is achieved using the accuracy of a recommender's past recommendations as well as the rank of the recommender in order to limit a recommender's influence. Especially, considering the rank of a recommender, i.e., its position in the group of recommenders, provides a means for limiting the influence of a potentially infinite number of malicious recommenders under certain circumstances.

Evaluation: The trust model has been evaluated in two user studies which support that users feel comfortable with the proposed graphical representation. Furthermore, in the simulation of collaboration in an opportunistic network, the model shows a good performance regarding the estimation of an entity's trustworthiness and regarding the average quality of interactions when using the trust model to find the best interaction partner. This results from the comparison to a state-of-the-art approach, as well as from a comparison to an artificial model that is initialized with the system variables of the simulation model, and therefore serves as perfect selection strategy. The simulation shows the results of the different approaches over a set of 15 populations, which have been canonically derived from the system model, modeling entities with different typical behaviors.

Zusammenfassung

Motivation: In der Vision des allgegenwärtigen Rechnens (engl. ubiquitous computing) wird der Mensch bei nahezu allen Tätigkeiten durch eine Vielzahl von Computern unterstützt. Hierfür ist eine nahtlose Zusammenarbeit aller Geräte der intelligenten Rechenumgebung nötig, welche auch neue Mechanismen zur Auswahl vertrauenswürdiger Interaktionspartner erfordert.

Ziel: In dieser Arbeit wird ein neuer Ansatz vorgestellt, um Entitäten, bspw. Nutzer oder deren Endgeräte, bei der Auswahl vertrauenswürdiger Interaktionspartner zu unterstützen. Das Ziel dieses Ansatzes ist es, die Vertrauenswürdigkeit potentieller Interaktionspartner möglichst gut zu schätzen. Durch die Auswahl vertrauenswürdiger Interaktionspartner soll die Zahl zufriedenstellender Interaktionen einer Entität erhöht werden.

Ansatz: In dieser Arbeit wird das Vertrauen in einen Interaktionspartner vor allem aus den bisherigen Erfahrungen aus vorangegangenen Interaktionen, bzw. genauer gesagt, aus den davon abgeleiteten *Hinweisen*, ermittelt. Hierzu wird auf sogenannten Bayes'schen Vertrauensmodellen aufgebaut, die aus der Literatur bekannt sind. Diese werden in mehrerlei Hinsicht erweitert: (i) um die Charakteristika des Anwendungskontextes, d.h. der Interaktionsdomäne, besser im Modell abzubilden; (ii) um den Nutzern einen intuitiveren Zugang zum Modell zu ermöglichen; und (iii) um Wissen Dritter besser einzubeziehen. Der letztgenannte Punkt ist wesentlich, da in vielen Situationen kein oder nur unzureichendes Wissen über potentielle Interaktionspartner vorliegt, welches direkt aus den Erfahrungen aus vorangegangenen Interaktionen abgeleitet werden kann. Im vorgestellten Ansatz werden neue Mechanismen entwickelt, um Wissen Dritter - sogenannte *Empfehlungen* - robust, d.h. unter Berücksichtigung möglicher Angriffe, zu integrieren.

Wissenschaftlicher Beitrag: Der wissenschaftliche Beitrag dieser Arbeit kann wie folgt zusammengefasst werden:

- Der entwickelte Ansatz erweitert Bayes'sche Vertrauensmodelle, um anwendungskontextabhängige Parameter wie bspw. das *Grundvertrauen* und das *Altern* von Hinweisen besser zu berücksichtigen. Insbesondere erlaubt es der Parameter *maximale Anzahl erwarteter Hinweise* festzulegen, wie viele Hinweise erwartet werden, um sie als repräsentativ für das Verhalten eines Interaktionspartners innerhalb des vorher

festgelegten Anwendungskontextes anzusehen. Dies ist ermöglicht das Berücksichtigen von Nutzerpräferenzen, wie auch eine Wahl dieses Parameters in Abhängigkeit vom Altern von Hinweisen. Darüberhinaus kann im entwickelten Modell das Grundvertrauen in Abhängigkeit von den Präferenzen des Nutzers gewählt oder aus dem Verhalten früherer Interaktionspartner innerhalb des betrachteten Anwendungskontextes abgeleitet werden.

- Im Rahmen dieser Arbeit wird zudem gezeigt, dass mit dem vorgestellten Ansatz Beschränkungen gegenwärtiger Vertrauensmodelle überwunden werden können, die sich bei der Berücksichtigung des Alters der gesammelten Hinweise ergeben. Es wird u.a. gezeigt, dass in bisherigen Arbeiten die Berücksichtigung des Alters von Hinweisen dazu führt, dass entweder der Vertrauenswert einer Entität bei Nichtaufkommen weiterer Hinweise unverändert bleibt oder der tatsächlich erreichbare Wertebereich des Vertrauenswertes eingeschränkt wird.
- Eine zweite, vereinfachte Repräsentation wird eingeführt, um die wesentlichen Modellparameter dem Nutzer intuitiver darzustellen, als dies im Rahmen bekannter Bayes'scher Vertrauensmodelle möglich ist. Diese Repräsentation ist auch die Grundlage für eine neue graphisch Darstellung, welche dem Nutzer die Interpretation der Vertrauenswerte und die Anpassung einstellbarer Parameter erleichtert.
- Da das entwickelte Modell zwei Repräsentationen der Vertrauenswürdigkeit einer Entität besitzt, wird eine Abbildung zwischen beiden definiert. Erst diese Abbildung macht es möglich, dass Nutzer und Entwickler von Vertrauensmodellen die Vorzüge beider Darstellungen simultan nutzen können.
- Für die Aggregation von direkten Hinweisen sowie Empfehlungen wird ein neuer Ansatz vorgeschlagen, welcher insbesondere die Robustheit gegenüber sogenannten Sybil-Angriffen (engl. Sybil attacks) verbessert. Dabei versucht der Angreifer, durch Aufbieten einer Vielzahl scheinbar unabhängiger Entitäten, gezielt die Auswahl eines Interaktionspartners zu beeinflussen. Die Verbesserung wird erreicht, indem bei der Gewichtung von Empfehlungen nicht nur die Richtigkeit früherer Empfehlungen, sondern auch der sogenannte Rang des Empfehlenden, d.h. dessen Einordnung in der Gruppe der Empfehlenden, berücksichtigt wird. Dies ermöglicht es unter bestimmten Bedingungen, den maximalen Einfluss einer potentiell unendlich großen Gruppe von Empfehlenden zu beschränken.

Evaluation: Das entwickelte Vertrauensmodell wurde einerseits in zwei Nutzerstudien evaluiert, welche die Hypothese unterstützen, dass Nutzer in-

tuitiv mit der entwickelten graphischen Darstellung umgehen können. Andererseits zeigt die Evaluation des Vertrauensmodelles in einer Simulation, dass sich mit dem Vertrauensmodell gute Ergebnisse hinsichtlich der Schätzung der Vertrauenswürdigkeit einer Entität und hinsichtlich der erreichten durchschnittlichen Qualität der Interaktionen erzielen lassen. Dies ergibt sich aus dem Vergleich zu konkurrierenden Ansätzen wie auch im Vergleich zum einem künstlichen Modell, welches die Systemvariablen der Simulationsumgebung kennt, und deshalb als sogenanntes perfektes Modell dient. Die Simulation zeigt die Ergebnisse der verschiedenen Ansätze für 15 Populationen, welche kanonisch aus dem Systemmodell abgeleitet wurden und sich hinsichtlich des typischen Verhaltens ihrer Entitäten unterscheiden.

Contents

Acknowledgements	iii
Abstract	v
Zusammenfassung	vii
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Goal	2
1.3 Object of Research	2
1.4 Scientific Contribution and Evaluation	3
1.4.1 Contributions	4
1.4.2 Evaluation	6
1.5 Publications	7
1.6 Thesis Structure	7
2 Background	9
2.1 Trust - A Social Concept	11
2.1.1 Definitions of Trust	11
2.1.2 Definition of Reputation	12
2.1.3 Properties of Trust	13
2.1.4 Categories and Qualities Relevant for Trust	13
2.2 Trust Establishment in Computer Mediated Interactions	14
2.2.1 Establishment of Personal Trust	15
2.2.2 Establishment of Structural and Dispositional Trust	17
2.3 Trust Based Decision Making	17
2.4 Conclusions	18

3	State-of-the-Art: Models of Trust	19
3.1	Representation and Computation of Trust	19
3.1.1	Aspects of the Computational Model	20
3.1.2	Aspects of the Representational Model	21
3.2	Trust Models by Semantics of the Trust Value	23
3.2.1	Commercial Feedback and Recommender Models . . .	23
3.2.2	Ranking Approach	25
3.2.3	Rating Approach	25
3.2.4	Probabilistic Approach	27
3.2.5	Belief Approach	31
3.2.6	Fuzzy Logic Approach	34
3.3	Analysis of the Trust Models	35
3.3.1	Analysis of the Representational Models	35
3.3.2	Analysis of the Computational Models	37
3.4	Conclusions	40
4	Concepts	43
4.1	Motivating Scenario in a Schematic View	44
4.2	System Model	45
4.2.1	Entities and Interactions	45
4.2.2	Roles	45
4.2.3	Process View	45
4.2.4	Definition of Trust	48
4.2.5	Trust Establishment	48
4.2.6	Basic Assumptions	54
4.2.7	Basic Attacks	57
4.3	Application Areas	58
4.3.1	Ubiquitous Computing - Opportunistic Networks . . .	58
4.3.2	Next-Generation Internet - Web Service Selection in Open SOA Market Places	60
4.3.3	Web 2.0 - Recommendations on Online Platforms . . .	62
4.4	Conclusions	63
5	Trust Model: CertainTrust	67
5.1	The Components of CertainTrust	67
5.2	Representational Model of Trust	70
5.2.1	Context-Dependent Parameters	71
5.2.2	Bayesian Representation	72
5.2.3	Aging of Evidence	77
5.2.4	Human Trust Interface (HTI)	80
5.2.5	Graphical Representation	82
5.2.6	Mapping between Representations	85
5.2.7	Evaluation of the Impacts of Aging and the Context- dependent Parameters	88

5.2.8	Summary	92
5.3	Computational Model of Trust	93
5.3.1	Basic Operators	94
5.3.2	Simple Trust Propagation	97
5.3.3	More Robust Trust Propagation	97
5.3.4	Sybil Attack-Resistant Trust Propagation	98
5.3.5	Evaluation of the Robustness to Sybil Attacks	102
5.3.6	Summary	106
5.4	Selection of an Interaction Partner and Update of Trust	107
5.4.1	Selection of a Candidate	107
5.4.2	Update Mechanism	108
5.4.3	Community-based Update of Dispositional Trust	111
5.4.4	Summary	113
5.5	Conclusions	114
6	Evaluation	117
6.1	Evaluation of CertainTrust in an Opportunistic Network	117
6.1.1	Basic Types of Behavior, Population Mixes & Settings	118
6.1.2	Simulation	119
6.1.3	Baselines and Models	120
6.1.4	Evaluation Metrics	122
6.1.5	Results	124
6.1.6	Summary	134
6.2	Evaluation of the Usability of the HTI	136
6.2.1	Evaluated Representations	136
6.2.2	User Study	139
6.2.3	Results	140
6.2.4	Discussion	143
6.2.5	Summary	144
6.3	Integration of CertainTrust in an Online Movie Recommendation Application	146
6.3.1	Description of the Application	146
6.3.2	User Study	147
6.3.3	Summary	150
6.4	Conclusions	151
7	Conclusions and Outlook	153
7.1	Conclusions	153
7.2	Outlook	155
	Bibliography	157

A Proofs	169
A.1 Proof for $E_{f,w,N}^{Beta} = E_{f,w,N}^{HTI}$	169
A.2 Proof for $E_{0.5,1,\infty}^{Beta} = E_{Simple}^{Beta}$	170
Erklärung	171
Wissenschaftlicher Werdegang des Verfassers	172

List of Figures

1.1	Representations of trust	3
1.2	Aggregation of direct evidence and recommendations	6
3.1	Trust network (Example 1)	21
3.2	Trust network (Example 2)	21
3.3	Opinion Triangle	32
4.1	Simple scheme of the system model	46
4.2	Main steps in establishing trust between entities and selecting entities	47
4.3	Basis of trust	49
4.4	Basis of trust (extended)	49
4.5	Context and sub-contexts	50
4.6	Simple trust network	52
4.7	More complex trust network	52
5.1	Components of a trust model	68
5.2	Representations of trust	71
5.3	Beta probability density function	73
5.4	Graphical representation of the HTI with labels (moderate strategy $f = 0.5$)	83
5.5	Graphical representation of the HTI with labels using different base trust values	84
5.6	Graphical representation of the HTI displaying the trust value	84
5.7	Alternative graphical representation ($f = 0.5$)	85
5.8	Alternative graphical representation ($f = 0.8$)	85
5.9	Comparison of the expectation values	89
5.10	Impact of the maximum number of expected evidence units N	89
5.11	Comparison of the certainty parameters	90
5.12	Impact of the parameters the base trust f and the weight w of the dispositional trust	90
5.13	Comparison of aging	91
5.14	Trust network	93
5.15	Effect of the discounting operator	95

5.16	Effect of the consensus operator	96
5.17	Discounting factor	100
5.18	Trust network - Sybil attack	104
5.19	Computation of trust: Aggregation of direct evidence and recommendations	105
5.20	Deriving binary evidence from continuous feedback	109
5.21	Determining the accuracy of a recommendation	110
6.1	Basic entity behaviors	118
6.2	Reputation evaluation over time in population <i>hmsw</i> using CT_C	125
6.3	Reputation evaluation over time in population <i>hmsw</i> using CT_None	125
6.4	Reputation evaluation over time in population <i>hmsw</i> using Beta_S	125
6.5	Reputation evaluation over time in population <i>hmsw</i> using Beta_D	126
6.6	Variants: Computational model	126
6.7	Variants: Update function	127
6.8	Variants: Maximum number of expected evidence units	127
6.9	Variants: Community-based update of base trust (1)	128
6.10	Variants: Community-based update of base trust (2)	129
6.11	Community factor in the context of interactions	130
6.12	Community factor in the context of interactions	130
6.13	Average error in estimating the trustworthiness of an entity	131
6.14	Average percentage of the accumulated sum of feedback	133
6.15	Representations evaluated in the user study	137
6.16	Example: CertainTrust - HTI	137
6.17	Example: Opinion Triangle (SL)	138
6.18	Example: Stars interface	139
6.19	Interpretation of the stars interface	139
6.20	Percentage of participants selecting interaction partner A per setting and per model	142
6.21	Mean values: Average percentage of participants selecting the same interaction partner as proposed by CT	142
6.22	Screenshots of TROP: Registration and login	147
6.23	Screenshots of TROP: Social network and rated movies	148
6.24	Screenshots of TROP: Actively recommended movies and directly rated movies	149
6.25	Screenshots of TROP: Rating of a movie	150

List of Tables

5.1	Relation of aging factor and E_{simple}^{Beta}	78
5.2	Example: Sybil attack	104
5.3	Influence of the factor d on ν for selected values of t and c . .	112
5.4	Influence of the factor d on ν in case of $t = c = 1$	113
6.1	Average error in estimating the trustworthiness across all populations	132
6.2	Average percentage of the accumulated sum of feedback for the selected populations	133
6.3	Number of evidence units per interaction partner and setting	141
6.4	Pairwise comparisons	141

Chapter 1

Introduction

Trust is a well-known concept in everyday life. In real life, trust can serve as the basis for decisions subject to risk and uncertainty. For the introduction of the common meaning of the concept, one may refer to the Merriam-Webster Online Dictionary. Among other statements, one find the following: trust is the “assured reliance on the character, ability, strength, or truth of someone or something”, and the “dependence on something future or contingent” [MW09]. In [BLRW04], Bhargava et al. point out that “trust [...] is pervasive in social systems” and that “socially based paradigms will play a big role in pervasive-computing environments”. This idea serves as the starting point for this thesis.

1.1 Motivation

The goal of ubiquitous computing is to support the users in their daily life. Supporting users anytime, anywhere requires giving up the desktop computer paradigm. Instead of using desktop computers, the user will be surrounded by a large number of loosely coupled, networked devices. Collaboration in ubiquitous computing environments can be predicted to increase substantially. Three major reasons for this trend can be given. First, due to increased integration of wireless technologies in mobile devices (e.g., most mobile phones are capable of exchanging data via Bluetooth or WiFi), the opportunities for spontaneous interaction increase. This brings about new paradigms for collaboration between users, the most prominent example being ‘Opportunistic Networks’ [Hei07]. Second, as ubiquitous computing devices will be heterogeneous and, hence, often limited regarding their capabilities, the collaboration between those devices will be indispensable to unfold the power of ubiquitous computing. Third, since everyday life will be more and more supported by or dependent on IT, users will want or will have to use more and more IT based services on the go. This may lead to a manifold set of applications where users are in the position to select their interaction

partners, e.g., service vendors or service providers, from a set of known and unknown interaction partners, which may depend on the time and location of the service request.

The success of collaboration is based on the selection of reliable partners. Relying only on traditional certificate based approaches does not seem to be appropriate as they have several shortcomings. First, an unmanaged domain, like ubiquitous computing, may lack a central authority [SH08], which is necessary for issuing and revoking certificates. Second, certificates providing only information about the identity of an entity may not be sufficient, as a unique identifier or pseudonym does not convey information about the behavior of this entity per se [CSG⁺03].

1.2 Goal

The goal of this thesis is the development of new techniques supporting entities, i.e., users or their devices, in selecting trustworthy interaction partners for making collaboration successful. In an unmanaged domain, such as ubiquitous computing, the challenges arise from the uncertainty about the identity and goals of the potential interaction partners.

According to the basic idea of ubiquitous computing, the solution needs to be non-intrusive and human-centered. The integration of these conflicting aspects can be achieved by a solution that is able to autonomously integrate relevant evidence about the trustworthiness of interaction partners collected within the system, as well as evidence available to the users based on “real world” experience.

1.3 Object of Research

The approach followed in this thesis is to improve the overall quality of interactions by using trust as a well-founded basis for the selection of an appropriate interaction partner. Trust between entities can be established based on direct evidence from past interactions and on recommendations from other entities (indirect evidence).

Representing trust based on evidence requires the definition of a relationship between trust and evidence. This relationship may depend on the application context the trust model is applied to, e.g., information exchange or recommendation of movies, and it needs to consider the age of the evidence, as well as the typical behavior of entities. Furthermore, an intuitive representation of trust is crucial, as the ease-of-use and the feeling of being in control are important aspects for gaining user acceptance. The provision of means that enable the users to control and to manipulate the trust assigned to other entities is especially important, as it allows for integration of “real world” experience and the intervention in the selection process by the user.

An expressive model needs to represent the level of the expected trustworthiness of an interaction partner, as well as the associated level of (un-)certainty. The latter can be expressed by reflecting the relation between the number of evidence units the expectation is based on and the number of evidence units that is assumed to be sufficient for a well-founded decision. In order to support ubiquitous computing in a non-intrusive and human-centered way, the model of trust needs to be appropriate to serve as a basis for autonomous decision making, as well as for an interface which is intuitively interpretable by humans.

The integration of recommendations is necessary for the establishment of trust in cases in which direct evidence is rare. As this enables foreign parties to provide misleading information, a robust integration of recommendations requires a carefully designed computational model of trust.

1.4 Scientific Contribution and Evaluation

This thesis proposes a new trust model called *CertainTrust*. It is based on a modified Bayesian approach for modeling trust. The approach overcomes limitations of current state-of-the-art approaches with respect to the representation of trust (see Figure 1.1) and the computation of trust (see Figure 1.2).

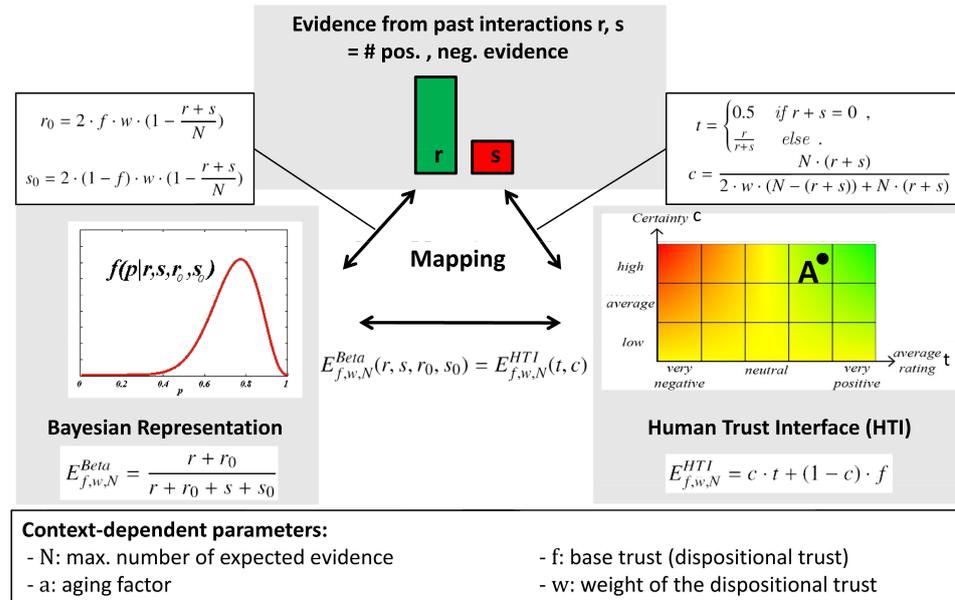


Figure 1.1: Representations of trust

1.4.1 Contributions

1. **Expressive trust model - deriving trust from evidence considering context-dependent parameters:** The model provides means for deriving trust from evidence from past interactions. It allows trust to be interpreted as a subjective probability and to consider the following context-dependent parameters:

First, the dispositional trust of an entity in an application context is expressed using two parameters. The *base trust* f specifies the trust value for unknown entities. It can be dynamically updated based on the experience with the encountered entities. The *weight* w of the base trust influences how quickly the trust value shifts from the base trust value to the average rating of the past interactions, when evidence is available.

Second, the *maximum number of expected evidence units* N is introduced to define a number of evidence units that is expected to be sufficient in order to consider the collected evidence as representative for the behavior of an entity in an application context.

Third, the *aging factor* a allows more recent evidence to be given a higher weight. It is proposed to choose the maximum number of expected evidence units based on the aging factor in order to take into account that aging may limit the number of evidence units that is used to derive the trust value.

2. **Extension to Bayesian trust models - overcoming limitations of aging:** Trust is modeled as a subjective probability and derived from evidence from past interactions, therefore, the model is based on state-of-the-art Bayesian trust models [JI02, Jøs01, WJI05, TPJL06, BLB04, MMH02a]. However, the introduction of the context-dependent parameters above required a modification to the state-of-the-art trust models. The integration of these parameters is achieved by a dynamic adaptation of the prior knowledge. Besides the introduction of dispositional trust, the proposed approach overcomes limitations of current approaches regarding the effects of introducing the aging of evidence. This thesis shows that in current Bayesian trust models, aging either does not have an impact on the expectation value in absence of evidence [BLB04], or aging narrows the range of the expectation value, e.g., [JI02, WJI05].
3. **Representation of trust supporting human users - providing a simple set of parameters and an intuitive graphical representation:** Supporting the user with an intuitive representation of trust that provides easy manipulation and interpretation of the trust information collected by the system requires an interface that is tailored

to the needs of human users. This issue is addressed by introducing a representation - called Human Trust Interface (HTI) - that derives the trust value of another entity based on two independent parameters (which in turn partly depend on the context-dependent parameters). The *average rating* expresses the average outcome of past interactions. The *certainty* increases with the number of collected evidence units. It defines the influence of the average rating on the trust value in relation to the base trust value. In the HTI, the trustworthiness of an entity is derived based on a simple formula that integrates the average rating, the certainty, and the base trust. The basic idea is that with increasing certainty, i.e., with an increasing number of collected evidence units, the final trust value shifts from the base trust value to the value of the average rating. The simple set of parameters is the basis for the graphical representation that has been designed in order to support human users when interpreting and adjusting trust values.

4. **Mapping between the proposed representations - to benefit from the advantages of each representation:** In order to benefit from the advantages of the Bayesian representation and the HTI, a mapping between both representations is provided. The challenge herein is to define this mapping in such a way that the derived trust values are independent from the representation, i.e., the expectation value in the Bayesian representation and the trustworthiness in the HTI are required to be equal. This is considered when deriving the representation dependent parameters from evidence and the context-dependent parameters.
5. **Robust computational model - limiting the impact of Sybil attacks:** The computational model allows for the aggregation of direct evidence and recommendations. The operators for the aggregation take their cues from the operators for *discounting* and *consensus* introduced in [Jøs01, JI02]. The discounting (weighting) of the recommendations is based on the trustworthiness of the recommenders. It is estimated according to the accuracy of a recommender's past recommendations. The consensus operator is extended to reduce the influence of recommenders based on the rank of their trustworthiness. Furthermore, the influence of recommenders is limited by considering only a maximum amount of evidence per recommender and the amount of direct evidence and recommendations available. Using the property of the convergence of a geometric series provides for limiting the maximum impact of an arbitrary high number of recommenders, as long as their trustworthiness is below a certain threshold t_s . This is especially important when facing Sybil attacks. Hereby, a single party can try to multiply the influence of its recommendations by creating a high number of

seemingly independent entities.

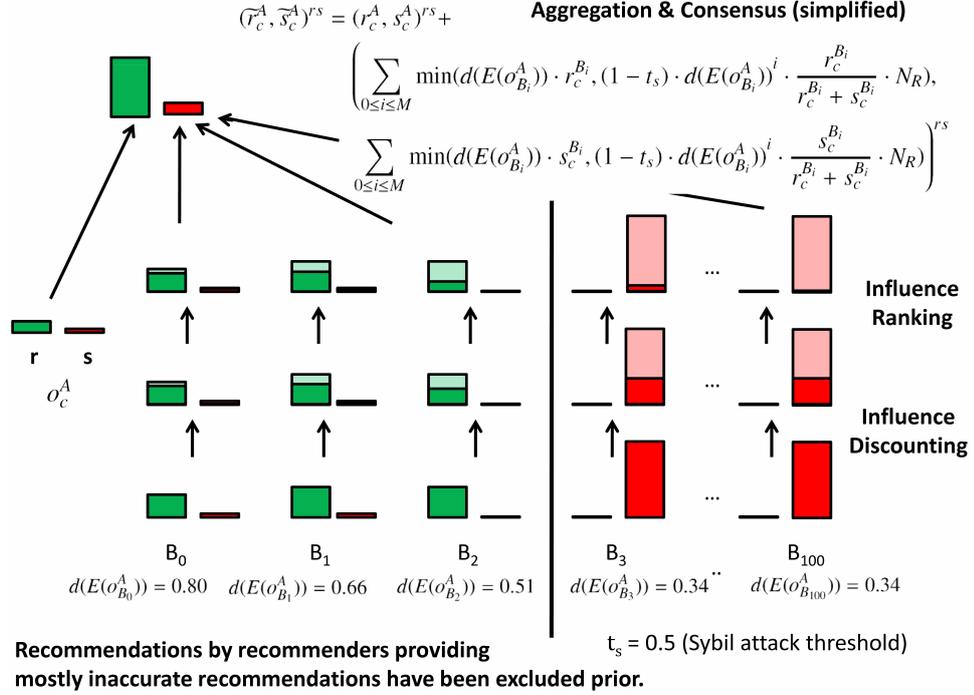


Figure 1.2: Aggregation of direct evidence and recommendations

1.4.2 Evaluation

The evaluation shows the improvements of the modified Bayesian approach in the face of aging and the enhancements of the model's robustness regarding attacks by entities providing misleading recommendations, especially with respect to Sybil attacks.

The impact of the trust model on the quality of interactions and the error in estimating the trustworthiness of an entity is based on simulations. The scenario for the simulation is a mobile file sharing scenario in an opportunistic network using user traces from the Reality Mining Project [EP06]. The results show how the model copes with different population mixes, regarding the distribution of a typical user's behavior and its stability.

The evaluation of the usability of the representation of trust for human users (HTI) is evaluated based on a user study with an online movie recommendation platform which was developed for this purpose. Furthermore, another user study was performed comparing the time users take for selecting the best interaction partner in the HTI, the Opinion Triangle [Jøs01], and an Amazon-like stars interface.

1.5 Publications

Parts of this thesis have been published in book chapters and in proceedings of international conferences and workshops. The state-of-the-art and challenges for modeling trust have been addressed in [Rie06, RKM06, Rie08a, Rie08b, ARSB⁺08]. The basic concepts for the proposed trust model and its evaluation have been published in [Rie07, RKM07, RH08, RS08, Rie09, RA09]. Furthermore, the collaboration in the related field of privacy led to an additional publication [WRH07].

1.6 Thesis Structure

The main part of this thesis is structured as follows. Chapter 2 provides background information about the concept of trust in general, and how trust can be established and used to support decisions in computer or network mediated interactions.

In Chapter 3, state-of-the-art trust models for deriving trust from evidence are introduced. In addition to introducing two commercial reputation or recommendation systems, the trust models are classified based on the semantics of their trust values. Afterwards, there is an analysis of the features of current trust models revealing the first issues which need to be tackled when developing a trust model for ubiquitous computing.

Chapter 4 introduces the basic concepts that are necessary for specifying a trust model. The chapter contains the definition of the system model and the basic assumptions that need to be fulfilled in order to apply an evidence based trust model. It shows that applications in the field of ubiquitous computing, open service platforms, and Web 2.0 can fulfill these assumptions. Based on the assumptions, the design goals for the new trust model are provided in the conclusions of the chapter.

Chapter 5 presents the developed trust model. This chapter shows how the trust model represents and computes trust values. It provides an extension to Bayesian trust models and it introduces a novel representation of trust for users. The computational model presents an extension to current trust models in order to improve the models robustness to Sybil attacks. Furthermore, mechanisms for assessing trust in recommenders based on a recommender's past recommendations are introduced.

In Chapter 6, the evaluation of the trust model is provided. The chapter presents the evaluation of the trust model in a simulated opportunistic network, as well as two user studies on the usage of the trust model.

Chapter 7 provides the conclusions recalling the major results of this thesis and an outlook.

In order to avoid confusion about the terminology used in Chapter 2 and Chapter 3, a short introduction of the most important terms is provided:

- *Entity*: An entity is an abstract concept that may refer to a user, an abstract service provider or interaction partner, a software agent or autonomous software component, a peer, or a computing device. Entities can establish trust between each other, and they are assumed to have a behavior. For example, an entity that shares files with others may provide good files, or it may provide corrupt files; entities that share information may offer accurate or misleading information. Entities offering accurate information can be considered to be trustworthy.
- *Interaction*: Interaction between entities is an abstract concept that is the basis for collaboration. An interaction can be information sharing, file sharing, or usage of a provided service, etc.
- *Trust and reputation*: In this thesis trust and reputation are introduced as two similar but distinct concepts; the definition for both terms are introduced at the beginning of Chapter 2. Yet, as in the current literature the terms *trust* and *reputation* as well as *trust model* and *reputation model* are partly used as synonyms and partly with different meanings, the distinction of these terms is not strictly maintained throughout this thesis.

Chapter 2

Background

The goal of this thesis is to support the selection of trustworthy interaction partners. In ubiquitous computing, as in real life, trust can serve as a basis for risky engagements in the presence of uncertainty. As successful collaboration depends on the selection of a trustworthy interaction partner, it is an interesting challenge to evaluate the trustworthiness of the entities that surround users in ubiquitous computing environments. If an entity is able to identify trustworthy interaction partners, it profits from the capabilities and services they offer. Thus, it takes advantage of the power of ubiquitous computing and avoids disappointments.

Ubiquitous computing is the application area that is the focus of this thesis. According to [SH08], ubiquitous computing may be an unmanaged domain. This means that in there are scenarios in which anybody and any device can participate in the ubiquitous computing environment. This leads to the threat that there are not only benevolent interaction partners that offer their services in order to contribute to the ubiquitous computing environment, but also malicious interaction partners that try to make others interact with them. For example, they might be interested in maximizing their profits by offering a low quality service for a high price or in distributing viruses and malware.

Trust and reputation systems have already been successfully applied in order to support users in finding trustworthy interaction partners in centralized and managed settings, e.g., on the auction platform eBay [eBa09b]. Yet, those approaches cannot be directly transferred to a distributed and potentially unmanaged domain.

In order to disburden the user and support autonomous decision making in the presence of risk, a trust model needs to provide measures for the trustworthiness of an interaction partner, i.e., a trust value, and for reasoning about the confidence or the certainty associated with this trust value. Furthermore, it needs to be suitable for integration in the decision making process of an autonomous software component or agent. In addition, as there

are cases in which the user wants to interact with the system, a trust model needs to provide an intuitive interface for human users. This need can arise, when a user has “real world” knowledge about the trustworthiness of another entity, e.g., if the user knows another entity from work or from school, then they might want to manually assess its trust value for this entity.

Furthermore, a trust model needs to provide parameters that can be adapted to the characteristics of the application context, i.e., the application area, it is used in. Here, e.g., aging provides a means for considering changes in an entity’s behavior. Furthermore, a user’s dispositional trust, that is their general attitude to trust in other entities, is an important factor that may vary from context to context.

Finally, another challenge arises when recommendations by other entities are considered. This is especially important in order to establish trust in cases when direct experience is rare. In unmanaged domains, this needs to be done quite carefully as it allows foreign parties, benevolent as well as malicious ones, to influence one’s decision making process.

The following provides an example showing how a trust model can support users in an opportunistic network scenario. The scenario is presented, as it has been presented to the participants of a user study that has been conducted as part of this thesis (see Section 6.2):

“You and your friends are on your way to a soccer match in the stadium in Frankfurt. As usual, you take your personal device - a next-generation mobile phone - with you. The week before, you informed your personal device that you are looking for a certain song (mp3) and that you want to buy an mp3 player. While moving through the crowd in front of the stadium, your device searches (wirelessly) for potential interaction partners who offer the song or the mp3 player. Shortly before passing the security check, one of your friends meets some of his colleagues. Your personal device discovers that a member of this group (Dirk) offers the song you are looking for. To reduce the risk of getting a file that is damaged or contains a virus, your personal device collects recommendations from the mobile devices of your friends, who either know Dirk or have had a number of interactions with him. As the recommendations are positive, your personal device downloads the song from Dirk’s device. Afterwards, it checks the file for noise and viruses. As the file is clean, your personal device generates a profile for Dirk, and notes that there has been a positive interaction in the context of mp3-exchange. All this has been done without your interaction; only after the successful exchange, a short vibration of your personal device indicates the positive interaction. In this scenario, the personal device made all decisions by itself. But in the case that your personal device

would have only collected a small number of evidence units, not sufficient for an autonomous decision, it could have notified you. Then, it would have been your choice whether to interact or not. In this example, the risk associated with the interaction is limited. However, in the case that someone offers a used mp3-player for 10 EUR, one will probably be glad about any evidence about the trustworthiness of this potential interaction partner.”

In the following parts of this chapter, first, a general notion of trust is introduced. Besides providing definitions for trust, the section introduces properties that are usually associated with trust and it introduces the main categories that are relevant for trust establishment. Second, it is shown how trust can be established in computer or network mediated interactions. Third, an approach for integrating trust in decision making is shown.

2.1 Trust - A Social Concept

Trust is a well-known concept in everyday life that simplifies many complex processes. On the one hand, trust in the social environment allows humans to delegate tasks and decisions to an appropriate person. On the other hand, trust facilitates an efficient rating of the quality of the information presented by a trusted party. There is much work on trust, not only in computer science, but also in other academic fields, e.g., sociology, economics [JIB07, GS00, AG07, WV07].

2.1.1 Definitions of Trust

Although trust is a well-known concept in everyday life and despite the fact that there is a set of properties of trust (see Section 2.1.3), on which most researchers agree, it is hard to define trust. Apart from the definition in the Merriam-Webster Online Dictionary stated in Chapter 1, there are a couple of definitions with different focuses [Mar94, AR04]. A definition, which is shared or at least adopted by many researchers [JIB07, ARH00, MMH02a, KR03, TPJL06], is the definition provided by the sociologist Diego Gambetta [Gam00, Gam90]:

“trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action.”

The most important points in this definition are: Trust is subjective, and it includes an element of prediction or expectation. Furthermore, trust

is tied to the performance of another agent which affects the action of the trusting agent.

In [JKD05], the term *trust* is differentiated in reliability trust - which may also be referred to as evaluation trust - and decision trust.

Reliability trust is defined as:

“Trust is the subjective probability by which an individual, *A*, expects that another individual, *B*, performs a given action on which its welfare depends.”

Decision trust is defined:

“Trust is the extent to which a given party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible.”

The definition of *reliability trust* is close to the definition *trust* provided by Gambetta. The definition of *decision trust* extends the previous definitions by implicitly introducing notions of utility, environmental factors and risk attitude [JKD05]. Furthermore, it states that trust is situation dependent. Assuming that the context of trust describes the field or the application area in which one entity assesses the trustworthiness of another one, e.g., splinting a fracture or providing a service for online banking, situation dependence is a step beyond context-dependence. For example, a person might not trust their ear, nose and throat doctor to splint a fracture (context), but make an exception in the situation of an emergency.

The definition provided in [MMA⁺01] also clearly states the idea that trust is a subjective expectation. Additionally, it introduces the information about how trust is established.

“Trust: a subjective expectation an agent has about another’s future behavior based on the history of their encounters.”

In this thesis, the definition of *reliability trust* provided by Jøsang et al. [JKD05] will also serve as basis for the definition of trust. The (final) definition will be provided in Chapter 4 together with the definition of the other elements of the proposed model. The aspects introduced in the definition of *decision trust* will only be taken into account in the face of decision making (see Section 2.3), but not in the face of evaluating the trust in another entity. The issue of trust establishment will be addressed separately in Section 2.2.

2.1.2 Definition of Reputation

A concept that is often mentioned together with trust is reputation. In order to avoid confusion, this section provides a definition for reputation and explains its relation to trust.

In [JIB07], there is the following definition for reputation:

“Reputation is what is generally said or believed about a person’s or thing’s character or standing.”

Although the definition only introduces an abstract notion of reputation, it allows one to easily differentiate between trust and reputation.

Trust describes a subjective relation between an entity and another entity (or group of entities). Reputation is what is generally said about an entity. Thus, the reputation of an entity is based on the opinions provided by all entities. Trust may be used to determine the reputation of an entity. The other way around, reputation may also be used to determine the trustworthiness of an entity [JIB07].

2.1.3 Properties of Trust

The following properties are usually assigned to trust [ARH00, Mar94] and relevant when transferring the concept to computer or network mediated interactions. Trust is *subjective*, i.e., the trust of an entity *A* in an entity *C* does not need to be the same as the trust of any other entity *B* in *C*. Furthermore, one cannot expect the trust of *A* towards *C* to be the same as the trust of *C* towards *A*, thus trust is *asymmetric*. Trust is *context-dependent* and *situation-dependent*. Obviously, there is a difference in trusting in another entity as provider of music files or as provider of an online banking service. It also is a difference in trusting in someone as service provider or as provider of recommendations. If *A* trusts *B* in the context of providing recommendations about a good service provider, e.g., for file-storing, this does not necessarily imply that *A* trusts in *B* as a good peer to store files at, and vice versa. Trust is *dynamic* and *non-monotonic*, i.e., experience can increase as well as decrease trust. Thus, it is necessary to model both positive and negative evidence. Trust is *not transitive* in a mathematical sense, but the concept of recommendations is very important. Particularly, as recommendations are necessary to establish trust in entities about which any or only little direct experience is available.

2.1.4 Categories and Qualities Relevant for Trust

The following provides an introduction of different categories of trust and qualities a trusted entity is considered to have. Both issues are important in order to establish trust.

McKnight and Chervany state in [MC96] that there are three principle categories of trust: *personal / interpersonal* trust, *impersonal / structural* trust, and *dispositional* trust. Personal (interpersonal) trust describes trust between two persons (groups of people) in a specific situation. Structural trust is not bound to a person but rises from the social or the organizational

situation. Dispositional trust can be explained as a person's general attitude towards the world or towards other people. It is cross-personal and cross-situational.

Furthermore, McKnight et al. introduce a concept called *Trusting Beliefs*, which “means the extent to which one believes (and feels confident in believing) that the other person is trustworthy in the situation” [MC96]. This is close to the definitions of trust introduced above. They found that *Trusting Beliefs* in another person usually means to expect the person to be *benevolent* (willing to serve another's interest), *honest* (proving the willingness by making and fulfilling agreements to do so), *competent* (able to serve another's interests) or/and *predictable* (one's willingness and ability to serve another's interests does not vary or change over time). If it is possible to find a person with these qualities, interaction with this person would be expected to have a positive outcome.

2.2 Trust Establishment in Computer Mediated Interactions

As introduced above, the concept of trust may be used for estimating the future behavior of an entity. In order to transfer this concept to computer or network mediated interactions, especially in the field of ubiquitous computing the following questions arise:

1. How to establish trust between entities in ubiquitous computing environments?
2. How to make trust based decisions in the face of interactions associated with risk?

In “real life”, trust can be established based on various cues. It may be based on personal encounters with another person or service provider. The appearance, the clothing, or the role in an organization allows one to estimate the competence, the honesty, the benevolence, or the predictability of a person or service provider regarding its future behavior.

In computer or network mediated interactions, these cues might not be available. Entities one may interact with, might be known only by their digital identifier. This can be a website or solely a unique pseudonym. While a website still might provide some cues that allow one to reason about the trustworthiness of an entity, e.g., based on the contact address, a simple pseudonym does not necessarily convey any information about the real world identity of the interaction partner. Thus, it might be impossible to directly evaluate the four qualities (see Section 2.1.4) of trust based on “real world” knowledge about the interaction partner. Therefore, computer mediated

interactions require new concepts that allow entities to establish trust and to reason about the trustworthiness of other entities.

Finally, in the case when a user A has “real world” knowledge about their interaction partner B , e.g., assume A and B are friends, there is the need for providing means that allow the user to make this information available to an autonomous software component or agent that is to make autonomous decisions on the behalf of its owner.

2.2.1 Establishment of Personal Trust

At first, the transfer of personal trust is considered. As shown in [AR04], much work is done on transferring this category of trust to computer sciences, whereas there is little work supporting the other categories. According to [AG07, BDOS05], there are currently two approaches to establish trust between entities in computer science.

In [BDOS05], the approaches have been referred to as “policy-based trust management” (relying on objective “strong security” mechanisms) and “reputation-based trust management” (based on direct experience and feedback provided by others). In order to not mix the semantics of the terms reputation and trust, the term “reputation-based trust management” is replaced by “evidence-based trust management” within this thesis.

2.2.1.1 Policy-based Trust Management

In this approach trust is stated implicitly in the form of credentials. A typical scenario is that an entity A wants to access the resources of another entity B . Entity B will only grant access rights to entity A , if entity A can provide the necessary credentials. Policies are used to state which credentials are necessary. The credentials are usually certificates, which have been signed by a trusted third party. The credentials may state information about the identity of the owner [ITU97] or information about the rights of the owner [BFL96]. The act of trust establishment, i.e., the evaluation whether an entity should obtain certain credentials - e.g., based on the evaluating the benevolence, competence, honesty, and predictability - is done by the party issuing the certificates. It is an external process.

In [BFIK99, BFL96], trust management is defined as

“a unified approach to specifying and interpreting security policies, credentials and relationships that allow direct authorization of security-critical actions.”

This definition is in the sense of policy-based trust management. It describes what can be considered to be a traditional approach to trust management, i.e., trust is only treated implicitly and in a rather static manner [CSG⁺03, Gra03].

Well-known examples of this approach are presented in [BFL96, BFK98]. A main drawback of both systems is that they treat trust as monotonic, i.e., additional credentials can only increase granted permissions.

Another shortcoming of the policy-based trust management is that it usually relies on a trusted third party that issues the certificate, stating that an entity is considered to be trustworthy. Thus, the process of trust establishment is external to policy-based trust management. Furthermore, certificate based approaches, e.g., [ITU97], that rely on a public key infrastructure require further means for the distribution, verification, and revocation of keys.

2.2.1.2 Evidence-based Trust Management

The second approach tries to establish trust without the need for an external source of trust. The trust is established based on evidence derived from past interactions. As direct evidence between entities may be rare, most evidence based approaches consider the exchange of recommendations between entities, i.e., an entity provides another entity with information about its previous experience. In case entity *A* (trusting entity) is to evaluate the trustworthiness of an entity *B* (trustee), a trust model provides a means for evaluating evidence about the trustworthiness entity *B*. In online communities, e.g., on auction platforms like eBay [eBa09b], the evidence is information about the quality of past interactions rated by former interaction partners of entity *B*. Like in policy-based trust management, the trusting entity does not directly evaluate the qualities (benevolence, competence, honesty, and predictability) of the trustee. Here, when an entity provides a sufficient number of interactions with positive outcome, it is assumed to have these qualities, and therefore it is assumed to be trustworthy.

The advantage of this approach is that it poses very little requirements to the environment it is applied to. The interaction between entities can be assumed to be the intrinsic purpose of the application, which requires the evaluation of the trustworthiness of the potential interaction partner. Furthermore, it can be expected that entities are capable of creating evidence that state the quality of an interaction, e.g., by providing ratings. After an interaction, this information can be used to update and re-evaluate the trustworthiness of the interaction partner. Finally, ubiquitous computing environments naturally provide capabilities for communication. These are necessary in order to find potential interaction partners and to exchange recommendations.

Thus, the evidence which is used to evaluate the trustworthiness of an entity is created by the participants of the system and distributed within the system. The approach does not require additional infrastructure or trusted third parties.

Yet, the approach has also obvious shortcomings. The approach provides

only evidence about the trustworthiness, but it does not directly state whether an entity is trustworthy or not. It is also necessary to adapt the trust value of an entity when its behavior changes. Furthermore, there is the need for a mechanism for aggregating one's direct experience with recommendations provided by others. This mechanism especially needs to be able to cope with misleading recommendations.

Another drawback is that the approach is implicitly based on the assumption that there has been interaction between at least some entities. The bootstrapping of an evidence based trust model may need some external information about trustworthiness of entities. The bootstrapping may be especially difficult in contexts in which the risk is too high to interact with unknown entities.

2.2.2 Establishment of Structural and Dispositional Trust

Besides personal trust, there are the categories containing structural trust and dispositional trust (see Section 2.1.4).

Structural trust rises from the social or the organizational situation. This requires that there is a community knowing each other (social situation) or having well-known rules for interaction with each other (organizational situation). Thus, this kind of trust may be hard to transfer to an unmanaged domain. Therefore, it is not considered in the following. Yet, there are several approaches that try to consider the establishment of trust based on the social relationship between agents or the owners, e.g., [SS02a, HJS06].

Dispositional trust is a person's general attitude towards the world. Thus, one way to initialize the dispositional trust would be to ask the user of the trust model. As dispositional trust has been introduced to be cross-personal and cross-situational, it should also be a rational approach to derive the dispositional trust from the behavior of all entities that have been encountered so far. This would only require the provision of a working mechanism for establishing personal trust.

2.3 Trust Based Decision Making

As motivated above, a trust based decision depends on the situation. Besides the trust in the interaction partner, an important aspect is the risk associated with an interaction. Up to the point of decision making, the information about trust and risk can be managed in arbitrarily representational structures. For the task of decision making, it is necessary to resolve all information provided in order to make a decision whether to interact or not.

In the case of trust based on certificates, the interaction is supported if the necessary credentials have been collected based on policies [BFL96, BFK98].

In other cases [Mar94, CSG⁺03, JHF03, Jøs99b], the decision making is done threshold based, i.e., there is a minimum value of trust (threshold) that

is required for the interaction.

The presence of uncertainty and contradicting evidence complicates decision making. A possible solution could be to integrate the user into a decision making process providing a preliminary decision and asking for commitment. Although this can lead to higher acceptance of trust-aided decision support by the user, it takes away the benefit of automation, and will not comply with the principles of a calm technology.

A well-founded approach of threshold based decision making in the face of risk is to calculate the expected utility of an interaction (according to [QH07] based on [Ber54, NM44, Sav54]).

Assume for an interaction a with possible outcomes o_1, \dots, o_n the utility of each outcome is given as $U_i(a)$, and the probability of o_i is given by $E_i(a)$. Then, the expected utility $EU(a)$ is defined as:

$$EU(a) = \sum_{i=1}^n U_i(a) \cdot E_i(a) \quad (2.1)$$

For an interaction with binary outcome, this can be further simplified. Let $U^+(a)$ denote the utility of a positive outcome (benefit) – occurring with probability $E^+(a)$ – and $U^-(a)$ the utility of a negative outcome (costs) – occurring with probability $E^-(a)$. Then, the expected utility $EU(a)$ can be calculated as:

$$EU(a) = E^+(a) \cdot U^+(a) + E^-(a) \cdot U^-(a) \quad (2.2)$$

The interaction takes place only if the expected utility $EU(a)$ is greater than a predefined threshold t . For rational entities the threshold is $t = 0$ [Jøs99b].

It is important to note that this approach requires that the trust model models the trustworthiness of an entity as a probability in order to use it in the equations above.

2.4 Conclusions

This section provided a basic introduction of the concept of trust and how trust establishment can be achieved in computer or network mediated interactions. As policy-based trust management may require a trusted third party in order to issue certificates and to introduce entities as trustworthy, and as the pure knowledge about the identity of an entity does not convey information about the behavior of the entity per se [CSG⁺03], this thesis focuses on deriving the trustworthiness of an entity based on evidence derived from past interactions.

Chapter 3

State-of-the-Art: Models of Trust

In the following, an overview of state-of-the-art trust models is provided. In current research, there is a large number of trust models. Trust models may be developed for an application area - e.g., electronic commerce [eBa09b], P2P [KSGM03, BLB04, AD01], web search [PBMW98], movie recommendations [GH06], public key authentication [Zim94], service-oriented computing [BHOC07] - or more generally for distributed environments like ubiquitous computing, virtual organizations, and agent societies [PTJL05, TPJL06, ARH00, YS02, SS02a, Sab03, Mar94, JI02, MMH02a, AMCG04, CSG⁺03].

The focus of this chapter is on models that evaluate the trustworthiness of an entity based on an entity's direct evidence or recommendations or both. After a short introduction of the general functionality of a trust model, and a side-look to commercial feedback or recommender systems, a selected set of trust models is introduced based on a classification of their trust values. Then, an analysis of the introduced models shows aspects with which current models deal.

3.1 Representation and Computation of Trust

In the following, it is proposed to make a separation between two aspects of trust models, namely between their *representational model* and their *computational model*.

The representational model defines how trust is represented and established, and the computational model defines how different sources of evidence are aggregated.

3.1.1 Aspects of the Computational Model

The computational model defines how the different sources of trust, i.e., direct evidence and recommendations, are integrated. Here, it is important to consider whether a trust value is supposed to be a subjective trust value, i.e., trust depends on the entity which evaluates its trust in another entity, as introduced in Section 2.1.1, or whether it is a global trust value, i.e., a reputation value according to the definition in Section 2.1.2.

According to [ZL04], the approaches in [KSGM03, PBMW98, RAD03] are characterized as follows: (1) they take into account all entities and trust links between them and (2) the calculated value is independent from the entity that evaluates the trust value. Thus, those models can be considered to compute global trust value.

In contrast, in trust models that provide means for the computation of subjective trust values, an entity usually only considers recommendations from a subset of all entities, and it may use subjective measures to define the impact of the collected recommendations [QHC06, TPJL06, BLB04, HJS06]. If a recommender does not have any knowledge about the interaction partner, it can either forward the request to other entities [Gol05, HJS06], or it may report that it does not have any experience [BLB04, TPJL06]. Finally, the models propose different mechanisms to filter and weight the recommendations before calculating the trust value.

The following shows a simple example illustrating how a subjective trust value can be derived given a network of entities providing recommendations. A simple trust network might look as shown in Figure 3.1 or 3.2.

In both examples, entity A wants to evaluate the trustworthiness of entity C based on its direct evidence and the recommendations provided by its neighbors (entities R_1 , R_2 , R_3 , and R_4). As the recommenders forward their recommendations along the edges of the graph, this is also called *chain-based* trust evaluation in this thesis.

In Figure 3.1, the recommendations by R_1 and R_2 are based on their direct evidence. In Figure 3.2, R_1 and R_2 do not have direct evidence but may forward the information received from R_3 . Furthermore, R_2 can also ask R_4 for recommendations. When receiving the recommendations, the entity A has to aggregate the direct evidence and the evidence provided by the recommenders in order to evaluate the trustworthiness of the C .

Having introduced the basic functionality of the computational model at a very high level of abstraction, the functionality of the representational model will be introduced next. Further details on the computational model, i.e., which mechanisms are applied in order to enhance a trust model's robustness in the face of misleading recommendations, are introduced with the description of the trust models in Section 3.2 and in Section 3.3.

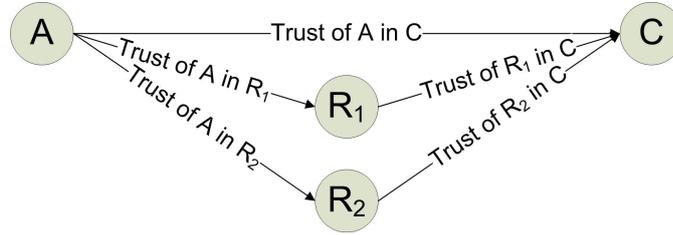


Figure 3.1: Trust network (Example 1)

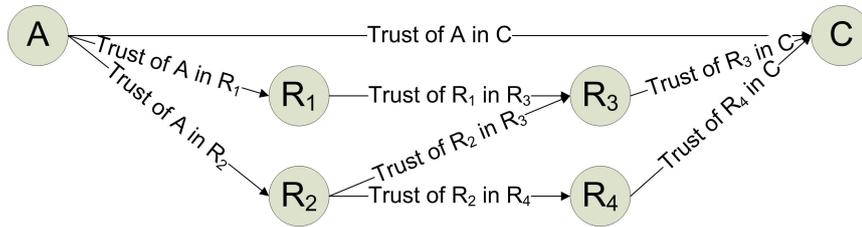


Figure 3.2: Trust network (Example 2)

3.1.2 Aspects of the Representational Model

The representational model of a trust or reputation model defines how trust is represented and established.

Differences in the representation of trust e.g., can be found in the *domain* of the trust value. A binary domain allows only the expression of the two states “trusted” and “untrusted”. This comes close to certificate or credential-based access control approaches (like in policy-based trust management), where a user is trustworthy and access is granted, if and only if the user presents the necessary credentials. Since trust is assumed to have several levels [AR04], binary models are considered to be insufficient. Trust can also be represented by more than two discrete values using either discrete or continuous numbers or labels. Regarding whether a trust model is designed for human users or software agents there can be arguments for all of these three representations. Discrete numbers, e.g., in a set of 1 to 10, and labels, can easily be assigned and understood by human users, where as continuous numbers allow the use of well-founded mathematical models.

Furthermore, the representation of trust can differ in the *dimension*, i.e., the number of its parameters. One-dimensional representations only allow the trustworthiness of an entity to be expressed in a single parameter, e.g., the trustworthiness of an entity is assumed to be “10”. In contrast, multi-dimensional representations can also express further influencing factors on the trust value. For example, the trust models presented in [Jøs01, JI02, MMH02a, SS02a, HJS04b, TPJL06] provide measures that express the uncertainty, reliability or confidence that is associated with a trust value. These measures are usually associated with the number of the collected

evidence units and possibly additional parameters.

Another important aspect is the interpretation or the *semantics* of a trust value. The *semantics* of trust values can be in the following set: rating, ranking, probability, belief, and fuzzy logic.

- The trust values which are computed in *ranking* based models, e.g., [KSGM03, PBMW98, Lev04, ZL04], are not directly associated with a meaningful semantics, but only in a relative way, i.e., a higher value means higher trustworthiness.
- The trust values which are directly linked with a trust related semantics may be referred to as *ratings*. For example, on a scale of natural numbers in the interval [1, 4], 1 can be linked to “very untrusted”,..., and 4 to “very trusted”. In current trust models [ARH00, Gol05] ratings are used to represent trust in a way which is easily accessible to humans.
- If trust is modeled as *probability*, the trust value expresses the probability that an entity will behave as expected. This comes closest to the definition introduced in Section 2.1.1. A critical issue when modeling trust as probability is how to derive trust from the collected evidence. Besides Bayesian or maximum likelihood based approaches, one finds mechanisms that propose their own approaches for deriving a probability from the collected evidence.
- Trust can be expressed using a *belief* based model. The major advantage of this approach is that it directly allows for expressing the uncertainty that is associated to a trust value. Details of this approach are introduced with the description of subjective logic in Section 3.2.5.1.
- Trust models based on *fuzzy logic* introduce their own semantics to the calculated trust values based on membership functions. In contrast to probabilistic models, trust is formally not treated as the subjective probability that an agent will behave as expected in the next encounter, but the interpretation of a fuzzy value like ‘very good’ is instead left up to the user or agent. Since fuzzy values are allowed to overlap, this also introduces a notion of fuzziness, i.e., an agent can be, for example, ‘good’ and ‘very good’ at the same time to a certain degree.

An aspect that is closely related with the semantics of the trust values is the *establishment of trust*. While label-based approaches are suitable when trust is directly assessed by human users, evidence based trust models need to define their own relation between evidence and the derived trust value. Here, it is also interesting whether the models consider further information, e.g., context-dependent information, like the user’s dispositional trust, or the entities’ trustworthiness in a related context. Furthermore, the domain of the considered evidence may be interesting. A number of trust models,

e.g., [TPJL06, BLB04, MMH02a, JI02] focus on binary evidence, i.e., the evidence derived from an interaction is expected to be either positive or negative. Other approaches allow multinomial evidence [JH07] and continuous evidence [JLC08] to be considered.

The classification of the trust models by the semantics of their trust value is used in the next section as the basis for the introduction of a set of trust models.

3.2 Trust Models by Semantics of the Trust Value

Having introduced the general concepts of the computation and the representation of trust, the following presents a selected set of current state-of-the-art trust models. Besides the two commercial feedback or recommender models of eBay [eBa09b] and Amazon [Ama09], the trust models are sorted according to the classification of the semantics of their trust value.

3.2.1 Commercial Feedback and Recommender Models

Probably the most well-known rating systems are currently hosted by eBay [eBa09b] and Amazon [Ama09], although they are simple models regarding the computation of their scores, and they do not provide a trust value in the sense of the definitions in Section 2.1.1. They are introduced to show how evidence from past interactions is currently presented to human users in online environments. On eBay the feedback score supports users when they have to find a reliable seller, on Amazon the ratings support users when evaluating the quality of offered goods, e.g., books.

3.2.1.1 eBay Feedback Forum

The eBay feedback forum [eBa09b] allows users of the auction platform to overcome the problem of missing traditional cues for assessing the trustworthiness of the seller or the quality of the offered goods - “The key to eBay’s success is trust” [eBa09a].

The basic idea is simple. After each transaction buyer and seller can give each other a rating, stating if they were satisfied with the transaction. The effect of the feedback forum is twofold.

For sellers, a high number of good transactions leads to a high feedback score or high reputation. Studies by [RZSL06] have shown that a high feedback score is rewarded, since well established sellers with a high reputation rating can achieve higher prices than new sellers with a low reputation rating. Thus, the feedback score helps to achieve accountable behavior of the sellers.

Additionally, it helps buyers to assess the trustworthiness of a seller, as a seller with a high feedback score can be supposed to deliver its offered goods according to the announced quality.

The details of the model (cf. [JIB07, RZ02, eBa]) are simple, too. Buyers and sellers are allowed to rate each other after each transaction. In general, this feedback can either be positive (+1), neutral (0) or negative (-1)¹. Furthermore, it may contain some additional free text information. Each user profile shows a feedback score as the total sum of feedback provided and the percentage of positive feedback in relation to the total feedback. Interpreting the sum of feedback alone can lead to misleading interpretations, since 100 positive and 0 negative ratings end in the same sum as 300 positive and 200 negative ratings. The percentage of positive values helps to overcome this problem. The interpretation of the percentage alone can again be misleading, as a high percentage not necessarily means a high number of total interactions. Therefore, the user has always to interpret both value together.

In the end, the decision whether a score of 80 and 99% positive ratings is trustworthy, or if another seller with a score of 5000 and 84% positive ratings is to be preferred, is left up to the user. The system does not provide a representation which integrates both values in a single measure.

In the case a user wants to see more details, they get information about the ratings in the last month, last 6 months, and last year, as well as short messages which the raters can provide with their rating. Furthermore, there are ratings for different aspects in the context of the transaction, e.g., shipping time.

Besides the feedback score, there are additional possibilities, which may help to assess the trustworthiness of a seller, e.g., a seller might be marked as a “power seller”²

3.2.1.2 Amazon Review Scheme

The review scheme provided by Amazon [Ama09] is not a trust model in the narrower sense. Like the rating system of eBay, it is well-known and it tries to support users in their decision making.

The idea is that users share their experience in order to support others, e.g., when evaluating whether a book is worth reading (and thus buying) or not. Therefore, users may provide reviews for articles which are sold on the Amazon website. The review may contain a rating on a range from one to five stars and additional text information.

Together with each article, its aggregated rating is presented, i.e., the average of the provided ratings and the number of contributing ratings.

¹A seller may only give positive ratings to the buyer.

²A user can become a PowerSeller when they fulfill the following requirements: “consistent sales volume, 98% total positive Feedback, eBay marketplace policy compliance, an account in good financial standing, and beginning in July 2008, detailed seller rating (DSRs) of 4.5 or higher in all four DSRs - item as described, communication, shipping time, and shipping and handling charges. If a seller no longer complies with any one of the above requirements, they are removed from the program.” [eBa09c]

Additional information is offered when the user clicks on the average rating with their mouse.

Like the eBay feedback forum the Amazon rating scheme does not provide a representation that aggregates the collected information in a single value. Furthermore, both systems do not calculate personalized ratings.

3.2.2 Ranking Approach

Typical approaches in this category are presented in [KSGM03, PBMW98, Lev04, ZL04]. As the semantics of ranking based trust values only states that a higher value is better, the trust value itself does not convey a real meaning. Thus, although ranking based approaches allow to identify the entity with the highest trust value, they are not directly applicable when an entity has to make a decision that is associated to risk, as the ranking does not allow to state “how good” the best entity is. Furthermore, the approaches in [KSGM03, PBMW98, Lev04] calculate global trust value, which does not fit to the proposed definitions of trust. For both reasons, these approaches are not further evaluated.

3.2.3 Rating Approach

In rating based approaches each trust value is associated with a certainty semantics, e.g., if 10 is the highest trust value, it can be associated with the semantics “high trust” [Gol05]. Thus, rating based approaches try to provide a simple access to the trust values for human users.

3.2.3.1 TidalTrust

In [Gol05], Golbeck provides a trust model that 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 [GH06] with about 400 users. In this network, the users can rate movies. Furthermore, one can rate other users, i.e., 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” [Gol05].

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

$$r_{sm} = \frac{\sum_{i \in S} t_{si} \cdot r_{im}}{\sum_{i \in S} t_{si}}, \quad (3.1)$$

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 maximum chain length or recursion depth can be limited. Based on the assumption that the opinions of the most trusted friends are the most similar to the 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 [Gol05] shows that it produces a relatively high accuracy, i.e. the ratings based on recommendation are close to the real ratings of the user. Yet, this approach does not allow one to state the reliability or the confidence that is associated with a rating. Furthermore, the approach does not deal with any form of decision making.

3.2.3.2 Model proposed by Abdul-Rahman and Hailes

The trust model presented by Abdul-Rahman and Hailes [ARH00] 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 [Gam90].

The model deals with direct trust and recommender trust. Direct trust is the trust of an agent in another agent 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 computed by discretely 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” experience 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 become 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.

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

3.2.4 Probabilistic Approach

In probabilistic approaches, trust values are represented in the range of $[0, 1]$, which allows these values to be interpreted as probabilities. Yet, those approaches still differ in the way the trust values, i.e., the probabilities, are computed.

3.2.4.1 UniTEC - A Generic Approach

In [KR03,KBR05], Kinateder et al. introduce a generic trust model. This approach identifies five relevant aspects for modeling trust between entities.

1. **Trust measure:** It describes the quality of the trust relationship ranging from complete distrust over neutral to complete trust.
2. **Trust certainty:** It describes the confidence of the trustor in her estimation of the trustee, i.e., the trust measure.
3. **Trust context:** The context in which entity A trusts in entity B , e.g., file sharing or online banking.
4. **Trust directness:** There are direct and indirect trust relationships. Direct trust refers to the trust in the context of providing a service. Indirect trust refers to trust in a recommender, i.e., in the recommendations about service providers in a certain context.
5. **Trust dynamics:** Trust is not static, it may dynamically change when new evidence is available or over time.

The trust measure is calculated based on an update rule that seems to be defined in an *ad hoc* manner. Yet, the trust measure is interpreted as probability. In [KBR05], it is proposed to calculate the certainty of the trust measure in similar manner as in ReGreT [SS02a,Sab03], yet, it is not stated whether and how the certainty parameter is to influence the trust measure. The approach is capable of considering that an entity's behavior

may dynamically change using two concepts called trust aging and trust fading.

Moreover, the approach allows trust to be transferred between similar contexts. This is done based on a “weighted direct trust context graph” [KBR05]. The weights between the different contexts, e.g., ‘cars’ and ‘limousines’, allow to define how trust and evidence that is collected in one context should be weighted when transferred to the other one.

Besides the trust model itself, the most interesting issue of the generic approach is that Kinatader et al. show how the representational models for trust presented in [ARH00,JI02,Sab03] can be integrated in this general model. For each model a mapping is provided introducing additional assumptions when necessary.

Although the inclusion of recommendations in the computation of trust values has been described, it is not evaluated. The evaluation in [KBR05] compares the performance of the UniTEC trust measure with others model only based on direct evidence.

3.2.4.2 Bayesian Trust Models

Bayesian trust or reputation models have been proposed, e.g., in [TPJL06, BLB04, MMH02a, JI02, QHC06, WJI05]. In the following, first, the representational model of Bayesian trust models is introduced as a common basis for these models. Afterwards, the differences regarding the computational models are described.

Representational Model

In Bayesian trust models, trust is interpreted as a subjective probability. This means that the expectation value, which is usually used to determine the trust value of an entity, is not only calculated based on the evidence that is available about an entity, but it also considers subjective prior knowledge, i.e., information that is independent from the analyzed data [Bol04]. The consideration of the prior knowledge provides the user with means to integrate their dispositional trust in the model, which is considered to be an advantage over maximum likelihood based trust models as proposed in [DA04, DA05]. For a formal comparison of a maximum likelihood based model to a Bayesian one see [NKS07].

In Bayesian trust models, the subjective probability for positive outcome in the next interaction is modeled as a random variable p using a beta probability density function h . The function is defined by the two parameters α , β :

$$h(p \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad (3.2)$$

where $0 \leq p \leq 1$, $\alpha > 0$, $\beta > 0$.

By defining $\alpha = r + r_0$ and $\beta = s + r_0$, it is possible to relate the probability function directly to the collected evidence, where r and s represent the number of positive and negative evidence units, and r_0 and s_0 define the prior knowledge. The expectation value is defined as $E = \frac{\alpha}{\alpha + \beta}$. The mathematical foundations of the Bayesian approach are described, e.g., in [Bo104].

In the approaches proposed in [TPJL06, BLB04, MMH02a, JI02], the prior information is used to statically initialize the model with a uniform distribution ($r_0 = s_0 = 1$). Although, this may be appropriate for a wide range of applications, it prevents the user from adjusting these parameters according to their dispositional trust or from adjusting the prior knowledge dynamically.

Assuming $r_0 = s_0 = 1$, the information of an entity A about the trustworthiness of an entity B can be expressed based on the aggregated positive and negative evidence. In the following, this is referred to as the opinion of entity A about (the trustworthiness of) entity B and denoted as $\sigma_B^A = (r, s)$.

Whenever new evidence is available, it can be integrated by updating the value of r , if there is new positive evidence, and the value of s in the case of negative evidence. For example, if $\sigma_B^A = (10, 3)$ and there is an additional piece of positive evidence available, the updated opinion is equivalent to $\sigma_B^A = (11, 3)$. Thus, it is not necessary to store the evidence per interaction per entity, as an entity only needs the information about the aggregated numbers of positive and negative evidence units. This is done in [BLB04, MMH02a, JI02], only TRAVOS [TPJL06] needs to store the evidence per interaction per entity for other reasons.

Although the approach is originally designed to only allow for binary feedback, that is either positive or negative feedback, it has also been shown how continuous feedback can be integrated in [JI02].

Furthermore, the trust models proposed in [BLB04, WJI05] are also capable of weighting evidence according to its recentness, which is referred to as *aging* of evidence in this thesis - in other work, the corresponding concept is called *forgetting* [JI02], or *longevity* [JHF03, Jøs07, JLC08]. In [JHF03], it has been shown that the aging of evidence has positive effects in an e-Market scenario. In contrast to the approach proposed by Buchegger et al. [BLB04], the approach proposed in [JHF03, WJI05] has the feature that in absence of evidence the expectation value moves back towards the initial expectation value that is defined by the prior knowledge. However, the latter group of models suffers from an effect that may be undesired. The aging as introduced in [JI02, JHF03, WJI05] can be shown to limit the number of evidence units $r + s$ that an entity can collect to a finite number. This leads to a narrowing

of the range of the values of the expectation value (see Section 5.2.3.1). The concept of a dynamic base rate, which has been proposed in [JLC08] allows to shift the (narrowed) interval depending on the dynamical base rate, which may soften the impact of this restriction.

Another disadvantage of current Bayesian trust models can be seen in the lack of suitable representation for human users. The representation misses an easily understandable set of parameters, which, e.g., allows for a simple explanation how the final trust value is influenced by the different parameters, and graphical representation. Only for the Beta Reputation System [JI02], which is based on “subjective logic”, a mapping to a more intuitive representation is presented (see Section 3.2.5).

Computational Models

Although Bayesian trust model have a common representation of trust and use the same mechanism for deriving trust from evidence, they provide different means to integrate direct evidence and recommendations, which are inspected in the following.

- **Beta Reputation System [JI02]:** The Beta Reputation System has been proposed as a centralized reputation system, but may also be applied to distributed environments. The ratings of all entities per interaction partner are aggregated by a so-called *reputation centre*. The model allows different weights to be given to a rating based on the value of the interaction.

Furthermore, it allows the impact of a rating to be weighted based on the trustworthiness of the entity that provides the rating. Hereby, the trustworthiness in providing ratings is assumed to be equal to the trustworthiness in providing interactions.

Finally, a number of extensions have been proposed: a filtering mechanism for misleading recommendations [WJI05], an extension for multinomial ratings based on the Dirichlet distribution [JH07, Jøs07], an extension for continuous ratings also based on the Dirichlet distribution and fuzzy set membership functions [JLC08], a community based update mechanism for the so-called *relative atomicity* or *base rate* [Jøs07, JLC08].

- **Mui et al. [MMH02b, MMH02a, MMA⁺01]:** The approach presented by Mui et al. supports a decentralized computation of trust. It provides means for deriving trust from recommendations [MMH02a]. However, the proposed method for the aggregation either considers recommendations only in absence of direct evidence [MMH02b] or it does not differentiate between direct evidence and recommendations. In [MMH02b], the authors also introduced an idea of evaluating the

reputation of groups of agents, which is used to evaluate the trustworthiness of an agent in absence of direct evidence. Furthermore, the approach introduces a measure for the reliability of a trust estimate. The measure depends on the number of evidence units that is necessary in order to reach a certain level of confidence and the number of collected evidence units. The parameter is used for weighting recommendations. The evaluation in [MMH02b] shows the performance of this approach a prisoner's dilemma game.

- **Buchegger et al. [BLB03, BLB04]:** Buchegger et al. provide a distributed trust model for peer-to-peer scenarios and mobile ad hoc networks. Recommendations by other entities are considered as long as they are similar to an entity's direct experience. This is done based on a so-called "deviation test" that calculates the absolute difference between the expectation value based on direct experience and the expectation value of the recommendation, in the case the difference is less than a pre-defined threshold ($d = 0.5$ in [BLB03]) the recommendation is considered. Furthermore, it is possible to reduce the influence of recommendations using a static weighting factor. The approach allows the entities to be classified regarding their behavior as interactors, as well as regarding their behavior as recommenders.
- **TRAVOS [TPJL06]:** Teacy et al. proposed their trust model for open dynamic systems. The approach tries to provide means to cope with inaccurate recommendation sources. Their computational model allows to weight the recommendations based on the accuracy of the past recommendations of each recommender. Especially, when an entity A evaluates the trustworthiness of an entity B , and A considers the recommendation of an entity C , then, the trustworthiness of C in the context of providing recommendations is only derived from the accuracy of C 's previous recommendations about the trustworthiness of B . The accuracy of C 's recommendations about other entities is not considered. Therefore, in TRAVOS each entity stores not only the evidence about the outcome of the past interactions per interactor (as introduced above), but also the recommendations per recommender per interaction partner. In an environment with a large number of entities, this needs to be considered as it increases the required amount of storage. Furthermore, the approach stops to evaluate further recommendations when a certain level of confidence is reached based on the direct evidence and the recommendations.

3.2.5 Belief Approach

In contrast to the probabilistic approach the belief approaches directly integrate a notion of uncertainty as shown in the following.

3.2.5.1 Subjective Logic

In [Jøs01], Jøsang presented “subjective logic”. It allows elements of Bayesian probability theory (referred to as ‘evidence space’) to be combined with a belief approach (referred to as ‘opinion space’). This approach can be used to model trust and serves as a basis for the Beta Reputation System [JI02].

In evidence space, which tempts find a simple intuitive representation of uncertain probabilities, an opinion can be expressed as a triple $o = (b, d, u)$, where $b \in [0; 1]$ represents the belief, $d \in [0; 1]$ the disbelief, and $u \in [0; 1]$ the uncertainty about a certain statement [Jøs99a]. The three parameters are interrelated by the equation $b + d + u = 1$. The expectation value $E(o)$ of an opinion o is defined as:

$$E(o) = E(b, d, u) = b + \frac{u}{2} \quad (3.3)$$

In [Jøs01], a parameter called relative atomicity a has been added. The expectation value then is calculated using $E(b, d, u, a) = b + u \cdot a$. Thus, the relative atomicity allows one to define the expectation value in absence of belief and disbelief.

Furthermore, Jøsang provides a graphical representation for the belief model (see Fig 3.3). The point A denotes an opinion with $(b, d, u) = (0.2, 0.1, 0.7)$. The point B denotes an opinion with $(b, d, u) = (0.2, 0.6, 0.2)$.

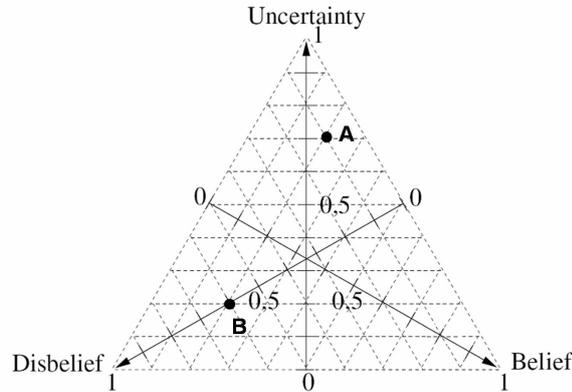


Figure 3.3: Opinion Triangle

Mapping between ‘opinion space’ and ‘evidence space’

In order to be able to benefit from the advantages of both representations, Jøsang provides a mapping between the Bayesian approach and the belief approach by defining the following equations. The opinion $o = (r, s)$ and $o = (b, d, u)$ are supposed to be equivalent if the following equations are true:

$$\begin{aligned}
b &= \frac{r}{r+s+2} , \\
d &= \frac{s}{r+s+2} , \\
u &= \frac{2}{r+s+2} \quad \text{where } u \neq 0 .
\end{aligned}
\tag{3.4}$$

This mapping allows an opinion to be transferred from the evidence space to the opinion space, e.g., in order to evaluate its uncertainty. The mapping can also be used the other way around. The expectation value is invariant to the representation. The mapping provided in [Jøs01] also takes the relative atomicity into account. In [JBXC08], it is shown how the mapping between the evidence space and the opinion space can be defined when considering multinomial ratings and a flexible choice of the parameters reflecting prior knowledge.

Operators

So far only the representation of an opinion has been described. In order to evaluate trust based on recommendations, Jøsang defines operators for combining opinions (consensus) and weighting recommendations (discounting) opinions. In contrast to the belief model presented in [YS02] the consensus operator is not based on Dempster's rule. The reason is that Dempster's rule may lead to counter-intuitive results [Jøs01]. It is worth noting that the model supports additional operators, e.g., for propositional conjunction, disjunction and negation. For the interpretation and justification of these operators see [Jøs97, Jøs01, JI02].

Finally, in [JBXC08] a new operator is introduced that allows an entity A to derive the trustworthiness of an entity B based on the similarity between their opinions about another entity C .

Subjective Logic as Basis for Trust and Reputation Models

Subjective logic provides a general basis for modeling trust based on evidence. It may be used in a wide range of applications. In [Jøs99a] 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 apply it in trust based decision making in electronic commerce [Jøs99b] and how the approach can be integrated in policy based trust management [JGA06]. The Beta Reputation System presented in [JI02] is also based on "subjective logic".

The approach of modeling trust in the SECURE Project [CSG⁺03, CNS03] is based on similar ideas - deriving trust from evidence and describing trust as belief. Yet, there are major differences in calculating trust. Another belief based approach to model trust is presented in [YS02].

3.2.6 Fuzzy Logic Approach

Trust models based on *fuzzy logic* introduce their own semantics to the calculated trust values based on membership functions. The final interpretation of a fuzzy value like 'very good' is left up to the user or agent.

3.2.6.1 ReGreT

ReGreT models trust for small and mid-size environments in electronic commerce [SS05]. The system is described in detail in [SS01, SS02a, SS02b, Sab03]. 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 (outcome reputation) is based on self-made experience of an agent. The trust values are called direct trust or outcome reputation. The calculation of the trust value considers the age of evidence, giving more relevance to more recent outcomes. However, in [HJS06], Huynh et al. show an effect of ReGreT's aging function that may not be desired.

The social dimension is based on recommendations (witness reputation), the social relationships between agents (neighborhood reputation), and the social role of the agents (system reputation). In order to avoid correlated evidence, it is proposed to select the recommenders based on a domain dependent sociogram, yet, as stated in [HJS06] it is not clear how the social network is built. Furthermore, it is proposed to weight recommendations based on social trust and outcome trust reputation.

Finally, the ontological dimension helps to transfer trust information between related contexts and aspects [SS01].

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.

Furthermore, for all trust values a measurement of reliability is introduced that depends on the number of past experience and expected experience (*intimate* level of interaction), and the variability of the ratings.

The inference of trustworthiness is supported by 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. The trust model that is proposed in [HJS04b, HJS04a, HJS06] seems to be similar to ReGreT. It additionally considers certified reputation information.

3.3 Analysis of the Trust Models

From the presentation of the different trust models one can see that current approaches address a wide range of aspects. The following provides an analysis of the different features that are supported by the different trust models.

3.3.1 Analysis of the Representational Models

As stated before, *ranking* approaches do not seem to be appropriate for modeling trust in ubiquitous computing as they do usually not calculate subjective trust values and it is unclear how they can be applied in the decision making process when the utility of an interaction needs to be considered.

The *rating* approaches have the advantage that they are based on simple trust values that easily can be assessed by human users, but it seems unclear how additional parameters, like a user's dispositional trust or the amount of information on which a trust value is based, can be integrated.

The *probabilistic* trust models have the advantage that they can easily be integrated in utility-based decision making as shown in Section 2.3. Within the probabilistic approach, the trust models that are based on the Bayesian approach are especially suitable. They allow trust to be calculated as a subjective probability since the expectation value is based on the collected evidence as well as on subjective prior knowledge. This goes along with the definition of trust provided in Section 2.1.1 and allows the integration of a component of dispositional trust.

The *belief* approach seems to be an interesting approach, too, as it allows to express the uncertainty of an opinion. Furthermore, a graphical representation has been introduced with "subjective logic". However, this representation is not very appropriate for intuitive usage by human users (as shown in the evaluation in Section 6.2).

The *fuzzy logic* based approaches seem to be close to the probabilistic approaches. Here, membership functions allows to introduce additional semantics.

In the following, the focus will be on trust models that follow a Bayesian approach as

- it allows trust to be modeled as subjective probability.
- it is based on solid mathematical foundations for deriving trust based on evidence and subjective prior knowledge.

- it may be easily integrated in utility-based decision making.

Furthermore, there has already been much work on Bayesian trust models showing how different parameters can be integrated, yet, there are still some shortcomings:

- The *dispositional trust* of a user can basically be integrated in the prior information. A shortcoming of the trust models within the Bayesian approach is that most trust models [TPJL06, MMH02a, JI02, BLB04] do not support the user to influence this dispositional trust component as the prior knowledge is set to $r_0 = s_0 = 1$ by the developers of these models.
- The weighting of evidence according to its recentness, which is referred to as *aging*, forgetting, or longevity, is considered in the approaches presented in [BLB04, JI02, Jøs07, WJI05, SS02a, HJS06]. Yet, they suffer from the limitation that either in absence of evidence the expectation value remains constant, or that the range of the expectation value is limited (see Section 5.2.3.1). The latter restriction can be softened using the concept of a dynamic base rate [JLC08] that allows the interval of reachable values to be shifted.
- A parameter that expresses the confidence or the certainty that is associated with the calculated trust value has already been introduced in [TPJL06, MMH02a]. In both trust models, this parameter is related to the number of evidence units which an opinion is based on. The model proposed in [MMH02a] computes the reliability of an opinion based on the number of collected evidence units and the number of evidence units that is necessary in order to reach a certain level of confidence. Similar measures are introduced in fuzzy logic approaches presented in [SS02a, HJS06].

Furthermore, an update mechanism of the models dispositional trust component based on the typical behavior of the entities within a community has been proposed in [Jøs07].

It has to be stated that although the Bayesian trust models have their strengths, they also have a major disadvantage. They do not provide an interface that has been designed for human users. According to Dingedine et al. [DFM00] it is important that trust models provide intuitive parameters which allow for an easy interpretation of the trust value by the user. This is important in order to provide the user with a feeling of being in control of the system, and it is especially necessary in order to support users in integrating their “real world” knowledge in the system. Here, it is the goal of this thesis to provide an intuitive graphical representation of trust for human users. Graphical representations are only provided using a star based

interface in [GH06, Ama09] and the belief representation of “subjective logic”. As the stars interfaces are very simple and the belief representation seems to be rather complex, it needs to be evaluated whether the advantages of both approaches could be combined. The aspect that the Bayesian approach of modeling trust is actually limited to binary evidence, i.e., the outcome of an interaction is supposed to be either positive or negative, is not seen as a major disadvantage, as this can be assumed to be sufficient for a wide range of applications and it reduces the subjective influence when rating the quality of an interaction. Furthermore, it has been shown in [JI02] how continuous ratings can be integrated in this approach.

Finally, it needs to be stated that some of the trust models, e.g. [BHOC07, SS02a, KR03], provide means in order to derive trust also from similar contexts. In this thesis, it is not considered to transfer trust between contexts, as a sound model for deriving trust from evidence and the parameters introduced above within a single context is considered to be an important and challenging step, which needs to be evaluated before transferring trust between contexts.

3.3.2 Analysis of the Computational Models

Regarding the *computational models*, the following features seem to be important for the application in a distributed and unmanaged environment as ubiquitous computing. In the following, the focus is on Bayesian models as they provide a solid way for deriving trust from evidence and allow one to integrate context-dependent parameters as a user’s dispositional trust.

When applying a trust model in a distributed environment, the integration of direct evidence and recommendations is especially important for two reasons. First, whenever direct evidence is rare, recommendations can provide a solid basis for the evaluation of the trustworthiness of an entity. Second, as recommendations allow foreign parties to take influence on an entity’s decision making process, the selection of the recommenders and the weighting of the influence of their recommendations have to be done carefully.

The critical issue with recommendations is that recommenders may intentionally or accidentally provide misleading recommendations. According to [JIB07, TPJL06] there are two approaches for dealing with recommendations in order to reduce the impact of misleading recommendations. They are called *endogenous* and *exogenous* filtering or discounting of recommendations.

1. The basic idea of *endogenous* handling of recommendations is that one can reduce the impact of misleading recommendations by considering the provided recommendations independent from the recommenders. This approach usually assumes that the majority of recommendations is accurate, and that misleading recommendations can be identified due to the statistical properties of the provided recommendations [TPJL06]. Approaches that fall in this class are presented in [QHC06, WJI05,

Del00].

2. *Exogenous* approaches consider additional information, e.g., the trustworthiness of a recommender. The trustworthiness of a recommender can be calculated based on different assumptions. In [JI02] the trustworthiness of a recommender is calculated using the assumption that an entity's behavior as recommender is the same as its behavior as interaction partner. In ReGret [SS02a], the assumption is the same, but additionally a social trust component may be considered. In [BLB04] recommendations are only considered if the expectation value of the recommendation (Bayesian approach) is close to the expectation value calculated based on the direct evidence.

The assumption that the majority of recommenders is honest may not be true in many cases. Therefore, it is considered to be not sufficient. Exogenous filtering of recommendations allows for a wide set of arrangements in order to reduce the impact of misleading recommendations.

In the exogenous approach the following mechanisms have been identified.

- *Filtering - exclusion of bad recommenders:* Probably the most simple variant of limiting the influence of bad recommenders is to exclude recommenders that are known to provide misleading recommendations. The approach presented in [BLB04] considers recommendations only if they are similar to the direct evidence based on a deviation test. Yet, the value threshold for this deviation test is chosen very tolerant (threshold $d = 0.5$ in [BLB03]). In contrast, the approaches presented in [JI02, SS02a, MMH02b, TPJL06] do not exclude recommendations by bad recommenders. Usually those approaches consider bad recommenders giving a low weight to their recommendations.
- *Weighting and limiting the influence of each recommender:* These mechanisms allow to limit the influence of a recommender by weighting the provided recommendations. The weighting of recommendations is also referred to as discounting. In general, a model has the following options for the discounting recommendations.
 - *No discounting:* The models do not weight recommendations at all. This implies that all recommendations have the same weight as direct experience. This is done in the most basic variant of the Beta Reputation System [JI02].
 - *Static discounting:* Static discounting means that the recommendations by each recommender are weighted independent from the recommender. The idea is to give a relatively higher weight to direct experience than to recommendations. An example can be found in [BLB04].

- *Dynamic discounting*: The intension of dynamic discounting is to give a higher weight to recommendations which are provided by recommenders that are considered to be more trustworthy than others. There are currently two approaches:
 - * Weighting based on the *behavior as interactor*: This approach assumes that the behavior of an entity as interaction partner is the same as the behavior when providing recommendation. This approach, implemented in [JI02,SS02a], has the shortcoming that the assumption may be wrong, e.g., an entity that offers a high quality service may interested in providing bad recommendations about competing service providers.
 - * Weighting based on the *behavior as recommender*: This approach weights recommendations according to a recommender's behavior in the past, giving a high weight to recommenders which provided a high number of accurate recommendations in the past, and low weight to recommenders providing less accurate recommendations, e.g., [TPJL06]. This is what comes closest to the intention of weighting recommendations. Yet, the approach comes with the cost of evaluating the accuracy of a recommendation and storing this information.

For all approaches of discounting, it is essential that intended effect cannot simply be circumvented by the recommender by providing a specially prepared recommendation.

Finally, the direct evidence and recommendations have to be aggregated. Within the considered approaches, the following mechanisms have been identified:

- *No aggregation*: The approach presented in [MMH02a] does not aggregate direct evidence with recommendations. Recommendations are only considered in complete absence of direct evidence. This seems to be not sufficient as recommendations are excluded as soon as a single piece of direct evidence is available.
- *Considering recommendations independent from the confidence in the direct evidence*: The approaches presented in [BLB04] provide means for statically weighting recommenders and for the exclusion of recommenders. However, the approach does not provide means for limiting the impact of recommendations when sufficient direct evidence is available.
- *Considering the confidence in the direct evidence and recommendations*: The approach presented in [SS02a] reduces the impact of recommendations when the reliability of the direct evidence increases. This is a

good choice, in the face of misleading recommendations and interaction partners possibly showing different behavior to different entities. TRAVOS [TPJL06] takes recommendations only into account as long as the aggregated direct evidence and the evidence of the recommenders considered so far does not reach a predefined level of confidence.

- *Considering the trustworthiness and the rank of recommenders:* The basic concept of this approach is to consider the best recommenders first and limit the influence of each recommender based on its trustworthiness and the number of recommenders that have been previously considered, i.e., based on their rank. This approach has not been proposed so far in Bayesian trust models, to the best of the author’s knowledge; yet, it seems to be the next step. In combination with the previous steps, this approach can be shown to cope with so-called Sybil attacks [Dou02] (see Section 5.3.4). This kind of attack allows an attacker to create an arbitrary number of seemingly independent entities, which may be used to multiply the attacker’s influence by creating an arbitrary number of recommenders providing misleading recommendations. Other trust models, e.g., the approach by Buchegger et al. [BLB04] try to address this attack relying on expensive pseudonyms that finally require some kind of access control, which is external to the trust model.

3.4 Conclusions

This chapter presented a survey of state-of-the-art trust models. It introduced a classification of trust model according to the semantics of their trust value and proposed to distinguish between the representational and the computational aspects of a trust model.

Based on the analysis of the representational models the Bayesian based trust models seem to be very promising as they allow trust to be derived based on evidence and subjective prior information. They also show how context-dependent parameters as dispositional trust and aging can be integrated.

Yet, current models do not use the full potentials of this approach. Most approaches treat the prior information to be static and aging is introduced with different limitations (see Section 5.2.3.1).

Furthermore, Bayesian approaches do currently not provide a representation that allows for intuitive use by humans. The belief based representation provided by “subjective logic” [Jøs01, Jøs99a] would be an approach to overcome this shortcoming as it also provides a mapping between the Bayesian representation and the belief representation. Furthermore, it supports a graphical representation. Yet, the author of this thesis aims to provide a representation that better fits the needs of the users.

Current computational models that use an *exogenous* approach for filtering and weighting recommendations provide a good basis for dealing with

misleading recommendations. Yet, most approaches [JI02, SS02a, BLB04] either do not weight recommendations or the weighting is based on an entity's behavior as interaction partner. In contrast, the approach proposed in [TPJL06] weights recommendations according to the accuracy of the recommender's past recommendations. This seems to be a better choice, but the proposed approach requires to store the recommendations per recommender and per interaction partner. Finally, an extension of current approaches in order to consider a recommender's rank to improve the model's resistance to Sybil attacks has been identified as a next step.

Chapter 4

Concepts

As introduced in the previous chapter, the development of new techniques supporting entities in selecting trustworthy interaction partners is in the focus of this thesis. Based on [DFM00, JIB07], a set of high-level criteria is provided that will be considered when introducing the concepts of the trust model.

1. **Accurate for long-term performance:** The model needs to be capable of expressing the confidence of the provided trust value. It must be capable of differentiating between known and unknown entities, especially between entities that are unknown and those who have a low trust or reputation value because of bad long term performance.
2. **Weighted toward current behavior:** The model is capable of reflecting recent trends of an entity's behavior.
3. **Robust against attacks:** The model should resist attacks on the manipulation of trust value as much as possible.
4. **Smoothness:** New evidence should have a limited impact on the change of the trust value.
5. **Understandable:** The parameters of the trust model should allow for easy interpretation. This is important to understand how a system works.

The following part of this chapter explains the main concepts that are necessary to use evidence based trust to achieve this goal. At first, a schematic description of the scenario introduced at the beginning of Chapter 2 is presented. It serves as a motivation for the concepts explained in this chapter. Second, the system model is presented. It provides the terms of the main building blocks of the model, the relations between them, and a set of basic assumptions that need to be fulfilled for the successful application of

an evidence based approach for trust is presented. Third, the author shows a selected set of application areas and how they meet these assumptions. Forth, the conceptual requirements for a trust model are derived. Fifth, in the conclusions of this chapter, the design goals for the trust model that is presented in Chapter 5 are summarized.

4.1 Motivating Scenario in a Schematic View

This section introduces a schematic view of the scenario that has been presented at the beginning of Chapter 2. In this scenario humans with mobile devices share music in an opportunistic network; it is inspired by an application called *musicClouds* [Hei07]. In the following, users and their mobile devices are referred to as entities. When these entities move around, they will meet other entities, and have the possibility to interact with each other, i.e., they exchange music (mp3 files) or information about the behavior of other entities (recommendations), in a spontaneous manner. When providing files, the entities may show different behaviors, e.g., providing mostly good or corrupted files, due to different goals or motivations. The goal of a typical user when selecting a candidate for an interaction is to select a trustworthy one, i.e., a candidate from which they expect to receive a good file (correct file, no viruses, complete song, and expected quality).

It is assumed that a user appreciates the support of a software component including a trust model, which supports them with information about the trustworthiness of the candidates available for an interaction, or which even is capable of making decisions and selecting an appropriate candidate on its own. This will be especially true, if it allows the quality of a user's interactions to increase, i.e., the number of received good files.

As opportunistic networks typically lack a central authority for the purposes of access control and identification of entities, the trustworthiness of entities is derived from evidence of past interactions. When users have knowledge about the capabilities of each other from real-life meetings (e.g., friends knowing that they will share files honestly) it is necessary to support that users can directly add this information to the system. Furthermore, as direct evidence may be rare, an important aspect is the robust integration of recommendations of other entities.

After an interaction, the *quality of the interaction* is determined by feedback. The generation of the feedback does not necessarily require user interaction. In some cases this can also be done automatically, e.g., by scanning the mp3 file for viruses, checking the size, the bit rate, and noise. This information can be used by the software component, i.e., the trust model, to create histories with evidence about interactors and recommenders, which are the basis for deriving the trustworthiness of corresponding entities.

4.2 System Model

Having introduced a motivating scenario, the next step is to formalize the basic building blocks of the system.

4.2.1 Entities and Interactions

The participants in the system are called *entities*. Entities can represent humans (also called *users*), their devices, autonomous software components or agents, an abstract service provider, or a web service.

When referring to a group of participants which take part in the system in a certain application context the term *community* is used.

Interactions are actions between entities. On an abstract level, an interaction can be the usage of a service or a capability that is offered by one entity to another. Concrete examples are downloading a file or buying goods offered by an online shop.

The focus of this thesis is on *uni-directional* interactions. A uni-directional interaction is an interaction, in which only one entity has an expectation about the outcome of the interaction. For example, in a file sharing scenario, only the entity that receives the file has an expectation about the quality of the expected file, e.g., bit-rate, noise, etc. The sender of the file does not have any expectations about the interaction. A *bi-directional* interaction is an interaction in which both entities have an expectation about the outcome. For example, entity *A* sells something to entity *B*. Here, in general entity *A* has an expectation about the provided service by entity *B*, and entity *B* has an expectation about receiving its money in return.

For dealing with bi-directional interactions within the developed model, a bi-directional interaction should be treated as two uni-directional interactions.

4.2.2 Roles

The next step is to define roles allowing for a description of the relationship between entities within the model (see Figure 4.1): *initiator* (*I*), *candidate* (*C*), and *recommender* (*R*). The initiator of an interaction is the entity that searches for an interaction partner in a certain application context. The candidates are the entities offering the possibility to interact in this context. The recommenders are the entities providing recommendations about the behavior of the candidates in this context.

4.2.3 Process View

The following provides an overview of the process of establishing trust between entities and selecting entities based on trust (see Figure 4.2). The process describes the steps the initiator performs when selecting a trustworthy candidate from a set of candidates and how, after an interaction, the feedback

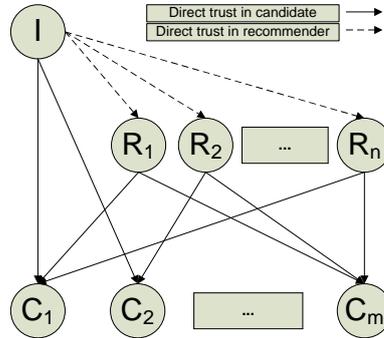


Figure 4.1: Simple scheme of the system model

on the quality of the interaction can be used to update the trust values of the initiator and the recommenders.

1. **Evaluation of the trustworthiness of the candidates:** This step includes the evaluation of direct evidence and recommendations about the candidates. The trust model, which is developed as the core of this thesis, is used to estimate the trustworthiness of each candidate based on this information.
2. **Evaluation of the utility of an interaction:** This step evaluates the utility of an interaction. This can be done by considering the expected costs and the expected benefits that are associated with the interaction by the initiator. This step is not considered in this thesis.
3. **Decision making:** In this step the initiator has to decide whether to interact at all, and if so, with which candidate. The decision whether to interact at all can be influenced by the expected utility of the interaction, the trustworthiness of the candidates and the risk attitude of the initiator. If those parameters are known they can be integrated in the decision making as shown in Section 2.3. The selection of an appropriate candidate may also be based on various strategies. In this thesis it is assumed that the initiator selects the candidate with the best trust value.
4. **Interaction:** When the initiator has decided to interact with the most trustworthy candidate, the interaction takes place.
5. **Evaluation and reporting of the quality of the interaction:** After an interaction, the *quality of the interaction*, and if necessary, the domain of the interaction, is evaluated and reported as feedback from the initiator. The generation of the feedback may require user interaction. Depending on the interaction, the provision of feedback may be based either on information by the user or on information

provided by an external software component capable of evaluating the quality of the interaction.

- 6. Update of opinions about the selected candidate and the recommenders:** In the last step, the provided feedback information is used to derive new evidence describing the behavior of the selected candidate and the recommenders. The new evidence is added to the evidence from past interactions, which allows to build a history from the behavior of an entity. In a distributed trust model, the initiator manages the histories of all the interactors and recommenders it has previously encountered by itself. Whenever, new evidence is available, the trust value of an entity may be updated. Furthermore, the initiator can use this step to update its expectation about unknown entities based on its experience with the encountered entities.

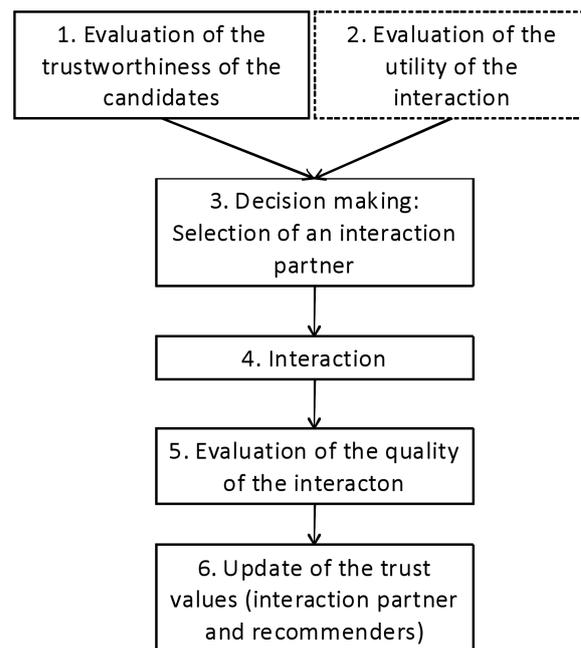


Figure 4.2: Main steps in establishing trust between entities and selecting entities

The steps that are considered in this process focus only on the trust establishment and the selection of a candidate. Before this process starts, one may need additional steps like finding candidates, i.e., potential interaction partners. Furthermore, one may specify an additional process for the distribution of recommendations.

4.2.4 Definition of Trust

The definition of trust for this thesis is derived from the definition of reliability trust provided by Jøsang et al. (see Section 2.1.1):

“Trust is the subjective probability by which entity expects that another entity performs a given action on which its welfare depends.”

It is worth noting that this thesis treats trust in the sense of *reliability trust* and not in the sense of *decision trust* (see Section 2.1.1), therefore the evaluation of trust is context-dependent, but not situation-dependent, e.g., trust is evaluated independent from the utility of an interaction.

Trust in Interactors: When an entity *A* assesses the trustworthiness of an entity *B* as interaction partner, then trust is the subjective probability with which entity *A* expects that the outcome of the next interaction in a certain context with entity *B* is positive.

Trust in Recommenders: When an entity *A* assesses the trustworthiness of an entity *B* regarding the provision of recommendations trust is the subjective probability with which entity *A* expects that the recommendation of entity *B* about another entity’s behavior as interactor in a certain context is accurate.

If an entity *A* receives a recommendation about another entity *B*, and the outcome of *A*’s interaction with entity *B* equates to the expectation based on the recommendation, then the recommendation is considered to be accurate.

4.2.5 Trust Establishment

The basis of trust is evidence derived from past interactions. There are basically two classes of evidence, i.e., *direct evidence* and *recommendations* (also called *indirect evidence*). Furthermore, parameters that are not directly assigned to each interaction partner or recommender, but depend on the application context (see next subsection) of the interaction or recommendation need to be considered. Figure 4.3 illustrates these basic influence factors.

If entity *A* has to estimate the trustworthiness of entity *B* and entity *A* has already interacted with entity *B*, and entity *A* remembers the evidence that it has derived from the outcome of these interactions, then entity *A* can use this evidence, called *direct evidence*, in order to estimate the trustworthiness of entity *B*.

Furthermore, an entity *A* may ask other entities, called *recommenders*, to provide information which helps estimate trustworthiness of entity *B*. This information is called *indirect evidence* or *recommendation*. The recommendation of a honest recommender is assumed to reflect the recommenders experience with the candidate. For example this is considered to be true, when the recommendation is equivalent to the direct evidence the recommender has derived from its previous interactions with the candidate.

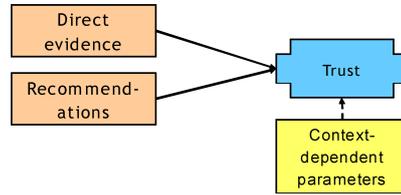


Figure 4.3: Basis of trust

Further sources of information can be the user knowledge from “real world” meetings and reputation information. The integration of user knowledge can be carried out by adjusting the direct evidence accordingly. The integration of reputation information can be done modeling the source of the reputation information as recommender, and the reputation information itself as recommendation. This allows one to focus on trust derived from direct evidence and recommendations in the following (see Figure 4.4).

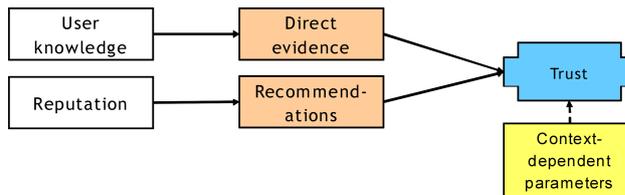


Figure 4.4: Basis of trust (extended)

4.2.5.1 Application Context

The application context describes the field in which the interaction takes place, e.g., exchange of mp3 files, selling fruits in an online shop, or provision of a web service for hotel reservations¹. In different application contexts, one has to deal with different conditions, e.g., regarding insurance, law enforcement, strength of identity, etc. This may lead to different preferences regarding how trust should be derived from the evidence from past interactions.

Furthermore, it is also important to note that within each application context C there are two sub-contexts (see Figure 4.5). The first sub-context $C(I)$ contains the interactions between entities that are in the focus of the main context. The second sub-context $C(R)$ contains the provision of recommendations about the behavior of other entities within the corresponding interaction sub-context. This is important, as the behavior of an entity when selected as interaction partner in a certain application context can differ for its behavior as recommender, e.g., as it might be interested in bad-mouthing other competitors.

¹In [JKD05], the term *scope* is used to refer to the domain or the purpose of the interaction.

Finally, it is worth noting that as evidence based trust models derive trust from evidence from past interactions, they may be assumed to work best, when the behavior of an entity, which goes along with its capabilities regarding the qualities of trust - benevolence, competence, honesty, and predictability - is assumed to be either stable or to change only slowly in each context.

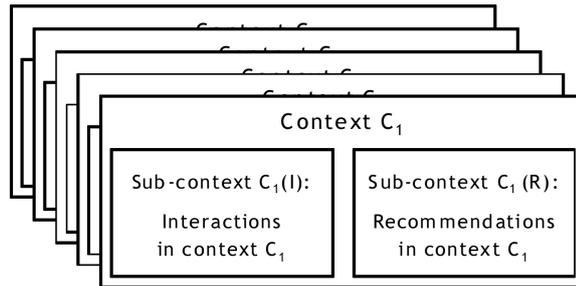


Figure 4.5: Context and sub-contexts

4.2.5.2 Context-Dependent Parameters

As introduced above, the basic idea of evidence based trust is to derive the trustworthiness of an entity based on the aggregated evidence from past interactions and recommendations. However, in different application contexts (see Section 4.2.5.1), such as sharing music files in an opportunistic network or buying goods in an online shop, it seems natural to consider context-dependent aspects that influence the relation between evidence and the derived level of trust. For example, in a file-sharing scenario, one might tend to give a high level of trust to unknown entities, but the same might be not true for online shops. The context-dependent parameters do not depend on the entity whose trustworthiness is evaluated, but only on the application context.

Based on the analysis of the state-of-the-art (see 3.3), the following aspects are addressed:

- **Expected behavior of unknown entities:** An entity's dispositional trust that is the general attitude of an entity to trust in others is used to express an entity's opinion on the expected behavior unknown entities. This opinion may change from application context to application context, e.g., depending on whether an entity expects that the majority of interaction partners to be benevolent or not, or depending on the degree of insurance that is associated with an interaction.

In the proposed model, dispositional trust is modeled using two parameters. The *base trust* value allows the expectation about the behavior of unknown entities to be defined. The parameter *weight* allows one to

influence how strongly new evidence influences the final trust value in relation to the base trust.

- **Dynamical change in an entity's behavior:** An entity's behavior can be expected to change over time, however, it may depend on the application context whether the behavior of the entities one encounters in this context is assumed to be rather stable or rather varying. For example, well-established online shops may be expected to show a rather stable behavior, whereas in a file sharing scenario the behavior of an entity may be expected to be not as continuous. Furthermore, it is necessary to consider that an entity, after building trust by providing a high number of positive interactions over a long time, may start to exploit the established trust long-lasting, when evidence from interactions far back in time has the same weight as evidence from recent interactions.

In this thesis, aging is introduced to take care of these aspects. It allows a higher weight to be given to evidence from more recent interactions as they may be more representative of an entity's future behavior and in order to prevent that an entity may exploit previously established trust in the long run. Furthermore, it allows one to consider that an entity's behavior may change over time, in the absence of interaction, by reducing the impact of older evidence. Positive effects of considering aging in trust models have been shown, e.g., in [JHF03].

- **Expected number of evidence units:** As a trust model is to reflect the certainty of its calculated trust values, a context-dependent parameter is introduced that allows the number of evidence units that is expected to be representative for an entity's behavior to be defined. The certainty of a trust value is to increase with the number of evidence units on which the trust value is based. When the maximum level of certainty is reached, the trust value of an entity will no longer depend on the dispositional trust, but solely on the evidence collected from past interactions, as it is expected to be representative.

There are two reasons for the introduction of this parameter. First, a user might decide that for a specific context, a certain number of evidence units is sufficient to believe that the collected information is representative for the future behavior of an entity. Second, in the face of aging, when modeled as presented in [JI02, JHF03, WJI05], one has to consider that even an infinite amount of observations from past interactions leads only to a finite amount of collected evidence (see Section 5.2.3). State-of-the-art approaches described in [JI02, JHF03, WJI05] do not take this into account (see Section 5.2.3.1).

4.2.5.3 Evaluation of Trust Chains

This section shows how recommendations will be propagated in given a network of recommenders (see also 3.1.1).

Figure 4.6 shows a simple example of a trust network. Here, entity A has direct evidence about the trustworthiness of the candidate C . Furthermore, entity A knows two recommenders R_1 and R_2 , who may provide recommendations about the trustworthiness of the candidate C based on their direct evidence. In this case, it is sufficient when entity A is provided with means for weighting recommendations based on the trustworthiness of the recommender and for aggregating its direct evidence with the weighted recommendations.

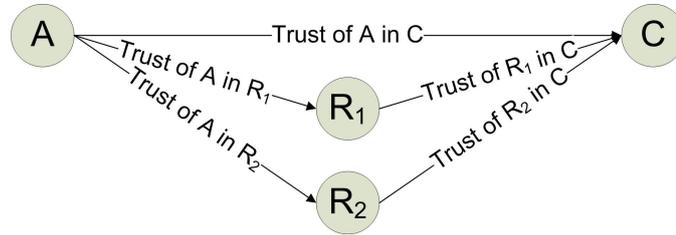


Figure 4.6: Simple trust network

Yet, when given more complex networks, it may be questionable if these means are sufficient for entity A to evaluate the trustworthiness of candidate C . Figure 4.7 shows a trust network in which entity A has direct evidence about the candidate C and knows two recommenders R_1 and R_2 . In this example, the recommender R_1 does not have direct evidence, but R_1 may forward the recommendation provided by R_3 . Furthermore, R_2 has direct evidence about the trustworthiness of C and may additionally forward the recommendations by R_3 and R_4 . As in a distributed environment, entity A does not know the structure of this network per se, this leads to two fundamental design choices regarding distribution of recommendations:

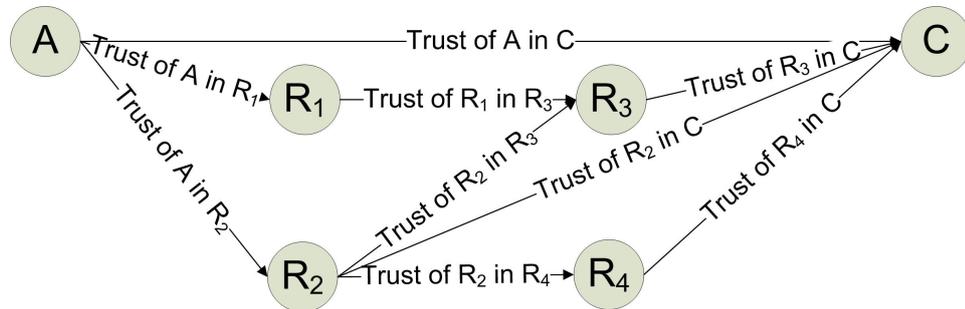


Figure 4.7: More complex trust network

1. Each recommender R forwards the single pieces of information that it

has collected and its information about the provider of these pieces of information.

2. Each recommender R forwards only a single trust value, which aggregates R 's direct evidence with the (weighted) recommendations by R 's recommenders.

In the first approach, each entity would forward its direct evidence and the collected recommendations together with the information about the origin of the recommendations and the information about the trustworthiness of the origin of recommendations. In the example, this would mean that the entities R_3 and R_4 would report their direct evidence about C together with the information that this is direct evidence. In the next step, the recommender R_2 would forward the recommendations from R_3 and R_4 to entity A . Furthermore, entity R_2 would forward the information about the trustworthiness of R_3 and R_4 and its direct evidence about the trustworthiness of C . When entity A has collected the information from R_1 and R_2 , it could reconstruct the trust network and include information about the graph in its analysis. In this approach, the complete analysis and computation has to be done by entity A , which may need additional computational power depending on the manner of analysis. Furthermore, it is not clear how to prevent an entity from providing an arbitrary amount of bogus information, which could multiply the complexity of analyzing the graph. Furthermore, this approach needs means preventing that an entity can deceptively claim that it forwards recommendations from another recommender in order to discredit the pretended recommender.

In the second approach, the computation of the trust values is distributed to all entities that provided information as each entity forwards only an aggregated value. Yet, none of the entities is able to rebuild the complete network, as none of the entities knows on which information the recommendations by its recommenders are founded. Additionally, this may lead to situations in which the information of a single recommender is unintentionally multiplied. For example, in Figure 4.7 the recommender R_3 may forward its trust value on candidate C to R_1 and R_2 . When R_1 and R_2 forward their information to the initiator A , then entity A cannot recognize that the information originally provided by R_3 is integrated in the recommendation by R_2 and in the recommendation by R_3 .

In this thesis, it is proposed to follow the second approach in distributed environments. Thus, the initiator does not know how the trust values of its recommenders have been calculated. In order to overcome the problem of unintentionally multiplying the impact of a single recommender, the assumption is introduced that honest recommenders only distribute trust information that is based on their direct evidence (Section 6.1). Furthermore, it is important to note that the recommendations provided by R_3 and R_4 (in

Figure 4.7) are not necessarily lost. In the case that R_3 and R_4 forward their recommendation about the trustworthiness of the candidate C directly to A , entity A can consider these recommendations as recommendations provided by unknown recommenders.

In Section 6.3, the proposed trust model is used in an application for trust-based movie recommendation in a central setting. In this setting, the whole network of recommendations is known when evaluating the rating of a movie (which then is the equivalent to the candidate) based on the recommendations of known users. This allows to remove edges between the nodes, which when considered would multiply the impact of a single recommender. For example, in Figure 4.7, the edge between R_2 and R_3 could be removed for this reason. After removing these edges, recommendations are allowed to include recommendations by other recommenders.

4.2.6 Basic Assumptions

This section introduces a set of basic assumptions that an application context must fulfill for an evidence based trust model to be successfully applied. Most of the trust models presented in the last chapter are based on these assumptions.

1. Distinguishability of entities:

The most fundamental assumption introduced is that entities can be distinguished. Based on this assumption entities may be recognized by others, and evidence from previous interactions can be linked to entities. It is a typical assumption in the field, e.g., [BLB04, QHC06, TPJL06, JI02].

To achieve the distinguishability of entities, each entity needs a unique digital representation or digital (re-)identifier, which cannot be easily forged by another entity. The digital representation of an entity does not need to reveal the “real world” identity of the entity. Therefore, the (real-world) identity and goals of an entity may be unknown. An entity recognition scheme that does not require the enrollment of entities is described in [SFJ⁺03, SJ04].

While within the model each entity maps to exactly one digital representation (and vice versa), a user or a service provider can be allowed to be represented by multiple entities. It is important to be aware of this fact, as unmanaged domains, e.g., ubiquitous computing environments [SH08], may lack a central authority providing reliable mechanisms for controlling the participation of entities in the environment.

Furthermore, this allows users or service providers to re-join the system, being represented by new entities, whenever they want to.

2. Interactions between entities:

Entities have an interest and the possibility to interact with others. This is a very basic assumption, as it is the reason for evaluating trust.

3. Behavior of entities:

An entity has a strategy about the quality of its offered services (e.g., [TPJL06, QHC06, JHF03]) - this strategy may be referred to as the behavior of an entity.

The behavior of an entity may depend on its motivation or its goals. Besides economical motivations behind a behavior, one can think of ideological ones. For example, in file sharing communities, there are entities providing good files for the sake of the community, although it may be not rational to spend resources for the upload of files. However, there might be also entities providing corrupt files, because they are interested in the downfall of the community, which may be ideologically or economically motivated.

Furthermore, the behavior of an entity can change over time [JHF03, KBR05]. This might have two reasons. First, an entity may build trust in order to misuse it later, which may be considered as an attack on a trust model. Second, a change of an entity's behavior over time may occur due to reasons coupled to the process for the provision of the interactions. For example, a service provider aims to provide a high quality service, however, the provider may only be able to satisfy 95% per cent of its customers in average, e.g., as the service may be delayed or of bad quality when under heavy load.

Besides the entity providing the interaction, the initiator of an interaction also shows a behavior when selecting a candidate. As described above, the initiator has the choice of either selecting one of the possible partners or to decide not to interact, based on the information about the expected trustworthiness, the estimate utility of the interaction, and the risk attitude. In the case, when selecting an interaction partner, it is assumed that the usual behavior of the initiator is to select the most trustworthy candidate (cf. [TPJL06]). If the information about the utility and the risk that is associated to an interaction is available, this information can be used to decide whether to interact or not (cf. [Jøs99b, JHF03]).

4. Creation of direct evidence:

It is assumed that after an interaction, the initiator is capable of rating the quality or the outcome of an interaction, e.g., [JI02, QHC06, BLB04]. This feedback is necessary in order to derive the evidence that is re-used later as a basis of trust.

5. Storing evidence:

An entity has the capabilities for storing the evidence that has been derived from past interactions. In centralized systems, e.g., [JI02], the data is centrally stored by a so-called *reputation centre*, in distributed approaches, e.g., [BLB04], the entity itself stores the information.

When storing evidence naively, i.e., per interaction per interaction partner, this may quickly lead to a large amount of data, as each interaction would require extra storage. Therefore, it is proposed to aggregate evidence from previous interactions per interaction partner and context as, e.g., presented in [JI02, BLB04], in order to reduce the amount of storage needed.

6. Opportunity for repeated interaction with the same entity in the same context:

As introduced above, a trust model provides the capabilities to derive an expectation about the behavior of another entity based on evidence derived from past interactions. Therefore, it is crucial that the environment allows for repeated encounters of entities.

Although, this may sound trivial at first glance, there are different scenarios in which this assumption may be hurt.

First, in a domain without access control or in a domain with cheap pseudonyms, there is the threat that users re-join the community whenever their representing entity is associated with a bad trust value [FC05].

Second, when thinking about spontaneous interaction in opportunistic networks, a basic requirement is that the persons repeatedly meet each other in order to have a benefit or building histories and trust.

Third, as trust has been introduced to be context-dependent, the concept of estimating the trustworthiness of an entity based on interactions can only be directly applied in applications in which the interactions can be treated as belonging to the same context, i.e., the behavior of an interaction partner is assumed to be homogeneous in this context. This may not be true for applications in which the interaction partners provide interactions that require different competence. An online shop, which sells all kind of goods and services, may provide non-homogeneous interactions. For example, the competence for providing a three-week holiday trip differs from the competence of selling fresh fruits.

7. Exchange of recommendations:

It is assumed that entities have capabilities to provide recommendations about other entities (cf. [BLB04, TPJL06]). A *recommendation*

- also referred to as *indirect evidence* - is information provided by a recommender about past interactions with another entity.

Entities may be willing to provide recommendations about entities they already have interacted with. If an entity is not willing to provide recommendations, it cannot be forced. If an entity is willing to do so, the recipient of the recommendation cannot be sure about the motivation of the recommender. Recommenders can be honest (providing evidence that truly reflects the quality of previous interactions) or malicious (providing a misleading evidence) [TPJL06].

4.2.7 Basic Attacks

Having introduced the system model and basic assumptions, this section shows several types of attacks, which are possible within this setting. As the trust model is to help entities in selecting the most trustworthy partner for interactions, it should be robust to these attacks as far as possible.

When the trust value of an entity is evaluated, the main factors that are considered are direct evidence and recommendations (see Section 4.2.5). This leads to two basic types of attacks. On the one hand, an entity can attack the model in the role of an interactor, i.e., it starts to build trust in order to exploit it later. On the other hand, an attacker can try to influence a trust value in the role of a recommender, i.e., by providing misleading recommendations.

1. Interactor-based attacks:

- **Cash in:** An entity shows good behavior in order to establish trust; later it changes its behavior and misuses the previously established trust.
- **Repeated cash in:** An entity establishes and misuses trust repeatedly.

2. Recommender-based attacks (based on [FC05]):

- **False praise:** An attacker provides overly positive recommendations about the trustworthiness of an interaction partner.
- **False accusation:** An attacker provides overly negative recommendations about the trustworthiness of an interaction partner
- **Sybil attack:** An attacker creates a group of seemingly independent entities that collude (see [Dou02]). In the context of influencing the selection of a certain entity, an attacker may try to multiply the influence of its recommendation(s) by creating a high number of seemingly independent recommenders and provide either overly positive or overly negative recommendations.

Furthermore, both kinds of attacks are susceptible to *whitewashing* [FC05], i.e., the attacker repeatedly joins the community as new entity in order to get rid of a bad history. Whitewashing and Sybil attacks need especially to be considered in domains which not apply access control, e.g., in ubiquitous computing, or where the barrier for joining is very low, e.g., based on the provision of an email address.

4.3 Application Areas

In this section, a selected set of application areas (or applications) is presented, in which users may benefit from a distributed trust model. This also means that the assumptions that have been introduced in Section 4.2.6 are considered to be true in these application areas.

4.3.1 Ubiquitous Computing - Opportunistic Networks

In Weiser's vision of ubiquitous computing, computers, as they are common today, vanish more and more [Wei91, WB97]. Pervasive or ubiquitous computing is characterized by a very large number of smart devices, e.g., PDAs, mobiles, intelligent clothes, etc., which come with different capabilities considering communication channels, storage or battery power. Both, the basic idea of ubiquitous computing and the heterogeneity of these devices enforce interaction between devices to unfold the complete power of a ubiquitous computing infrastructure. In [BLRW04], Bhargava et al. point out that "socially based paradigms will play a big role in pervasive-computing environments".

On the one hand, the interactions with devices, which are possessed or controlled by foreign parties, include uncertainty and risk, since the identity and the goal of their owners may be unknown. On the other hand, the interactions with reliable partners are the basis for the services ubiquitous computing environments can provide.

A set of typical ubiquitous computing settings, namely mobile computing, ad hoc interaction, smart spaces, and real-time enterprises, is presented in [SH08]. The authors of [SH08] associate the ad hoc interaction setting with the following characteristics:

1. *Spontaneous interaction*: The entities establish temporary, wireless, and ad hoc communication links between each other, which allow for spontaneous interaction on the application layer.
2. *Lack of an infrastructure provider*: There is not any central authority allowing or restricting participation of entities.
3. *Anonymity of users or devices*: The entities per se might not have any information about the real-world identity of their potential interaction

partners, as they might interact anonymously.

4. *A priori, no user groups:* Devices can join and leave without restrictions.

Although the setting supports anonymity, it is not a requirement for ad hoc interaction. Basically, the user can decide whether they want to be anonymous, recognizable by a pseudonym, or even reveal their “real world” identity.

Example Opportunistic Networks - musicClouds: Within this setting one can think of a scenario in which mobile users are willing to exchange information amongst each other (see 4.1). This is known as passive collaboration in opportunistic networks [SH08, Hei07].

The scenario that has been introduced at the beginning of Chapter 2 and in Section 4.1 is set in this setting of ad hoc interaction. More specifically, the scenario is an example of collaboration in opportunistic networks. The basic idea of collaboration in an opportunistic network is that users may exchange information with users in proximity in a spontaneous manner [Hei07]. Thus, opportunistic networks do not rely on the presence of internet connectivity or any other infrastructure. On the one hand, this may have advantages. In [HCS⁺05], Hui et al. argue that there are numerous scenarios in which local connectivity might be preferred over an internet-based connection, due to bandwidth, latency or costs. On the other hand, it may also have disadvantages, as traditional identification and authentication relies on infrastructure, e.g, a public key infrastructure [ITU97].

In [Hei07], Heinemann studied and evaluated the capabilities of information dissemination in opportunistic networks with encouraging results. The evaluation was done based on one-hop routing scheme allowing for anonymous communication. Furthermore, Heinemann described an application called *musicClouds* [Hei07, HKLM03], which allows autonomous sharing of information, e.g., music files, in opportunistic networks. This application focuses on defining filters for specifying the meta-information of the files of interest, and the exchange of information. Yet, the application does not consider that users may (intentionally or accidentally) provide corrupted files, e.g., containing malware or viruses. Here, the application of a trust model can be the next step in order to support users not only in exchanging information but in finding reliable partners for the exchange.

A project that currently works on a concept to encourage people to exchange music in the London underground is called *undersound* [und09, BBM06].

Fulfillment of the basic assumptions: In this scenario, it may be assumed that entities interact with each other. It is also easily possible to evaluate the quality of an interaction after a file exchange took place. This can be done either by asking the user for feedback, or by a software component,

that checks the file for viruses, noise, etc. The behavior of an entity may depend on different aspects. Entities might be willing to share good files with other, they might be interested in distributing malware, or entities might participate without taking care on the quality of the information they share.

Furthermore, this scenario shows that it is desirable to support the user with an intuitive interface that allows them to check the trustworthiness of possible interaction partners, but also to adjust trust values based on the “real world” knowledge of the users. Assuming two persons know each other, e.g., they are friends or they work together, then they might want to directly assess the trustworthiness of the each other; without waiting until they have sufficiently often interacted with each other, so that the system could derive the trustworthiness based on the outcome of the interactions.

As ubiquitous computing environments may be unmanaged and many entities may only be locally or spontaneously available, it may be non-trivial to fulfill the assumption of distinguishability of entities. A solution may be the entity recognition scheme referred to in [SFJ⁺03,SJ04]. It does not require an enrollment of entities, but provides a reliable mechanism to recognize entities – if they want to be recognized. The authors of [SFJ⁺03] also provide the idea that recognition schemes are more suitable approach for representing entities in ubiquitous computing environments as traditional authentication schemes as PKI, e.g., [ITU97], or Kerberos [KN93].

Furthermore, the opportunity for repeated interactions with the same entity depends on the movement of the users carrying the devices. It can be shown that people tend to move in regular patterns [GHB08,MMC08], e.g., they work together, or they regularly take the same bus or underground. The term *familiar stranger* has been coined for people that meet periodically but who do not interact with each other [PG04]. The Reality Mining Project [EP06] collected data on the mobility patterns and spatial proximity of 100 human users tracking the mobile phones of the users. This data can be used evaluate whether there usually are opportunities for repeated interactions, and serve as a basis for mobility patterns in a simulation environment.

Moreover, one may expect that people are interested in regularly sharing information with the same people, e.g., their friends, and also to provide recommendations. They also should be able to rate the quality of their interactions, e.g., whether they received the correct file or not, and they should be able to store information about the outcome of their interactions on their personal devices, e.g., their mobile phone.

4.3.2 Next-Generation Internet - Web Service Selection in Open SOA Market Places

Besides ubiquitous computing, trust based selection of an interaction partner may be applied to electronic market places where anybody can offer web

services. On the technical level the so-called Service Oriented Architecture (SOA) provides the concepts that are necessary for description of services (e.g., WSDL), communication (e.g., SOAP), and discovery (e.g., UDDI) [WV07,Aus08]. In an electronic market place some entities (service providers) offer services, and other entities (service clients or customer) look for services providing the functionality they are interested in.

Whenever a customer has a choice between multiple services providing the same functionality, the quality of the provided services becomes important. For example, a customer who looks for a service providing a weather forecast, may select an unreliable service provider, providing a bad quality of service, or a service provider with high response time. Even if the service providers offers information about the quality of its service in the service description, the customer cannot rely on this information, as it is not an obligation [WV07]. According to [WV07] there are three ways to improve the customer's choice:

1. *Service Level Agreement (SLA)*: Customer and service provider may negotiate an agreement on the quality of the service (SLA). If the service provider does not comply to the agreed service level, there may be a penalty. A service level agreement comes with the cost of time and expense. Furthermore, it requires a common ontology for quality of service metrics.
2. *Monitoring the quality of services by a trusted third party*: This approach requires a trusted third party, and monitors (or sensors) to measure the quality of the services, e.g. execution time, which might be costly and result in a big overhead in a dynamic environment with many leaving and upcoming services.
3. *Customer feedback*: Feedback from customers can be collected and published by a trusted third party or in distributed manner. It is worth noting, that if there is a trusted third party, this approach disburdens the trusted third party from monitoring the services itself. Furthermore, it brings the advantage of quality of service information that cannot be directly collected by monitors.

The advantages of the last approach may be the reason for the upcoming research to apply trust and reputation mechanisms to web services. This approach can be applied to business-to-business (B2B), business-to-consumer (B2C), and consumer-to-consumer (C2C) environments. According to [WV07] current research focuses on reputation and trust mechanisms relying on a trusted third party (central node) to collect and publish the customers' feedback information. Decentralized systems, in which entities store their feedback on their own and may provide this information to other on request, are also applicable in this domain.

Fulfillment of the basic assumptions: As service providers have to register and to announce their offered service they are assumed to be distinguishable entities. For consumers it is not obvious that they need unique identifiers, yet, in case they want to exchange recommendations these identifiers are necessary in order to create a history about the accuracy of the recommendations of a recommender. If the SOA market place requires some kind of registration of consumers, or if the provision of recommendations itself is treated as a service, it should be possible to recognize entities that previously provided recommendations.

Furthermore, one can expect that the entities that join the market place are willing to interact with each other, and that they are able to create and store the evidence that describes the outcome of their past interactions. It may also be assumed that entities have the possibility to repeatedly interact with each other, e.g., as service providers may gain higher prices when more trusted.

Service providers can be expected to have a strategy about the quality of service they offer, e.g., they offer a service with high quality in order to build long-term customer relationship, or they provide a low quality service or do not delivery any valuable results at all in order to quickly maximize their profits.

4.3.3 Web 2.0 - Recommendations on Online Platforms

A growing number of users takes part in Web 2.0 applications on the internet. The term Web 2.0 is associated to wikis and blogs, but also to popular websites, Flickr (www.flickr.com), Wikipedia (www.en.wikipedia.org), YouTube (www.youtube.com), MySpace (www.myspace.com), and Facebook (www.facebook.com) [OL08]. The aspects that users can produce or publish information *user generated content* or that they can build *social networks*, are vital to these applications.

An upcoming trend is that users are not only encouraged to rate products, e.g., Amazon (www.amazon.com), hotels, e.g., (www.mytravelguide.com), or movies, e.g., IMDb (www.imdb.com), but to combine the information from social networks and user ratings in order to derive personalized recommendations. FilmTrust [Gol05] (trust.mindswap.org/FilmTrust) allows users to rate movies and their friends, if they have joined the network. Rating a friend means to rate the trust in a friend's ability to recommend movies, or more strictly speaking movie ratings. When a user wants to get a recommendation for a film they havenot seen (or rated), yet, the rating is calculated giving a higher weight to more trusted friends. Another approach that tries to integrate trust mechanisms in a recommendation system can be found in [LHC08].

In general, trust based approaches can support the user when they have to evaluate the value of an information or whether the information

is correct. Furthermore, the approach can be used to calculate subjective ratings, e.g., for movies, using the trust mechanism as a basis for weighting the recommendations (ratings) of different users.

Fulfillment of the basic assumptions: In order to evaluate whether Web 2.0 applications fulfill the basic assumptions that are necessary to successfully use a trust based approach, a similar argumentation as for Web services can be provided. On platforms that are based on social networks or on which users have to register it may be assumed that entities can be distinguished. Interaction between users is inherent to Web 2.0 applications, e.g., on platforms like YouTube and in blogs, users regularly offer content to others and search for information. Selecting the information of a user can be considered to be an interaction in this context. Afterwards, the user can rate whether the interaction, i.e., the information, was satisfying. The rating can be stored in a user's profiles and it can also be send to others as recommendation. The users that provide information have different interests and competence on different topics. Thus, it can be assumed that they provide information with different quality, and the trust mechanism may serve to find high quality information.

4.4 Conclusions

Based on the evaluation of the current state-of-the-art and the concepts that have been presented in this chapter, the author proposes a new distributed trust model in the next chapter.

The trust model is especially designed to cover the following aspects that are addressed by none of the state-of-the-art models in their entirety:

1. **Modeling trust as a subjective probability as basis for autonomous decision making:** Modeling trust as a subjective probability directly derives from the definition of trust that has been introduced in Section 4.2.4. If the utility of the outcomes of an interaction are known this allows the trust value to be applied not only in order to select an interaction partner, but also to reason about whether it is a rational choice to interact (see Section 2.3).
2. **Deriving trust from evidence considering context-dependent parameters:** As in an unmanaged domain identity and the goals of interaction partners may be unknown, trust is derived from the evidence from past interactions. When deriving trust from evidence, it is important that the relation between trust and evidence considers context-dependent parameters as introduced in (see Section 4.2.5.2). Based on the analysis of the state-of-the-art, the following *context-dependent* parameters are considered:

- *Dispositional trust*: The dispositional trust expresses a user's general attitude to trust in other entities. The dispositional trust may be assessed by the user, using the parameters *base trust* and *weight*; in addition, a mechanism to derive the base trust value from the behavior of the entities that have previously been encountered will be presented.
 - *Aging*: Aging is introduced as an entity's behavior may dynamically change. Aging allows more recent interactions to be given a higher weight and to prevent an entity that established a high trust value along time ago from misusing the previously established trust long-lasting. Furthermore, as an entity's behavior may also change over time, aging also allows to reduce the impact of older evidence in absence of interactions.
 - *Maximum number of expected evidence units*: The maximum number of expected evidence units allows the user to define how much evidence is necessary in order to believe that it is representative for the behavior of an entity.
3. **Representation of trust for human users:** As motivated in the scenario at the beginning of Chapter 2 and in Section 4.1, the users need to be able to control and manipulate the parameters of a trust model, either to intervene in the decision making process or to adjust the trustworthiness of an entity according to their "real world" knowledge. Therefore, a trust model needs provide a simple set of parameters that is easily understandable. Since the trust model is to express not only a trust value, but at least the confidence (or certainty) associated with this trust value (cf. beginning of Chapter 4), too, a conflict between simplicity and expressiveness arises. In the following, it is proposed to integrate the considered parameters in an intuitive graphical representation.
4. **Providing a robust mechanism for the integration of recommendations:** Recommendations are an additional source of evidence that is introduced in order to improve the selection of the interaction partner. However, this potentially allows other entities to maliciously influence the selection process. Based on the analysis of the computational models (see Section 3.3.2), this is addressed by proposing a new approach for weighting and filtering recommendations. The approach assesses the trustworthiness of a recommender based on the accuracy of its past recommendations. The influence of a recommender is not only limited by its trustworthiness in the context of providing recommendations, but also by its rank, i.e., the maximal influence of the best recommender is higher than the maximal influence of the second one, etc. The approach will be introduced in Section 5.3.4 can be shown to

improve the trust model's robustness to Sybil attacks.

After presenting the proposed model in Chapter 5, the trust model is evaluated in Chapter 6. The evaluation shows the performance of the trust model in a distributed setup for a large set of different populations regarding the typical behavior of an entity. Furthermore, it presents a user study evaluating the proposed graphical representation, and an application which shows how the trust model integrates in an online movie recommendation platform.

Chapter 5

Trust Model: CertainTrust

The concepts that have been presented in the previous chapters require the extension of the state-of-the-art trust models. This chapter presents the trust model that is developed in this thesis.

The chapter is structured as follows: Section 5.1 provides a short introduction of the basic components of the proposed trust model. The subsequent sections are devoted to the details of the trust model. The representational and the computational model are introduced in Section 5.2 and Section 5.3, respectively. The update mechanisms and a simple strategy for selecting an interaction partner are presented in Section 5.4. Finally, the conclusions of the chapter are presented in Section 5.5.

5.1 The Components of CertainTrust

The model can be split into three main components, as shown in Figure 5.1.

- The *representational model* addresses the issues how trust is derived from evidence, and how trust is represented.
- The *computational model* addresses the aggregation of direct and indirect evidence.
- The component called *update mechanisms* provides means for deriving new evidence after an interaction. While the update of the evidence of the interactor can be done straight-forward, one needs to take care when updating evidence collected about the trustworthiness of the recommenders based on the accuracy of their recommendations. Furthermore, this component allows the dispositional trust to be updated.

This section presents a short overview of the components, before their functionality is explained in detail in the subsequent sections.

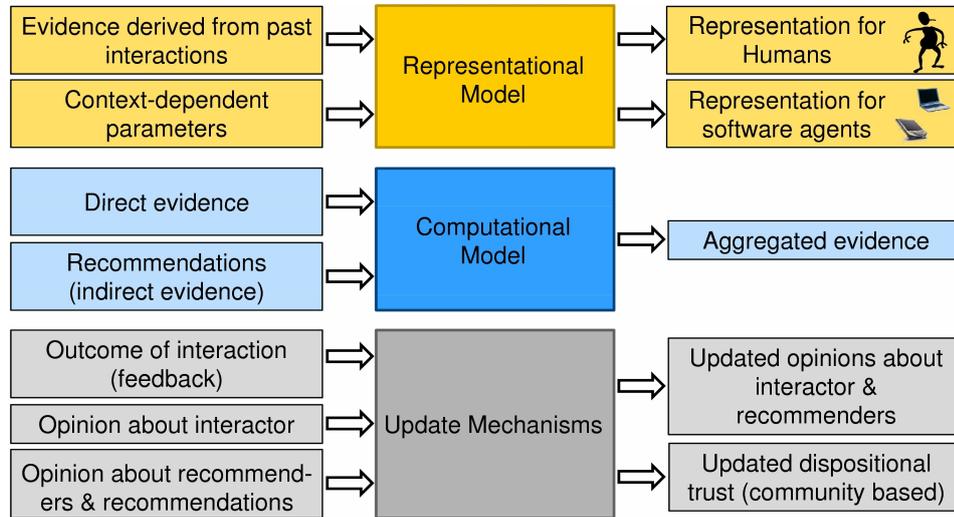


Figure 5.1: Components of a trust model

Representational model

The representational model focuses on two crucial issues.

1. How can trust be derived from evidence considering context-dependent parameters?
2. How can trust be represented to software agents and human users?

In order to address the first issue, a *relationship between trust and evidence* is required. Here, a Bayesian approach is chosen, as it provides means for deriving a subjective probability from prior information and collected evidence (see Section 3.2.4.2). This goes along with the definition of trust given in Section 4.2.4. Thus, the approach seems to be a good basis to start from in order to provide a mapping between trust and evidence.

Yet, there are additional aspects that need to be taken into account depending on the trust model's *application context* (see Section 4.2.5.1). In different application contexts, one has to deal with different general conditions, e.g., regarding insurance, law enforcement, strength of identity, etc. This may lead to different preferences regarding how trust should be derived from the evidence from past interactions, e.g., whether a few pieces of evidence are expected to be representative for the future behavior of an entity. Furthermore, one might have different expectations about the typical behavior of an entity, i.e., whether the behavior of unknown entities is assumed to be rather benevolent or rather malicious, and one might have different expectations about how the behavior of entities changes over time. For example, one can expect that well-known online shops try to keep their quality of service rather stable, while in peer-to-peer networks each peer's

quality of service may vary over time. These aspects are dealt with in the proposed model.

Furthermore, when developing a representation of trust, it is necessary to consider to whom trust is represented. A software component or a software agent can easily handle mathematical representations of trust. Thus, the Bayesian representation of trust is appropriate. However, human users are usually not used to interpreting the parameters of a beta probability density function or its graphical representation. If the user is given only the expectation value, they cannot figure out the number of evidence units the expectation value is based on or the quality of the observed evidence.

Therefore, a new *representation of trust designed for human users* is introduced. It provides a simple set of parameters reflecting the amount of information, the average quality of the collected evidence, and the base trust (dispositional trust). The trust value of an entity is derived from these parameters in an easily interpretable way. Finally, these parameters and the trust value are integrated in an intuitive graphical representation. Users can use this graphical representation in order to inform themselves about the trust value of an entity or to manipulate the trust value of an entity.

As this leads to two representations of trust, the Bayesian one and the one for human users, a mapping between both representations is required. The evidence from past interactions is a common basis for both representations. The challenge is to define the mapping between both representations - within the given constraints - that takes care that the derived trust value is independent of the representation, i.e., given the same information both representations are required to calculate the same trust value.

Computational Model

The computational model proposes a new approach for *aggregating direct evidence and recommendations*.

In general, recommendations are collected to increase the amount of information available about the candidates in order to improve the estimate of their trustworthiness. Yet, as recommenders may be interested in influencing the choice of the interaction partner, they cannot be assumed to provide recommendations that accurately describe the trustworthiness of the candidates. Therefore, the recommendations need to be integrated carefully, in order to really improve the estimate of the trustworthiness of the candidates. This is called robust integration of recommendations.

In order to achieve this goal, the proposed computational model introduces a new approach for weighting recommendations. The approach has two main features. First, it weights recommendations according to the trustworthiness of the recommender in context of providing recommendations. Second, it considers a recommender's rank in the set of the available recommenders, i.e., it reduces the influence of the recommenders with lower rank. Especially,

the approach can be shown to limit the impact of Sybil attacks (see Section 5.3.5).

The presentation of the computational model that is proposed in this thesis starts with an introduction of its basic operators which are based on operators for *consensus* and *discounting* (see Section 3.2.5). Afterwards, those operators are extended in order to improve the resistance to attacks.

Update mechanisms

The last component, *update mechanisms*, provides means to update the trust values after an interaction.

The initiator of the interaction uses the *outcome of the interaction* to derive new evidence that contributes to the history of the interactor and the recommenders. Thus, the outcome of the interaction is the basis for the update of the initiator's *opinion about the interactor* and the initiator's *opinions about the recommenders*.

For the update mechanism it is important to distinguish between the context of interactions and the context of providing recommendations.

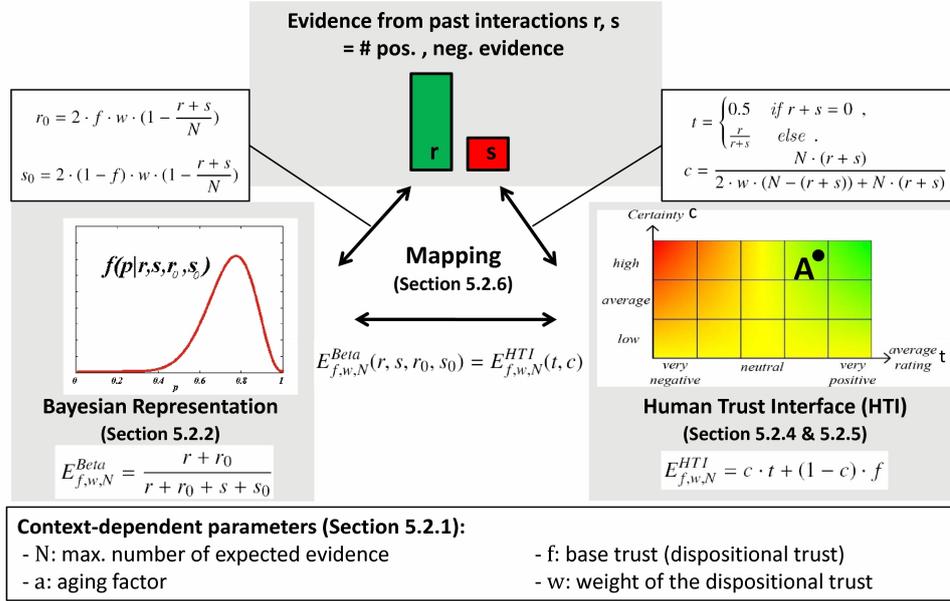
Regarding the update of the opinion about the interactor, it is important whether outcome of the interaction was positive or negative. The trust value of the interactor increases if the outcome of the interaction was positive, and it decreases if the outcome was negative.

Regarding the update of the opinion about the recommenders, the accuracy of the recommendations important. Here, the update mechanism is designed to increase the trust value of a recommender (in the context of providing recommendations), when the provided recommendation about interactor is considered to be accurate (see 4.2).

Besides the update of the opinion about the interactor and the recommenders, this component provides means to derive the expectation about the trustworthiness of unknown entities (*dispositional trust*) based on the behavior of the encountered entities.

5.2 Representational Model of Trust

This section presents the representational model of the proposed approach. It introduces the context-dependent parameters that are integrated in the trust model and two representations of trust. The first one - the Bayesian representation - provides a sound mechanism for deriving an estimate from prior information and collected evidence. Section 5.2.2 shows an extension of this approach considering context-dependent parameters in a way that especially allows the limitations of state-of-the-art approaches to be overcome. The second one - the Human Trust Interface (see Section 5.2.4) - is tailored for intuitively representing trust to human users. It is based on a simple set of parameters and it provides a graphical representation of trust. Afterwards,

Figure 5.2: Representations of trust¹

a mapping between both representations is provided (see Section 5.2.6). It allows one to benefit from the advantages of both representations. For an overview of the presented representations and their interrelation, see Figure 5.2.

5.2.1 Context-Dependent Parameters

The context-dependent parameters that are considered in the trust model have already been presented in Section 4.2.5.2. They do not depend on the entity whose trustworthiness is evaluated, but only on the application context.

- The dispositional trust, describing the expectation about the typical behavior of unknown entities, is modeled using the parameters *base trust* and *weight*.
 - The *base trust* $f \in [0; 1]$ expresses the trust assigned to unknown entities. A higher base trust value means that entities are expected to be more trustworthy (see Section 5.2.4.1).
 - The *weight* $w \in \mathbb{R}^+$ of the dispositional trust influences how quickly the final trust value of an entity shifts from the base trust value

¹Although Figure 5.2 introduces the aging factor a as a context-dependent parameter in its lower part, the aging factor does not appear in the formulas in the upper part. This is due to the fact that the aging factor is not directly considered in the representational model, but only indirectly, as it is proposed to choose $N = \frac{1}{1-a}$ in Section 5.2.3.2.

to the relative frequency of positive outcomes when evidence is available. Its impact is shown in Section 5.2.7.

- The parameter for *aging* allows one to consider that the behavior of an entity may dynamically change. The *aging* factor $a \in [0; 1[$ allows more recent evidence to be given a higher weight. This is necessary to prevent entities from misusing high trust values established a long time ago. Furthermore, aging provides means to reduce the influence of previously collected evidence in absence of new interactions. The limit of $a \rightarrow 1$ will lead to not considering aging; the smaller the value of a , the higher is the influence of recent evidence. Using $a = 0$ leads to only considering the evidence from the interaction in the last time slot. The mechanism for introducing aging is presented in Section 5.2.3.
- The *maximum number of expected evidence units* $N \in \mathbb{R}^+$ is introduced to provide a means for modeling the certainty of a trust value context-dependent. The parameter allows one to define a (finite) number of evidence units that is expected to be sufficient in order to consider the collected evidence as representative for the behavior of an entity. The higher N the more evidence is necessary for the trust value to be associated with the maximum level of certainty.

The parameters will influence the Bayesian representation (see Section 5.2.2), and the representation for human users (see Section 5.2.4). Especially, it is worth noting, that it is proposed to chose the maximum number of expected evidence units depending on the aging factor a , as in the face of aging, one has to consider that even an infinite amount of observations from past interactions leads only to a finite amount of collected evidence (see Section 5.2.3). State-of-the-art approaches described in [JI02, JHF03, WJI05] do not take this into account (see Section 5.2.3.1).

5.2.2 Bayesian Representation

The background of the Bayesian approach for deriving trust from evidence has already been presented in Section 3.2.4.2. This section shortly summarizes the notation and the concepts necessary for understanding the model provided.

The main parameters used to derive the trustworthiness of an entity based on evidence are the numbers r of positive and s of negative evidence, respectively, that have been collected based on direct evidence and recommendations. Within a given application context, the opinion about the trustworthiness of an entity derived from the past experience is denoted as $o = (r, s)^{rs}$. Note that the superscript refers only to the notation. Furthermore, the parameters r_0 and s_0 are introduced to reflect the prior knowledge.

For the parameters α and β , the beta probability density function for a random variable p is given as $h(p | \alpha, \beta)$ (see Section 3.2.4.2):

$$h(p | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad (5.1)$$

where $0 \leq p \leq 1$, $\alpha > 0$, $\beta > 0$

Figure 5.3 shows some examples of beta probability density functions for different values of α and β . In this thesis, the beta probability density function $h(p | \alpha, \beta)$ is referred to as the Bayesian representation of trust.

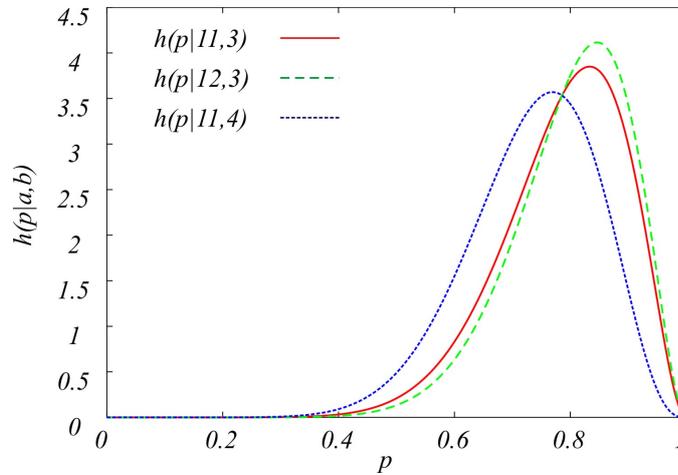


Figure 5.3: Beta probability density function

The interrelation between the parameters of the beta probability density function, the collected evidence, and the prior knowledge is given by defining $\alpha = r + r_0$ and $\beta = s + s_0$ (see Section 3.2.4.2).

As the approach for utility-based decision making introduced in Section 2.3 does not require the distribution itself, but only a point estimate, the distribution is summarized using its expectation value. In [Bol04], it is explained that the mean value of the distribution is a good choice to summarize the distribution in a point estimate.

The expectation value, i.e., the mean value of the beta distribution $Beta(\alpha, \beta)$, is given as:

$$E^{Beta}(r, s, r_0, s_0) = E^{Beta}(\alpha, \beta) = \frac{\alpha}{\alpha + \beta} = \frac{r + r_0}{r + r_0 + s + s_0} . \quad (5.2)$$

In the following, the expectation value is considered to characterize the trustworthiness of an entity.

5.2.2.1 Evidence based Update

The Bayesian representation provides for an easy integration of new evidence. After a new interaction, it is assumed that the user or an autonomous

software component (external to the proposed model) rates the outcome of an interaction. The rating of the outcome needs to be a value v in $[-1; 1]$, where -1 expresses a negative outcome and 1 a positive outcome. New evidence r_{new} and s_{new} is derived from this outcome as introduced in [JI02] (see Equation 5.3).

An additional parameter g allows different weights to be given to the evidence derived from an interaction, e.g., the parameter g can be used to reflect the value of an interaction [JI02]. However, in the following it is assumed, that all interactions are weighted equally using $g = 1$, i.e., interactions are assumed to be homogeneous within an application context.

$$\begin{aligned} r_{new} &= g \cdot 0.5 \cdot (1 + v) , \\ s_{new} &= g \cdot 0.5 \cdot (1 - v) \end{aligned} \tag{5.3}$$

Assuming the opinion before the interaction is $o = (r, s)^{rs}$, then the new opinion after the interaction is given by $o = (r + r_{new}, s + s_{new})^{rs}$.

5.2.2.2 Integration of the Context-Dependent Parameters

In current state-of-the-art approaches [TPJL06, BLB04, MMH02a], the prior knowledge is usually set to $r_0 = s_0 = 1$. Although this leads to a uniform distribution for the expected behavior of unknown entities, it is just an assumption introduced by the developers of those models and it prevents that the user introduces their own preferences. The expectation value as calculated by these approaches will be referred to as E_{simple}^{Beta} or as simple approach. It is defined in Equation 5.4.

$$E_{simple}^{Beta}((r, s)^{rs}) = \frac{r + 1}{r + s + 2} \tag{5.4}$$

The section shows how the context-dependent parameters for *base trust*, *weight*, and *maximum number of expected evidence units*, are integrated in the Bayesian representation. *Aging* will be introduced separately in next section.

Integration of Dispositional Trust

In contrast to the approaches proposed in [TPJL06, BLB04, MMH02a], the proposed approach allows the user to adjust the model according to their preferences. This section introduces a relation between a users dispositional trust and parameters r_0 and s_0 of the prior knowledge of the Bayesian representation. The parameters r_0 and s_0 can be assessed independently of each other. This is also true for the parameters base trust f and weight w that have been introduce to model a user's dispositional trust. As dispositional trust may primarily be used to estimate the trustworthiness of entities in absence of evidence, the following constraints are defined:

$$\begin{aligned}
w &= \frac{1}{2}(r_0 + s_0) && \text{if } r + s = 0 \\
f &= E^{Beta}(r, s, r_0, s_0) && \text{if } r + s = 0
\end{aligned} \tag{5.5}$$

Given these constraints, one sees that choosing $f = 0.5$ and $w = 2$ leads to $r_0 = s_0 = 1$. It is easy to verify that it holds $2 = \frac{1}{2}(1+1)$ (see Equation 5.5 upper part); furthermore, the lower equation can be shown to be true: In absence of evidence, i.e., $r = s = 0$, it holds $E^{Beta}(r, s, r_0, s_0) = \frac{0+1}{0+0+1+1} = 0.5 = f$.

Finally, increasing the value of w increases the influence of the prior knowledge on the expectation value (see Equation 5.2), as it increases the value of $r_0 + s_0$. For example, $f = 0.5$ and $w = 4$ leads to $r_0 = s_0 = 2$.

Thus, the w reflects the weight of the dispositional trust, and f defines the expectation value in absence of evidence.

Integration of the Parameter N

The parameter N is introduced to allow for the definition of a number of evidence units that is considered to be sufficient to be representative for the behavior of an entity. Thus, in case of $r + s = N$, the relative frequency of positive evidence should be equal to the derived expectation value.

Using $E_{simple}^{Beta}((r, s)^{rs})$ (see Equation 5.4), this can only be achieved for an infinite amount of evidence $r + s \rightarrow \infty$, as the prior knowledge is always considered in expectation value.

Therefore, it is proposed that the values of r_0 and s_0 also depend on the number of collected evidence units $r + s$ in relation to the maximum number of expected evidence units N . In Definition 5.2.1, a linear fade out of the prior information with increasing number of collected evidence units is proposed.

Definition 5.2.1 (Linear fade out of prior knowledge)

Given the context-dependent parameters for base trust f , weight w , and the maximum number of expected evidence units N , and given the numbers of positive r and negative s evidence units, the parameters of the prior knowledge r_0 and s_0 are defined by Equation 5.6.

$$\begin{aligned}
r_0 &= 2 \cdot f \cdot w \cdot \left(1 - \frac{r + s}{N}\right) \\
s_0 &= 2 \cdot (1 - f) \cdot w \cdot \left(1 - \frac{r + s}{N}\right)
\end{aligned} \tag{5.6}$$

Thus, r_0 and s_0 become functions depending on the parameters f , w , N , r , and s . Furthermore, when calculating r_0 and s_0 from the values of f and w using Equation 5.6, the constraints defined in Equation 5.5 are fulfilled. The proposed mechanism focuses on adjusting the calculated expectation value in order to reflect the ideas introduced above; it does not consider

other parameters of the distribution, e.g., the variance. A linear fade out is proposed as it was the simplest variant fulfilling all previously defined constraints².

This also allows for the introduction of a new notation of the expectation value that refers to the context-dependent parameters, as defined in Equation 5.7.

$$E_{f,w,N}^{Beta}(r, s) = E^{Beta}(r, s, r_0, s_0) \quad (5.7)$$

where r_0 and s_0 are calculated based on N , f , w , r and s using Equation 5.6. The expectation value that is calculated for an opinion $o = (r, s)^{rs}$ may also be denoted as $E_{f,w,N}^{Beta}(o)$.

Based on these constraints, the expectation value for an unknown entity is $E_{f,w,N}^{Beta}((0, 0)^{rs}) = f$; thus, for $r + s = N$, it holds $E_{f,w,N}^{Beta}((r, s)^{rs}) = \frac{r}{r+s}$ as intended.

Example: The example shows how recalculating the values of r_0 and s_0 depending on f , w , N , r , and s reduces the impact of the prior knowledge on the expectation value when the number of the collected evidence units increases (for simplicity aging of evidence is not considered).

Assuming an entity expects that 10 pieces of evidence are sufficient in order to belief that they are representative for an entity's behavior ($N = 10$). This means, e.g., after this entity has had 10 positive interactions with a certain interactor (leading to $r = 10$ and $s = 0$), it would expect that all interactions with this interactor will have a positive outcome.

When the expectation value is calculated as $E_{simple}^{Beta}((r, s)^{rs})$, i.e., the prior knowledge is statically set to $r_0 = s_0 = 1$, after 10 positive interaction the expectation value evaluates to $E_{simple}^{Beta}((10, 0)^{rs}) = \frac{10+1}{10+0+2} \approx 0.92$. Here, the expectation value is not only determined by the collected evidence, but also by the prior knowledge, as the model is not capable of considering the parameter N .

In contrast in the proposed approach, the parameters r_0 and s_0 are dynamically evaluated using Equation 5.6. From this equation, one can easily see that in case of $r + s = N$ it holds $r_0 = s_0 = 0$, independent from the initial setup of the prior knowledge. Thus, when setting $N = 10$ the expectation value is calculated as $\frac{10+0}{10+0+0+0} = 1$. Here, the expectation value is equivalent to the relative frequency of positive outcomes, reflecting that the collected evidence is expected to be representative.

²The author of this thesis has also proposed alternative mapping together with the mapping presented above in [Rie09]. The alternative mapping provides a more complex way for deriving the parameters r_0 and s_0 that not only considers that the expectation value approaches the relative frequency when $r + s$ approaches the value of N , but that additional takes care how the expectation value evolves for $0 < r + s < N$. As this approach does not introduce major difference regarding the design goals introduced in Section 4.4, it is not presented as a part of this thesis.

Beyond introducing a context-dependent measure of certainty, the parameter N can be used to consider that the introduction of aging may reduce the number of evidence units $r + s$ that is collected based on an infinite amount of interactions to finite number as shown in the next section. Then, choosing the parameter N depending on the aging factor a prevents that aging unintendedly introduces an estimation error.

5.2.3 Aging of Evidence

The core approach for *aging*, given in Definition 5.2.2, is based on [JI02, JHF03, WJI05].

Definition 5.2.2 (Aging)

Let the aging factor be denoted by $a \in [0; 1[$. The opinion at time $t - 1$ is $o_{t-1} = (r_{t-1}, s_{t-1})^{rs}$. The opinion $o_t = (r_t, s_t)^{rs}$ at time t is calculated using:

$$r_t = a \cdot r_{t-1} s_t = a \cdot s_{t-1} \quad (5.8)$$

In the face of new evidence (r_{new}, s_{new}) within time slot t , it holds at the end of time slot t :

$$r_t = a \cdot r_{t-1} + r_{new} s_t = a \cdot s_{t-1} + s_{new} \quad (5.9)$$

5.2.3.1 Limitations of Aging When Using E_{simple}^{Beta}

Assuming there is no evidence available at time $t = 0$, and at each point in time $t > 0$ there is exactly one interaction with weight $g = 1$, leading to either positive or negative evidence, i.e., $r_{new} + s_{new} = 1$. Then the sum of the collected evidence $r_t + s_t$ at time $t > 0$ can be calculated as a geometric sum.

$$r_t + s_t = \sum_{i=0}^{t-1} a^i = \frac{1 - a^t}{1 - a} \quad (5.10)$$

An infinite amount of time and interactions leads to (for $a \in [0; 1[$):

$$\lim_{t \rightarrow \infty} r_t + s_t = \lim_{t \rightarrow \infty} \frac{1 - a^t}{1 - a} = \frac{1}{1 - a} \quad (5.11)$$

Thus, the described aging limits the amount of evidence derived from an infinite amount of interactions to a finite number. This is not considered when calculating the expectation value using $E_{simple}^{Beta}((r, s)^{rs}) = \frac{r+1}{r+s+2}$.

Assuming there is no aging of evidence and at each point in time an entity collects a new positive evidence, then the expectation value derived from the evidence converges to 1. If there is an infinite number of negative experience, the expectation value converges to 0. Yet, when introducing aging, this is no longer true when using E_{simple}^{Beta} as proposed in [JI02, JHF03, WJI05]. Table 5.1

shows that the minimum E_{min} and maximum E_{max} expectation values that are calculated using E_{simple}^{Beta} depend on the aging factor.

Aging factor a	Max. amount of evidence	E_{max}	E_{min}
1	∞	1	0
0.99	100	0.990196	0.0098039
0.95	20	0.954545	0.0454545
0.9	10	0.916666	0.083333
0.8	5	0.8571428	0.1428571
0.7	3.33...	0.8125	0.1875
0.6	2.5	0.77777	0.2222222
0.5	2	0.75	0.25
0.0	1	0.66666	0.33333

Table 5.1: Relation of aging factor and E_{simple}^{Beta}

Table 5.1 shows that aging narrows the interval $[E_{min}; E_{max}]$ for the possible expectation values. As the basic idea when introducing aging of evidence was to give more recent evidence a higher weight in relation to older evidence, the effect that the introduction of aging narrows the range of the expectation value seems to be unintended. Furthermore, it may lead to an estimation error. For example, assume there is an aging factor of $a = 0.8$ and one has a large (infinite) amount of interactions with an entity providing only interactions with positive outcome. The derived expectation value based on these observations is $E_{simple}^{Beta} \approx 0.86$. This leads to an estimation error of about 14% compared to an expectation value based on the same observations without considering aging.

In contrast, Buchegger et al. [BLB04] introduced a slightly different approach for aging. They apply aging not only to the collected evidence but also to the prior knowledge. In their approach, aging does not narrow the range of the expectation value as presented above. Yet, Section 5.2.7 shows that their approach has another drawback. The expectation value in their approach does not move back to the initial expectation value in absence of evidence; instead, in absence of evidence, the expectation value does not change. Thus, the approach does not properly reflect that an entity's behavior may change over time without interaction.

5.2.3.2 Overcoming those Limitations

The reason why aging narrows the range of the expectation value is that the expectation value E_{simple}^{Beta} requires an infinite number of collected evidence units in order to approach the relative frequency of positive evidence.

Section 5.2.3.1 shows that depending on the aging factor, the number of collected evidence units $r + s$ is limited by $1/(1 - a)$. The introduction of

the context-dependent parameter N allows this to be considered, choosing $N = \frac{1}{1-a}$.

Then, it holds that the evidence derived from an infinite amount of interactions is equal to N ($r + s = N$). This leads to $r_0 = s_0 = 0$ (using $N = \frac{1}{1-a}$ in Equation 5.6) and $E_{f,w,N}^{Beta} = \frac{r}{r+s}$ (Equation 5.2). This means that the prior information does no longer influence the expectation value as the number of collected evidence units is supposed to be representative. Thus, the expectation value can reach the complete interval $[0; 1]$. The evaluation of the modified expectation value $E_{f,w,N}^{Beta}$ is shown in Section 5.2.7 together with a comparison to state-of-the-art approaches.

Using the aging factor $a \rightarrow 1$ leads to $N \rightarrow \infty$, this is equal to not considering the age of evidence. Although it is proposed to choose $N = \frac{1}{1-a}$, the user is free to choose N depending on their preferences. When choosing $N < \frac{1}{1-a}$, then the maximum level of certainty will be reached after a finite number of interactions. When choosing $N > \frac{1}{1-a}$, then the maximum level of certainty cannot be reached. In the latter case, the expectation value is always influenced by the prior value and does not use the full range of the interval $[0; 1]$.

5.2.3.3 Normalization as Implicit Aging

The normalization of an opinion is introduced to ensure that the evidence on which an opinion $o = (r, s)^{rs}$ is based does not exceed the maximum number of expected evidence units ($r + s \leq N$). This may occur for finite values of N , e.g., when a user decides to choose $N \leq \frac{1}{1-a}$, when it is possible to provide ratings with a higher weight than $g = 1$ (see Section 5.2.2), or when aggregating evidence from multiple parties (see Section 5.3). Thus, the normalization is more a technical necessity to ensure that the equations defined in the previous section are always applicable.

Whenever an opinion $o = (r, s)^{rs}$ of an entity is based on a greater number of evidence units than the maximum number of expected evidence units, the collected number of evidence units will be scaled to the allowed maximum (see Equation 5.12). The normalization preserves the relative frequency of positive evidence.

$$\text{norm}((r, s)^{rs}) = \begin{cases} (r, s)^{rs} & \text{if } r + s \leq N \text{ ,} \\ (\frac{N}{r+s} \cdot r, \frac{N}{r+s} \cdot s)^{rs} & \text{else .} \end{cases} \quad (5.12)$$

Assume the opinion about an entity is given as $o = (r, s)^{rs}$ with $r + s = N + \delta$. For example, this may occur when setting $N = 20$ without aging of evidence, i.e., $a \rightarrow 1$. When collecting the 21th piece of evidence the resulting opinion will be normalized according to Equation 5.12.

As the normalization may also be denoted as shown in Equation 5.13, the normalization may be interpreted as an implicit aging of the evidence in

the case of $r + s \geq N$ using an aging factor $a = \frac{N}{N+\delta}$ (see Equation 5.8).

$$\begin{aligned} r_{norm} &= \frac{N}{N + \delta} \cdot r \\ s_{norm} &= \frac{N}{N + \delta} \cdot s \end{aligned} \tag{5.13}$$

Note that in the case of setting $N < \frac{1}{1-a}$, normalization of an opinion should be done after the aging of the evidence in order to avoid unintended interference.

5.2.4 Human Trust Interface (HTI)

The Human Trust Interface (HTI) is new representation of trust that is designed to be intuitively interpretable by human users. Like the Bayesian representation, the HTI is based on the idea that trust between entities can be established based on past experience. Based on this experience, a user can associate an average value of the outcomes of those past interactions, and a value describing a notion of (un-)certainty (certainty in the following). The latter expresses the user's estimate of how reliable or representative the collected information is, in order to derive an expectation about the outcome of future interactions. Similar to ideas presented in [TPJL06,Jøs01,MMH02a], in this thesis the certainty is modeled to increase with the number of collected evidence units.

The main parameters of the HTI are called *average rating*, *certainty*, and *trust value*. The following introduces only the basic semantics of the parameters, the formulas are presented in Section 5.2.6.

The *average rating* $t \in [0; 1]$ ³ expresses the average outcome of the past interactions. It is calculated as the relative frequency of interactions with positive outcome. This value indicates the past behavior of an entity. The extreme values can be interpreted as follows:

- average rating = 0: There have been only bad interactions (very negative)
- average rating = 1: There have been only good interactions (very positive)

The *certainty* $c \in [0; 1]$ increases with the number of collected evidence units. Based on the idea that a trust value that is based on a higher number of evidence units is more representative for an entity, it expresses the influence of the average rating on the trust value in relation to the base trust value. The maximum level of certainty ($c = 1$) is reached if the number of collected evidence units is equal to the number of expected evidence units N . The extreme values can be interpreted as follows:

³The *average rating* was formerly called *trust value* in [Rie07,RS08,RH08].

- certainty = 0: There is no evidence available.
- certainty = 1: The collected evidence is considered to be representative.

The context-dependent parameters are also integrated. As before, the *base trust value* f expresses the trust value for unknown entities. It influences the final trust value directly. The *weight* w of the dispositional trust and the *maximum number of expected evidence units* N are integrated in the certainty parameter (see Equation 5.19).

Finally, the *trust value* $E_{f,w,N}^{HTI}(t,c)$ expresses the trust of the owner of the opinion in another entity. It is derived from the other parameters. Assuming that an entity has an initial expectation about the trustworthiness of unknown entities (base trust), when the number of collected evidence units increases, the expectation about the outcome in the next interaction shifts from this initial expectation to the average rating. This means that in absence of evidence (certainty = 0), it holds that the trust value is equal to the base trust value; in presence of sufficient evidence (certainty = 1), it holds that the trust value is equivalent to the average rating. This consideration is the basis for the calculation of the expectation value in the HTI (see Equation 5.14). Thus, the parameters of the HTI and their interrelation are easy to explain. Furthermore, they are the basis for the graphical representation shown that is introduced in Section 5.2.5.

It is important to note that the interpretation of the trust value in the HTI is equal to the interpretation of the expectation value in the Bayesian representation. Therefore, when based on the same numbers of evidence units the trust value in the HTI is supposed to be equivalent to the expectation value in the Bayesian representation.

5.2.4.1 Trust Value and Dispositional Trust

The basic concept that describes how the trust value is derived from the other parameters has already been explained above. The opinion about the trustworthiness of an entity is based on information collected from past interactions. As the trust model is to support users in future interactions and the certainty of an opinion is to indicate whether the average rating is expected to be a good prediction or not, both values need to be included in the expectation value.

According to the naming convention introduced in Section 5.2.2, the superscript *HTI* will be used when denoting an opinion $o = (t,c)^{HTI}$ using the parameters of the HTI. For an opinion $o = (t,c)^{HTI}$ and the base trust value f , the trust value $E_{f,w,N}^{HTI}(o)$ is defined as

$$E_{f,w,N}^{HTI}(o) = c \cdot t + (1 - c) \cdot f \quad (5.14)$$

This equation describes the consideration introduced above. The parameter f is used to determine the base trust value in case of complete uncertainty

and influences the expectation value until complete certainty $c = 1$ is reached.

Thus, f can be used to express a user's general attitude (dispositional trust) or depend on additional knowledge about the distribution of trustworthy and untrustworthy entities. The following briefly presents a few selected values to initialize the base trust value f in order to introduce some examples for the interpretation of the base trust value. However, the user is free to choose any other value in the range of $[0; 1]$ depending on their preferences. In addition, Section 5.4.3 presents an approach for dynamically deriving the base trust value from the behavior of the encountered entities.

Pessimistic strategy ($f = 0$):

The pessimistic strategy is based on the assumption that unknown entities are expected to provide interactions with negative outcome. According to this strategy, the trust value is 0, if no evidence has been collected (complete uncertainty). This reflects a user's attitude such as "I believe that entities are untrustworthy, unless I know the opposite with high certainty".

$$E_{0,w,N}^{HTI}(o) = t \cdot c \quad (5.15)$$

Moderate strategy ($f = 0.5$):

The moderate strategy is appropriate for binary decisions when entities providing interactions with positive and negative outcomes occur with the same probability. According to this strategy, the trust value is 0.5, if no evidence has been collected (complete uncertainty). This reflects a user's attitude such as "I believe that unknown entities to provide interactions with negative and positive outcomes with equal probabilities."

$$E_{0.5,w,N}^{HTI}(o) = t \cdot c + (1 - c) \cdot 0.5 \quad (5.16)$$

Optimistic strategy ($f = 1$):

The optimistic strategy reflects a user's attitude such as "I believe that entities are trustworthy, unless I know the opposite with high certainty".

$$E_{1,w,N}^{HTI}(o) = t \cdot c + (1 - c) \quad (5.17)$$

5.2.5 Graphical Representation

Given the probabilistic trust models presented in the state-of-the-art (see Chapter 3), one can see that the focus often is only on deriving trust from (direct and indirect) evidence considering context-dependent parameters. A representation designed for human users is usually beyond the scope of those approaches. In contrast, the approaches that are designed for human usage

do usually not consider context-dependent parameters. Furthermore, those approaches tend to model trust using discrete numbers (e.g., TidalTrust [Gol05]), or a small set of labels, e.g., [ARH00]. However, as in the proposed model trust is to be interpreted as probability, it needs to be considered that continuous values cannot be represented in a small set of labels or numbers without a loss of information.

Furthermore, the user might not only be interested in the trust value, but also in the certainty of the trust value, or the average rating of the past interaction. This requires that the representation integrates these parameters. Therefore, this section proposes a new multi-dimensional graphical representation that is designed for the interpretation and manipulation of trust values by human users.

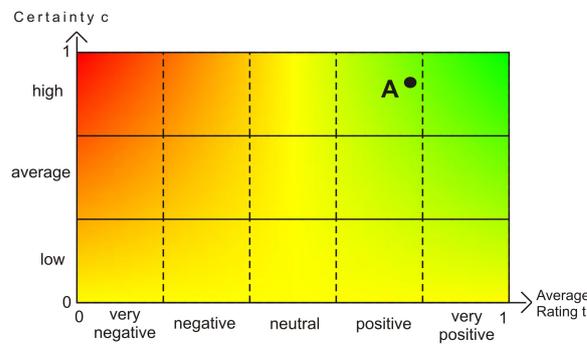


Figure 5.4: Graphical representation of the HTI with labels (moderate strategy $f = 0.5$)

The representation is based on the parameters introduced in the last section. The two parameters *average rating* and *certainty* are the basis for the two-dimensional layout of the graphical representation. In the Figure 5.4, the average rating is shown on the x-axis and the certainty on the y-axis. The parameter *trust value* is indicated by a red-yellow-green color gradient. Here, red indicates a low trust value ($E_{f,w,N}^{HTI}(o) = 0$), yellow a medium one ($E_{f,w,N}^{HTI}(o) = 0.5$), and green a high one ($E_{f,w,N}^{HTI}(o) = 1$). In-between these values, the color gradient is calculated depending on the expectation value using a linear combination of the colors red, yellow, and green in the RGB color model.

As the colors red, yellow, green allow for intuitively linked to the semantics of the trust value, this representation can be used to integrate representations of trust in applications in order to support a user's decisions and to allow the users to express opinions about the trustworthiness of interaction partners. Thus, the representation allows for an easy manipulation and interpretation of trust by users. In addition, the user may also be supported with labels introducing a coarse-grain semantics of the average rating and the certainty (see Figure 5.4).

Figure 5.4 shows the trust information for an entity A . In the example, the experience with entity A is given as $opinion = (10, 3)^{rs} = (0.77, 0.87)^{HTI}$. The context-dependent parameters are $N = \infty$, $w = 2$, and $f = 0.5$. The user may change the indicated trustworthiness of entity A by moving the spot marking the trustworthiness to another position.

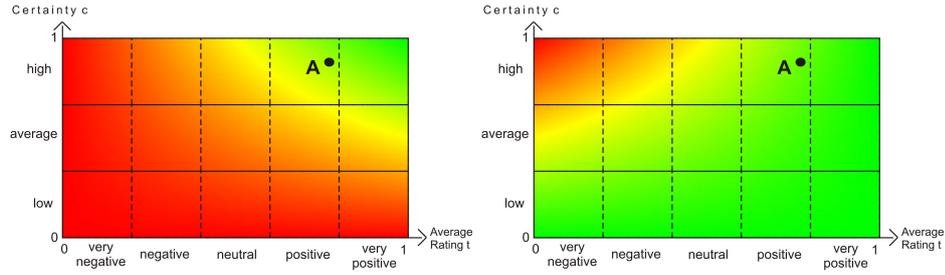


Figure 5.5: Graphical representation of the HTI with labels using different base trust values – pessimistic strategy $f = 0$ (left) and optimistic strategy $f = 1$ (right)

Furthermore, Figure 5.5 shows how different values of the base influence the trust value. In both example, the experience with entity A is given as $opinion = (10, 3)^{rs} = (0.77, 0.87)^{HTI}$ (as in Figure 5.4); only the values of the base trust are chosen differently, i.e., $f = 0$ (left) and $f = 1$ (right).

In Figures 5.4 and 5.5, entity A 's trust value is not explicitly shown. This can be overcome by simply displaying the trust value as shown in Figure 5.6.

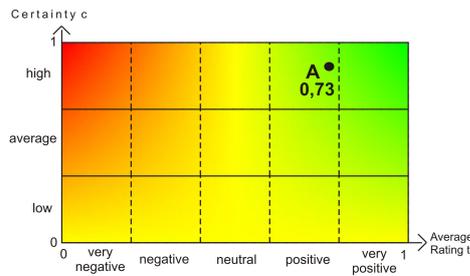


Figure 5.6: Graphical representation of the HTI displaying the trust value ($f = 0.5$)

In addition, for users that prefer to have the trust value shown on a separate axis, a graphical representation that is based on the trust value and the certainty parameter has been designed. In Figure 5.8, the x-axis represents the trust value and the y-axis shows the certainty value. Furthermore, the value of the average rating is indicated by the color gradient. The areas that are left white do not represent valid opinions. The representations only differ in the base trust value f .

The advantage of the latter representation is that its axis are devote to

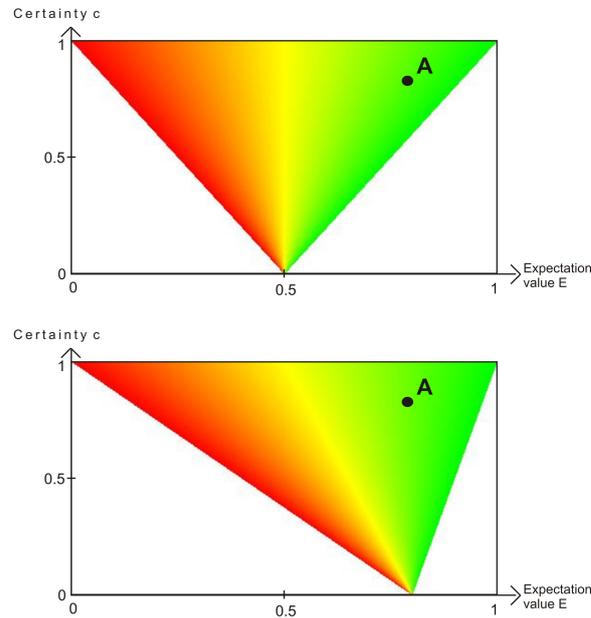


Figure 5.8: Alternative graphical representation ($f = 0.8$)

the two parameters, which might be the most important one for the decision making, i.e., the trust value and its certainty. However, users might feel more comfortable with the rectangular layout of the first proposal of the representation.

The evaluation of the intuitivity of the representation is presented in Section 6.2. It is based on a user study comparing the proposed representation and two graphical representations, namely the one introduced with “subjective logic” (see Section 3.2.5.1) and an Amazon-like stars interface (see Section 3.2.1.2). The first one is especially interesting as subjective logic provides a mapping between the “belief space” and the “evidence space” of the Bayesian approach. The latter one is a much simpler representation. However, it seems to be a good base line, as it is used in a commercial application.

5.2.6 Mapping between Representations

Having introduced both representations the question arises, how the parameters that have been defined for the HTI can be set in relation to the parameters of Bayesian representation within the provided specification. Here, it is important that the trust value of the HTI is equivalent to the expectation value in the Bayesian representation as both parameters have the same interpretation. This section provides a mapping between the parameters of both representations taking care of these aspects. Thus, it allows the advantages of both representations to be combined - easily interpretable

parameters (HTI) and well-founded, evidence based trust establishment (Bayes).

The mapping is only defined for opinions $o = (r, s)^{rs}$ with $0 \leq r + s \leq N$. Opinions that are based on a higher number of evidence units need to be normalized first (see Section 5.2.3.3).

Based on the introduction of the semantics of parameters of the HTI, the constraints for the mapping can informally be described as follows:

- The trust value is equivalent to the relative frequency of positive evidence.
- The certainty parameter is 0 in absence of evidence, and it is 1, if the collected evidence is supposed to be representative, i.e., it is equal to the maximum number of expected evidence units.
- The expectation value of the Bayesian representation needs to be equivalent to the trust value defined in the HTI.

The formal constraints for the mapping are given in Equation 5.18.

$$t = \frac{r}{r+s} \quad \text{if } r+s > 0$$

$$c = \begin{cases} 0 & \text{if } r+s = 0, \\ 1 & \text{if } r+s \geq N. \end{cases} \quad (5.18)$$

$$E_{f,w,N}^{Beta}(r, s) = E_{f,w,N}^{HTI}(t, c)$$

The mapping is defined in Definition 5.2.3.

Definition 5.2.3 (Mapping Bayesian Representation to HTI)

Given the parameters f , w , and N , the constraints given in Equation 5.18, and the proposed fade out of r_0 and s_0 in Equation 5.6, the mapping of an opinion $(r, s)^{rs}$ in the Bayesian representation to an opinion $(t, c)^{HTI}$ in the HTI is defined Equation 5.19.

$$t = \begin{cases} 0.5 & \text{if } r+s = 0, \\ \frac{r}{r+s} & \text{else.} \end{cases} \quad (5.19)$$

$$c = \frac{N \cdot (r+s)}{2 \cdot w \cdot (N - (r+s)) + N \cdot (r+s)}$$

The mapping fulfills the properties defined in Equation 5.18. As one can see, it holds $t = \frac{r}{r+s}$ if $r+s > 0$. Furthermore, it holds $c = 0$ for $r+s = 0$ and $c = 1$ for $r+s = N$. In absence of evidence, it is also easy to see that it holds $E_{f,w,N}^{Beta}(r, s) = f = E_{f,w,N}^{HTI}(t, c)$ using the Equations 5.5 and 5.14. As the fulfillment of the equality of the expectation value and the trust value in

presence of evidence, i.e., $0 < r + s \leq N$, is not considered to be this trivial, it is shown in Appendix A. The proof shows the equivalence using only simple algebraic manipulations.

The inverse mapping from the parameters of the HTI to the Bayesian representation is defined in Definition 5.2.4.

Definition 5.2.4 (Mapping HTI to Bayesian Representation)

Given the parameters f , w , and N , the constraints given in Equation 5.18, and the proposed fade out of r_0 and s_0 in Equation 5.6, the mapping of an opinion $(t, c)^{HTI}$ in the HTI to an opinion $(r, s)^{rs}$ in the Bayesian representation is defined in Equation 5.20.

$$\begin{aligned} r &= \frac{2 \cdot c \cdot w \cdot N \cdot t}{2 \cdot c \cdot w + \cdot N \cdot (1 - c)} \\ s &= \frac{2 \cdot c \cdot w \cdot N \cdot (1 - t)}{2 \cdot c \cdot w + \cdot N \cdot (1 - c)} \end{aligned} \quad (5.20)$$

The correctness of the inverse mapping can be verified, when replacing the values of r and s in Equation 5.19 using Equation 5.20.

5.2.6.1 Integration of E_{Simple}^{Beta}

The extension of the Bayesian representation is considered to be conservative in the sense that, for finite values of r and s , the expectation value is equivalent to E_{simple}^{Beta} (see Equation 5.4) when using $w = 1$, $f = 0.5$, and $N \rightarrow \infty$. Assuming that $r + s$ is a finite, then it holds for any opinion $o = (r, s)^{rs}$ with $0 \leq r + s \leq N$

$$E_{0.5,1,\infty}^{Beta}((r, s)^{rs}) = E_{simple}^{Beta}((r, s)^{rs}) \quad (5.21)$$

The proof is shown in the Appendix A. The proof is carried out straightforward using only simple algebraic manipulations and it uses that for a fixed and finite value of $r + s$ holds $\lim_{N \rightarrow \infty} \frac{r+s}{N} = 0$.

5.2.6.2 Integration of the ‘‘Opinion Space’’

The ‘‘opinion space’’ of subjective logic has already been introduced in Section 3.2.5.1. It is based on the parameters *belief* b , *disbelief* d , *uncertainty* u , and an additional parameter *atomicity* a that has an impact similar to the initial trust value f . The expectation value of a triple (b, d, u) in the opinion space is defined as $E(b, d, u) = b + u/2$ (using atomicity $a = \frac{1}{2}$). When the atomicity may be freely chosen, it holds $E(b, d, u, a) = b + u \cdot a$.

As the belief space is an interesting and important concept a direct mapping between the parameters of the opinion space and the parameter of the HTI is provided. As presented in Section 3.2.5.1, there is a mapping between the ‘‘evidence space’’, i.e., the Bayesian representation, and the

“opinion space” fulfilling $E(b, d, u, a) = E_{Simple}^{Beta}(r, s)$. This is taken into account when defining the mapping between the “opinion space” and the HTI as additional constraint, i.e., $E(b, d, u, a) = E_{Simple}^{Beta}(r, s) = E_{f,0.5,\infty}^{HTI}(t, c)$.

Definition 5.2.5 (Mapping Opinion Space to HTI)

Given parameters $w = 1$ and $N \rightarrow \infty$, the mapping between the parameters of the opinion space to the parameters of the HTI is defined in Equation 5.22.

$$\begin{aligned} f &= a \\ t &= \frac{b}{b+d} \text{ if } b+d \neq 0 \\ c &= 1-u \end{aligned} \tag{5.22}$$

The definition is based on the observation that for $w = 1$, and $N \rightarrow \infty$ it holds:

$$\begin{aligned} c &= \lim_{N \rightarrow \infty} \frac{N \cdot (r+s)}{2 \cdot w \cdot (N-r-s) + N \cdot (r+s)} \\ &= \frac{r+s}{r+s+2} \end{aligned} \tag{5.23}$$

The equation for the inverse mapping is given in Definition 5.2.6.

Definition 5.2.6 (Mapping HTI to Opinion Space)

Given parameters $w = 1$ and $N \rightarrow \infty$, the mapping between the parameters of the HTI to the parameters of the opinion space is defined in Equation 5.22.

$$\begin{aligned} a &= f \\ b &= t \cdot c \\ d &= (1-t) \cdot c \\ u &= 1-c \end{aligned} \tag{5.24}$$

Both mappings should be easy to verify. Based on these mappings it is possible to directly switch between the graphical representation of the opinion space and the one provided with the HTI.

5.2.7 Evaluation of the Impacts of Aging and the Context-dependent Parameters

This section compares the expectation values calculated by the proposed approach $E_{f,w,N}^{Beta}$ and the simple approach E_{Simple}^{Beta} .

The two diagrams in Figure 5.9 show the evaluation of the expectation value for different aging factors, using $N = \frac{1}{1-a}$ as proposed in Section 5.2.3.2. At time $t = 0$, no evidence is available. For the next 50 steps in time ($1 \leq t \leq 50$), there is one additional positive evidence per time step. For ($51 \leq t \leq 100$) there is one additional negative evidence per time step.

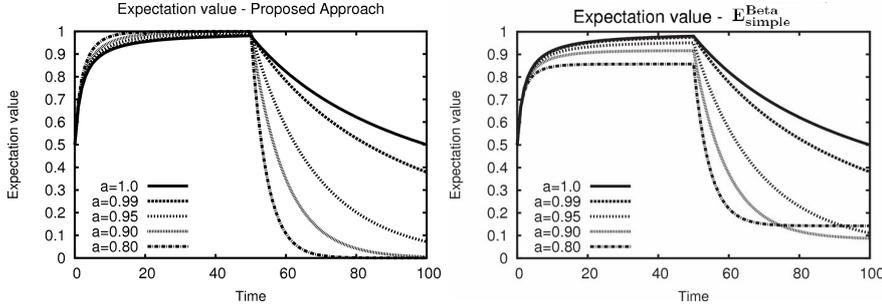


Figure 5.9: Comparison of the expectation values

The transition from the first interval to second one shows that in both approaches the aging allows for weighting toward the interactors recent behavior. It holds the lower the value of a the faster the expectation value reacts on the change of the behavior.

However, at the end of the first interval ($t = 50$), one can clearly see that the maximum of expectation value E_{simple}^{Beta} depends on the aging factor (as described in Table 5.1), whereas the proposed approach $E_{f,w,N}^{Beta}$ reaches values close to 1. This is also true the minimum of the expectation value in in the second interval ($51 \leq t \leq 100$). The example in Figure 5.9 shows that the introduction of aging narrows the range of the expectation value E_{simple}^{Beta} depending on the value of the aging factor a . Therefore, when estimating the trustworthiness of an entity A that is known to provide only interactions with positive outcome, i.e., $p(\text{"positive interaction"}) = 1$, the estimation error would depend on the aging factor a . For example, after an infinite amount of (positive) interactions using $a = 0.8$ leads to $E_{simple}^{Beta} \approx 0.86$ as shown in Section 5.2.3.1, i.e., an estimation error of about 14%. In contrast in the proposed approach, the expectation value would have been assess correctly, i.e., $E_{f,w,5}^{Beta} \rightarrow 1$. Thus, the proposed approach allows the limitation of the simple approach to be overcome.

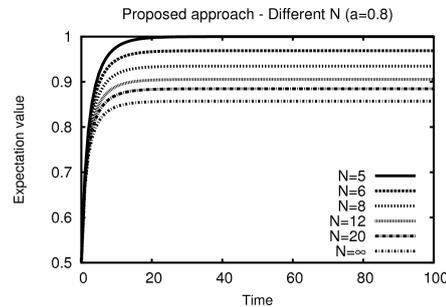


Figure 5.10: Impact of the maximum number of expected evidence units N

Finally, it is possible to argue that aging is introduced as it is assumed that the behavior of an entity may change and that the narrowing of the range of

the expectation value allows this to be considered, as values close to 0 or close 1 cannot be reached anymore. However, in the example above has just been shown that this may lead to an estimation error. Furthermore, if intended, this consideration may also be reflected in the proposed approach choosing $\frac{1}{1-a} \leq N \leq \infty$. When the value of N is chosen in $\frac{1}{1-a} \leq N \leq \infty$, the effect of narrowing the range of the expectation value becomes parameterizable even for a given value of a as shown in Figure 5.10. This can be explained, as it holds that $E_{simple}^{Beta} = E_{f,w,\infty}^{Beta}$. Thus, the presented approach is more expressive than previous ones.

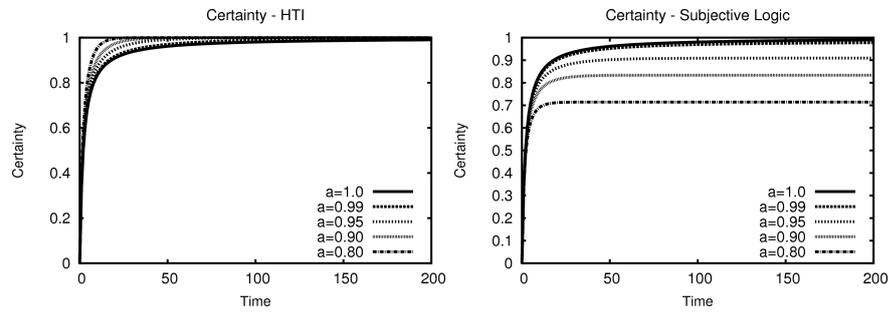


Figure 5.11: Comparison of the certainty parameters

The two diagrams in Figure 5.11 show the evaluation of the certainty parameter in the HTI and subjective logic. In the latter the certainty parameter is derived from the uncertainty parameter using $c = 1 - u$ (see Definition 5.2.5). In both figures, there is no evidence available at time $t = 0$, and at each point in time one additional piece of evidence is added. While the certainty parameter derived from subjective logic depends on the aging parameter, the certainty in the HTI is (almost) independent of aging.

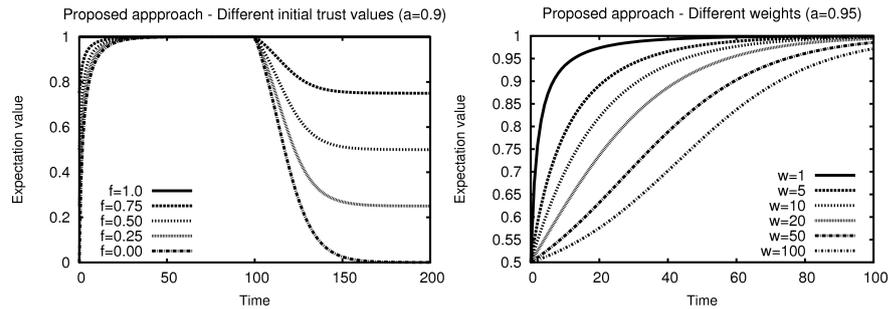


Figure 5.12: Impact of the parameters the base trust f and the weight w of the dispositional trust

The diagrams in Figure 5.12 show the impact of the base trust f and weight w .

In the diagram on the left side, the aging factor is set to $a = 0.9$. There is one new positive evidence for each point in time for $0 \leq t \leq 100$ and no

evidence for $101 \leq t \leq 200$. As one can see, for $t = 0$ the expectation value is equal to f , then the expectation value shifts towards 1. In absence of evidence, aging shifts the expectation value to its initial value f .

In the diagram on the right side, the aging factor is set to $a = 0.95$. There is one new positive evidence at each point in time. As one can see, the different values of the parameter weight influence how quickly the expectation value converges towards the relative frequency.

Thus, both parameters allow the expectation value to be influenced as intended (see Section 4.2.5.2). Both effects cannot be achieved using E_{Simple}^{Beta} . Furthermore, the example shows that the expectation value also reacts on the absence of evidence. When no further evidence is collected and older evidence is considered to be less representative, the expectation value moves back to the base trust value. This models that in absence of evidence an entity again becomes unknown.

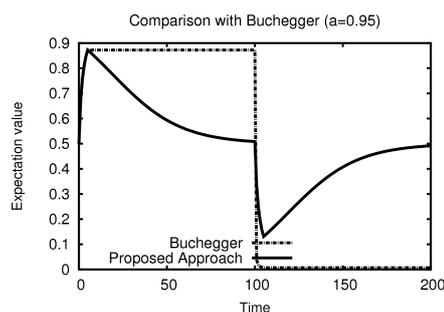


Figure 5.13: Comparison of aging

The diagram in Figure 5.13 shows the comparison of the proposed approach compared to the aging proposed by Buchegger et al. [BLB04]. In contrast to the aging proposed in Definition 5.2.2, Buchegger et al. apply aging not only to the collected evidence, but also to the prior information, i.e., $\alpha_t = a \cdot \alpha_{t-1}$ and $\beta_t = a \cdot \beta_{t-1}$. For the comparison, the following parameters are used: $f = 0.5$, $w = 1$, $a = 0.95$. For $1 \leq t \leq 5$, there is one positive evidence at each point in time, for $6 \leq t \leq 100$ there is no additional evidence; then, for $101 \leq t \leq 105$, there is one negative evidence at each point in time, and for $106 \leq t \leq 200$ there is again no additional evidence. One can see that the aging proposed by Buchegger et al. does not influence the expectation value in absence of evidence; after a period of absence of evidence new evidence has a very high impact. Thus, this approach does not fulfill function of the parameters as introduced in Section 4.2.5.2 and Section 4.4. In contrast, the expectation value shifts towards the base trust value f in absence of evidence and smoothly integrates new evidence that might become available afterwards in the proposed approach.

5.2.8 Summary

The section presented a novel approach for integrating context-dependent parameters for dispositional trust (base trust f , weight w), a number of maximum expected evidence units N , and aging a in Bayesian trust models. The novel model was developed in such a manner that a considerable number of advantages over the state-of-the-art become possible. The scientific contributions can be summarized as follows:

- In the proposed trust model, the Bayesian representation is the basis for deriving trust, as it provides an evidence based update mechanism and a well-founded way for justifying the derived expectation value. The proposed approach extends existing Bayesian trust models to allow for a more flexible, context-dependent derivation of trust. Especially, it provides means to integrate a context-dependent notion of certainty by introducing the parameter *maximum number of expected evidence* in the expectation value. This has not been proposed in Bayesian trust models before, at the best of the authors knowledge.
- The proposed approach covers the state-of-the-art Bayesian approach using the parameters $w = 1$, $f = 0.5$, and $N \rightarrow \infty$. As the context-dependent parameters for base trust f , weight w , the number of maximum expected evidence units N are integrated in the parameters reflecting prior knowledge, which are usually treated as static in state-of-the-art trust models, e.g., in [BLB04, TPJL06, JI02], the proposed approach should easily be applicable to extend those models.
- The HTI is a novel representation of trust for human users. It allows for a simple explanation of its parameters and for an intuitive graphical representation. This may be considered as a major advantage over state-of-the-art models. Furthermore, the HTI also support the introduced context-dependent parameters.
- The proposed mapping allows switching between the Bayesian representation and the HTI. This may provide means for a simpler interpretation of the concepts - for developers of trust models as well as for users. Furthermore, it allows developers of trust models to transfer their knowledge of the Bayesian representation to a more intuitive representation, which is more appropriate for the users of a trust model. The mapping provided in this section has the following property that the expectation value of an opinion in the Bayesian representation is equivalent to the trust value of the opinion in the HTI when both opinions are based on the same number of evidence units.
- The mapping to the “opinion space” of subjective logic above shows that when using the parameters $w = 1$, $f = 0.5$, and $N \rightarrow \infty$, the

representations are interchangeable. Yet, the HTI has the advantage that it is based on only two main parameters (average rating and certainty) that can be assigned independent from each other in contrast to the parameters of belief, disbelief, and uncertainty.

- The evaluation of the impact of aging has also shown that aging in the proposed approach has the features introduced in Section 4.4. It allows the consideration that an entities behavior may change by giving a higher weight to more recent evidence and, in absence of evidence, it allows the reduction the impact of older evidence on the expectation value. Especially, both properties are realized without limiting the range of the expectation value, which may introduce an unintended estimation error.

5.3 Computational Model of Trust

The computational trust model provides means for aggregating the direct evidence of the initiator and recommendations by third parties. In this thesis, this is also referred to as trust propagation. The basic concepts for trust propagation in the proposed approach have been inspired by work presented in [Jøs01, JI02] (also see Section 3.2.4.2 and Section 3.2.5.1). For easier comparison, the operators for the trust propagation are given the same names. The *consensus* operator provides a means for aggregating several opinions to a single one, and the *discounting* operator allows weighting recommendations based on the opinion about the recommender.

For the explanation of the trust propagation, a simple network is given as example (see Figure 5.14). In this example, entity *A* plays the role of the initiator of an interaction. As introduced in Section 4.2, the initiator evaluates the trustworthiness of a set of available candidates. In order to evaluate the trustworthiness of an arbitrary candidate *C* (who is part of the before mentioned set), entity *A* uses both its direct evidence and recommendations by third parties. In the example, entity *A* receives recommendations from the recommenders R_1 and R_2 .

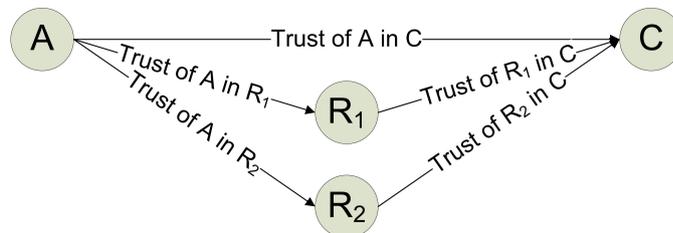


Figure 5.14: Trust network

As described in Section 4.2.5.1, it is important to distinguish between the

different application contexts in which an entity gained trust. As introduced there, within each context C that refers to an application, like file-sharing or the provision of a certain service, there are two sub-contexts: the sub-context referring to the interactions $C(I)$ in context C , and the sub-context referring to the recommendations $C(R)$ in the context C .

An entity gains (or losses) trust in the context $C(I)$ when it interacts with other entities. Yet, the behavior of an entity when providing interactions does not necessarily convey information about its behavior as a recommender, and vice versa. It is important to note that both sub-contexts refer to different capabilities of an entity. Therefore, trust is derived differently in both sub-contexts. In the context of interactions, a candidate is trusted based on the quality of the interactions, and in the context of providing recommendations, an entity is trusted when providing accurate recommendations. The proposed approach for considering recommendations weights the recommendations based on the trustworthiness of the recommender in the context of providing recommendations.

In order to keep the notation simple, it is assumed that there is a fixed, but arbitrary context C , e.g., file-sharing. The opinion of an entity A about an entity B in the context of interactions $C(I)$ is denoted as o_b^A (lowercase b), the opinion of an entity A about an entity B in the context of providing recommendations $C(R)$ is denoted as o_B^A (uppercase B). Note that the letter A will usually be used for the initiator of an interaction, the letters B and R (or B_i and R_i) for recommenders, and the letter C for the candidate. Furthermore, the expectation value of an opinion is denoted as $E(o_{b_i}^A)$ or $E(o_{B_i}^A)$, respectively, leaving out the super- and sub-scripts referring to the representation and the context-dependent parameters that have been introduced in the previous section.

5.3.1 Basic Operators

The aggregation of direct evidence and recommendations is realized as the aggregation of

- the evidence reflecting the direct evidence and
- the evidence provided in the recommendations.

The (basic) operators that are used for the aggregation are called *discounting* and *consensus*.

5.3.1.1 Discounting

The discounting operator weights the evidence provided by a recommender according to the trustworthiness of the recommender in the context of providing recommendations. The weight of the recommendation is given by the discounting factor in Definition 5.3.1.

Definition 5.3.1 (Discounting factor - basic)

Let $o_{B_i}^A$ denote the opinion of entity A about the trustworthiness of entity B_i as a recommender. The (basic) discounting factor $d(E(o_{B_i}^A))$ is defined as:

$$d(E(o_{B_i}^A)) = E(o_{B_i}^A) \quad (5.25)$$

Thus, the discounting factor $d(E(o_{B_i}^A))$ is defined to be equivalent to the expectation value $E(o_{B_i}^A)$, which describes entity A 's trust in entity B in the context of providing recommendations.

Definition 5.3.1 shows how the discounting factor is used in order to weight recommendations.

Definition 5.3.2 (Discounting) 5.3.1

Let $o_{B_i}^A$ denote the opinion of entity A about the trustworthiness of entity B_i as a recommender, and let $o_c^{B_i}$ denote B_i 's recommendation about entity C as candidate. Assuming that for all i holds $0 \leq r_c^{B_i} + s_c^{B_i} \leq N$ (otherwise an opinion will be normalized first using Equation 5.12), the discounting operator is defined as:

$$\begin{aligned} \text{discounting}(o_{B_i}^A, o_c^{B_i}) &= o_{B_i}^A \otimes o_c^{B_i} \\ &= d(E(o_{B_i}^A)) \cdot (r_c^{B_i}, s_c^{B_i})^{r_s} \\ &= (d(E(o_{B_i}^A)) \cdot r_c^{B_i}, d(E(o_{B_i}^A)) \cdot s_c^{B_i})^{r_s} \end{aligned} \quad (5.26)$$

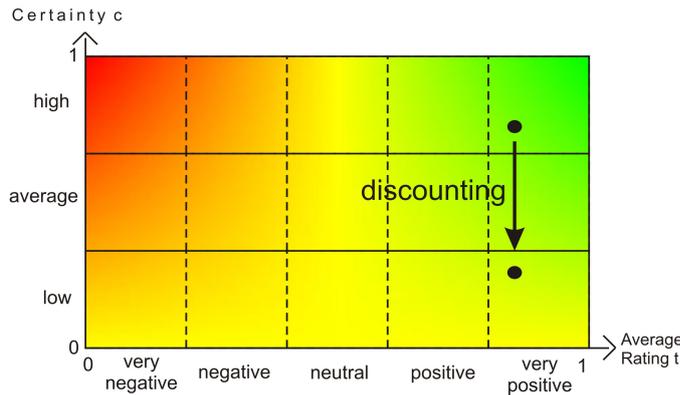


Figure 5.15: Effect of the discounting operator

The discounting operator reduces the number of evidence units of a recommendation according to the trustworthiness of the recommender. Furthermore, it keeps the value of the average rating. The effect of discounting is shown in Figure 5.15.

5.3.1.2 Consensus

The consensus operator provides for the aggregation of different opinions. The resulting opinion is the one of an entity that collected the evidence contributing to the consensus itself.

Definition 5.3.3 (Consensus - basic)

The consensus of the opinions $o_c^{B_1}, \dots, o_c^{B_n}$ is defined as:

$$\begin{aligned} \text{consensus}(o_c^{B_1}, \dots, o_c^{B_n}) &= \sum_{i=1}^n o_c^{B_i} \\ &= o_c^{B_1} \oplus \dots \oplus o_c^{B_n} \\ &= \left(\sum_{i=1}^n r_c^{B_i}, \sum_{i=1}^n s_c^{B_i} \right)^{rs} \end{aligned} \quad (5.27)$$

If the number of evidence units of the resulting opinion is beyond N , it will be normalized using Equation 5.12.

In general, if the input of the consensus operator consists of two (or more) opinions, then the certainty associated with the resulting opinion will be higher than (or equal to) the certainty of the contributing ones. This is true, as the resulting opinion will be based on more evidence, in general⁴. Furthermore, the average rating of the resulting opinion is not simply the average of the average rating $t_c^{B_i}$ provided by the recommenders B_i , as the consensus also considers the number of evidence units of each recommendation. The effect of the consensus operation is shown in Figure 5.16.

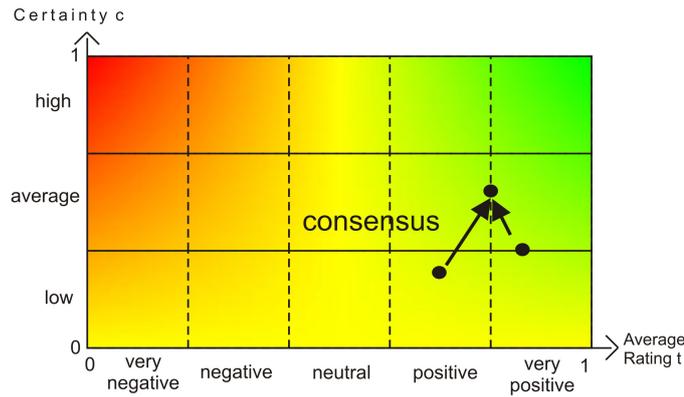


Figure 5.16: Effect of the consensus operator

⁴The certainty of the resulting opinion will not exceed the certainty of the contributing ones, if there is none or only a single opinion with $r_c^{B_i} + s_c^{B_i} > 0$, or if there is an opinion with $r_c^{B_i} + s_c^{B_i} = N$.

5.3.2 Simple Trust Propagation

The aggregated opinion \bar{o}_c^A of an entity A about a candidate C aggregates A 's opinion based on direct experience and the weighted recommendations that A has received. The (basic) aggregation is defined as follows.

Definition 5.3.4 (Aggregation - basic)

Let o_c^A denote the direct evidence of entity A for candidate C and let A collect recommendations from the recommenders B_1, \dots, B_n . The (basic) aggregated opinion of A about candidate C denoted as \bar{o}_c^A is defined as:

$$\bar{o}_c^A = o_c^A \oplus \sum_{i=1}^n o_{B_i}^A \otimes o_c^{B_i} \quad (5.28)$$

The basic aggregation presented in Definition 5.3.4 is also called *simple trust propagation* in this thesis. It has been introduced to provide the basic ideas for a robust trust propagation, i.e.,

- to weight a recommendation according to the trustworthiness of its recommender in the context of providing recommendations.
- to use recommendations for increasing the certainty of an opinion.

Furthermore, this variant of the trust propagation limits the influence of a single recommender as all recommendations are normalized if necessary. Thus, a recommender can not outweigh the discounting of its provided recommendation by providing an arbitrary high number of evidence units. Yet, this variant of trust propagation has some shortcomings:

1. The opinions of recommenders that are known to provide bad recommendations are still considered.
2. All available recommendations are used. Thus, if an attacker can create an arbitrary high number of entities, the attacker can use these entities to provide misleading recommendations. As all available recommendations are considered (except the ones provided from recommenders B_i with $d(E(o_{B_i}^A)) = 0$), the attacker can dominate the aggregated opinion even if the weight of a single recommendation is very low.

5.3.3 More Robust Trust Propagation - Limiting and Filtering

The more robust variant of the trust propagation provides a few enhancements to overcome the shortcomings pointed out above. To deal with the first issue, it seems reasonable that the initiator A only considers recommendations from recommenders that have provided mostly accurate recommendations in

the past, i.e., the average rating of σ_B^A is greater than or equal to 0.5. The approach still considers recommendations of unknown recommenders.

To overcome the second issue, another feature of the trust model is used to limit the considered recommendations. The recommendations are sorted in descending order according to their trustworthiness in the context of providing recommendations. In addition, recommendations are only considered as long as the certainty of the aggregated opinion is less than or equal to 1. Thus, this approach only uses the best recommendations, until the sum of direct evidence and weighted indirect evidence is equal to the maximum number of expected evidence units.

These arrangements together are supposed to improve the robustness of the model against misleading recommendations (either false praise or false accusation), since the approach uses only those recommenders that have been known to be the best recommenders from their past recommendations. Furthermore, if there are sufficient direct evidence and recommendations by highly trusted recommenders, the model is also quite robust to Sybil attacks (see Section 4.2.7), since it is no longer possible to overtake an opinion based on sufficient direct evidence and good recommendations by simply providing an arbitrary huge number of recommendations using specially created recommenders.

Although this variant of trust propagation overcomes the shortcoming introduced above, it still may be improved in the following aspects:

1. Unknown entities are always considered.
2. Sybil attacks will still be successful by little trusted recommenders B_i (average rating greater than or equal to 0.5 ($r_{B_i}^A \geq 0.5$) and expectation value at least slightly greater than 0 ($E_{B_i}^A \geq 0$)), when aiming on pushing the trustworthiness of candidates that are unknown to the rest of the community.

5.3.4 Sybil Attack-Resistant Trust Propagation

In order to overcome the shortcomings pointed out above, it is necessary to slightly adapt the operators of consensus and discounting. The goal of the adaption is described as follows:

1. Prevent that recommenders that are “little trusted”, but considered, may provide sufficient evidence to boost the certainty of the influenced opinion to 1.
2. Do not overly reduce the impact of recommendations by highly trusted recommenders.
3. Do not exclude unknown entities per se.

The proposed solution is based on the following ideas:

1. Introduce a threshold t_e that specifies a minimal expectation value necessary for recommenders to be considered. Recommenders with an expectation value lower than t_e are excluded. As the expectation value of a recommender depends on the accuracy of its past recommendations as well as on the dispositional trust of the entity evaluating the recommendations, this allows one to dynamically include or exclude recommendations of unknown recommenders. Additionally, recommenders providing mostly bad recommendations are excluded as introduced above.
2. Increase the influence of the recommendations with the trustworthiness of the recommender as above. The calculation of the discounting factor is adjusted in order to prevent an erratic increase of the impact of recommendations by recommenders that have just crossed the threshold t_e .
3. Limit the maximum influence of a single recommender. Therefore, a new parameter *maximum number of recommendable evidence* N_R is introduced. The parameter N_R defines the maximum number of evidence units that is considered per recommendation. Whenever a recommendation is based on a higher number of evidence units, it will be normalized as proposed in Section 5.2.3.3 before it is considered. It is proposed to choose $N_R \leq N$.
4. Limit the maximum influence of a recommendation based on the rank of its recommender. The steps above only reduce the impact of a single recommender. This step provides a means for taking control of the aggregated impact of all recommenders. Therefore, a threshold t_s is introduced. The threshold t_s defines the minimal trustworthiness that is necessary for recommenders to be able to influence the certainty of an aggregated opinion to reach 1 in absence of other recommenders or direct evidence.

The discounting is done as in Equation 5.26, but the discounting factor is redefined in Definition 5.3.5.

Definition 5.3.5 (Discounting factor - extended)

Let $o_{B_i}^A$ denote the opinion of entity A about the trustworthiness of entity B_i as recommender. Furthermore, let the threshold for the minimal trustworthiness of a recommender necessary to be considered be given as t_e . The (extended) discounting factor $d(E(o_{B_i}^A))$ is defined as:

$$d(E(o_{B_i}^A)) = \begin{cases} 0 & \text{if } E(o_{B_i}^A) \leq t_e \text{ ,} \\ \frac{1}{1-t_e} \cdot (E(o_{B_i}^A) - t_e) & \text{else .} \end{cases} \quad (5.29)$$

The influence of Equation 5.3.5 is explained in the following example. Assuming the value of the threshold is $t_e = 0.5$, then the discounting factor reduces the influence of all recommenders which have a trustworthiness of at most 0.5 to 0. The relation between the expectation value and the discounting factor using $t_e = 0.5$ is shown in Figure 5.17.

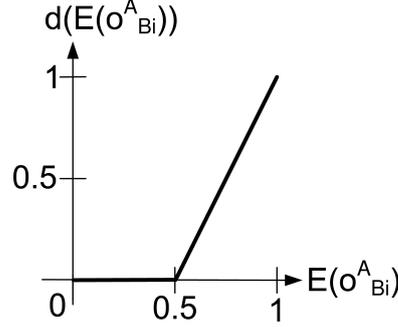


Figure 5.17: Discounting factor

The extension of the consensus operator is more complex. The operator is extended to limit the impact of a recommendation based on the rank of the trustworthiness of its recommender.

Definition 5.3.6 (Consensus - extended)

Let the trust for the recommenders B_0, \dots, B_n be given by the opinions $o_{B_0}^A, \dots, o_{B_n}^A$. Furthermore, let those recommenders provide the recommendations $o_c^{B_0}, \dots, o_c^{B_n}$. Let $0 \leq r_c^{B_i} + s_c^{B_i} \leq N_R$ for any i (recommendations based on a higher amount of evidence are normalized first, adapting Equation 5.12 using N_R instead of N) and let the ordering of the recommendations be according to $E(o_{B_j}^A) \geq E(o_{B_k}^A)$ for any $j < k$. Let t_s denote the threshold for Sybil attacks, then the (extended) consensus operator for these opinions is defined as:

$$\begin{aligned}
 \text{consensus}_{t_s}(o_{B_0}^A, \dots, o_{B_n}^A; o_c^{B_0}, \dots, o_c^{B_n}) &= [o_{B_0}^A, o_c^{B_0}] \hat{\oplus} \dots \hat{\oplus} [o_{B_n}^A, o_c^{B_n}] \\
 &= \left(\sum_{i=0}^n \min(d(E(o_{B_i}^A)) \cdot r_c^{B_i}, (1 - t_s) \cdot d(E(o_{B_i}^A)))^i \cdot \frac{N_R}{r_c^{B_i} + s_c^{B_i}} \cdot r_c^{B_i}, \right. \\
 &\quad \left. \sum_{i=0}^n \min(d(E(o_{B_i}^A)) \cdot s_c^{B_i}, (1 - t_s) \cdot d(E(o_{B_i}^A)))^i \cdot \frac{N_R}{r_c^{B_i} + s_c^{B_i}} \cdot s_c^{B_i} \right)^{rs}
 \end{aligned} \tag{5.30}$$

The influence of the extended consensus operator (see Definition 5.3.6) can be explained as follows. The first term in the min expression, i.e., $d(E(o_{B_i}^A)) \cdot r_c^{B_i}$ or $d(E(o_{B_i}^A)) \cdot s_c^{B_i}$, respectively, expresses the positive or the negative evidence that are calculated by the extended discounting operator. The latter term,

i.e., $(1 - t_s) \cdot d(E(o_{B_i}^A))^i \cdot \frac{N_R}{r_c^{B_i} + s_c^{B_i}} \cdot r_c^{B_i}$ or $(1 - t_s) \cdot d(E(o_{B_i}^A))^i \cdot \frac{N_R}{r_c^{B_i} + s_c^{B_i}} \cdot s_c^{B_i}$, respectively, expresses the maximum number of positive or negative evidence that the recommender is allowed to provide according to its discounting value and its rank i .

Using the steps proposed for filtering and limiting in Section 5.3.3 and the extended operators, the aggregation of direct evidence and recommendations is calculated as defined in Definition 5.3.7.

Definition 5.3.7 (Aggregation - extended)

Let $o_c^A = (r_c^A, s_c^A)$ denote the opinion of entity A about the candidate C , whose trustworthiness is being evaluated. Let $\mathcal{R} = \{R_0, \dots, R_k\}$ denote the set containing all available recommenders. The recommenders that provided mostly misleading recommendations to A are excluded; the remaining recommenders are given as $\{B_0, \dots, B_n\} = \{B_i \mid B_i \in \mathcal{R} \wedge t_{B_i}^A \geq 0.5\}$. Furthermore, the elements in this set are sorted in way that holds $E(o_{B_j}^A) \geq E(o_{B_k}^A)$ for any $j < k$. Let $M = \max\{m \in \{-1, \dots, n\} \mid \hat{r}_c^A + \hat{s}_c^A \leq N_R\}$. The (extended) aggregated opinion $(\bar{r}_c^A, \bar{s}_c^A)^{rs}$ is defined as:

$$(\hat{r}_c^A, \hat{s}_c^A)^{rs} = (r_c^A, s_c^A)^{rs} + \text{consensus}_{t_s}(o_{B_0}^A, \dots, o_{B_M}^A; o_c^{B_0}, \dots, o_c^{B_M}) \quad (5.31)$$

$$(\bar{r}_c^A, \bar{s}_c^A)^{rs} = \begin{cases} (\hat{r}_c^A, \hat{s}_c^A)^{rs} & \text{if } \hat{r}_c^A + \hat{s}_c^A = N_R \vee M = n, \\ \left(\hat{r}_c^A + \frac{N_R - (\hat{r}_c^A + \hat{s}_c^A)}{r_c^{B_{M+1}} + s_c^{B_{M+1}}} \cdot r_c^{B_{M+1}}, \right. & \\ \left. \hat{s}_c^A + \frac{N_R - (\hat{r}_c^A + \hat{s}_c^A)}{r_c^{B_{M+1}} + s_c^{B_{M+1}}} \cdot s_c^{B_{M+1}} \right)^{rs} & \text{else.} \end{cases} \quad (5.32)$$

If the resulting opinion $(\bar{r}_c^A, \bar{s}_c^A)^{rs}$ is based on more than N units of evidence, it is normalized using Equation 5.12.

The influence of the aggregation proposed in Definition 5.3.7 is shown in an example provided in the next section. The variable M is introduced to limit the number of recommenders that is considered based on the maximum number of evidence units recommenders are allowed to provide. In the proposed approach, it has been decided that all recommenders together may at maximum provide N_R units of evidence. Furthermore, the aggregation considers the case that the opinion resulting from the aggregation of the direct evidence and the first M recommendations as input may be based on less than N_R evidence, i.e., $\hat{r}_c^A + \hat{s}_c^A \leq N_R$. In this case, the recommendation by recommender $M + 1$ (if $M + 1 \leq n$) is normalized and used to fill the gap (see Equation 5.32).

5.3.5 Evaluation of the Robustness to Sybil Attacks

Let the threshold for Sybil attacks be t_s , and the threshold for the expectation value be t_e . Assume there is a Sybil attack by an attacker who created a group of recommenders B_0, \dots, B_n that have established little trustworthiness in the context of providing recommendations to entity A , i.e., $d(E(o_{B_i}^A)) < t_s$. These recommenders provide arbitrary recommendations $o_c^{B_i} = (r_c^{B_i}, s_c^{B_i})^{rs}$ about a candidate C .

If the expectation value of the all recommenders B_i is below the threshold t_e , i.e., $E(o_{B_i}^A) \leq t_e$, the recommenders are not considered at all. This is also true if they have provided mostly misleading recommendations.

If entity A has direct evidence from past interactions with candidate C , i.e., it holds $r_c^A > 0$ or $s_c^A > 0$ for $o_c^A = (r_c^A, s_c^A)^{rs}$, or there are recommendations by higher trusted recommenders, the influence of the recommenders B_i is reduced, as the direct evidence and the recommendations by more trusted recommenders are considered first.

Thus, the group of recommenders B_i has the maximum impact on the aggregated opinion \tilde{o}_c^A if they are the only entities able to provide evidence about the candidate C . This case is considered in the following.

Based on these assumptions and Equations 5.30, 5.31, and 5.32, the result of the extended aggregation of the direct evidence and recommendations by all recommenders is the result of the consensus of the recommendations provide by the group of recommenders B_0, \dots, B_n . For the evidence of the aggregated opinion $(\tilde{r}_c^A, \tilde{s}_c^A)$ holds⁵:

$$\begin{aligned}
\tilde{r}_c^A &\leq \sum_{i=0}^n (1 - t_s) \cdot d(E(o_{B_i}^A))^i \cdot \frac{r_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \cdot N_R \\
&< \sum_{i=0}^n (1 - t_s) \cdot t_s^i \cdot \frac{r_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \cdot N_R \\
&< (1 - t_s) \cdot N_R \cdot \sum_{i=0}^n t_s^i \cdot \frac{r_c^{B_i}}{r_c^{B_i} + s_c^{B_i}}
\end{aligned} \tag{5.33}$$

Analogue for \tilde{s}_c^A :

⁵Based on the assumptions introduced before, it holds $(r_c^A, s_c^A)^{rs} = (0, 0)^{rs}$, i.e., no direct evidence. Thus, the result of the aggregation (Equation 5.31) depends on the extended consensus operator (Equation 5.30). Here, the minimum operator defines the maximum outcome for r and s , respectively.

$$\begin{aligned}
\widetilde{s}_c^A &\leq \sum_{i=0}^n (1-t_s) \cdot d(E(o_{B_i}^A))^i \cdot \frac{s_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \cdot N_R \\
&< \sum_{i=0}^n (1-t_s) \cdot t_s^i \cdot \frac{s_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \cdot N_R \\
&< (1-t_s) \cdot N_R \cdot \sum_{i=0}^n t_s^i \cdot \frac{s_c^{B_i}}{r_c^{B_i} + s_c^{B_i}}
\end{aligned} \tag{5.34}$$

For the number of positive and negative evidence of the aggregated opinion $(\widetilde{r}_c^A, \widetilde{s}_c^A)$ holds:

$$\begin{aligned}
\widetilde{r}_c^A + \widetilde{s}_c^A &< (1-t_s) \cdot N_R \cdot \left(\sum_{i=0}^n t_s^i \cdot \frac{r_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} + \sum_{i=0}^n t_s^i \cdot \frac{s_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \right) \\
&< (1-t_s) \cdot N_R \cdot \sum_{i=0}^n t_s^i \cdot \frac{r_c^{B_i} + s_c^{B_i}}{r_c^{B_i} + s_c^{B_i}} \\
&< (1-t_s) \cdot N_R \cdot \sum_{i=0}^n t_s^i
\end{aligned} \tag{5.35}$$

For an arbitrary high number of recommenders ($n \rightarrow \infty$) holds:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \widetilde{r}_c^A + \widetilde{s}_c^A &< \lim_{n \rightarrow \infty} (1-t_s) \cdot N_R \cdot \sum_{i=0}^n t_s^i \\
&< (1-t_s) \cdot N_R \cdot \frac{1}{1-t_s} \\
&< N_R
\end{aligned} \tag{5.36}$$

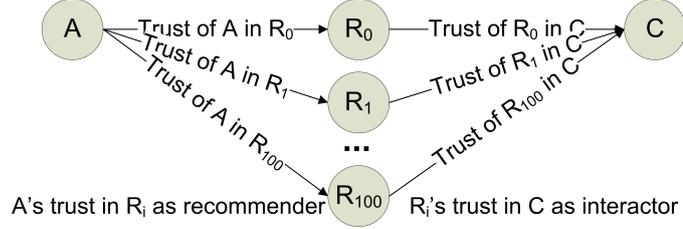
Thus, the resulting opinion is based on less than N_R evidence units. As it has been proposed to chose $N_R \leq N$, the certainty of this opinion is less than 1. Thus, the final trust value is still influenced by the entity's base trust and the group of attackers has only limited influence on the aggregated opinion o_c^A (and on the derived trust value).

This means especially, that an attacker cannot arbitrarily increase its influence on the aggregated opinion by simply increasing the number of recommenders.

However, if the attacker first manages that its recommenders become highly trusted by the initiator that will be attacked later, then the attack can still be successful, as trusted entities are considered to be benevolent. Yet, when the attacker takes the time or the costs to establish trust, the attack needs no longer to be carried out as Sybil attack, as trustworthy entities have a high influence.

Example

In this example, entity A has to evaluate the trustworthiness of the interactor C . Entity A has some direct evidence $o_c^A = (3, 1)^{rs}$. Furthermore, entity A receives recommendations from B_0, \dots, B_{100} (see Figure 5.18).

**Figure 5.18:** Trust network - Sybil attack

Entity A 's trust in the recommenders in the context of providing recommendations $o_{B_i}^A$ and the recommendations $o_c^{B_i}$ by each recommender B_i is given in Table 5.2.

i	0	1	2
$o_{B_i}^A$	$(12, 1)^{rs}$	$(11, 2)^{rs}$	$(10, 3)^{rs}$
$o_c^{B_i}$	$(6, 1)^{rs}$	$(8, 1)^{rs}$	$(8, 0)^{rs}$
$E(o_{B_i}^A)$	0.90	0.83	0.76
$d(E(o_{B_i}^A))$	0.45	0.66	0.51
X	$(4.82, 0.80)^{rs}$	$(5.26, 0.66)^{rs}$	$(4.09, 0.00)^{rs}$
Y	$(8.57, 1.43)^{rs}$	$(5.84, 0.73)^{rs}$	$(2.61, 0.00)^{rs}$
$\min(X, Y)$	$(4.82, 0.80)^{rs}$	$(5.26, 0.66)^{rs}$	$(2.61, 0.00)^{rs}$

i	3	4	5	...	100
$o_{B_i}^A$	$(1, 0)^{rs}$	$(1, 0)^{rs}$	$(1, 0)^{rs}$...	$(1, 0)^{rs}$
$o_c^{B_i}$	$(0, 20)^{rs}$	$(0, 20)^{rs}$	$(0, 20)^{rs}$...	$(0, 20)^{rs}$
$E(o_{B_i}^A)$	0.67	0.67	0.67	...	0.67
$d(E(o_{B_i}^A))$	0.34	0.34	0.34	...	0.34
X	$(0.00, 6.90)^{rs}$	$(0.00, 6.90)^{rs}$	$(0.00, 6.90)^{rs}$...	$(0.00, 6.90)^{rs}$
Y	$(0.00, 0.41)^{rs}$	$(0.00, 0.14)^{rs}$	$(0.00, 0.05)^{rs}$...	$(0.00, 5.8E - 46)^{rs}$
$\min(X, Y)$	$(0.00, 0.41)^{rs}$	$(0.00, 0.14)^{rs}$	$(0.00, 0.05)^{rs}$...	$(0.00, 5.8E - 46)^{rs}$

Table 5.2: Example: Sybil attack - it holds $X := d(E(o_{B_i}^A)) \cdot o_c^{B_i}$ and $Y := (1 - t_s) \cdot d(E(o_{B_i}^A))^i \cdot \frac{N_R}{r_c^{B_i} + s_c} \cdot r_c^{B_i}$

In the table, the recommenders have already been sorted according to their trustworthiness and recommenders that provided mostly misleading recommendations have already been excluded. The table shows that three recommenders (B_0 , B_1 , and B_2) provided a higher number (13) of mostly accurate recommendations to entity A . Furthermore, it shows an attack on

the trustworthiness of candidate C . The attacker created 97 Sybils B_3, \dots, B_{100} . In the past, each Sybil provided a single accurate recommendation in order to get considered in the evaluation of the trustworthiness of entity C . In the attack, each Sybil tries to provide a bad recommendation about entity C in order to reduce the calculated trust value. The example is based on the following parameters: $N = N_R = 20$, $t_e = t_s = 0.5$, $f = 0.5$, $w = 1$.

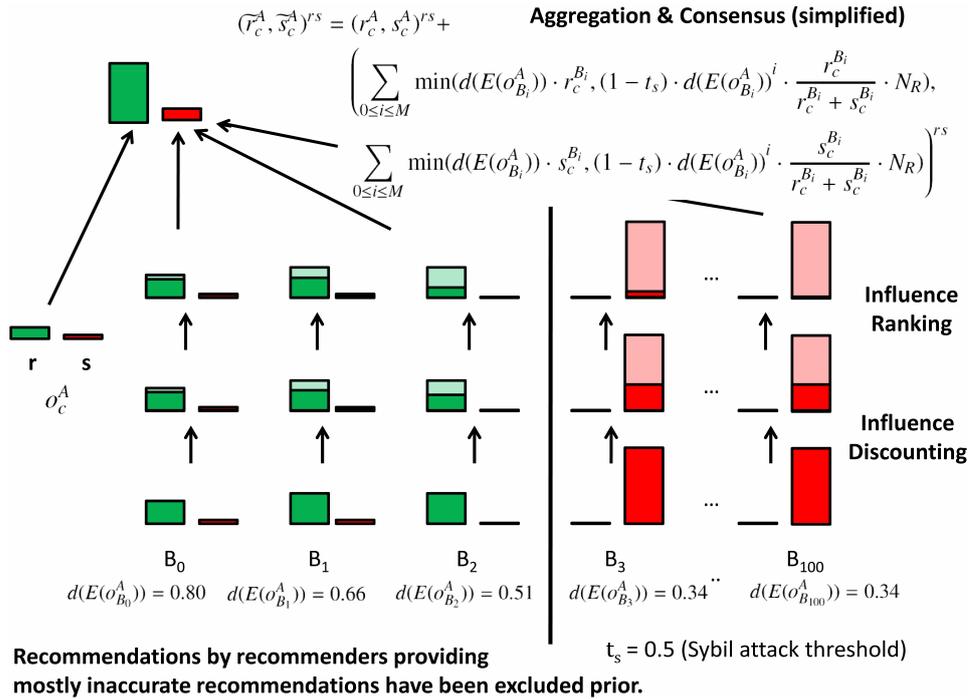


Figure 5.19: Computation of trust: Aggregation of direct evidence and recommendations

Figure 5.19 visualizes the impact of the extended aggregation mechanism. In this Figure, the number of positive and negative evidence, which is provided by each entity, is visualized by the height of the green and red bars, respectively. In the lower left side of the Figure, one sees the direct evidence of entity A . Furthermore, the lower part of the diagram shows the numbers of positive and negative evidence that are provided by the recommenders $B_0, B_1, B_2, B_3, \dots, B_{100}$. The row above shows reduced numbers of evidence units after applying the extended discounting operator on the opinion provided by each recommender. The light shaded bars show the number of evidence units that has originally been provided by the recommender, the opaque bars in the foreground show the number of evidence units that is left after applying the operator. Here, it is important to note, that the reduced numbers of evidence units are equal for the recommenders B_3, B_4, \dots, B_{100} . Thus, the influence of the recommenders B_3, B_4, \dots, B_{100} would still be equivalent. As

in the extended mechanism the rank of each recommender is considered, the influence of the recommenders may be further reduced. This is shown in the upper row. While the evidence provided by the recommenders B_0 , B_1 , and B_2 , is not strongly reduced, as they have a high trust value and a high rank, the influence of the attackers is strongly reduced. Finally, the top left corner shows the aggregated opinion that is calculated by entity A.

From the table, one can see how the new ranking-based approach reduces the impact of recommenders with lower rank. Using the consensus operator directly to aggregate the direct evidence and the discounted opinions (see the row with the X) would lead to an aggregated opinion $(\bar{r}_c^A, \bar{s}_c^A)^{rs} = (17.16, 678.3)^{rs}$. This opinion is strongly influence by the lowly trusted Sybil attackers. Using the extended aggregation mechanism (based on the row with $\min(X, Y)$) leads to $(\bar{r}_c^A, \bar{s}_c^A)^{rs} = (15.68, 3.09)^{rs}$.

As the trust models proposed in [JI02, BLB04, TPJL06] use the basic consensus operator that has been defined in Equation 5.27 for the aggregation of discounted evidence, they are susceptible for this kind of attack based on the design of their aggregation mechanism. In contrast, in the proposed approach, the influence of a Sybil attacker is strongly reduced, especially, the aggregated opinion is not simply dominated when increasing the number of attackers.

5.3.6 Summary

This section provided a new computational model of trust. The main features of this model are:

1. The discounting (weighting) of recommendations considers the trustworthiness of recommenders in the context of providing recommendations. Thus, the discounting is based on the right type of trust.
2. The influence of bad recommenders is reduced based on two mechanisms. First, recommendations by recommenders providing mostly misleading recommendations are excluded. Second, recommendations by recommenders with an expectation value for providing accurate recommendations lower than or equal to t_e are not considered, as the corresponding discount factor is 0.
3. Recommendations by unknown recommenders are considered if the base trust value $f^{C(R)}$ of A in the context of recommendations $C(R)$ in context C is above t_e . This is important as in contexts in which one expects the recommendations by unknown entities to be accurate those recommendations can be included in the aggregated opinion. Then, in absence of recommendations provided by trusted entities, recommendations by unknown entities can be a valuable contribution.

4. The aggregated opinion, which is derived from direct evidence and collected recommendations, favors direct evidence and the recommendations by the best recommenders. Thus, in the presence of sufficient direct evidence or recommendations by highly trusted recommenders, recommendations by less trusted recommenders (potential attackers) do not have any influence.
5. The aggregation mechanism is robust against Sybil attacks in presence of sufficient direct evidence or better recommenders. In this case Sybil attacks have little or no influence as the Sybils might not be considered at all. Furthermore, the impact of Sybil attacks on candidates that are mostly unknown is improved. Especially, the aggregation mechanism is robust to Sybil attacks in the sense that an attacker cannot arbitrarily increase its influence on the aggregated opinion by simply increasing the number of recommenders. The novelty of the extended mechanism for the aggregation of evidence is that the trustworthiness and the rank of a recommender are considered in order to limit its maximum influence.

The evaluation of the trust model in Section 6.1 shows that the trust model improves the estimate of the trustworthiness of entities and the average quality of interactions in the simulation of a distributed environment.

5.4 Selection of an Interaction Partner and Update of Trust

Having evaluated the trustworthiness of the candidates, the next step is to select an appropriate candidate and to decide whether the trust in the candidate is sufficient in order to interact. After the interaction, the trustworthiness of the candidate as well as the recommenders needs to be updated.

5.4.1 Selection of a Candidate

The selection of a candidate for an interaction of a set of candidates C_1, \dots, C_n is done based on their trustworthiness in the context of interactions that is calculated by the initiator.

If a decision making component does not have information about the utility of the interaction, i.e., the possible benefit in case of success, or the loss in case of a failure, one should select the most trustworthy candidate, i.e., the entity for which the maximum expectation value for positive outcome in the next interaction is calculated. In case of equal expectation values, the one with the higher certainty is selected.

In the case of presence of information about the utility of the interaction and possible information on the risk attitude, then the trust values can be used in the utility-based decision making introduced in Section 2.3.

5.4.2 Update Mechanism

After each interaction, the entity that initiated the interaction updates the opinions about its interaction partner (selected candidate) and the recommenders (see Figure 4.2). If the base trust value is based on the behavior of the encountered entities, the base trust is updated after an interaction, too. The update is carried out using the feedback fb provided after an interaction. The feedback may be binary in $\{-1; 1\}$ or continuous in $[-1; 1]$. In both cases, -1 is the worst possible feedback and 1 is the best one.

5.4.2.1 Updating the Opinion about the Selected Candidate

Let the opinion (direct evidence) of entity A about the selected candidate C before the interaction be given as $o_c^A = (r_{c_{old}}^A, s_{c_{old}}^A)^{rs}$. After the interaction the feedback fb , which describes the quality of the interaction, is used to update the evidence collected by A about C . The opinion of entity A about interactor C after the interaction, including the feedback fb , is denoted as $(r_{c_{new}}^A, s_{c_{new}}^A)^{rs}$.

The update of the opinion is computed as introduced in Section 5.2.2.1 and Section 5.2.3. In order to use the update mechanism proposed in Equation 5.3, it is necessary to define a mapping between fb and v , and to define the weight of the interaction. As proposed in [JI02], the weight g could be chosen based on the value of the interaction. However, as stated in Section 5.2.2.1, a weight of $g = 1$ is assumed in this thesis.

If the domain of the values of fb and v are both binary or both continuous, the mapping between fb and v can simply be defined as $v := fb$. If fb allows for continuous values and v is supposed to be binary, it is proposed to define the following mapping (also see Fig. 5.20):

$$v := \begin{cases} 1 & \text{if } fb \geq 0, \\ 0 & \text{else.} \end{cases} \quad (5.37)$$

Without aging, the update of the opinion about the candidate is described by the following Equation (based on Equation 5.3):

$$(r_{c_{new}}^A, s_{c_{new}}^A)^{rs} = (r_{c_{old}}^A + (v + 1)/2), s_{c_{old}}^A + (1 - v)/2)^{rs} \quad (5.38)$$

For example, assume that the opinion of entity A about the candidate C before the interaction was $(r_{c_{old}}^A, s_{c_{old}}^A)^{rs} = (16, 8)^{rs}$, and the feedback for

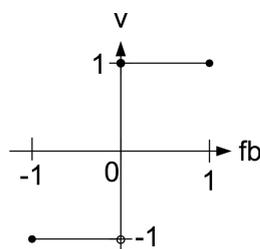


Figure 5.20: Deriving binary evidence from continuous feedback

the interaction is negative $fb = -1$. Then, the updated opinion after the interaction is $(r_{c_{new}}^A, s_{c_{new}}^A)^{rs} = (16, 9)^{rs}$.

5.4.2.2 Updating the Opinions about the Recommenders

The update of the opinions about the recommenders is performed according to the accuracy of their recommendations. As there are multiple ways to define the accuracy of a recommendation, two approaches are proposed.

Considering only the last interaction

In order to evaluate the accuracy of a recommendation, the first approach considers only the outcome of the last interaction with the selected candidate. Therefore, the average rating t_c^B of the recommendation of a recommender B about the candidate C is compared with the feedback with which A rated the interaction. If both have the “same tendency”, then the recommendation is supposed to be accurate and the opinion of A about B as recommender is updated positively; otherwise, there is a negative update.

This can be described more formally as follows: If the opinion of A about entity B as recommender was $o_B^A = (r_{B_{old}}^A, s_{B_{old}}^A)^{rs}$ before the interaction and the recommendation by B about the candidate C was $o_c^B = (t_c^B, c_c^B)^{HTI}$ with $(c_c^B > 0)$, and A 's feedback for the interaction with C is fb , the value of v is calculated as:

$$v := \begin{cases} 1 & \text{if } (2 * t_c^B - 1) * fb > 0 , \\ -1 & \text{if } (2 * t_c^B - 1) * fb < 0 , \\ 0 & \text{else .} \end{cases} \quad (5.39)$$

The update of o_B^A is done using v in Equation 5.3 (with $g = 1$). For example, if the average rating t_c^B of the recommendation is in $]0, 1]$ and the interaction was positive ($fb > 0$), then the recommendation is considered to be accurate ($v = 1$), and the positive evidence of the opinion about the recommender is increased by 1; the negative evidence is kept unchanged.

If the behavior of C as interactor depends on the initiator of the interaction, and C shows different interaction behavior towards B and A , the recommendations of B will be misleading, and A will negatively update the recommender trust for B . This is due to the fact that A is not capable of distinguishing between whether B intentionally provided misleading information or C 's interaction behavior is interactor dependent.

Considering the direct evidence

The update function above might still be considered to be rather simplistic, as a recommendation is considered to be accurate when it has the same tendency as the last interaction. It does neither consider whether the recommendation has the same tendency as the complete history the initiator has collected about the candidate nor the certainty of the recommendation. This can be overcome when considering the difference between the average rating $t_{c_{new}}^A$ of entity A 's direct evidence after the interaction and the average rating t_c^B provided by B . The value of v is then calculated as:

$$v = -2 \cdot |t_{c_{new}}^A - t_c^B| + 1 \quad (5.40)$$

The update is computed as introduced in Section 5.2.2.1 using v and $w = c_c^B$ in Equation 5.3. Thus, the update also reflects the certainty of the provided opinion.

Figure 5.21 shows the relation between the parameter v and the recommendation provided for both approaches proposed above.

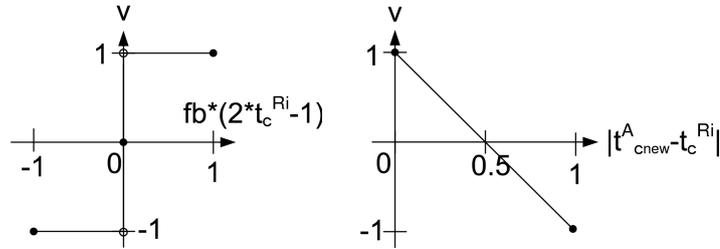


Figure 5.21: Determining the accuracy of a recommendation: Considering only the last interaction (left); considering the direct evidence (right)

The difference between both update mechanisms can be shown when continuing the example that has been introduced in Section 5.4.2.1. Before the interaction the opinion of entity A about the candidate C was $(r_{c_{old}}^A, s_{c_{old}}^A)^{rs} = (16, 8)^{rs}$, and the feedback for the interaction is negative ($fb = -1$). The updated opinion after the interaction is $(r_{c_{new}}^A, s_{c_{new}}^A)^{rs} = (16, 9)^{rs}$. Assume the recommender B provided the recommendation $(r_c^B, s_c^B)^{rs} = (16, 9)^{rs}$ and the trust of A in B in the context of providing recommendations is $(r_B^A, s_B^A)^{rs} = (6, 1)^{rs}$. If the opinion of the recommender is updated only based on the outcome of the last interaction, then the update will be negative, i.e., the

new opinion of A about B is given as $(6, 2)^{rs}$. This seems counter-intuitive as the recommendation provided by B about the behavior of C is equal to the direct evidence of A . This is overcome by the second approach. Here, as the direct evidence of A is equal to the recommendation provided by B , it holds $v = 1$. This leads to a positive update of A 's trust in B as recommender.

5.4.3 Community-based Update of Dispositional Trust

As introduced in Section 5.2, the dispositional trust of an entity is supposed to be a context-dependent parameter. The parameter can either be initialized according to an entity's preferences or it might be dynamically assessed according to the behavior of the entities encountered in the considered context.

In the following, a mechanism for assessing the base trust f that specifies the trust value of unknown entities is proposed. The weight w of the dispositional trust, e.g., $w = 1$, is assumed to be static.

The basic idea of the proposed update mechanism is that as an entity's experience within an application context grows, it is reasonable that it dynamically updates its initial expectation about the typical behavior of entities within the context. As the entities that interact with each other within a specific context may be referred to as a community, the approach is referred to as *community-based dispositional trust*. The update mechanism is applied to the base trust $f^{C(I)}$ of an entity within the context of interactions and to the base trust $f^{C(R)}$ within the context of recommendations in a similar manner.

The following shows how the *community factor* cf is derived in the context of recommendations and interactions. The value of the community factor can be used to replace the static value of the base trust f in the considered context.

5.4.3.1 Updating the Dispositional Trust for Recommenders

The update mechanism for the community factor in the context of recommendations is provided in pseudo code.

```

r = 0;
s = 0;
for (Entity B: KnownRecommenders) {
    v = t_B^A;
    r = r + v;
    s = s + (1 - v);
}
o_cf = (r, s)^{r,s};
cf^{C(R)} = E_{0.5,w,N}^{Beta}(o_cf);
f^{C(R)} = cf^{C(R)}

```

The term “KnownRecommenders” refers to all recommenders that are known to entity A , i.e., the group of recommenders that provided recommendations. The specification of the algorithm shows that the experience with each recommender has limited influence on the community factor $cf^{C(R)}$, i.e., as it holds $r + s = 1$ per recommender. The parameter N is used to express how many recommenders must contribute to the value of $cf^{C(R)}$ that the value is supposed to be representative and no longer influenced by the prior.

5.4.3.2 Updating the Dispositional Trust for Interactors

The update mechanism for the interactors is similar. It is also given in pseudo code.

```

r = 0;
s = 0;
for (Entity c: known Interactors){
    v = t_c^A · (1 - c_c^A · d);
    r = r + v;
    s = s + (1 - v);
}
o_cf = (r, s)^{r,s};
cf^{C(I)} = E_{0.5,w,N}^{Beta}(o_cf);
f^{C(I)} = cf^{C(I)}

```

The factor d is introduced to counterbalance the selection strategy. For example, assume an entity chooses $N = 20$ and knows 20 entities that only provided interactions with positive outcomes. As the entity would prefer to only interact with those 20 entities, the community factor $cf^{C(I)}$ of this entity would reach 1. This might not be representative for the typical behavior of the entities in the community, as the selection of the interaction partner is not random but strongly depends on the selection strategy. The factor d reduces the positive influence of opinions with increasing certainty. Using the proposed strategy to counterbalance the selection strategy (see forth line in the pseudo code above), the value v depends on the values of t and c as shown in Table 5.3. The table shows the relation only for the extreme values of t and c .

		c	
	t	0	1
0		0	0
1		1	$v(d)$

Table 5.3: Influence of the factor d on v for selected values of t and c using $v(d) = t \cdot (1 - c \cdot d)$

The value $v(d)$ indicates that the value of v not only depends on the values of t and c , but also on the choice of $d \in [0; 1]$ as shown in Table 5.4. The lower the value of d the higher is the positive influence of an interactor that has a high average rating t and a high certainty c ; using $d = 0$ is equivalent to not counterbalancing the selection strategy.

d	$v(d)$
1	0
$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{4}$	$\frac{3}{4}$
0	1

Table 5.4: Influence of the factor d on v in case of $t = c = 1$ using $v(d) = t \cdot (1 - c \cdot d)$

5.4.4 Summary

This section presented update mechanisms for the trust in the interactor, the recommenders, and for the dispositional trust.

- The update of the trust in the interactor is carried out using the feedback provided after an interaction for deriving binary evidence.
- The update of the trust in the recommenders is done based on the accuracy of their recommendations. Here, two approaches have been proposed. The first one evaluates the accuracy of the recommendation only by considering the outcome of the past interaction. The second one considers a recommendation to be accurate based on its similarity to the initiator's direct evidence. The evaluation in Section 6.1.5.2, shows the comparison of both approaches. The approaches for deriving evidence from the accuracy of a recommendation are similar to the ones proposed in [BLB04, TPJL06].
- The approach for a community-based calculation of dispositional trust is similar to the approach presented in [Jøs07, JLC08]. However, the approaches differ as in the approach proposed in this thesis, the community factor is calculated based on the average rating, whereas in [Jøs07, JLC08], the *community base rate* is calculated based on the expectation values of the encountered entities. The reason for calculating the community factor based on the average ratings of the encountered entities is that the average ratings are not influenced when the community factor is updated. This is considered to be a significant advantage, as when calculating the community factor based on the

expectation values (i.e., the trust values in the HTI), a change of the community factor would result in a change of the expectation values. Thus, the community factor and the expectation value would amplify each other.

- The update mechanism of the dispositional trust for interactors has additionally been extended in order to counterbalance the selection strategy. As the initiator decides to select the best interaction partners, the simple average of the average ratings of the known interactors might not be representative for the behavior of unknown entities. The evaluation of the impact of the property is shown in Section 6.1.5.2.

5.5 Conclusions

This section provided a new trust model that addresses the most important aspects regarding the representation, computation, establishment and update of trust.

- The *representational model* provides means for deriving trust based on evidence from past interactions and context-dependent parameters. Especially, the Bayesian representation, which is the basis for a number of trust models as described in Section 3.2.4.2, has been extended in order to overcome limitations that are usually introduced with aging (see Section 5.2.3.1). Furthermore, a new representational model, the Human Trust Interface (HTI), provides for a simple set of parameters and a graphical representation of trust. Along with the provided mapping, the provided representations are major contributions as they provide means for overcoming limitations of state-of-the-art trust models, and they allow users and developers to think about trust in the representation they prefer.
- The *computational model* provides means for robustly aggregating direct evidence and recommendations. A major concept here is to derive the trustworthiness of recommenders based on the accuracy of their past recommendations. The proposed approach is especially robust to Sybil attacks as it prefers direct evidence and recommendations by more trusted recommenders over the recommendations by lower trusted recommenders and limits the influence of a single recommender. Finally, it provides a new approach for limiting the influence of recommenders based on their ranking.
- The *update mechanisms* provided for deriving evidence about the trust in the selected candidate, as well as in the recommenders, are based on the feedback after an interaction. The mechanism for calculating the base trust value based on the encountered entities provides a means to

consider the number of encountered entities as well as counterbalancing the effects of the selection strategy.

As the trust model allows for the interpretation of trust as probability, it can easily be integrated in decision making. Therefore, beyond simply choosing the best candidate available, the integration in utility based decision making is possible. Thus, the proposed model complies with the design goals that have been introduced at the end of the concept chapter in Section 4.4.

The major limitations of the proposed approach can be described as follows. As interactions within an application context are assumed to be homogenous, the presented model does not focus on interactions with different values. In [JI02], it has been proposed to model interactions with different values using different weights (see also Section 5.2.2.1), however, this needs especially to be considered when choosing the parameters of aging and the maximum number of expected evidence units in order to prevent unintended effects. Furthermore, the proposed approach only considers a single feedback value per interaction that describes the overall outcome. It does not propose means for the integration of fine-grained ratings that refer to different aspects of the interaction, e.g., in the context of information exchange, one might think about rating aspects separately, like correctness of the information and the time needed for providing the information. Finally, although the model has been shown to cope with different kinds of attacks, the approach is susceptible when trusted entities suddenly turn malicious. This is not a shortcoming that is special to the proposed approach, but it is typical for evidence based trust models. However, the introduction of aging prevents long-term exploitation of previously established trust.

The evaluation of the trust model's performance in a distributed environment and of the intuitivity of the HTI is presented in the next chapter.

Chapter 6

Evaluation

This chapter shows the evaluation of the trust model that has been motivated and introduced in the previous chapters. Note that the evaluation of improvements regarding aging has already been presented in Section 5.2.7, and the argumentation for the improved robustness to Sybil attacks has been presented in Section 5.3.4. This chapter is organized as follows:

1. Section 6.1 shows the impact of the proposed trust model in a simulation of an opportunistic network using real world user traces as a basis of the mobility model.
2. Section 6.2 presents the results of a user study evaluating the usability of the graphical representation in comparison with an Amazon-like stars interface (see Section 3.2.1.2) and the Opinon Triangle (see Section 3.2.5.1).
3. Section 6.3 presents the results of a user study evaluating the application of the trust model in an online movie recommendation platform.

6.1 Evaluation of CertainTrust in an Opportunistic Network

The evaluation of the impact of the proposed trust model CertainTrust is based on a scenario in which humans with mobile devices share music in an opportunistic network as introduced at the beginning of Chapter 2 and in Section 4.1. In this scenario, users with their mobile devices are referred to as entities. As these entities move, they will meet other entities, and interact with each other, e.g., they exchange music (mp3 files) or recommendations, in a spontaneous manner. Due to different goals or motivations, the users will show different behaviors when providing files to others. The goal of a typical user is to interact only with trustworthy interactors, i.e., interactors from

which they expect to receive a good file (correct file, no viruses, complete song, and expected quality).

It is assumed that a user appreciates the support of a trust model, which supports them with information about the trustworthiness of the available candidates for an interaction, or is even capable of making decisions and interacting on its own. This will be especially true, if it allows the quality of a user's interactions, i.e., the number of received good files, to increase.

After an interaction, the *quality of the interaction* is reported through *feedback*. The determination of the feedback values does not necessarily require user interaction. In some cases this could be done automatically, e.g., by scanning the mp3 file for viruses, checking the size, the bit rate, and noise.

6.1.1 Basic Types of Behavior, Population Mixes & Settings

According to the system model that has been introduced in Section 4.2 entities may be recommenders or interactors. In both roles, an entity can be good (+) or bad (-). A good interactor provides good interactions, leading to positive feedback ($fb = 1$), a bad interactor provides interactions leading to negative feedback ($fb = -1$). A good recommender provides recommendations that reflect its real experience. The model for bad (lying) recommenders is derived from [TPJL06]. Bad recommenders try to provide recommendations with a maximum misleading expectation value, i.e., if $E_{simple}^{Beta}(o_c^B)$ is the expectation value calculated by recommender B for interactor C based on its direct evidence, the recommendation of B would be an opinion with the expectation value $1 - E_{simple}^{Beta}(o_c^B)$. This can be achieved by switching the positive and negative evidence. Thus, four basic types of behaviors are identified, see Figure 6.1.

Basic entity behaviors		Recommendation behavior	
		+	-
Interaction behavior	+	honest (h)	selfish (s)
	-	malicious (m)	worst (w)

Figure 6.1: Basic entity behaviors

Combining these basic types of behaviors leads to 15 canonical population mixes: h , m , s , w , hm , hs , hw , ms , mw , sw , hms , hsw , hmw , msw , and $hmsw$. The percentage of entities with a specific behavior within a population is set to be equal. For example, the population mix h contains only entities with honest behavior; the population hm contains 50% entities with honest behavior and 50% malicious, and so on.

The assumption that the interaction behavior of an entity is stable, in the sense that it is either only positive or only negative may be too simplistic. Therefore, an additional parameter called *stability* y is introduced. This parameter allows the adherence of an entity to its assigned behavior to be described. In the case of stability $y = 1$ an entity totally adheres to its assigned interaction behavior. In the case the stability of entity is set to 0.9 it adheres only in 90% of its interactions to the behavior it has been assigned, in the other 10% it will do the opposite. Given the stability of an entity and its assigned behavior, one can derive the probability with which an entity provides interactions with positive outcomes. For simplicity, it is assumed that the stability only influences the interaction behavior, the behavior in the context of providing recommendations is assumed to be stable.

Finally, the evaluation is based on two different settings per population mix. In the first setting, called the *deterministic setting*, the stability factor is $y = 1$ for all entities. Thus, for each entity A in a population holds $y = y_A = 1$. In the second one, called the *probabilistic setting*, a randomly chosen stability parameter y_A is assigned to each entity A in the population. For each entity A the parameter y_A is randomly and uniformly distributed chosen from the interval $y_A \in [0.5; 1]$.

For example, assuming the *deterministic* setting and the population *hm* leads to a population in which 50% of all entities provide only good interactions and 50% provide only bad interactions. Using the same population but the *probabilistic* setting, the probabilities for good interactions over all entities are uniformly distributed in $[0; 1]$.

6.1.2 Simulation

As the goal of this thesis is to develop a trust model for ubiquitous computing, the scenario of the evaluation comes from this field - to be more specific from the field of opportunistic networks. In this scenario the possibility of interactions depends on the spatial proximity of people. Therefore, it is important to have realistic user traces, i.e., a realistic mobility model, as this is a basic influence factor on the results of the simulation.

6.1.2.1 User Traces

The presented simulation is based on user traces which have been collected in the Reality Mining project [EP06]. The data provides information about 97 users of mobile phones and their location. The latter is given as the ID of the cell tower the mobile phones were connected to.

The data used for the simulation is only a subset of the complete data set. It has been collected in a week in which a big number of users were connected to a small number of cell towers. Thus, it is expected to have a big number of possible interactions. Based on [Hei07], it is assumed that a group of users

is in proximity to each other if the users are connected to the same cell tower within a 15 minute time interval. For the evaluation, a so-called *meeting* happens when six or more users are connected to the same cell tower in the same time interval. This allows the trust model's capabilities in selecting the most trustworthy candidate from a set of candidates to be evaluated. The set of candidates is determined randomly as half of the available entities, i.e., an initiator has at least 3 candidates for an interaction. In the restricted data set, there are 68 distinct users (entities), which met each other in 556 meetings. In average an entity took part in 59.94 meetings, and met 46.76 distinct entities. The average number of entities per meeting is 7.33. In one run of the simulation, the meetings of this week are consecutively repeated 3 times in a row, in order to evaluate the performance of the trust model over a longer period.

The simulation is done for all 15 populations introduced in Section 6.1.1, each in the deterministic (stability $y = 1$) and the probabilistic setting (stability $y \in [0.5; 1]$). Each simulation was repeated 20 times per trust model and population mix using the same seeds for the comparison of the different models and baselines.

6.1.2.2 Meeting Procedure

A time interval in which a group of people meet is called a *meeting*. During a meeting entities may interact with each other and provide recommendations to others. Each meeting proceeds as follows: In each meeting each participating entity has to interact with one candidate, i.e., each entity is the initiator of one interaction. The candidates for an interaction are randomly chosen from half of the entities which are part of the meeting, i.e., half of the entities in the meeting can provide a specific mp3-file. If the trust model includes recommendations, the initiator asks all entities that are part of the meeting for providing recommendations about the candidates. Then, the initiator evaluates the trustworthiness of the candidates, and selects the most trustworthy one, i.e., the one with the greatest trust value. This setting was chosen in contrast to a setting in which each entity has the choice whether to interact or not, since the evaluation is to show the impact of the trust model without the additional influence of a decision making component. After each interaction, the initiator updates the opinions about its interaction partner (selected candidate) and, in the case this is part of the model, opinions about the recommenders.

6.1.3 Baselines and Models

The first baseline is the *Random* strategy. This strategy selects the partner for the interaction randomly. Furthermore, it assigns a trustworthiness to each entity that is randomly (and uniformly distributed) chosen from the

interval $[0; 1]$.

The *Perfect* strategy always selects the best candidate based on the behavior (and stability) an entity has been assigned by the simulation environment. In a way, this is similar to a “best possible” selection process that one could apply for the selection of an interaction partner in a hypothetic world, in which all entities have labels on their foreheads stating their behavior (and the probability for providing a good interaction).

Furthermore, the evaluation compares the results of several variants of CertainTrust in order to show the impact of different parameters and features, and it compares the results of CertainTrust with variants of the Beta Reputation System (see 3.2.4.2 and [JI02]). The trust models are denoted and configured as follows:

1. There are different variants for configuring the proposed trust model CertainTrust (CT). In the following the notation *CT_C* is used to refer to the trust model that uses the following parameters:
 - Representational model:
 - Dispositional trust:
 - * Base trust value $f = 0.5$
 - * Weight of dispositional trust $w = 1$
 - Aging factor $a = 1$ (no aging - as the entities do not change their behavior over time)
 - maximum number of expected evidence units $N = 20$
 - Computational model with improvements regarding the robustness to Sybil attacks (as proposed in Section 5.3.4) using $t_e = t_s = 0.5$ and $N_R = N$
 - Update mechanism: Simple update mechanism (considering only the last interaction) for trust in recommenders as proposed in Section 5.4.2.2
 - Community-based update of the dispositional trust without a bias (i.e., $d = 0$).

Whenever the evaluation compares variants of CT, the differences from the configuration of *CT_C* are explained and a notation is introduced.

2. *Beta_S*: The Beta Reputation System was proposed in [JI02] (see also Section 3.2.4.2 and Section 3.2.5). Since the goal of this thesis is to provide a trust model for ubiquitous computing, a distributed variant of this reputation system is used in which each entity is its own reputation centre. The reputation centre stores only direct experience. Yet, entities can exchange recommendations with all entities that are part of the current meeting. The expectation value for an interaction

partner is calculated using the consensus operator as proposed in [JI02] for combining the direct experience with the available recommendations. This variant of the Beta Reputation System does not discount recommendations.

3. *Beta_D*: This variant of the Beta Reputation System differs from the Beta_S in the point that it discounts recommendations. The discounting is done as proposed in [JI02]. In this approach the discounting uses the trustworthiness of an entity in the context of interactions for the discounting (weighting) of the provided recommendations.
4. *Ext_Beta*: Ext_Beta provides an extension by the author of this thesis to the Beta Reputation System that has not been published before (to the best of the author's knowledge). The extension provides for an evaluation of the trustworthiness of an entity in the context of recommendations as proposed in Section 5.4.2.2, i.e., based on the accuracy of its past recommendations. Then, the discounting operator that has been proposed in [JI02] is used to weight the recommendation of an entity according to its trustworthiness in the context of providing recommendations. This variant is introduced to compare the effects of weighting recommendations according to the trustworthiness of the recommender's behavior as interactor as proposed in [JI02] and weighting recommendations according to the trustworthiness of the recommender's behavior as recommender (based on the accuracy of its past recommendations) as proposed in this thesis.

6.1.4 Evaluation Metrics

For the evaluation, the following notation is introduced. The set of entities in a population is denoted by P . The number of entities in the population P is $|P|$. For an entity $B \in P$ the characteristic probability for providing a good interaction is denoted as p_B . This probability can be derived from the behavior of an entity and the stability factor that has been assigned to each entity at the beginning of the simulation (as described in Section 6.1.1). There are four types of different behaviors $Behavior \in \{honest, malicious, selfish, worst\}$. For an entity A to whom behavior $Behavior$ was assigned holds $A \in Behavior$, e.g., for an entity A to whom the behavior "honest" was assigned holds $A \in honest$. The evaluation is done using the following metrics (based on [TPJL06,SVB06]):

1. In order to evaluate the performance of a trust model when estimating the trustworthiness of an entity in the context of interactions, the first metric measures the *average error in estimating the trustworthiness*.

For entity A the mean absolute error $err(A)$ that entity A makes when estimating p_B using $E(\sigma_b^A)$ for all entities B in the population P is

defined as:

$$err(A) = \frac{\sum_{B \in P} |E(o_b^A) - p_B|}{|P|} \quad (6.1)$$

For the calculation of $E(o_b^A)$ entity A may ask all entities in P for recommendations. The *average error in estimating the trustworthiness* avg_{err} is defined as:

$$avg_{err} = \frac{\sum_{A \in P} err(A)}{|P|} \quad (6.2)$$

The average error should be close to 0.

2. The second metric allows one to measure the average reputation of entities. The reputation $R(A)$ of an entity A is defined as the average of the expectation value calculated by each entity B in the population P for entity A :

$$R(A) = \frac{\sum_{B \in P} E(o_a^B)}{|P|} \quad (6.3)$$

Again, an entity may ask all entities in P for recommendations. As the average reputation over all entities in the population depends on the population mix, it is only calculated for the entities that have been assigned the same type of behavior (*Behavior*), e.g., honest or malicious.

The *average reputation* of entities of the same type of behavior is defined as:

$$avg_R(Behavior) = \frac{\sum_{A \in Behavior} R(A)}{|A \in Behavior|} \quad (6.4)$$

3. The third metric provides information about the (perceived) quality of interactions by an entity. Note that in the simulation the feedback, which may be seen as the perceived quality of an interaction, is equal to the outcome of the interaction. As it is assumed that positive outcome always leads to positive feedback (+1) and negative outcome to negative feedback (-1), the sum of the collected feedback indicates the average quality of the interactions an entity has achieved in the simulation.

The accumulated sum of feedback (*acc_sum*) is calculated for each entity as sum of the feedback over its past interactions. For an entity A that uses trust model *model_X* and has had n interactions with the feedback $(fb(1), \dots, fb(n))$ the accumulated sum of feedback $acc_sum(A, model_X)$ is calculated as:

$$acc_sum(A, model_X) = \sum_{i=1}^n fb(i) \quad (6.5)$$

This value strongly depends on the population mix. In a population with stability $y = 1$ and only honest entities, there are only positive interactions; in a population with only malicious entities, there are only negative ones. Therefore, the *average percentage of the accumulated sum of feedback* is introduced. It is defined as the portion of the accumulated sum of feedback that has been achieved using the considered trust model relative to the accumulated sum achieved using the *Perfect* selection strategy:

$$avg_{acc_sum}(model_X) = \frac{\sum_{A \in P} acc_sum(A, model_X)}{\sum_{A \in P} acc_sum(A, Perfect\ strategy)} \quad (6.6)$$

The closer the *average percentage of the accumulated sum of feedback* is to 1.0 (assuming that 1.0 is its maximum value), the more positive interactions an entity has had, and the quality of interactions that has been achieved using the trust model *model_X* is the closer to the result that has been achieved by the *Perfect* selection strategy.

6.1.5 Results

The evaluation shows the impact of the parameters on the proposed model, as well as the comparison to distributed variants of the Beta Reputation System and two baselines, i.e., the *Random* strategy and the *Perfect* strategy.

Besides the first example that shows the evaluation of the reputation over time in a selected population mix, the evaluation presents the results that have been calculated at the end of the simulation for each population mix. The evaluation of the trust models across different population mixes allows one to evaluate whether the trust model is appropriate for a wide range of populations with entities showing different behaviors or whether it may only be applied to a few population mixes.

6.1.5.1 Average Reputation Evaluation over Time

The Figures 6.2, 6.3, 6.4, and 6.5 show the evaluation of the reputation in the population *hmsw* for different trust models and settings. In the *deterministic* setting (stability $y = 1$) the *true*¹ average reputation of honest and selfish entities would be 1, for malicious and worst entities it would be 0. In the *probabilistic* setting (stability $y \in [0.5; 1]$) the *true* average reputation of honest and selfish entities would be 0.75 and for malicious and worst entities it would be 0.25.

As one sees from Figure 6.2 the proposed trust model CT_C is capable of detecting the different behaviors of the entities in the context of interactions,

¹The *true* average reputation refers to the average reputation that would have been calculated based on the probabilities for providing interactions with positive outcomes that have been assigned to each entity by the simulation environment.

i.e., good interactors and bad interactors. This is also true for the variant of the proposed model CT_None (see Figure 6.3). In contrast to CT_C the variant CT_None does not use the community-based update of the base trust value f . It uses a static base trust value $f = 0.5$. Note that in both figures the graphs of *honest* and *selfish* entities and of *malicious* and *worst* entities are very similar, as they have the same behavior in the context of interactions. Therefore, they are hardly distinguishable in the figures.

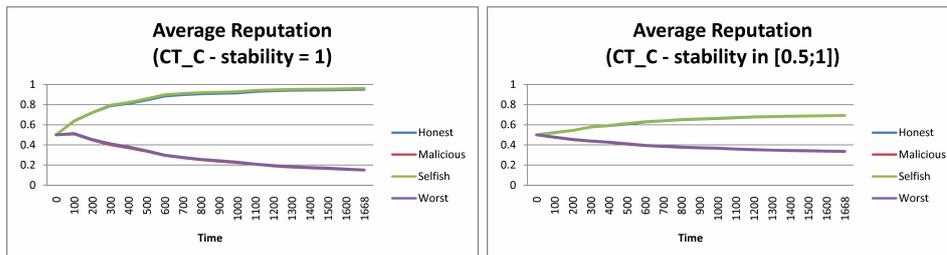


Figure 6.2: Reputation evaluation over time in population *hmsw* using CT_C

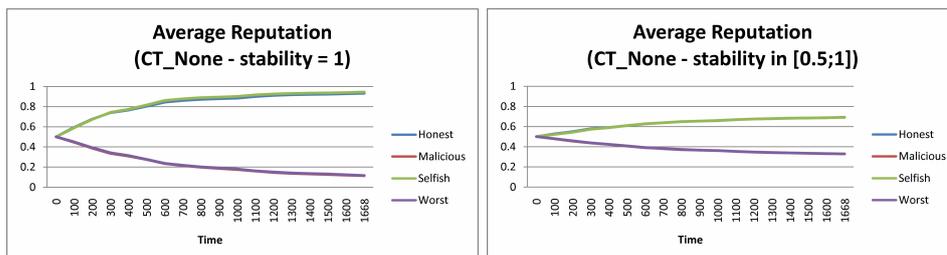


Figure 6.3: Reputation evaluation over time in population *hmsw* using CT_None

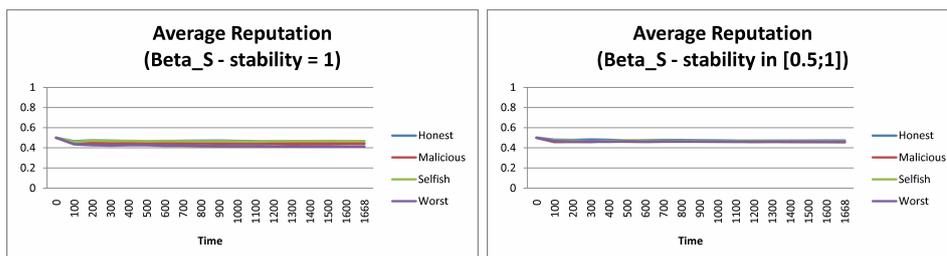


Figure 6.4: Reputation evaluation over time in population *hmsw* using Beta_S

In contrast, the reputation values calculated by Beta_S and Beta_D do not allow for a clear distinction between entities with different behavior. This can be explained as the population contains 50% entities providing misleading recommendations and Beta_S (see Figure 6.4) gives the same weight to all recommendations.

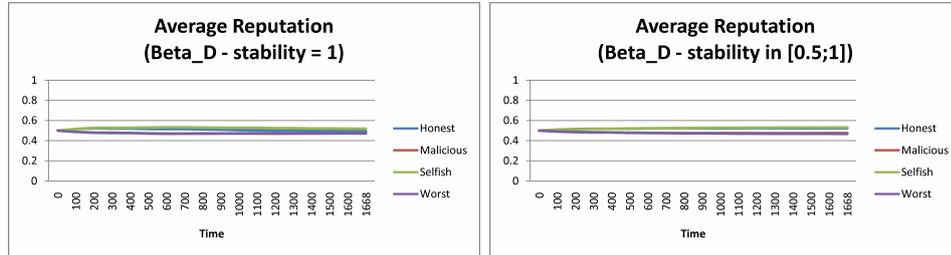


Figure 6.5: Reputation evaluation over time in population *hmsw* using Beta_D

The variant Beta_D (see Figure 6.5) suffers from the fact that its discounting mechanism is based on the assumption that an entity’s behavior in the context of interactions is the same as in the context of recommendations. The assumption does not hold in this population.

6.1.5.2 Evaluation of Variants of CertainTrust

In this section, the performance in estimating the trustworthiness of different variants of CertainTrust is evaluated across the different population mixes and settings.

Computational Model

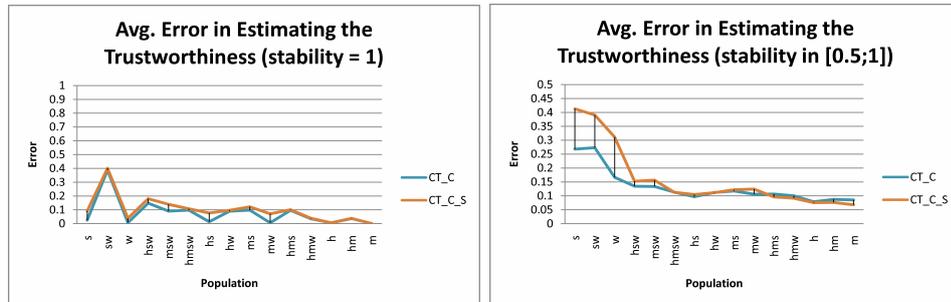


Figure 6.6: Variants: Computational model - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

Figure 6.6 shows the evaluation of the average error in estimating the trustworthiness of an entity for the standard variant CT_C and a variant CT_{C_S} using the (“more robust”) variant of the computational model that has been introduced in Section 5.3.3. CT_C has advantages when the percentage of accurate recommenders is 50% or less. This may be expected as the computational model of CT_C uses additional mechanisms for dealing with misleading recommendations. In the probabilistic setting there are

improvements in the range of 10% to 20% in the populations s , sw , and w . It is especially worth to note that there is no decline that is similarly significant in other populations.

Update Mechanism

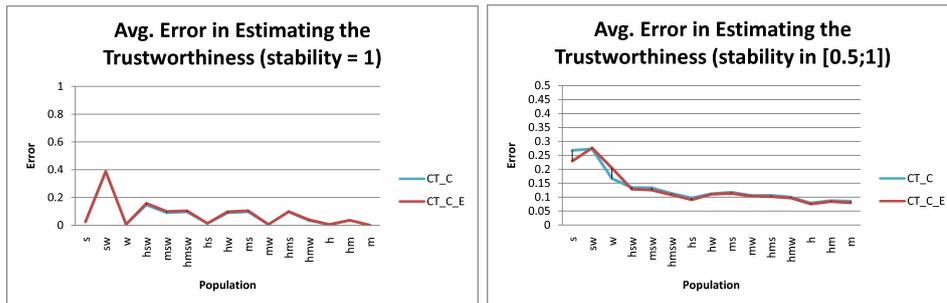


Figure 6.7: Variants: Update function - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

Figure 6.7 shows the evaluation of the average error in estimating the trustworthiness of an entity for the standard variant CT_C and a variant CT_C_E using the alternative approach (“considering the direct evidence”) for the update of the trust in recommenders as proposed in Section 5.4.2.2.

The evaluation shows that the exchange of the update mechanism does not lead to major differences. This may be due to the fact that entities try to interact with the best entities, i.e., entities that provide interactions with positive outcomes with a probability close to 1.

Maximum Number of Expected Evidence Units

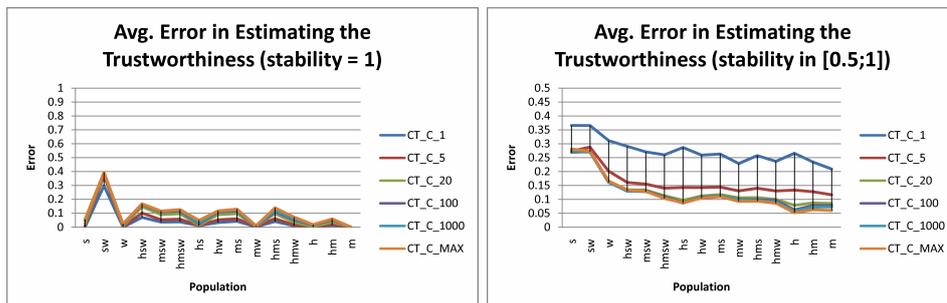


Figure 6.8: Variants: Maximum number of expected evidence units - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

Figure 6.8 shows the evaluation of the average error in estimating the trustworthiness of an entity using different parameters for the maximum

number of expected evidence units N . The standard variant CT_C is equivalent to CT_C_20. The variant CT_C_1 uses $N = 1$, CT_C_5 uses $N = 5$, and so on. The variant CT_C_MAX models $N \rightarrow \infty$ by setting $N = 2147483647$ (which is equivalent to Integer.MAX_VALUE in Java). As the simulation does not consider aging, the choice of $N \rightarrow \infty$ would be justified by the representational model. However, as in the current implementation N is assumed to be equivalent to N_R , the usage of $N = N_R = 20$ has been chosen for the evaluation.

The evaluation shows that in the deterministic setting (stability = 1) smaller values of N have advantages over higher values. In the probabilistic setting, it is the other way around as expected. Yet, it is interesting to note that besides the two smallest values $N = 1$ and $N = 5$, the results of the different variants are quite similar.

The comparison overall populations in both settings shows that the average difference in the error in estimating the trustworthiness between CT_C_20 and CT_C_MAX is less than 1% - (average error CT_C_20=10.37% and CT_C_MAX=11.06%), which may be due to the limitation of $N = N_R$ of the current implementation.

Community-based Update of Dispositional Trust Compared to Static Value of f

Figure 6.9 shows the evaluation of the average error in estimating the trustworthiness of an entity for the standard variant CT_C and a variant CT_None that does not update the dispositional trust based on the experienced behavior. Instead CT_None uses a static value for the base trust value $f = 0.5$. CT_None_S refers to the variant using a static value $f = 0.5$ and the computational model that was proposed in Section 5.3.3.

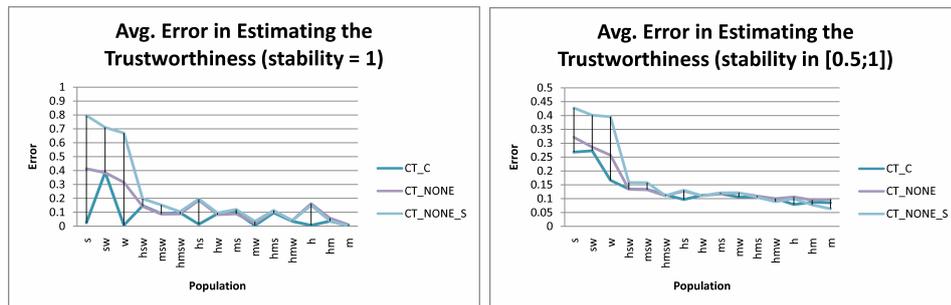


Figure 6.9: Variants: Community-based update of base trust f compared to static value of $f = 0.5$ - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

The evaluation shows that the variant CT_C has advantages in populations that provide mostly misleading recommendations (populations s , sw ,

and w). Furthermore, it has advantages the populations hs and h in which interaction partners usually provide good interactions.

It is worth noting that the static value of $f = 0.5$ is the optimal choice for the community factor in the populations hm , $hmsw$, hw , ms , and sw (see “Expected Com_Factor” in Figure 6.11).

Community-based Update of Dispositional Trust Using Different Values d to Counterbalance the Selection Strategy

Figure 6.10 shows the evaluation of the average error in estimating the trustworthiness of an entity for variants using different values for the factor d that is introduced to counterbalance the selection strategy (see Section 5.4.3.2). The variants are referred to as CT_C_d , where CT_C_0 is equal to CT_C .

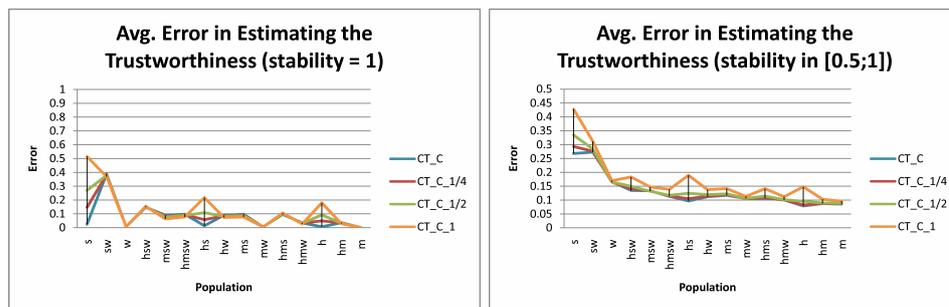


Figure 6.10: Variants: Community-based update of base trust - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

The evaluation shows that the most noticeable differences occur in the populations s , hs , and h . In these populations, there are mostly entities providing interactions with positive outcomes. As the mechanism that has been proposed in order to counterbalance the selection strategy has its major impact when an entity interacts repeatedly with a good interactor, this was expected.

When comparing the other populations one sees that the different variants CT_C (CT_C_0), $CT_C_1/4$, and $CT_C_1/2$ provide similar results. Only the variant CT_C_1 provides significantly worse results in the probabilistic setting.

For a deeper analysis the average community factor that is calculated by all entities is analyzed. The average community factor is equal to the average base trust value (see Section 5.4.3). It is calculated as the average of the community factor over all entities in the population.

Figure 6.11 shows the average community factor in the context of interactions and Figure 6.12 shows the average community factor in the context

of recommendations. The value “Expected Com_factor” is the community factor that is expected based on the knowledge of the population.

Figure 6.11 clearly shows the impact of the different variants on the average community factor in the context of interactions. The variant CT_C_1 obviously produces too negative values. Although the average community factor that is calculated using CT_C (CT_C_0) is too positive in many populations in the deterministic setting, it achieves good results in the probabilistic setting. All in all, the community factor calculated by CT_C_0 approximates the value “Expected Com_factor” best.

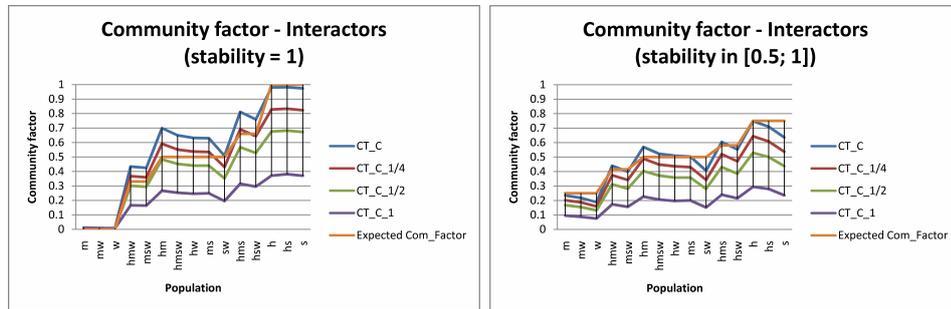


Figure 6.11: Community factor in the context of interactions - Populations are sorted according to the percentage of good interactors (the lines are only for ease of reading)

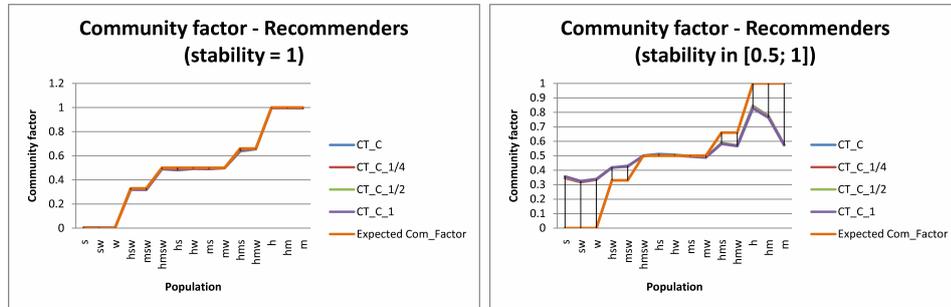


Figure 6.12: Community factor in the context of recommendations - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

Figure 6.12 shows the evaluation of the average community factor in the context of recommendations across the different population. From the figure, one sees that in the deterministic setting the expected community factor is well approximated. In the probabilistic setting the difference between the expected value and the calculated value is noticeable in most populations. This is due to the fact that the behavior of the interactors is probabilistic and the simple variant for the update of the trustworthiness of recommenders (see Section 5.4.2.2) does not account for this. Although the recommenders

in the population h and m provide accurate recommendations, there is a difference in the calculated average community factor. This is probably due to the fact that in the population h there are mostly good interaction partners and accurate recommenders. Thus, there are many recommenders that provide accurate recommendations for the selected interaction partner. This leads to a high value of the community factor. In the population m there are primarily bad interaction partners and accurate recommenders. As recommenders provide accurate but bad recommendations about the entities they know, the initiator usually selects an entity about which only little is known. Thus, there is only little information that may be used in order to update the trust in recommenders.

6.1.5.3 Comparison to Baselines and Other Models

The Figures 6.13 and 6.14 show the comparison of the proposed approach (CT_C) compared to other models and baselines.

Comparison: Error in Estimating the Trustworthiness of an Entity

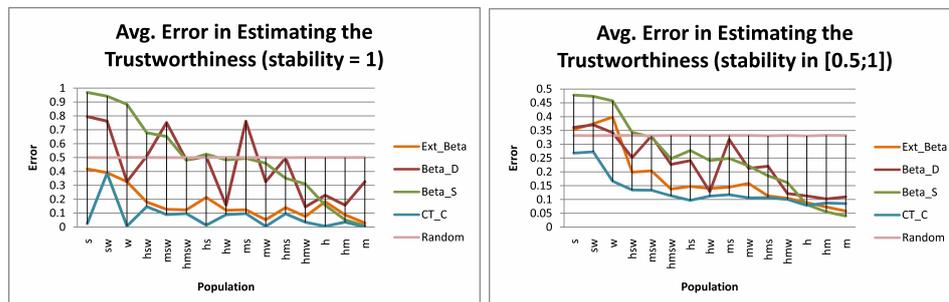


Figure 6.13: Average error in estimating the trustworthiness of an entity - Populations are sorted according to the percentage of good recommenders (the lines are only for ease of reading)

The results of the evaluation are presented in Figure 6.13. The results of the *Perfect* selection strategy are not shown in the graph. As the *Perfect* selection strategy knows the true value of the parameter that defines the behavior of an entity, the error of estimating this parameter is 0.

The *Random* strategy estimates the trustworthiness of an entity randomly. The evaluation shows that the estimation error is independent from the population, but depends on the setting (either deterministic or probabilistic). The results achieved by the *Random* strategy are usually worse than the results by other strategies.

From the comparison of CT_C and Beta_S one can see that CT_C achieves considerable better performance than the Beta_S whenever there are 33% or more entities providing misleading recommendations, i.e., in the populations

s , sw , w , hsw , msw , $hmsw$, hs , hw , ms , mw , hms , and hmw . Only in the probabilistic setting in populations with only good recommenders Beta_S has slight advantages, i.e., in the populations h , hm , and m .

From the comparison of CT_C and Beta_D one can see, that CT_C out-performs in all populations and settings, although Beta_D weights recommendations. The assumption that the trustworthiness of an entity as interactor can be used to weight its provided recommendations does not perform well. Even in populations where the assumption is true the results of Beta_D are worse than the ones produced by CT_C.

In order to evaluate the impact of the concept of deriving the trustworthiness of a recommender from the accuracy of its previous recommendations, the Beta Reputation System has been extended (Ext_Beta) with the update mechanism for the trustworthiness of recommenders as proposed in Section 5.4.2.2. The results show that this extension clearly improves the performance compared to Beta_S and Beta_D. Yet, the extension alone is not sufficient for the achievement of the performance of CT_C.

Table 6.1 shows average error in estimating the trustworthiness across all populations.

Model	Deterministic setting	Probabilistic setting	Average over both settings
CT_C	7.56%	13.19%	10.37%
Ext_Beta	17.32%	17.95%	17.64%
Beta_D	44.93%	22.99%	33.96%
Beta_S	49.66%	25.59%	37.63%
Random	49.99%	33.14%	41.56%

Table 6.1: Average error in estimating the trustworthiness across all populations

Comparison: Average Percentage of the Accumulated Sum of Feedback

Figure 6.14 shows the results of the average percentage of the accumulated sum at the end of each simulation (averaged over all runs) per population and per setting. The populations m , w , and mw are not considered in the figure, as in these populations the results of the *Perfect* selection strategy were negative which would lead to a misleading interpretation.

The *Perfect* selection strategy achieves 100% in each population. The results that are achieved by the *Random* strategy depend on the population as expected. The results are usually worse than the results achieved by the

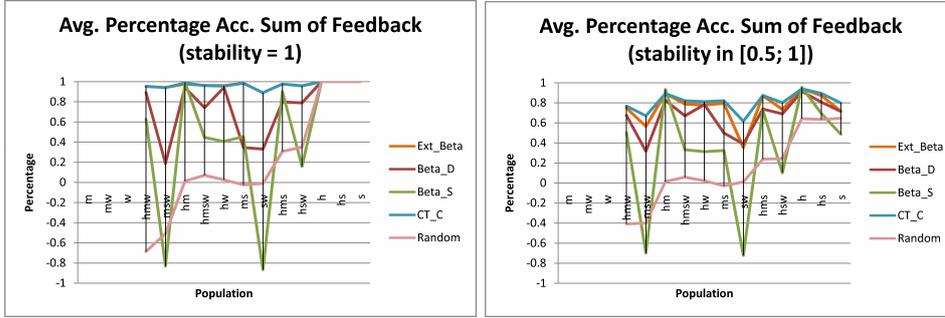


Figure 6.14: Average percentage of the accumulated sum of feedback - Populations are sorted according to the percentage of good interactors (the lines are only for ease of reading)

other models, yet, the *Random* model serves as a baseline which shows the results that are achieved by an entity that does not use any trust model.

Furthermore, one can see that CT_C mostly outperforms Beta_S and Beta_D. In the *deterministic* setting the results of Ext_Beta are very similar to the results of CT_C, therefore it is hard to distinguish the graphs. In the *probabilistic* setting CT_C especially has advantages in the populations *msw* and *sw*.

All in all, CT_C achieves positive results in all populations and settings. It is worth noting that CT_C achieves 89% of the results of the *Perfect* selection strategy, when taking the average of the 24 populations (12 per setting). The worst result of CT_C (61%) is in the population *sw* in the probabilistic setting; considering that all entities in this population try to provide misleading recommendations this is a good result.

The Tables 6.2 show the average accumulated sum of feedback for the selected populations.

Model	Deterministic setting	Probabilistic setting	Average over both settings
CT_C	96.98%	81.97%	89.48%
Ext_Beta	96.97%	76.79%	86.88%
Beta_D	77.10%	69.03%	73.07%
Beta_S	49.44%	38.34%	43.89%
Random	30.00	21.05%	25.52%

Table 6.2: Average percentage of the accumulated sum of feedback for the selected populations, i.e., all populations exclusive the populations *w*, *mw*, and

6.1.6 Summary

In contrast to the simulations in [QHC06, TPJL06, JI02], the presented simulation evaluates the results of different trust models over a wide set of populations and uses a mobility model for the movement of entities. The population mixes are derived from the classification of the basic types of behaviors. As all kinds of combinations are considered the results present a good overview of the performance of the considered models. The mobility model, which determines when entities are in proximity and may interact, is based on real world user traces. This is especially important as the developed trust model is to support collaboration in opportunistic networks and ubiquitous computing applications.

- The results of the average percentage of the accumulated sum of feedback (see Figure 6.13) show that the proposed trust model achieves good results in all considered populations. Compared to the perfect selection strategy *CT_C* is capable of achieving more than 75% of the results reached by the *Perfect* selection strategy in 22 of 24 populations and 89% in average.
- The comparison of *CT_C* to *Beta_S* and *Beta_D* shows that *CT_C* outperforms in most population mixes. This is especially important as a user will evaluate not the trust models itself, but the improvement of the quality of her interactions. The results for average error in estimating the trustworthiness across all populations and both setting are given by *CT_C* 10.37%, *Beta_D* 33.96%, and *Beta_S* 37.63% (see Table 6.1). The results for the average percentage of the accumulated sum of feedback across all populations, but m , w , and mw are given by *CT_C* 89.48%, *Beta_D* 73.07%, and *Beta_S* 43.89% (see Table 6.2). The reason for the improved overall quality of interactions can be seen in the smaller error in the estimated trustworthiness (see Figure 6.14) for most of the populations.
- The extension of the Beta Reputation System that has been introduced by the author of this thesis in *Ext_Beta* shows that significant improvements can be achieved when weighting recommendations according to the trustworthiness of its recommender in the context of providing recommendations, as opposed to not weighting recommendations or using the trustworthiness of an entity in the context of interactions.
- The results indicate that using the community-based update of the base trust value f and the extension of the computational trust model improve the results when estimating the trustworthiness of interactors. The extension of the update mechanism as proposed in Section 5.4.2.2 and the adaption of the update mechanism for base trust value f in

order to counterbalance the selection strategy did not lead to significant improvements.

All in all, the evaluation supports the claim that the proposed trust model improves the quality of interactions especially in the presence of misleading recommendations. Thus, the results show the robustness of the proposed approach with respect to different populations and varying stability of the users' interaction behaviors.

6.2 Evaluation of the Usability of the HTI

As introduced in the previous chapters, it is important that a trust model that is to support users in their decision making provides a representation which is appropriate for integration in an intuitive graphical representation. Trust models that have only been developed for use by software agents, e.g., [BLB04, QHC06, TPJL06], usually do not consider this aspect. In this section, the usability of the graphical representation of CertainTrust (CT) - called Human Trust Interface (HTI) - that has been proposed in Section 5.2.4 is evaluated. The evaluation has been carried out in the form of a user study comparing the HTI with a graphical representation that has been proposed for “subjective logic” (SL) [Jøs01], called “Opinion Triangle” (see also Section 3.2.5), and “Stars interface”, that mimics the rating interface used by Amazon [Ama09].

Although the “Stars interface” does not come from the domain of trust models in the closer sense (see the definition of trust 4.2.4), it comes from a related domain - online recommender systems - which also deals with representing recommendations based on previously collected ratings. In contrast to trust models that are currently under research where the computation of a trust value seems more important than an intuitive representation for users, an intuitive representation of the ratings of offered products is vital for selling platforms. Therefore, a comparison to the “Stars interface” seems to be a good choice.

The user study has been based on the same scenario as the simulation in Section 6.1. The scenario that has been presented to the participants of the user study has already been introduced at the beginning of Chapter 2.

In the following sections, first, the evaluated representations are briefly presented. Second, the procedure and the design of the user study are explained. Third, the results are shown, and discussed (fourth). Finally, there is a summary of this section.

6.2.1 Evaluated Representations

This section briefly introduces the necessary details of each representation. Figure 6.15 shows the three representations that have been compared in the user study.

6.2.1.1 CertainTrust - Human Trust Interface

The details of the graphical representation of the Human Trust Interface (HTI) have already been introduced in Section 5.2.4. In the user study, the base trust value is assumed to have moderate value, i.e., $f = 0.5$; and the maximum number of expected evidence units has been set to $N = 20$. The relation between the numbers of collected positive and negative evidence units

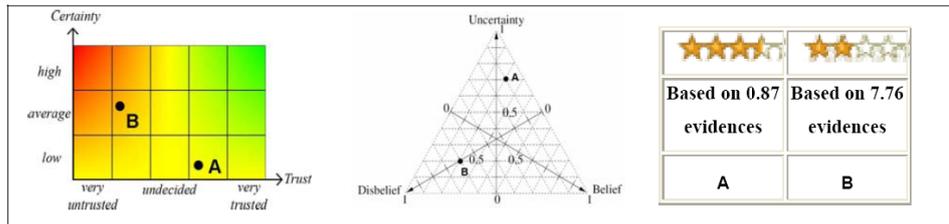


Figure 6.15: Representations evaluated in the user study: Each representation shows the opinion about the trustworthiness of 2 interactors based on the same number of evidence units

and the formula for the certainty parameter that was used when the user study was conducted is presented in [Rie07]; it can be roughly approximated using $w = 4$. Figure 6.16 shows an example of an opinion about an entity A that is based mostly positive evidence with a high certainty value.

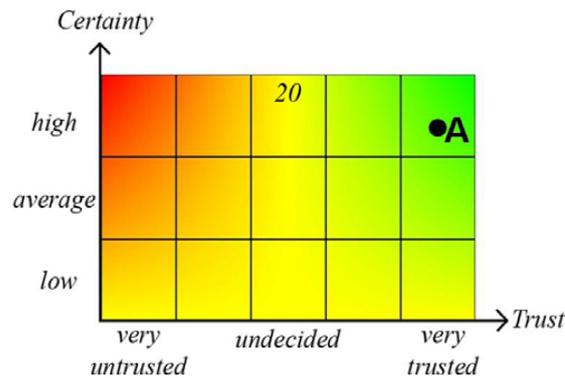


Figure 6.16: Example: CertainTrust - HTI

6.2.1.2 Subjective Logic (SL) - Opinion Triangle

The second representation uses a triple of parameters to represent opinions about the trustworthiness of an interactor (b, d, u) (b = belief, d = disbelief, u = uncertainty) - assuming *atomicity* $a = 0.5$ (see Section 3.2.5.1). The relation between these parameters and the collected evidence can be explained as follows:

- uncertainty: depends on the number of collected evidence units
 - uncertainty = 0: infinite number of evidence units collected.
 - uncertainty = 1: not any collected evidence.
- belief: increases with the relative frequency of collected positive evidence:

- belief = 1: infinite number of evidence units collected, which have all been positive
- belief = 0: not any positive evidence collected (only negative once, if any)
- disbelief: increases with the relative frequency of collected negative evidence:
 - disbelief = 1: infinite number of evidence units collected, which all have been negative
 - disbelief = 0: not any negative evidence collected (only positive once, if any)

The axes for belief, disbelief, and uncertainty are indicated by the corresponding labels at the end of each axis (see Figure 6.17). The interpretation of the opinion represented by the point *A* in Figure 6.17 is similar to the one explained in the example presented with *CertainTrust* in Figure 6.16. This interface does not integrate a representation of the expectation value in contrast to the interface of *CertainTrust*.

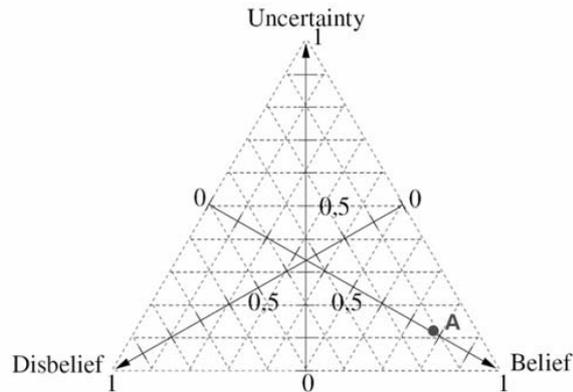


Figure 6.17: Example: Opinion Triangle (SL)

6.2.1.3 Stars

The interface “Stars” represents the average number of positive ratings of interactions as stars in the range of one to five stars (in half steps). Additionally, the interface shows the number evidence which contributed to the rating. For an example and the interpretation of the stars see Figure 6.18 and Figure 6.19. In the example, the decimal places of Figure 6.18 may originate from weighted evidence from recommendations.

A difference to the representations introduced before is that this interface does not integrate both values (stars and number of evidence units) to a

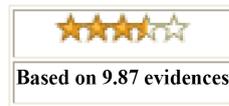


Figure 6.18: Example: Stars interface

single value or point, but leaves the interpretation up to the user. This makes it also harder to manipulate opinions, as both parameters have to be manipulated separately. This representation does not provide a means to derive a trust (or expectation value) that integrates both parameters.



Figure 6.19: Interpretation of the stars interface - taken from Amazon (www.amazon.com)

6.2.2 User Study

As pointed out above, the user study is to evaluate how intuitive users can interact with those representations. Therefore, the experiment was conducted with three conditions (CT, SL, Stars) corresponding to the three interfaces described above.

The experiment is to provide insight into how well the interfaces support the user in making a decision on the trustworthiness of a potential interaction partner. Furthermore, it will be evaluated whether and how often the decision that would automatically be taken by CertainTrust matches the decisions the users took. A good prediction of the user's choice would allow for the automation of the interaction process in cases as presented in the scenario.

This study did not evaluate how the interfaces perform with regard to manipulating opinions. The Stars interface was included as a baseline as it is already widely used in internet sites on the web.

The hypotheses for the experiment were as follows:

Hypothesis 1: The users will be faster in the CT Interface than in the Opinion Triangle (from Subjective Logic) to decide on the trustworthiness of interaction partners. In the Stars interface, the user were to be faster, as they were used to this interface.

Hypothesis 2: The participants' decisions in the Stars and SL interface will be the same as the decisions suggested by CertainTrust.

6.2.2.1 Design

The experiment used a within-subject factorial design, with the interface type as primary factor. Each participant completed a series of eight tasks with each interface. The task was to pick one of two potential interaction partners which were displayed in the same interface. The same eight pairs in the same order were used for all conditions; however carry over effects have not been expected and evaluated, as the representations were supposed to be sufficiently distinctive. In order to counterbalance learning effects in the general setup the order of interfaces between subjects was varied. The within subject design does account for variability between subjects.

The study was conducted as an online survey, so that participants could take part remotely to prevent experimenter effects.

6.2.2.2 Participants

Thirty two subjects took part in the study. All but one did have prior experience with online shopping. 4 females and 27 males completed the study (one participant did not reveal the gender). They were in the age range of 21 to 40 years. Participants were not paid for taking part in the study.

6.2.2.3 Procedure

Participants were presented the scenario that is described at the beginning of Chapter 2. Afterwards, they answered a questionnaire regarding their opinion about the relationship between trust and evidence.

The procedure for the experiment was as follows. The interface of the first condition was explained to the subjects. They were then sequentially presented with the 8 pairs of potential interaction partners (named *A* and *B*) using the same trust representation. The users were to select the best interaction partner from each pair. The time they took to decide between the two marked by a click on a button and the interaction partner they preferred was logged for analysis.

The same was done for the other two conditions, so that every user judged on every pair of interaction partners three times, once in every interface. In Figure 6.16 there is an example how the same example (setting) is presented using different interfaces.

The evaluation was done using only 6 of 8 settings, as the number of evidence units presented in the three interfaces was identical for those (for the numbers see Table 6.3).

6.2.3 Results

Hypothesis 1: For the analysis of the first hypothesis the time the users took for their decisions was aggregated over the 6 settings per model (CT, SL,

Setting	Evidence for Interactor A		Evidence for Interactor B	
	pos.	neg.	pos.	neg.
1	0.5903	0.2883	1.868	5.8968
2	1.9294	1.9294	3.8816	3.8816
3	0.9623	0.0385	9.8276	2.0319
4	8.7332	2.036	1.9467	8.8225
5	1.9089	1.9499	8.7332	2.036
6	2.0119	9.4476	0.4129	0.4328

Table 6.3: Number of evidence units per interaction partner and setting

Stars). The mean times (in ms) per model are given in ascending order: 44635.094 (CT), 46486.250 (Stars), and 74324.250 (SL). The Kolmogorov-Smirnov test indicated that the data is normally distributed. For the further analysis one-way repeated measures ANOVA was done: Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 24.0$, $p < .01$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .645$). The results show that the choice of the representation has significantly affected the time a user needs to select an interaction partner, $F(1.3, 40.0)$, $p < .01$, $\omega = 0.56$. Using the benchmarks for effect size this represents a strong effect. Bonferroni post hoc tests revealed that the mean time of the participants was significantly higher than when using the Opinion Triangle of Subjective Logic as in both other interfaces. For details see Table 6.4.

(I) Model	(J) Model	Mean Differ- ence (I-J)	Std. Error	Sig. (a)	99% Confidence Interval for Difference (a)	
					Lower Bound	Upper Bound
Stars	SL	-27838.000*	7057.201	.001	-50280.028	-5395.972
	CT	1851.156	3044.874	1.000	-7831.598	11533.910
SL	Stars	27838.000*	7057.201	.001	5395.972	50280.028
	CT	29689.156*	6967.931	.001	7531.001	51847.302
CT	Stars	-1851.156	3044.874	1.000	-11533.910	7831.598
	SL	-29689.156*	6967.931	.001	-51847.302	-7531.011

Based on estimated marginal means
 *. The mean difference is significant at the .01 level.
 a. Adjustment for multiple comparisons: Bonferroni.

Table 6.4: Pairwise comparisons (time in milliseconds)

Hypothesis 2: As the selected interaction partner in the experiment leads to nominal, non-parametric data - a participant either selected to interact with *A* or *B* - the frequencies of the selection of *A* and *B* have been counted. Figure 6.20 shows the corresponding the percentages. One sees that the majority of the participants has selected the same interaction partner independently of the model. Furthermore, the interaction partner which has been selected by the majority of the users was in all settings the one which would have been autonomously selected by CertainTrust. Thus, this hypothesis is considered to be true.

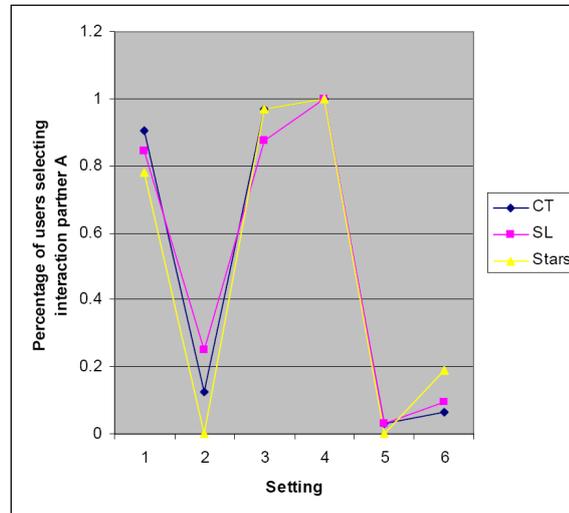


Figure 6.20: Percentage of participants selecting interaction partner A per setting and per model (the lines are only for ease of reading)

For a further analysis, whether the representation has an effect on the choice of the interaction partner, the percentage of participants which have chosen the same interaction partner as the one proposed by CertainTrust per setting and per model was calculated. The results show that there are not any significant differences ($p > .05$) between the participants choices in the different models. The mean values are: .943 (CT), .891 (SL), and .927 (Stars); the confidence intervals are given in Figure 6.21.

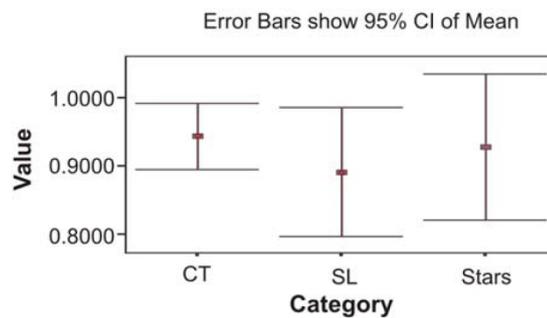


Figure 6.21: Mean values: Average percentage of participants selecting the same interaction partner as proposed by CT

6.2.4 Discussion

6.2.4.1 Effects of the Representation

From the results, one can learn that the representation has a significant effect on the time the participants needed to select an interaction partner. Although, it was expected that most participants are faster in the Stars interface as they were supposed to be used to this representation (all but one had experience with online shopping), the participants were slightly, non-significantly slower than in the HTI. In both interfaces the participants were significantly faster than in the Opinion Triangle. There may be two reasons to explain these effects. First, users are not used to interpreting triangular representations with three axes, as orthogonal two axis layouts are more common in everyday life. On the website on Subjective Logic (<http://sky.fit.qut.edu.au/josang/sl/demo/BV.html>), one can find the statement that the Opinion Triangle is a more mathematical representation. That page offers further representations for opinions, but as to the author's knowledge the details have neither been evaluated nor published, it has been decided to use the representation which is usually used as graphical representation of opinions for Subjective Logic. Second, the HTI tries to support the user with a green-yellow-red color gradient and is capable of integrating the average rating and the certainty in one graphical representation as opposed to the Stars interface.

The answers given by the participants on the questions at the end of the experiment also indicated that they felt comfortable with the interface of CertainTrust (small sample from the questions):

Question A: Which interface would you prefer in the described scenario: (CT 68.8%, SL 3.1%, Stars 28.1%)

Question B: The color gradient supported your decision for an interaction partner. Do you agree? The mean value of the answers on a scale from 0 (strongly disagree), 1 (disagree), 2 (slightly disagree), 3 (slightly agree), 4 (agree), 5 (strongly agree) is 3.81.

Question C: Can you imagine using the Human Trust Interface without the labels after a short time of familiarization? mean value (scale as above): 3.19.

6.2.4.2 Rationality of the Choices

At this point, it has still to be discussed, whether the decisions of the participants were rational or not. The definition of the "best choice" is not trivial, as the choice is influenced at least by the relative frequency of positive interactions, as well as the total number of evidence units, and the risk which is associated with an interaction.

In settings 1, 4, and 5 the decision goes along with the relative frequency of positive and negative interactions and the amount of collected evidence.

The more interesting cases are the settings 2, 3, and 6. For the exact numbers of evidence units see Table 6.3.

In setting 2, there are two candidates for the interaction having the same of ratio of positive to total evidence, but the opinion about *B* is based on more evidence. Using the Stars interface (3 stars for both) all users decided to interact with *B* - which is the candidate about whom more information is available. In both remaining interfaces, a small number of participants selected the candidate about whom less information is available. In both cases, the expected trustworthiness calculated by CertainTrust is 0.5. Although in this case, the expectation values are identical, the selection algorithm favors the one with the higher number of evidence units; thus, going along with the majority of the participants.

In setting 3, there is candidate *A* about whom one has very little, but very positive information (mean: 0.96) information; and candidate *B* about whom one has a higher amount of information, mostly positive (mean: 0.82). In all three interfaces the majority of the participants selected user *A*. In the face that interactions are associated with a certain risk and a little amount of evidence may be quite misleading, therefore this choice is considered to be rational.

In setting 6, there are six participants who selected candidate *A* in the Stars interface, while in the other interfaces only one (CT) or two (SL) participants selected *A*. In short, about candidate *A*, there is very little information available (similar amount of positive and negative evidence), while about candidate *B*, there is more information available, which is mostly negative. In this case, one might say that the participants made a better choice in the CertainTrust and Subjective Logic interface, than in the Stars interface.

At last, one can ask, if there is a number of maximum expected evidence units, which allows to interpret the average rating as representative expectation for future interactions. At the beginning of the experiment, after the participants were given the scenario and before they were introduced to the different representations, the following statement was presented to the participants: "Having collected a certain number of evidence units, you are able to properly estimate the trustworthiness of your interaction partner." Most users agreed (all, but 5). On a scale from 0 to 5 (as above) the mean was 3.41. For the mp3-exchange the majority of users expected 6-10 pieces of evidence, for buying an mp3-player (used, 10 EUR) the majority expected 21-50 pieces of evidence.

6.2.5 Summary

In this section, the usability of the graphical representation of the HTI has been evaluated. The results may be summarized as follows:

- In the HTI the participants have been significantly faster than in the Opinion Triangle when selecting an interaction partner.
- Comparing the HTI and “Stars interface”, the performance of the participants is similar in both interfaces. The participants have been slightly, non-significantly faster in the HTI than “Stars interface”.
- In all representations, the users selected the candidate that would have been preferred by the selection strategy proposed in Section 5.4.1.

Thus, the results of the user study support the hypothesis that the graphical representation of the HTI is appropriate for human users.

6.3 Integration of CertainTrust in an Online Movie Recommendation Application

In order to evaluate whether users feel comfortable with the concept of a trust model, the trust model has been additionally evaluated in an online movie recommendation application.

6.3.1 Description of the Application

The application, called “trust-aided online recommendation platform for movies” (TORP) [Win08], has been developed and evaluated in a Master thesis. It allows users to receive personalized recommendations for movies based on their social network.

TORP is basically a web application - for screenshots see Figures 6.22 (Registration and login), 6.23 (Social network and rated movies), 6.24 (Actively recommended movies and directly rated movies), and 6.25 (Rating of a movie). Users can register at TORP using a freely chosen nickname. After the registration a user can start adding other users as friends to their social network. Furthermore, a user can rate movies and their friends. The movies are rated according to the user’s personal preference; friends are rated according to their expected trustworthiness in the context of providing good recommendations for movies. Both ratings are presented and manipulated in the graphical representation provided by the Human Trust Interface (HTI) (see Section 5.2.4). Thus, the user can provide a rating that includes a statement about the certainty of this rating².

The application provides two mechanisms for providing personalized movie recommendations.

- The *active* mechanism provides users with a personalized list of the best movies based on weighted recommendations derived from the social network. In this list a user will only find movies they have not rated themselves.
- The *passive* mechanism calculates a personalized rating for any movie a user wants to get a personalized rating for.

In order to overcome the bootstrapping problem - at the beginning a user might have a very small network and look for movies no one in their social networks has rated before - the Internet Movie Data base (www.imdb.com) is integrated. The IMDb is added to each user social network as a virtual friend.

Furthermore, TORP incorporates an update mechanism for the trustworthiness of one’s friends. Each time a user rates a movie that a friend

²The relation between the collected number of evidence units and certainty was the same as used for the user study in the Section 6.2.1.1.

has rated before, the trustworthiness of the friend is updated according to the similarity of the ratings. This update mechanism is also applied to personalized the rating of the trustworthiness of the IMDb.

The application is similar to FilmTrust [Gol05], yet, in the presented approach the ratings are presented in a two-dimensional layout - including the certainty of a rating - and the trustworthiness of friends is automatically updated as described above.



Figure 6.22: Screenshots of TROP: Registration and login

6.3.2 User Study

The section briefly presents the design and the result of the user study that was carried out after developing the application.

6.3.2.1 Participants and Design

In the user study with 26 participants, mostly students of computer science, it has been evaluated whether the users felt comfortable with this application. In order to overcome the lack of real users, a set of artificially created predefined users was added to TORP prior to the study. The preferences of the predefined users were associated to different genres of movies, e.g., comedy or action. A list with the predefined users and their preferences was given to the participants of the user study.

In the experiment, the users had to create their own account for TORP and build a social network from a set of predefined users. Afterwards, the participants could rate movies and adjust the ratings of the users they previously added to their social network according to their preferences. They were told especially to look at how the new rating of their friend changes when providing ratings on movies.

After the experiment, the participants had to fill out a questionnaire. The questions focused on the overall usability of the application, as well as

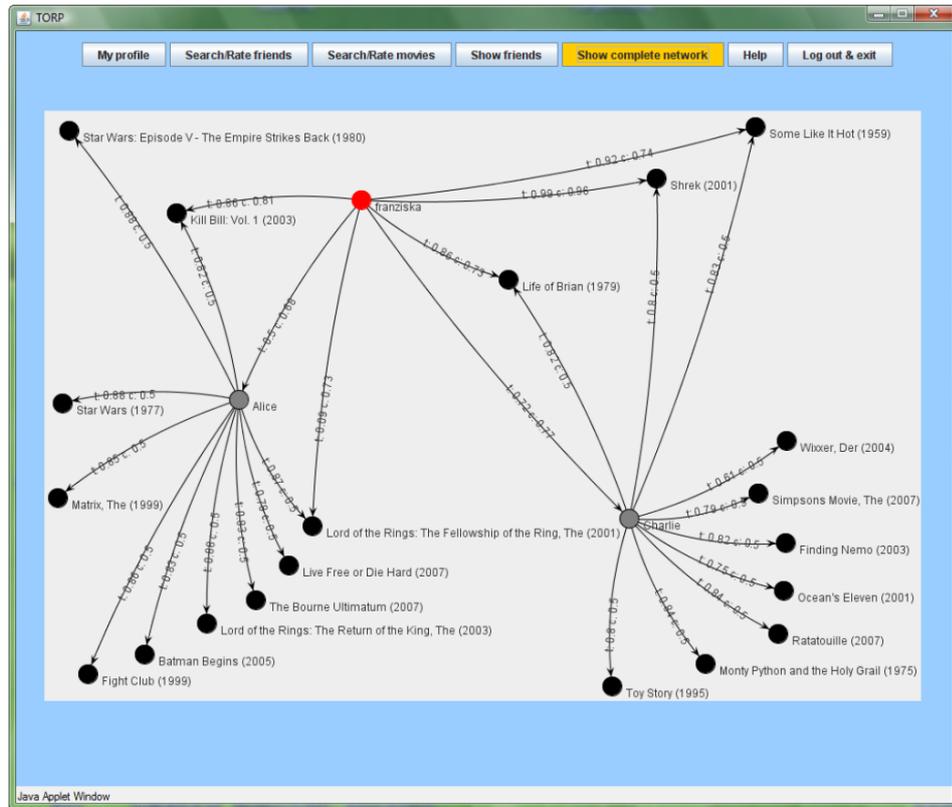


Figure 6.23: Screenshots of TROP: Social network and rated movies

on the aspects relevant to the trust model.

6.3.2.2 Results

The results of the user study supported that the users feel comfortable with two-dimensional interface of the HTI (see below questions A - D). Furthermore, the results of the user study support that the users like the idea of receiving personalized recommendations based on their social network and they feel comfortable with the integrated update mechanism for the trustworthiness of recommenders (see below questions E - I).

- Question A: I felt comfortable with the interpretation of the two-dimensional interface. Do you agree? The mean value of the answers on a scale from 1 (strongly disagree), 2 (disagree), 3 (slightly disagree), 4 (slightly agree), 5 (agree), 6 (strongly agree) is 5.04.
- Question B: I felt comfortable using the two-dimensional interface to provide ratings on my friends. Do you agree? The mean value of the answers is (scale as above): 4.58

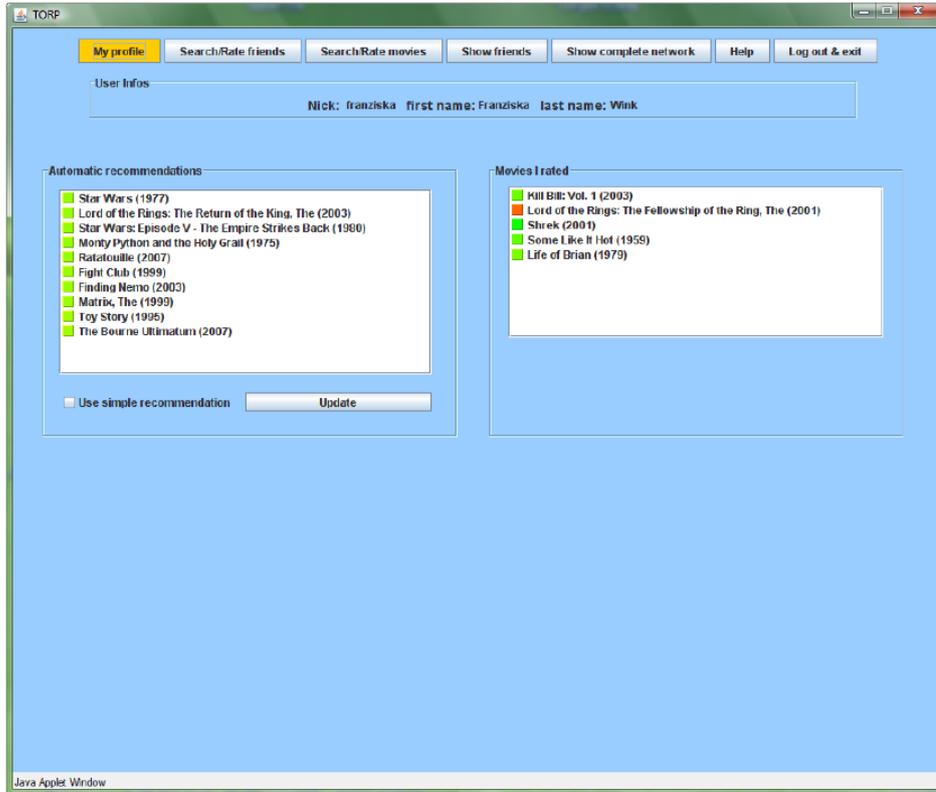


Figure 6.24: Screenshots of TROP: Actively recommended movies and directly rated movies

- Question C: I felt comfortable using the two-dimensional interface in order to provide ratings on movies. Do you agree? The mean value of the answers is (scale as above): 5.27
- Question D: I felt comfortable with the possibility of providing and receiving ratings in the two-dimensional interface. Do you agree? The mean value of the answers is (scale as above): 4.92
- Question E: I would like to use recommendations by friends in addition to other internet-based sources of movie ratings, like IMDb. Do you agree? The mean value of the answers is (scale as above): 5
- Question F: I think it is reasonable to rate friends according to their preferences of movies. Do you agree? The mean value of the answers is (scale as above): 4.62
- Question G: I think it is reasonable to automatically update the ratings of my friends. Do you agree? The mean value of the answers is (scale

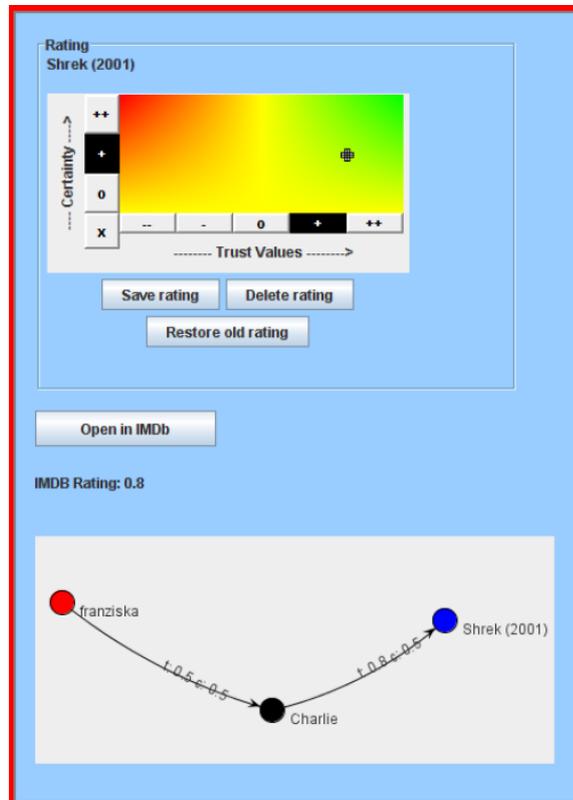


Figure 6.25: Screenshots of TROP: Rating of a movie

as above): 5.19

- Question H: The quality of the recommended ratings increased after having rated some movies. Do you agree? The mean value of the answers is (scale as above): 5.04
- Question I: I would like to use this application on another domain, e.g., recommendation of clips on YouTube (www.youtube.com). Do you agree? The mean value of the answers is (scale as above): 4.88

6.3.3 Summary

The section may be summarized by following statements:

- This section introduced an application of CertainTrust as part of a trust-aided online recommendation platform for movies. The details of the application can be found in [Win08].
- The results of the conducted user study support the argument that the users feel comfortable when using the graphical representation of the

HTI in the context of movie recommendation application. Furthermore, the results encourage the integration social networks together with means that allow the trustworthiness of users in this network to be rated.

6.4 Conclusions

This chapter evaluated the performance of the proposed trust model in an opportunistic network scenario. Furthermore, the usability of the proposed graphical representation for users has been evaluated.

- The results of the simulated opportunistic network in Section 6.1 clearly show the positive impact of the proposed trust model. In order to show that the trust model may be applied over a wide set of applications in which the typical user behavior may differ, the simulation presents the results over a canonical set of 15 populations that are derived from the system model which has been proposed in Section 4.2.
 - The results of the simulation support that assess the trustworthiness of a recommender based on the accuracy of its past recommendations leads to an improvement of the performance of the trust model.
 - The results shows that the dynamic community-based re-evaluation of the base trust may lead to improvements in the model's performance.
 - The overall performance regarding the two introduced measures show that the proposed trust model has not only a superior performance when compared to a state-of-the-art approach, but also that the model performs well when compared to the introduced *Perfect* selection strategy.
 - * The results for the *average error in estimating the trustworthiness* across all populations and both settings are given by CT_C 10.37%, Beta_D 33.96%, and Beta_S 37.63% (see Table 6.1).
 - * The results for the *average percentage of the accumulated sum of feedback* across all populations, but m , w , and mw are given by CT_C 89.48%, Beta_D 73.07%, and Beta_S 43.89% (see Table 6.2).
- The user studies that have been performed support the claim that the proposed graphical representation allows for an intuitive interpretation by human users.

- It has been shown that users are significantly faster when selecting an interaction partner in the HTI than in the Opinion Triangle.
 - In the provided examples, the majority of users selected the interaction partner that would have been preferred by the selection strategy proposed in Section 5.4.1.
- The integration of CertainTrust in an online movie recommendation platform (TORP) and evaluation of TORP have shown that users feel comfortable with the integration of the trust model in a social network in the context of movie recommendations. Furthermore, users provided positive feedback on the graphical representation of the HTI.

Chapter 7

Conclusions and Outlook

Having presented and evaluated the proposed trust model, this chapter summarizes the main contributions and findings of the thesis and presents an outlook with issues that may be addressed in future work.

7.1 Conclusions

In this thesis, a new distributed trust model has been provided. The requirements for the trust model have been inspired by the ideas of unmanaged ubiquitous computing environments. The trust model supports users in finding trustworthy interaction partners, which is a fundamental requirement for successful collaboration not only in ubiquitous computing but also in open service market places and on Web 2.0 platforms. The trust model supports users as it allows the decision making to be delegated to an autonomous software component or the collected trust information to be presented in an appropriate representation to the user.

Representational Model and Update Mechanisms

- In the presented approach trust is interpreted as a subjective probability. The trust model extends Bayesian trust models in order to improve the integration of the context-dependent parameters. The dispositional trust of users can be directly mapped to the prior knowledge using the parameters *base trust* f and *weight* w . The *maximum number of expected evidence units* N is introduced to allow for the definition of a number of evidence units that is expected to be sufficient in order to consider the collected evidence as representative for the behavior of an entity.
- The introduction of the parameter N has been shown to be especially helpful when aging of evidence is introduced. Choosing the parameter N depending on the *aging* factor a as proposed ($N = \frac{1}{1-a}$), the approach

combines the features that the trust value can reach the complete range of values in $[0; 1]$ and the trust value moves back to the base trust value in absence of evidence.

- In order to support human users, a new representational model of trust, called Human Trust Interface (HTI), has been developed. It is based on a simple set of parameters and provides an intuitive graphical representation. The HTI is considered to be especially helpful when users want to control trust values, intervene in the decision making process, or adjust trust values according to their “real world” experience.
- The evaluation of the HTI in two user studies supports the intuitiveness of the HTI and that the users feel comfortable with the graphical representation.
- The provided mapping makes both representations, the Bayesian representation and the HTI, interchangeable. As the calculated trust value or expectation value is independent from the choice of the representational model, the developers and users of trust models can benefit from the advantages of both representations. Furthermore, a mapping to the belief representation of “subjective logic” is provided.
- Regarding evidence based trust establishment, two update mechanisms for deriving a recommender’s trustworthiness based on the accuracy of the recommender’s past recommendations have been provided. Furthermore, the model supports the derivation of the base trust value from the typical behavior of the entities that have been met within a certain context.

Computational model

Beyond the advances in the representational trust model, an extended computational model is proposed.

- The computational model is based on the two operators called *consensus* and *discounting*, and it provides new features for *exogenous* filtering of recommendations:
 - It excludes recommenders that are known to provide mostly bad recommendations.
 - It considers the recommendations by the best recommenders first.
 - It considers only recommendations until the number of evidence units that is calculated from the direct evidence and the recommendations considered so far is below a certain threshold.

- Recommendations are weighted by the trustworthiness of its recommender and by the rank of the recommender.
- It has been shown that those steps together allow the robustness of the trust model to increase regarding Sybil attacks for two reasons. First, in the presence of sufficient direct evidence or well-known good recommenders, recommendations by other recommenders are not considered at all. Second, in the absence of sufficient direct evidence or good recommenders, the impact of a Sybil attack is limited by the trustworthiness of the Sybils in the context of providing recommendations. This is particularly achieved by considering the rank of a recommender.
- Furthermore, the robustness of the trust model has been evaluated in the simulation of an opportunistic network scenario. The evaluation shows the performance of the trust model over a large set of populations that have canonically been derived from the proposed system model. The results from the comparison to a state-of-the-art approach as well as from a comparison to the introduced perfect selection strategy show that the proposed model reaches good results regarding the estimation of the trustworthiness of an entity and regarding the average quality of interactions.

As the evaluation of the representational model as well as the evaluation of the computational model has been shown to lead to improvements in comparison to the state-of-the-art, the concepts that have been presented and evaluated in this thesis are considered to be valuable and significant contributions in field of evidence based trust models.

7.2 Outlook

Beyond the issues that have been evaluated in the focus of this thesis, there are still numerous aspects for further research.

As a next step it could be evaluated how the proposed extensions of the Bayesian trust models can be applied in Dirichlet based trust models that allow multinomial ratings [JH07]. It would be especially interesting to transfer the concepts that have been proposed for aging and the introduction of the maximum number of expected evidence units.

A further extension of the current work could be the evaluation of the proposed approach in utility based decision making, which has been introduced in Section 2.3. Hereby, the assessment of the utility of non-monetary interactions and the evaluation of the risk attitude of users in different application contexts seems to be challenging.

More generally, the evaluation of the performance of the different trust models in further scenarios is interesting. Beyond ubiquitous computing, the

integration of trust models in other domains as open service market places and Web 2.0 applications seems to be promising in order to evaluate the trust models with real users and real attacks. Hereby, one can also analyze the typical interaction patterns between entities in different domains, e.g., how often entities interact with each other, what kinds of attacks occur most, and how often entities change their pseudonyms. This would be especially helpful as currently much evaluation is done based on simulations. An analysis of real world data would also be important in order to further evaluate a rational choice of the values of the trust model's parameters.

Furthermore, it would be interesting to evaluate strategies that support the evaluation of the trustworthiness of groups of entities. This may be reasonable in a case when a group of entities, e.g., the web services of one company, offer the same quality of service, but there is only little experience with a single service. This also leads to approaches that try to transfer trust between different application contexts. For example, how can the trustworthiness of a service provider in the context of hotel reservations be transferred to the context of car reservations or online banking.

Moreover, it should be investigated how evidence based approaches can be extended to application fields in which direct feedback can only be expected in the case that an interaction partner does not meet its obligations. For example, when assessing the trustworthiness of a service provider regarding the enforcement of the advertised privacy policies, users may hardly give positive feedback, as they simply do not know whether the service provider keeps their data secretly. However, users can provide negative feedback when they find out that their data has been misused. This lack of positive feedback is not reflected in the current approach.

Finally, the integration of evidence-based trust management and policy-based trust management is an obvious challenge. Here, it would be interesting to develop approaches which support the evaluation of the trustworthiness of certificates using evidence based models and approaches that are capable of assessing the trustworthiness of entities based on certificates and evidence.

Bibliography

- [AD01] Karl Aberer and Zoran Despotovic. Managing Trust in a Peer-2-Peer Information System. In *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM '01)*, pages 310–317. ACM Press, 2001.
- [AG07] Donovan Artz and Yolanda Gil. A Survey of Trust in Computer Science and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):58–71, 2007.
- [Ama09] Amazon. Amazon homepage [online]. 2009. Available from: www.amazon.com [seen 2009].
- [AMCG04] Florina Almenarez, Andres Marn, Celeste Campo, and Carlos Garcia. PTM: A Pervasive Trust Management Model for Dynamic Open Environments. In *Proceedings of Workshop on Pervasive Security, Privacy and Trust (PSPT)*, 2004.
- [AR04] Alfarez Abdul-Rahman. *A Framework for Decentralised Trust Reasoning*. PhD thesis, University College London, UK, 2004.
- [ARH00] Alfarez Abdul-Rahman and Stephen Hailes. Supporting Trust in Virtual Communities. In *Proceedings of Hawaii International Conference on System Sciences*, 2000.
- [ARSB⁺08] Erwin Aitenbichler, Sebastian Ries, Julian Schröder-Bernhardi, Georg Turban, Stephan Borgert, Dirk Bradler, Michael Hartle, and Gina Häussge. Smart Products: Integration Challenges. In *Proceedings of Smart Products: Building Blocks of Ambient Intelligence (AmI-Blocks'08)*, 2008.
- [Aus08] Gerhard Austaller. Service Discovery. In Max Mühlhäuser and Iryna Gurevych, editors, *Ubiquitous Computing Technology for Real Time Enterprises*, chapter 5, pages 107–127. Information Science Reference, 2008.
- [BBM06] Arianna Bassoli, Johanna Brewer, and Karen Martin. under-sound: Music and Mobility under the City. In *International Conference on Ubiquitous Computing*, 2006.

- [BDOS05] Piero A. Bonatti, Claudiu Duma, Daniel Olmedilla, and Nahid Shahmehri. An Integration of Reputation-based and Policy-based Trust Management. In *Semantic Web Policy Workshop in conjunction with 4th International Semantic Web Conference*, 2005.
- [Ber54] Daniel Bernoulli. Exposition of a New Theory on the Measurement of Risk. *Econometrica*, 22(1):23–36, 1954.
- [BFIK99] Matt Blaze, Joan Feigenbaum, John Ioannidis, and Angelos D. Keromytis. *Secure Internet Programming: Security Issues for Mobile and Distributed Objects*, chapter The Role of Trust Management in Distributed Systems Security, pages 185–210. Springer, 1999.
- [BFK98] Matt Blaze, Joan Feigenbaum, and Angelos D. Keromytis. Keynote: Trust Management for Public-key Infrastructures. In *Security Protocols Workshop*, pages 59–63, 1998.
- [BFL96] Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized Trust Management. In *Proceedings of the 1996 IEEE Symposium on Security and Privacy (SP '96)*, pages 164 – 173. IEEE Computer Society, 1996.
- [BHOC07] Holger Billhardt, Ramón Hermoso, Sascha Ossowski, and Roberto Centeno. Trust-based Service Provider Selection in Open Environments. In *SAC '07: Proceedings of the 2007 ACM Symposium on Applied Computing*, pages 1375 – 1380. ACM Press, 2007.
- [BLB03] Sonja Buchegger and Jean-Yves Le Boudec. A Robust Reputation System for Mobile Ad-hoc Networks. Technical report, École Polytechnique Fédérale de Lausanne, Suisse, 2003.
- [BLB04] Sonja Buchegger and Jean-Yves Le Boudec. A Robust Reputation System for Peer-to-Peer and Mobile Ad-hoc Networks. In *P2PEcon 2004*, 2004.
- [BLRW04] Bharat Bhargava, Leszek Lilien, Arnon Rosenthal, and Marianne Winslet. Pervasive Trust. *IEEE Intelligent Systems*, 19(5):74–88, 2004.
- [Bol04] William M. Bolstad. *Introduction to Bayesian Statistics*. John Wiley & Sons, Inc, 2004.
- [CNS03] Marco Carbone, Mogens Nielsen, and Vladimiro Sassone. A Formal Model for Trust in Dynamic Networks. In *Proceedings*

of IEEE International Conference on Software Engineering and Formal Methods. IEEE Computer Society, 2003.

- [CSG⁺03] Vinny Cahill, Brian Shand, Elizabeth Gray, Nathan Dimmock, Andy Twigg, Jean Bacon, Colin English, Waleed Wagealla, Sotirios Terzis, Paddy Nixon, Ciaran Bryce, Giovanna di Marzo Serugendo, Jean-Marc Seigneur, Marco Carbone, Karl Krukow, Christian Jensen, Yong Chen, and Mogens Nielsen. Using Trust for Secure Collaboration in Uncertain Environments. *IEEE Pervasive Computing*, 2/3:52–61, 2003.
- [DA04] Zoran Despotovic and Karl Aberer. Maximum Likelihood Estimation of Peers' Performance in P2P Networks. In *The Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.
- [DA05] Zoran Despotovic and Karl Aberer. Probabilistic Prediction of Peers' Performances in P2P Networks. *Engineering Applications of Artificial Intelligence*, 18(7):771–780. Elsevier, 2005.
- [Del00] Chrysanthos Dellarocas. Mechanisms for Coping with Unfair Ratings and Discriminatory Behavior in Online Reputation Reporting Systems. In *ICIS '00: Proceedings of the Twenty First International Conference on Information Systems*, pages 520–525. Association for Information Systems, 2000.
- [DFM00] Roger Dingledine, Michael J. Freedman, and David Molnar. Accountability Measures for Peer-to-Peer Systems. In Andy Oram, editor, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, pages 271–340. O'Reilly, 2000.
- [Dou02] John R. Douceur. The Sybil Attack. In *IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems*, pages 251–260. Springer-Verlag, 2002.
- [eBa] eBay Inc. So funktioniert das Bewertungssystem [online]. Available from: <http://pages.ebay.de/help/feedback/howitworks.html> [seen 2009].
- [eBa09a] eBay Inc. Company overview [online]. 2009. Available from: <http://pages.ebay.com/aol/community/aboutebay/overview/trust.html> [seen 2009].
- [eBa09b] eBay Inc. eBay homepage [online]. 2009. Available from: www.eBay.com [seen 2009].
- [eBa09c] eBay Inc. What is a powerseller? [online]. 2009. Available from: <http://pages.ebay.com/services/buyandsell/welcome.html> [seen 2009].

- [EP06] Nathan Eagle and Alex (Sandy) Pentland. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Computing*, 10(4):255–268, 2006.
- [FC05] Michal Feldman and John Chuang. Overcoming Free-riding Behavior in Peer-to-Peer Systems. *SIGecom Exchanges*, 5(4):41–50, 2005.
- [Gam90] Diego Gambetta. Can we Trust Trust? In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, New York, 1990.
- [Gam00] Diego Gambetta. Can we Trust Trust? In Diego Gambetta, editor, *Trust: Making and Breaking Cooperative Relations, electronic edition*, chapter 13, pages 213–237. 2000.
- [GH06] Jennifer Golbeck and James Hendler. FilmTrust: Movie Recommendations using Trust in Web-based Social Networks. In *Proceedings of the Consumer Communications and Networking Conference*, 2006.
- [GHB08] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453(7196):779–782, 2008.
- [Gol05] J. Golbeck. *Computing and Applying Trust in Web-Based Social Networks*. PhD thesis, University of Maryland, USA, 2005.
- [Gra03] Tyrone Grandison. *Trust Management for Internet Applications*. PhD thesis, Imperial College London, UK, 2003.
- [GS00] Tyrone Grandison and Morris Sloman. A Survey of Trust in Internet Applications. *IEEE Communications Surveys and Tutorials*, 3(4):2—16, 2000.
- [HCS⁺05] Pan Hui, Augustin Chaintreau, James Scott, Richard Gass, Jon Crowcroft, and Christophe Diot. Pocket Switched Networks and Human Mobility in Conference Environments. In *WDTN '05: Proceeding of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking*, pages 244–251. ACM Press, 2005.
- [Hei07] Andreas Heinemann. *Collaboration in Opportunistic Networks*. PhD thesis, Technische Universität Darmstadt, Germany, 2007.
- [HJS04a] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. Developing an Integrated Trust and Reputation Model for Open Multi-agent Systems. In *7th International Workshop on Trust in Agent Societies*, pages 65–74, 2004.

- [HJS04b] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. Fire: An Integrated Trust and Reputation Model for Open Multi-agent Systems. In Ramon López de Mántaras and Lorenza Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI)*, pages 18–22. IOS Press, 2004.
- [HJS06] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.
- [HKLM03] Andreas Heinemann, Jussi Kangasharju, Fernando Lyardet, and Max Mühlhäuser. iClouds – Peer-to-Peer Information Sharing in Mobile Environments. In *Euro-Par 2003. Parallel Processing, 9th International Euro-Par Conference*, volume 2790, pages 1038–1045. Springer, 2003.
- [ITU97] ITU-T. ITU-T Recommendation X.509, the Directory: Authentication Framework. <http://www.itu.int/rec/T-REC-X.509-199708-S/e>, June 1997.
- [JBXC08] Audun Jøsang, Touhid Bhuiyan, Yue Xu, and Clive Cox. Combining Trust and Reputation Management for Web-based Services. In *TrustBus '08: Proceedings of the 5th International Conference on Trust, Privacy and Security in Digital Business*, pages 90–99. Springer-Verlag, 2008.
- [JGA06] Audun Jøsang, Dieter Gollmann, and Richard Au. A Method for Access Authorisation through Delegation Networks. In Rei Safavi-Naini, Chris Steketee, and Willy Susilo, editors, *ACSW Frontiers '06: Proceedings of the 2006 Australasian workshops on Grid computing and e-research*, pages 165–174. Australian Computer Society, 2006.
- [JH07] Audun Josang and Jochen Haller. Dirichlet reputation systems. In *ARES '07: Proceedings of the The Second International Conference on Availability, Reliability and Security*, pages 112–119, 2007. IEEE Computer Society.
- [JHF03] Audun Jøsang, Shane Hird, and Eric Faccor. Simulating the Effect of Reputation Systems on E-markets. In *Proceedings of the First International Conference on Trust Management (iTrust'03)*, pages 179–194, 2003.
- [JI02] Audun Jøsang and Roslan Ismail. The Beta Reputation System. In *Proceedings of the 15th Bled Conference on Electronic Commerce*, 2002.

- [JIB07] Audun Jøsang, Roslan Ismail, and Colin Boyd. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems*, 43(2):618–644, 2007.
- [JKD05] Audun Jøsang, Claudia Keser, and Theodosios Dimitrakos. Can we manage trust? In Peter Herrmann, Valérie Issarny, and Simon Shiu, editors, *Trust Management, Third International Conference, iTrust 2005, Paris, France, May 23-26, 2005, Proceedings*, pages 93–107. Springer, 2005.
- [JLC08] Audun Jøsang, Xixi Luo, and Xiaowu Chen. Continuous Ratings in Discrete Bayesian Reputation Systems. In Yücel Karabulut, John C. Mitchell, Peter Herrmann, and Christian Damsgaard Jensen, editors, *2nd Joint iTrust and PST Conference on Privacy, Trust Management and Security*, pages 151–166. Springer Boston, 2008.
- [Jøs97] Audun Jøsang. Artificial Reasoning with Subjective Logic. In *Proceedings of the Second Australian Workshop on Commonsense Reasoning*, 1997.
- [Jøs99a] Audun Jøsang. An Algebra for Assessing Trust in Certification Chains. In *Proceedings of the Network and Distributed System Security Symposium*. The Internet Society, 1999.
- [Jøs99b] Audun Jøsang. Trust-based Decision Making for Electronic Transactions. In L. Yngström and T. Svensson, editors, *Proceedings of the Fourth Nordic Workshop on Secure IT Systems (NORDSEC'99)*, 1999.
- [Jøs01] Audun Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–212, 2001.
- [Jøs07] Audun Jøsang. Trust and Reputation Systems. In Alessandro Aldini and Roberto Gorrieri, editors, *Foundations of Security Analysis and Design IV, FOSAD 2006/2007 Tutorial Lectures*, volume 4677 of *Lecture Notes in Computer Science*, pages 209–245. Springer, 2007.
- [KBR05] Michael Kinateder, Ernesto Baschny, and Kurt Rothermel. Towards a Generic Trust Model. In *Proceedings of the Third International Conference on Trust Management: iTrust'05; Rocquencourt, France, May 23-26, 2005*, pages 177–192. Springer, 2005.

- [KN93] John Kohl and Clifford Neuman. RFC 1510: The Kerberos Network Authentication Service (Version 5). <http://www.ietf.org/rfc/rfc1510.txt>, September 1993.
- [KR03] Michael Kinateder and Kurt Rothermel. Architecture and Algorithms for a Distributed Reputation System. In P. Nixon and S. Terzis, editors, *Proceedings of the First International Conference on Trust Management*, pages 1–16. Springer, 2003.
- [KSGM03] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The Eigentrust Algorithm for Reputation Management in P2P Networks. In *Proc. of the 12th International Conference on World Wide Web*, pages 640–651. ACM Press, 2003.
- [Lev04] Ralph Levien. Attack Resistant Trust Metrics (Draft Version of Ph.d. Thesis). <http://www.levien.com/thesis/compact.pdf>, July 2004.
- [LHC08] Neal Lathia, Stephen Hailes, and Licia Capra. Trust Based Collaborative Filtering. In *IFIPTM 2008: Joint iTrust and PST Conferences on Privacy, Trust Management and Security*, 2008.
- [Mar94] Stephen Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, UK, 1994.
- [MC96] D. Harrison McKnight and Norman L. Chervany. The Meanings of Trust. Technical report, Management Information Systems Research Center, University of Minnesota, USA, 1996.
- [MMA⁺01] Lik Mui, Mojdeh Mohtashemi, Cheewee Ang, Peter Szolovits, and Ari Halberstadt. Ratings in Distributed Systems: A Bayesian Approach. In *Workshop on Information Technologies and Systems*, 2001.
- [MMC08] Liam McNamara, Cecilia Mascolo, and Licia Capra. Media Sharing based on Colocation Prediction in Urban Transport. In *MobiCom '08: Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, pages 58–69. ACM Press, 2008.
- [MMH02a] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A Computational Model of Trust and Reputation for E-businesses. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7*. IEEE Computer Society, 2002.

- [MMH02b] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of Reputation in Multi-agent Systems: A Review. In *International Conference on Autonomous Agents and Multi-Agent Systems*, pages 280–287. ACM Press, 2002.
- [MW09] Incorporated Merriam-Webster. Merriam-Webster Online Dictionary [online]. 2009. Available from: www.m-w.com [seen 2009].
- [NKS07] Mogens Nielsen, Karl Krukow, and Vladimiro Sassone. A Bayesian Model for Event-based Trust. *Electronic Notes in Theoretical Computer Science (ENTCS)*, 172:499–521, 2007.
- [NM44] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [OL08] Saila Ovaska and Juha Leino. A Survey on Web 2.0. Technical report, University of Tampere, Finland, 2008.
- [PBMW98] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The Pagerank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, USA, 1998.
- [PG04] Eric Paulos and Elizabeth Goodman. The Familiar Stranger: Anxiety, Comfort, and Play in Public Places. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 223–230. ACM Press, 2004.
- [PTJL05] Jigar Patel, W. T. Luke Teacy, Nicholas R. Jennings, and Michael Luck. A Probabilistic Trust Model for Handling Inaccurate Reputation Sources. In Peter Herrmann, Valérie Issarny, and Simon Shiu, editors, *iTrust*, volume 3477 of *Lecture Notes in Computer Science*, pages 193–209. Springer, 2005.
- [QH07] Daniele Quercia and Stephen Hailes. MATE: Mobility and Adaptation with Trust and Expected-utility. *International Journal of Internet Technology and Secured Transactions (IJITST)*, 1:43–53, 2007.
- [QHC06] Daniele Quercia, Stephen Hailes, and Licia Capra. B-Trust: Bayesian Trust Framework for Pervasive Computing. In Ketil Stølen, William H. Winsborough, Fabio Martinelli, and Fabio Massacci, editors, *4th International Conference on Trust Management (iTrust)*, volume 3986 of *Lecture Notes in Computer Science*, pages 298–312. Springer, 2006.

- [RA09] Sebastian Ries and Erwin Aitenbichler. Limiting Sybil Attacks on Bayesian Trust Models in Open SOA Environments. In *Proceedings of the The First International Symposium on Cyber-Physical Intelligence (CPI-09)*, 2009.
- [RAD03] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust Management for the Semantic Web. In *The Semantic Web - ISWC 2003*, pages 351–368. Springer, 2003.
- [RH08] Sebastian Ries and Andreas Heinemann. Analyzing the Robustness of CertainTrust. In Yücel Karabulut, John C. Mitchell, Peter Herrmann, and Christian Damsgaard Jensen, editors, *2nd Joint iTrust and PST Conference on Privacy, Trust Management and Security*, pages 51 – 67. Springer, 2008.
- [Rie06] Sebastian Ries. Engineering Trust in Ubiquitous Computing. In *Proc. of Workshop on Software Engineering Challenges for Ubiquitous Computing*, 2006.
- [Rie07] Sebastian Ries. CertainTrust: A Trust Model for Users and Agents. In *Proceedings of the 2007 ACM Symposium on Applied Computing*, pages 1599 – 1604. ACM Press, 2007.
- [Rie08a] Sebastian Ries. Engineering Trust in Ubiquitous Computing. In Ravi Kumar Jain B, editor, *Trust Management in Virtual Environment*, pages 65 – 70. Icfai Books, 2008.
- [Rie08b] Sebastian Ries. Trust and Accountability. In Max Mühlhäuser and Iryna Gurevych, editors, *Handbook of Research on Ubiquitous Computing Technology for Real Enterprises*, chapter 16, pages 363 – 389. IGI Global, 2008.
- [Rie09] Sebastian Ries. Extending Bayesian Trust Models regarding Context-dependence and User Friendly Representation. In *Proceedings of the 2009 ACM Symposium on Applied Computing*. ACM Press, 2009.
- [RKM06] Sebastian Ries, Jussi Kangasharju, and Max Mühlhäuser. A Classification of Trust Systems. In R. Meersman, Z. Tari, and P. Herrero, editors, *On the Move to Meaningful Internet Systems 2006: OTM Workshops*, pages 894 – 903. Springer, 2006.
- [RKM07] Sebastian Ries, Jussi Kangasharju, and Max Mühlhäuser. Modeling Trust for Users and Agents in Ubiquitous Computing. In T. Braun, G. Carle, and B. Stiller, editors, *Kommunikation in Verteilten Systemen (KiVS)*, pages 51 – 62. Springer, 2007.

- [RS08] Sebastian Ries and Daniel Schreiber. Evaluating User Representations for the Trustworthiness of Interaction Partners. In *International Workshop on Recommendation and Collaboration (ReColl '08) in conjunction with the International Conference on Intelligent User Interfaces (IUI'08)*. ACM Press, 2008.
- [RZ02] Paul Resnick and Richard Zeckhauser. Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In Michael R. Baye, editor, *The Economics of the Internet and E-Commerce*, volume 11 of *Advances in Applied Microeconomics*, pages 127–157. Elsevier Science, 2002.
- [RZSL06] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The Value of Reputation on eBay: A Controlled Experiment. *Experimental Economics*, 9(2):79–101, June 2006.
- [Sab03] Jordi Sabater. *Trust and Reputation for Agent Societies*. PhD thesis, Universitat Autnoma de Barcelona, Spain, 2003.
- [Sav54] Leonard J. Savage. *The Foundations of Statistics*. J. Wiley, New York, 1954.
- [SFJ⁺03] Jean-Marc Seigneur, Stephen Farrell, Christian Damsgaard Jensen, Elizabeth Gray, and Yong Chen. End-to-End Trust Starts with Recognition. In *First International Conference on Security in Pervasive Computing*, 2003.
- [SH08] Tobias Straub and Andreas Heinemann. Security for Ubiquitous Computing. In Max Mühlhäuser and Iryna Gurevych, editors, *Handbook of Research on Ubiquitous Computing Technology for Real Time Enterprises*, chapter 15, pages 337 – 362. IGI Global, 2008.
- [SJ04] Jean-Marc Seigneur and Christian D. Jensen. Trading Privacy for Trust. In *Second International Conference on Trust Management (iTrust 2004)*, pages 93–107, 2004.
- [SS01] Jordi Sabater and Carles Sierra. REGRET: Reputation in Gregarious Societies. In Jörg P. Müller, Elisabeth Andre, Sandip Sen, and Claude Frasson, editors, *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 194–195. ACM Press, 2001.
- [SS02a] Jordi Sabater and Carles Sierra. Reputation and Social Network Analysis in Multi-agent Systems. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 475–482. ACM Press, 2002.

- [SS02b] Jordi Sabater and Carles Sierra. Social ReGreT, a Reputation Model based on Social Relations. *SIGecom Exchanges*, 3(1):44–56, 2002.
- [SS05] Jordi Sabater and Carles Sierra. Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, 24(1):33–60, 2005.
- [SVB06] Andreas Schlosser, Marco Voss, and Lars Brückner. On the Simulation of Global Reputation Systems. *Journal of Artificial Societies and Social Simulation*, 9(1), 2006.
- [TPJL06] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. TRAVOS: Trust and Reputation in the Context of Inaccurate Information Sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.
- [und09] undersound. undersound website [online]. 2009. Available from: www.undersound.org [seen 2009].
- [WB97] Mark Weiser and John Seely Brown. *Beyond Calculation: The Next 50 Years of Computing*, chapter The Coming Age of Calm Technology, pages 75–85. Copernicus, 1997.
- [Wei91] Mark Weiser. The Computer for the 21st Century. *Scientific American*, 265(3):94–110, 1991.
- [Win08] Franziska Wink. Trust-aided Online Recommendation Platform for Movies. Master’s thesis, Technische Universität Darmstadt, Germany, 2008.
- [WJI05] Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out Unfair Ratings in Bayesian Reputation Systems. *The ICFAIN Journal of Management Research*, 4(2):48 – 64, 2005.
- [WRH07] Stefan G. Weber, Sebastian Ries, and Andreas Heinemann. Inherent Tradeoffs in Ubiquitous Computing Services. In Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, and Marc Ronthaler, editors, *INFORMATIK 2007 - Informatik trifft Logistik, Band 1 der Proceedings zur 37. Jahrestagung der Gesellschaft für Informatik GI e. V.*, pages 364–368. GI, 2007.
- [WV07] Yao Wang and Julita Vassileva. Toward Trust and Reputation based Web Service Selection: A Survey. *International Transactions on Systems Science and Applications (ITSSA) Journal, Special Issue on New Tendencies on Web Services and Multi-agent Systems (WS-MAS)*, Vol 3(No. 2):118–132, 2007.

- [YS02] Bin Yu and Munindar P. Singh. An Evidential Model of Distributed Reputation Management. In *Proc. of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 294–301. ACM Press, 2002.
- [Zim94] Phil Zimmermann. *PGP User's Guide*. M.I.T. Press, 1994.
- [ZL04] Cai-Nicolas Ziegler and Georg Lausen. Spreading Activation Models for Trust Propagation. In *EEE '04: Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)*, pages 83–97. IEEE Computer Society, 2004.

Appendix A

Proofs

A.1 Proof for $E_{f,w,N}^{Beta} = E_{f,w,N}^{HTI}$

This proof is referred to Section 5.2.6.

$$\begin{aligned}
E_{f,w,N}^{Beta} &= \frac{r + r_0}{r + s + r_0 + s_0} \\
&= \frac{r + 2 \cdot f \cdot w \cdot (1 - \frac{r+s}{N})}{r + s + 2 \cdot f \cdot w \cdot (1 - \frac{r+s}{N}) + 2 \cdot (1-f) \cdot w \cdot (1 - \frac{r+s}{N})} \\
&= \frac{r \cdot N + 2 \cdot f \cdot w \cdot (N - (r+s))}{(r+s) \cdot N + 2 \cdot f \cdot w \cdot (N - (r+s)) + 2 \cdot (1-f) \cdot w \cdot (N - (r+s))} \\
&= \frac{r \cdot N + 2 \cdot f \cdot w \cdot (N - (r+s))}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \\
&= \frac{r \cdot N}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \\
&\quad + \frac{2 \cdot f \cdot w \cdot (N - (r+s))}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \\
&= \frac{r \cdot N \cdot (r+s)}{[(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))] \cdot (r+s)} \\
&\quad + \left(\frac{2 \cdot w \cdot (N - (r+s)) + N \cdot (r+s)}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \right. \\
&\quad \left. - \frac{N \cdot (r+s)}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \right) \cdot f \\
&= \frac{r}{r+s} \cdot \frac{(r+s) \cdot N}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \\
&\quad + \left(1 - \frac{(r+s) \cdot N}{(r+s) \cdot N + 2 \cdot w \cdot (N - (r+s))} \right) \cdot f \\
&= t \cdot c + (1 - c) \cdot f \\
&= E_{f,w,N}^{HTI} \blacksquare
\end{aligned} \tag{A.1}$$

A.2 Proof for $E_{0.5,1,\infty}^{Beta} = E_{Simple}^{Beta}$

This proof is referred to Section 5.2.6.2.

Assumption: The values of r and s are fixed and finite.

$$\begin{aligned}
E_{0.5,1,\infty}^{Beta} &= \lim_{N \rightarrow \infty} \frac{r + r_0}{r + s + r_0 + s_0} \\
&= \lim_{N \rightarrow \infty} \frac{r + 2 \cdot f \cdot w \cdot \left(1 - \frac{r+s}{N}\right)}{r + s + 2 \cdot f \cdot w \cdot \left(1 - \frac{r+s}{N}\right) + 2 \cdot (1-f) \cdot w \cdot \left(1 - \frac{r+s}{N}\right)} \\
&= \lim_{N \rightarrow \infty} \frac{r + 2 \cdot f \cdot w \cdot \left(1 - \frac{r+s}{N}\right)}{r + s + 2 \cdot w \cdot \left(1 - \frac{r+s}{N}\right)} \\
&= \frac{r + 2 \cdot f \cdot w \cdot \left(1 - \lim_{N \rightarrow \infty} \frac{r+s}{N}\right)}{r + s + 2 \cdot w \cdot \left(1 - \lim_{N \rightarrow \infty} \frac{r+s}{N}\right)} \tag{A.2} \\
&= \frac{r + 2 \cdot f \cdot w}{r + s + 2 \cdot w} \\
&= \frac{r + 1}{r + s + 2} \\
&= E_{Simple}^{Beta} \blacksquare
\end{aligned}$$

Erklärung¹

Hiermit erkläre ich, die vorgelegte Arbeit zur Erlangung des akademischen Grades “Dr. rer. nat.” mit dem Titel *Trust in Ubiquitous Computing* selbstständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel erstellt zu haben. Ich habe bisher noch keinen Promotionsversuch unternommen.

Darmstadt, 06. Mai 2009

Sebastian Ries

¹gemäß §9 Abs. 1 der Promotionsordnung der TU Darmstadt

Wissenschaftlicher Werdegang des Verfassers²

09/2000 – 09/2005	Studium der Informatik, Nebenfach Betriebswirtschaftslehre Technische Universität Darmstadt Abschluss: Diplom-Informatiker Diplomarbeitsthema: <i>Entwicklung und Implementierung einer Steuerungsarchitektur für einen autonomen mobilen Roboter in Java am Beispiel eines modifizierten Pioneer 2DX</i>
10/2005 – 09/2008	Stipendiat im DFG Graduiertenkolleg 492 <i>Infrastruktur für den elektronischen Markt</i> Technische Universität Darmstadt
10/2008 – 12/2008	Stipendiat im Center for Advanced Security Research Darmstadt (CASED) Technische Universität Darmstadt
seit 01/2009	Wiss. Mitarbeiter im Center for Advanced Security Research Darmstadt (CASED) Technische Universität Darmstadt

²gemäß §20 Abs. 3 der Promotionsordnung der TU Darmstadt