Volume 2 No. 3, March 2012 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Survey of Computer Trust and Reputation Models – The Literature Overview

Adis Medić
InfoSys LTD, Kolodvorska bb
77240 Bosanska Krupa,Una Sana Canton,Bosnia and Herzegovina
Ceravačka brda 106, 77000 Bihać, Una Sana Canton, Bosna and Herzegovina

ABSTRACT

P2P networks have been rapidly spread in the last few years. Nevertheless, together with its fast development, many security threats have also appeared, compromisingsensitive information and promoting frauds in electronic transactions Trust and Reputation management has arisen as one of the most innovative and accurate solutions to most of these threats. By using a trust and reputation model apeer who wants to interact with another peer in the community has more informationand, therefore, more opportunities to select the right partner to have a transaction with, rather than with a fraudulent one. The paper presents a description of some of the most representative trust and reputation models for P2P networks. It is explained how each of them works, how they manage the concepts of trust and reputation or how they gather information about other peers in the network. If we consider to overcome the uncertainty of risk on the open market, such as market information, we must establish a relationship of trust between users and service providers.

Keywords: *Trust and reputation, trustee, trust model*

INTRODUCTION

It is out of discussion the importance of trust and reputation in human societies. In this review, however, we will focus our attention on another discipline where the study of trust and reputation has acquired a great relevance in the last few years. We are talking about computer science and specifically about the area of distributed Artificial Intelligence. Two elements have contributed to substantially increase the interest on trust and reputation in this area: the multi-agent system paradigm and the spectacular evolution of e-commerce. The study of trust and reputation has many applications in Information and Communication technologies [17]. Trust and reputation systems have been recognized as key factors for successful electronic commerce adoption. These systems are used by intelligent software agents both as a mechanism to search for trustworthy exchange partners and as an incentive in decision-making about whether or not to honor conctracts. Reputation is used in electronic markets as a trust-enforcing, deterrent, and incentive mechanism to avoid cheaters and frauds. E-markets are not the single field of application, for example in [11], Barber and Kim use trust to improve the performance of belief revision mechanisms. Another important area of application in agent technology is teamwork and cooperation [12]. There are not many works that give a general view of trust and reputation from the point of view of computer science. Dellarocas in his article "The digitalization of Word-Of-Mouth: Promise and Challenges of Online Reputation Mechanisms" [13] presents an overview of online reputation mechanisms that are currently used in commercial web sites. In the area of trust, Grandison et al. in their work "A survey of trust in Internet application" [14] examine the various definitions of trust in the literature and provide a working definition of trust for Internet applications. There are also some proposals to establish a typology for reputation [15] and trust [16].

In this article we present the most popular and widely used computational models of trust and reputation.

This article needs to present an aspect of the inspection area of computational trust and reputation. In first row, we take a coherent approach to study current literature in trust and reputation systems, trust models in a first row. We believe that the proposed recommendations can be used as foundation for advancing the research framework agenda in Trust and Reputation systems.
We begin with an extensive overview of several most known trust and reputation systems. Subsequently, we provide the detailed description of the framework overview and its respective dimensions and also the detailed description of most populr and most used trust and reputation models. We also analyze their strengths and weak sides by effectively addressing some advanced features of the framework. Finally, we conclude the paper by explaining some of the open problems in this field by giving a conclusive comments and remarks. Most current models of trust focus on the first or the last of the above categories.

1. What is Trust? (Definition of trust!)

Gambetta (1990) gives the following definition of trust, which is commonly accepted: …trust, (or symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before he can monitor such action

(or independently of his capacity to monitor it) and in a context in which it affects his own action [18].

The term "subjective probability" is important in the above definition, because it points to a certain amount of arbitrariness in the trust metric. Thus, trust is not something that can be captured fully using objective measures. Trust is not an objective property, but a subjective degree of belief about others' competence and disposition. So, as in [19] points out, "however indispensable trust may be as a device for coping with the freedom of others, it is a device with a permanent and built-in possibility of failure". In [20] points out that "trusting a person means the trustor takes a chance that the trustee will not behave in a way that is damaging to the truster, given that choice". In general, "trust …presupposes a situation of risk."[21], [20] integrating various views, argues that trust is:

- A means of understanding and adapting to the complexity of the environment
- A means of providing added robustness to independent agents
- A useful judgement in the light of experience of the behaviour of others
- Applicable to inanimate others (including artificial agents)

Because the goal of this work is to find a definition that facilitates making computations with trust in social networks, it is natural to turn to the computer science literature. One of the most widely cited works is Marsh's unpublished PhD dissertation from the University of Stirling, "Formalising Trust as a Computational Concept" from 1994. In this work, Marsh gives careful attention to many facets of trust, from the biological to the sociological, in order to develop a model for trust among agents interacting in a distributed way. His model is complex and highly theoretical. Aside from the difficulties with implementation, it is particularly inappropriate for use in social networks because his focus was on interacting agents that could maintain information about history and observed behaviors. In social networks, users assign a trust as a single rating describing their connection to others, without explicit context or history. Thus much of the information necessary for a system like Marsh's is missing [22]. Web-based social networks are tools for the average web user. The definition of trust must be uncomplicated and straightforward enough than average web users understand what they are expressing, so they can express it accurately.

1.1. Kinds of Trust

The label "Trust" is quite amorphous, and is applied to a range of phenomena, involving objects, processes, and people. Three general types of trust have been identified. Dispositional trust describes an internal state of the trustor, a basic trusting attitude. This is "a sense of basic trust, which is a pervasive attitude towards oneself and the world" [23]. This trust is extremely open-ended and independent of any party or context. Dispositional trust has been further divided into two – type A concerns the truster's belief on others' benevolence, type B is the "disposition that irrespective of the potential trustee's benevolence, a more positive outcome can be persuaded by acting "as if' we trusted her" [23].

Impersonal trust refers to trust on perceived properties or reliance on the system or institution within which the trust exists. An example is the monetary system [23]. This can also be seen as dispositional trust directed towards an inanimate system. Impersonal trust is related to the notion of trust involved in learning, where Person A, while learning something from Person B, trusts that the facts s/he learned are true. Part of this trust is based on experience, part of it on institutional settings.

Interpersonal trust refers to the trust one agent has on another agent directly. This can be seen as dispositional trust directed towards an animate system. This trust is agent and context specific. For instance Person A might trust Person B in the context of fixing a furnace, but not for fixing a car.

Sometimes the word "trust" is used interchangeably with "faith". In this sense, "I trust him", implies that "I have an unjustified (unjustifiable?) belief that he will do the right thing." [24]. This is closely connected to dispositional trust and another notion of trust, where Person A has known Person B for a long time, and has interacted with him/her extensively. A now trusts B in a "non-specified" manner. That is, the set of situations for which A trusts B is an open set. The responsibilities of B are not set out in advance, and sometimes B may not even know what his/her responsibilities are. This is a complicated notion and we will call this Open Trust. This is person-specific and not as open-ended as dispositional trust.

Finally, the word "trust", as used in common parlance, implies something people have inside them — a fluctuating internal state, a sort of meter that goes up or down, depending on the situation and people involved. There is talk about "trust levels". Trust is also considered to have qualia, a phenomenal feeling of trusting or being trusted, associated with it. A breach of trust results in emotional changes, both in the trustor and the trustee. It is interesting to note that "the loss or pain attendant to unfulfillment of the trust is sometimes seen as greater than the reward or pleasure deriving from fulfilled trust" [20].

1.2. Characteristics of Trust

As pointed out earlier, trust is not an objective property, but a subjective degree of belief about a person, process or object. The degree can vary from complete trust

to complete distrust. There is also a situation where a person does not have an opinion on the trustworthiness of another person— i.e. the person is ignorant of the other person's trustworthiness.

Trust is not a blind guess or a game of chance, even though it involves a decision taken on the face of uncertainty. Trust involves a decision taken in anticipation of a positive outcome, and the decision is based on the knowledge and experiences of the trustor. It is this knowledge and experience that makes trust more than a blind guess.

[23] points out that trust reasoning is inductive. It is also dynamic and non-monotonic – additional evidence or experience at a later time may increase or decrease our degree of trust in a person.

An important feature of trust is that it cannot be brought about by will. The statement "trust me" does not work unless trust is present in the first place [20]. "I cannot will myself to believe that X is my friend, I can only believe that he is." [25] argues that trust is innate in children, and hence it is a fundamental mental state, and related to rationality and doubt. Drawing on Wittgenstein, he says that "one must begin somewhere, begin with not-doubting. This is not hasty and excusable, but is part of the process of judgement." Doubt comes after belief. All judgements must be seen in the context of an initial belief. One must first have faith to be able to lose it later.

Trust has the interesting property that the seeking of evidence for trust affects the evidence. One distrust has set in, it is difficult to know if such distrust is justified, because such experiments will not be carried out. Trust is thus capable of spiraling dramatically downwards [20]. On the other side, it is also capable of spiraling dramatically upwards, and can be self-reinforcing.

Trust also has the property that it grows with use, and decays with disuse. [26] points out that this property of trust makes it similar to other moral resources. Trust is closely related to confidence. The difference between the two is that trust presupposes an element of risk, while confidence dos not. In a situation of confidence, alternatives are not considered. As [20] points out, leaving the house without a gun every morning shows confidence in not needing to use the gun. However, leaving the house every morning without a gun, after considering the probability of having to use the gun that day, shows trust.

2. TRUST MODELS

2.2. Stephen Paul Marsh

One of the oldest proposed trust models is Marsh Trust model from 1994. The model is established with only direct interaction in trust process. Marsh proposed trust model with tree types of trust: Basic trust, General trust and Situational trust; but it is very important to put the accent on temporal notation (time variable) that is critical factor in these three types of trust:

- **Basic trust:** Models the general trusting disposition independently of who is the agent that is in front. It is calculated from all the experiences accumulated by the agent. Good experiences lead to a greater disposition to trust, and vice versa. The author uses the notation T_x^t to represent the trust disposition of agent x at time t.

- **General trust:** This is the trust that one agent has on another without taking into account any specific situation. It simply represents general trust on the other agent. It is noted as $T_x(y)^t$ representing the general trust that agent x has on agent y at time t.

- **Situational trust:** This is the amount of trust that one agent has in another taking into account a specific situation. The utility of the situation, its importance and the 'General trust' are the elements considered in order to calculate the 'Situational trust' [17].

Limitations

- The model considers the trustee to be a passive entity. All the action happens at the trustor's end. S/he takes the decision to delegate after a series of calculations, based on experience and other such factors. However, in the real world, the trustee is never passive. S/he constantly sends out signals to the truster, either positive or negative.
- The role of the environment is not captured. Though trust is considered to be 'situated', the notion of situatedness here is a limited one — used in the sense that the trust decision happens in a 'situation' and can vary across 'situations'. A situation is considered as something like a "box" or a "framework" within which the trusting decision is made. Also, for the model to work, situations need to be identifiable as similar or dissimilar. This situatedness is different from the larger notion of 'being in the world', where events are continuous. In the real world, a situation is not a slice of time and space, but a broader intermingling of contexts. To understand this notion better, think of A considering entering the taxi of B, who has just come out from a bar. Suppose that A has been driven safely by B for a number of times before, and B doesn't show any outward symptoms of being drunk. Should, or would, A trust B in this particular situation? Most likely not, even though B probably meets all the criteria set out in the model. It is interesting to note here that there is a possibility of B not being drunk, maybe s/he went in to check on a friend. But, in spite of that, A probably would not trust B. What complicates the trust situation here is

the context, which is external to the driving/driven situation, and part of a larger worldview.

- As observed earlier, trust is closely connected to reputation and social institutions. The role of these institutions are assumed, particularly in the Perceived_Competence variable, but the roles are not captured formally by the model. The Perceived_Competence variable is too broadly defined, and the mechanisms and the parameters that govern the perception is not specified.
- There is considered to be a distinguishable and independent state of mind called trust. However, there is a set of mental states (like beliefs) that contribute to trust (Castelfranchi, 1999). The role played by background beliefs in trust is not explored. Particularly, how beliefs are revised after a trusting (or symmetrically, non-trusting) decision.
- The crucial role played by communication in trust is not captured.

The notion of relevant information is a bottleneck. It introduces the frame problem.

2.3. OnLine reputation models

eBay [27], Amazon Auctions [28] are good examples of online marketplaces that use reputation mechanisms. eBay [27] is one of the world's largest online marketplace with a community of over 50 million registered users. Marketplace at local level is represented by ProdajIKupi [29] Web portal that uses eBay reputation mechanisms. Most items on eBay are sold through English auctions and the reputation mechanism used is based on the ratings that users perform after the completion of a transaction. The user can give three possible values: positive(1), negative(-1) or neutral(0). The reputation value is computed as the sum of those ratings over the last six months. Similarly, Amazon Auctions [28] use also a mean (in this case of all ratings) to assign a reputation value. All these models consider reputation as a global property and use a single value that is not dependent on the context. The information source used to build the reputation value is the information that comes from other agents that previously interacted with the target agent (witness information). They do not provide explicit mechanisms to deal with users that provide false information. A great number of opinions that "dilute" false or biased information is the only way to increase the reliability of the reputation value. In [13], Dellarocas points out that the commercial success of online electronic markets suggest the models have achieved their primary objective: 'generate sufficient trust among buyers to persuade them to assume the risk of transacting with complete strangers'. Certainly these reputation mechanisms have contributed to the success of e-markets like eBay but what is not clear is to which extend. There are several studies that try to analyze the properties of these models specially based on eBay data sets [13], [17].

2.3.1. Sporas model

Sporas [30] is an evolved version of the online reputation model. In this model, only the most recent rating between two users is considered. Another important characteristic is that users with very high reputation values experience much smaller rating changes after each update than users with a low reputation. Using a similar approach to the Glicko [31] system —a computational method used to evaluate the player's relative strengths in pairwise games, Sporas incorporates a measure of the reliability of the users' reputation based on the standard deviation of reputation values. This model has the same general characteristics as the previously commented online reputation mechanisms. However, it is more robust to changes in the behavior of a user and the reliability measure improves the usability of the reputation value.

2.3.2. Histos model

Histos [30] was designed as a response to the lack of personalization that Sporas reputation values have. The model can deal with direct information (although in a very simple way) and witness information. Contrary to Sporas, the reputation value is a subjective property assigned particularly by each individual. The treatment of direct interaction in this reputation model is limited to the use of the most recent experience with the agent that is being evaluated. The strength of the model relies on its use of witness information. Pairwise ratings are represented as a directed graph where nodes represent agents and edges carry information on the most recent reputation rating given by one agent to another. The root node represents the agent owner of the graph. This structure is similar to the TrustNet used by Schillo et al. [32]. The reputation of an agent at level X of the graph (with $X > 0$) is calculated recursively as a weighted mean of the rating values that agents in level X-1 gave to that agent. The weights are the reputations of the agents that rate the target agent. As we have seen, the agents who have been rated directly by the agent owner of the graph have a reputation value equal to the rating value. This is the base case of the recursion. The model also limits the length and number of paths that are taken into account for the calculation. The reputation value does not depend on the context and no special mechanisms are provided to deal with cheaters. A drawback of this model is the use of the reputation value assigned to a witness also as a measure of its reliability. If an agent is a good seller, this does not mean that it has to be also a reliable witness.

2.4. FIRE model

In the FIRE model [3], trust is evaluated within the context of a different number of information components:

- Interaction Trust (IT) that is built from the direct self experience of an agent with the other agents;
- Witness Reputation (WR) that is based on the direct observation of an agent's behavior by some third-party agent;
- Certified Reputation (CR), being one of the novelties in the FIRE model, consists of certified references disclosed by third-party agents. Such information is made available upon request of an inquiring agent. The CR component is desirable in the absence of direct interaction and when witnesses are self-interested and reluctant to share their experiences. Moreover, the use of CR enables agents to be freed from the cost of locating witnesses while their confidence rate of the anticipated trust value in not compromised.
- the last component is Role-based Trust (RT), which models the trust across predefined role-based relationships between two agents, e.g., (owned by the same company, friendship relationship, team-mate relationship) [33].

In this case, by defining and updating these roles in open Multi-Agent Systems (MAS) as well as assigning the expected trust value and belief strength (of relying agent) on them, RT is able to contribute in trustworthiness prediction for future interactions. It is worthwhile to mention that the significance of each component in the composite trust formula is adjusted automatically according to unforeseen changes in the environment. In this trust model, each component owns a trust formula with relevant rating weight function to determine the quality of ratings tailored to its responsibility. For instance, it seems sufficient for IT to design the weight function solely based on the recency of ratings whereas WR and CR have to take the credibility of rating into account as well. To address this requisite, FIRE has developed a mechanism to filter out the inaccurate reports revealed by unfaithful witnesses and penalises them accordingly. In so doing, it defines an inaccuracy tolerance threshold (L) to specify the maximal permitted differences between the actual performance and witness rating. Credibility of each rating is tuned to be inversely proportional to the differences, i.e., the higher the differences are, the lower the credibility [2]. Furthermore, the FIRE model defines a reliability measure to calculate the confidence level of an agent in believing that another agent can perform as expected. In general, it provides two types of reliability: rating reliability, which depends on the number of available ratings with high values, which depict the expected performance of the target agent. The other type is deviation reliability, which intends to examine the volatility of the target agent in accomplishing an agreement. Basically, it calculates the deviation of ratings around the produced expected value [2]. Intuitively, if the target agent showed an inconsistent behavior while countering a different requesting agent, its reliability value will be gradually affected negatively.

2.5. Anticipatory trust model

Typically, a trust model considers two main sources of information to estimate trust: direct experience, sometimes referred to as direct trust or interaction trust, and recommendations, often called witness-information or \word of mouth". In our model we keep this distinction between direct experience and recommendations, but in addition, we distinguish between the recommendations about third party agents and the recommendations provided by an agent about itself, what we call advertisements. All in all, our model builds trust upon three components, namely: Direct Trust (DT), Advertisements-based Trust (AT), and Recommendations-based Trust (RT). In order to adapt quicker to the dynamic and uncertain nature of an open environment, an agent can anticipate or have expectations (not necessarily rational) about the possible consequences of its actions, therefore, we distinguish between the historic components of trust, based on past information only, and the anticipatory components. In our model, only the Advertisements-based Trust and the Recommendations-based trust are anticipatory, while trust by direct experience is purely an historic belief. To simplify the dynamics of a multi-agent system, we use a discrete time model made up of time steps. A time step represents the minimal time period an agent requires to take decisions, act, and perceive the result of its actions [34].

2.6. OpenPGP

The PGP trust model has some particular characteristics. First of all, (only) three levels of trust are supported: complete trust, marginal trust, and no trust. The owner of the key ring, who needs to manually assign these trust values for all other users, automatically receives full trust (also called implicit or ultimate trust). When a user places trust in an introducer, implicitly it means that the user possesses a certain amount of confidence in the introducer's capability to issue valid certificates, i.e. correct bindings between users and public keys. This is the general intuition, but the actual meaning of the three trust levels in PGP is not clearly defined [35].

2.7. SOLAR trust model

The Solar Trust Model [36] overcomes many of the limitations inherent in the designs of the other trust

models. It does this by providing a simple and efficient method by which many levels of trust can be implemented, by permitting an unlimited number of independent (certificate authorities) CAs with no requirement for a central PCA, and by defining a procedure for determining the trustworthiness of a message that has been signed by a CA with whom the receiver of the message has no direct relationship. To demonstrate how the Solar Trust Model works, let a CA be defined as any entity which issues digital certificates. If CA_1 and CA_2 are two certificate authorities, then CA_2 can establish a set of rules to determine how much it trusts messages signed by CA_1. This set of rules is called a trust relationship. For example, suppose that Bob wants to read a document sent by Alice and signed by CA_1, which uses a fixed procedure to ensure Alice's identity. Now, if CA_1 can prove to CA_2 that the procedure used to verify Alice's identity meets CA_2's criteria, then CA_2 can be certain that Alice's document is indeed Alice's. On the other hand, if CA_3 is another certificate authority, and its policy for verifying the identity of a second person, say Ted, is unknown to CA_2, and Ted is sending a message to Bob, then CA_2 will not have adequate assurance to believe any claims about the identity of Ted made by CA_3. If CA_2 has a set of rules that say that any certificate authority that uses the same procedures as itself for verifying the identity of the sender of a message can be trusted more than a certificate authority that does not use those procedures, and if CA_2 can verify that CA_1 uses these procedures, then CA_2 can say that it has a stronger trust relationship with CA_1 than with CA_3. Let a solar system be defined as the representation of an ordering of trust relationships with respect to a specific certificate authority. It is helpful to think of a solar system as a series of objects that exist within concentric orbits around a central body, much as the planets in a solar system orbit around the sun. For any set of certificate authorities CA_1 through CA_N, CAI_i is the central body or primary in its own solar system, and all of the other certificate authorities with which CA_I has established a trust relationship are objects or planets in orbit around the primary. An ordering of trust can now be established for all certificate authorities which are planets in a solar system, with distance from the solar system's primary indicating the level of the trust relationship between the primary and a planet. A certificate authority places itself in the 0'th orbit of itself, because it trusts itself completely. Orbits 1 through n are occupied by all other certificate authorities in the solar system, where a certificate authority in orbit m is more trusted than a certificate authority in orbit m+1. It is possible for two or more certificate authorities to share the same orbit, or for orbits to be empty. If certificate authority CA_I does not have a trust relationship with certificate authority CA_X, then CA_X is not a planet in CA_I's solar system. Note that since every certificate authority has its own solar system, the primary certificate authority in one solar system can (and often will) be a planet in another solar system, and that a certificate authority can be a planet in many different solar systems. It is also important to recognize that two certificate authorities do not necessarily have the same trust relationship with each other. Since a trust relationship is derived from a set of rules which each certificate authority independently establishes (although common rule sets can be established), there is no guarantee that two certificate authorities will ever have the same trust relationship with each other. Furthermore, if one certificate authority has a trust relationship with a second certificate authority, then it is not guaranteed nor is it necessary that the second certificate authority has a trust relationship with the first. Although a certificate authority can establish direct trust relationships with many other certificate authorities, it is infeasible that it will establish such relationships with all other certificate authorities. The solar trust model solves this problem by establishing indirect trust relationships. For example, if CA_2 is a planet in the solar system of CA_1, and CA_3 is a planet in the solar system of CA_2, then CA_3 is a "moon" of CA_2 in the solar system of CA_1. This can be extended through any number of iterations. Since a certificate authority can be a planet in many different solar systems, that certificate authority can be a moon of many different certificate authorities in the same solar system. Although the moon of a planet is regarded as being in the same orbit as the planet, the moon is not considered to be the same entity as the planet. It is important to understand that an indirect trust relationship does not imply a transitivity of trust. When trust is transitive, then if CA_1 trusts CA_2, and CA_2 trusts CA_3, then CA_1 must trust CA_3 in the same way that CA_2 trusts CA_3. In an indirect trust relationship, if CA_1 trusts CA_2, and CA_2 trusts CA_3, then CA_1 may or may not trust CA_3. Furthermore, since CA_1 relates to CA_3 indirectly, CA_1 is unlikely to trust CA_3 as much as it does CA_2. At this point, it should be noted that if a message comes from a moon, it may appear to come from many different orbits. To resolve this issue, we define the path of trust taken by a message as the set of certificate authorities that sign the message in the order in which they are signed. For example, if a message is signed first by CA_3, then by CA_2, then by CA_1, the path taken by the message is CA_3, CA_2, CA_1. Given any two certificate authorities with an indirect trust relationship, it is likely that there is more than one path that a message could take between the two certificate authorities. However, the only path that counts is the one that the message actually takes. If a CA appears more than once on the same path of trust, it is regarded as a different CA each time that it appears. In order to improve efficiency, it may be desirable for paths of trust between CA_S to be computed in advance. There are several methods by which this may be implemented. In the first method, a three way handshake is used to send a trusted path to a CA that sends a message. For example, if CA_2 wished to send a message to CA_1, CA_2 would send a

request for a trusted path to CA_1. CA_1 would then send an acceptable path of trust back to CA_2. CA_2 would then send its message to CA_1 along the path of trust. Note that CA_1 does not have to believe the origin of the path request from CA_2, since CA_1 can determine whether or not the final message came from CA_2, and can determine whether or not it trusts messages from CA_2. Another method would involve the computation of trust tables, which would be similar in form to the routing tables used in IP protocol routers.

Finally the ICE-TEL trust model proposes publishing paths using public forums. [39] In addition to establishing the orbit from which a message derives, the concept of a path of trust also allows the determination of the levels of trust for messages that are sent between certificate authorities that do not have direct relationships. When a message is first signed by a certificate authority, that certificate authority can attach a copy of its rule set to the message. As the message is passed from certificate authority to certificate authority, each certificate authority concatenates its rule set to the rule set that is passed to it, forming a composite rule set. The certificate authority that ultimately receives the message applies all of the rules in the composite rule set, until the message is either rejected as untrustworthy, or is accepted after meeting the requirements of all of the rules. The rule set for each individual certificate authority is represented using the Solar Trust Model Rule Set Header shown below.

Field A: Local Direct Range
Field B: Local Indirect Range (m)
Field C:Maximum Path Lenght(p)
Field D:Premited Local Range for Next CA (q)
DATA

Figure 1 The Solar Trust Model Rule Set Header

The following fields are represented in the Solar Trust Model Rule Set Header:

- Field A: Local Direct Range: Trust all messages that have been directly signed by a CA in an orbit with a number (k) no greater than this value.
- Field B: Local Indirect Range: Trust messages that have been indirectly signed by a CA in an orbit with a number that is less than or equal to this value. Do not trust any messages that have been signed by a CA in an orbit with a number that is greater than this value.
- Field C: Maximum Path Length: A message can be trusted only if the total number of CA_S that have signed the message is no greater than this value (p).
- Field D: Permitted Local Range for Next CA: A message that has been indirectly signed by a CA inside the local indirect range can be trusted only if it came from an orbit in that CA's system with a number no greater than this value (q).

2.8. Schillo et al.

The trust model proposed by Schillo et al. [32] is intended for scenarios where the result of an interaction between two agents (from the point of view of trust) is a boolean impression: good or bad; there are no degrees of satisfaction. More concretely, to make the experiments they propose a Prisoner's dilemma set of games with a partner selection phase. Each agent receives the results of the game it has played plus the information about the games played by a subset of all players (its neighbors). The result of an interaction in this scenario is an impression on the honesty of the partner (if she did what she claimed in the partner selection phase) and which was the behavior she had according to the normal prisoner's dilemma actions (cooperation or defection). The model is based on probability theory. The formula to calculate the trust that an agent Q deserves to an agent A (that is, the probability that the agent A be honest in the next interaction) is T(A,Q) = e n where n is the number of observed situations and e the number of times that the target agent was honest. Complementing the information that results from direct interaction/ observation, an agent can interview other agents that it has met before. Each agent uses a different TrustNet data structure. A TrustNet is a directed graph where nodes represent witnesses and edges carry information on the observations that the parent node agent told the owner of the net (the root node of the TrustNet) about the child node agents. In this model, testimonial evidence from interviews may be , as witnesses may have different motives and may try to deceive agents about their true observation. Thus, every agent is confronted with noise in the information and also with the possibility that the source of information itself is biasing the data. The answer of witnesses to a query is the set of observed experiences (and not a summary of them). Given that, the authors assume that it is not worth it for witnesses to give false information. A witness will not say that a target agent has played dishonest in game x if this was not the case because the inquirer could have observed the same game and, therefore, notice that the witness is lying. Witnesses do not want to be uncovered by obviously betraying. Therefore, the model assumes that witnesses never lie but that can hide (positive) information in order to make other agents appear less trustworthy. Assuming that negative information will be always reported by witnesses, the problem is reduced to know to what extend those witnesses have biased the reported data (hiding positive observations). To do that, betraying (hiding information) is modelled as a stochastic process where an agent decides to inform about a positive fact of another agent with probability p and hide that information with probability (1 − p). The application of this process

can be seen as a Bernoulli-experiment and the repetition of the experiment as a Bernoulli-chain. Probability theory is then used to estimate the hidden amount of positive information. This process can be applied recursively from the target agent through all its ancestors up to the root node of the TrustNet. With all this process, the agent is building for each piece of information an approximation of what the witnesses would have said if they had been completely honest about their information. As the information from the witnesses comprises the list of observations it can be collated to eliminate the "correlated evidence" problem [38]. This, however, cannot be done for the hidden information. The proposed solution in this case is based on the assumption that the relation of overlapping of the data in reported and non reported (hidden) information is constant. No information is given about how to combine direct experiences with information coming from witnesses. The trust value is a subjective property assigned particularly by each individual and it does not depend on the context.

2.9. Abdul-Rahman and Hailes

This trust model [23] uses four degrees of belief to typify agent trustworthiness: vt (very trustworthy), t (trustworthy), u (untrustworthy) and vu (very untrustworthy). For each partner and context, the agent maintains a tuple with the number of past experiences in each category. Then, from the point of view of direct interaction, the trust on a partner in a given context is equal to the degree that corresponds to the maximum value in the tuple. For instance, if the associated tuple of a partner in a given context is (0, 0, 4, 3) the trust assigned to that partner will be t (trustworthy) that corresponds to the third position in the tuple. If there is more than one position in the tuple with the maximum value, the model gives an uncertainty trust degree according to a table of pattern situations that cover this cases. There are three possible uncertainty values (and the corresponding patterns) to cover the situations where there are mostly good and some bad, mostly bad and some good and equal amount of good and bad experiences. This is the only model analyzed where before combining the information that comes from witnesses, the information is adjusted according to previous information coming from that witness and the consequent outcomes that validate that information. For example, suppose a informs to x that b is vt and x's evaluation of its experience with b is merely t. Next time that a gives information to x, x will adjust the information accordingly before taking it into account. The problem of this approach is that it is not possible to differentiate those agents that are lying from those agents that are telling the truth but "think" different. Although there are scenarios where this is not important (like the scenario suggested by the authors where agents recommend goods to other agents) it can be a limitation in some scenarios. In order to combine information, the model gives more relevance to the information coming from those agents with a more similar point of view. That is, it gives more importance to the information that needs to be adjusted very little or, even better, does not need to be adjusted at all because it comes from agents that have a similar perspective in a given context. Contrarily to other trust models where witness information is merged with direct information to obtain the trust on the specific subject, this model is intended to evaluate only the trust on the information given by witnesses. Direct experiences are used to compare the point of view of these witnesses with the direct perception of the agent and then be able to adjust the information coming from them accordingly.

2.10. Esfandiary and Chandrasekharan

In the trust model proposed by Esfandiari and Chandrasekharan [39], two one-on-one trust acquisition mechanisms are proposed. The first is based on observation. They propose the use of Bayesian networks and to perform the trust acquisition by Bayesian learning. In the simplest case of a known structure and a fully observable Bayesian network, the learning task is reduced to statistical considerations. The second trust acquisition mechanism is based on interaction. The approach is the same used in [40]. There are two main protocols of interaction, the exploratory protocol where the agent asks the others about known things to evaluate their degree of trust and the query protocol where the agent asks for advice from trusted agents. A simple way to calculate the interaction-based trust during the exploratory stage is using the formula Tinter(A,B) = number of correct replies total number of replies .

To deal with witness information, each agent builds a directed labeled graph where nodes represent agents and where an (a,b) edge represents the trust value that a has on b. Edges are absent if the trust value is unknown. In such a graph, there is the possibility of having cycles that artificially decrease the trust value and different paths that give contradictory values. To solve this problem, instead of using a single value for trust the model uses a trust interval determined by the minimum and maximum value of all paths without cycles that connect two agents. The authors claim that the calculation of this trust interval is equivalent to the problem of routing in a communication network and, therefore, known distributed algorithms used to solve that problem can be successfully applied to this situation. To allow a multi-context notion of trust (see section 2.4) the authors propose the use of colored edges, with a color per task or type of trust. Trust would only propagate through edges of the same color. Finally, the authors propose a trust acquisition mechanism using institutions, what they call institutionalized trust. This is similar to the concept of system reputation in the ReGreT in section 2.17. [41], [42] model. The idea is to

exploit the structure in the environment to determine trust values. No information is given about how to combine the different trust acquisition mechanisms.

2.11. Yu and Singh

In the model proposed by Yu and Singh [43], [44], [45], the information stored by an agent about direct interactions is a set of values that reflect the quality of these interactions (Quality of Service – QoS). Only the most recent experiences with each concrete partner are considered for the calculations. Each agent defines an upper and lower threshold that define the frontier between what are considered QoSs ascribed to trustworthy agents, QoSs with no clear classification and QoSs ascribed to non trustworthy agents. Then, using the historic information together with Dempster-Shafer theory of evidence, an agent can calculate the probability that its partner gives a service ascribed to each one of these groups. If the difference between the probability that the service belongs to the first and latest group is greater than a threshold for trustworthiness, the agent being evaluated is considered a trusty agent. There are two kinds of information that a witness can provide when it is queried about a target agent. If the target agent is one of its acquaintances it will return the information about it. If not, it will return referrals to the target agent that can be queried to obtain the information. These referrals, when queried, can provide the desired information or provide again new referrals. If the referral that finally gives the information is not far away to a depth limit in the chain, its information will be taken into account. The set of referral chains generated due to a query is a TrustNet similar to that used by Schillo et al. [32] and in the Histos [31] model (sections 2.7 and 2.8). As we have said this model uses Dempster-Shafer theory of evidence as the underlying computational framework. In this case, to aggregate the information from different witnesses they use Dempster's rule of combination. This model does not combine direct information with witness information (the two sources of information that takes into account). If direct information is available, that's the only source that is considered to determine the trust of the target agent. Only when direct information is not available the model appeals to witness information.

The Trust and Reputation model designed by Yu and Singh [10], [45] contains various distinctive features which surpass other available models in some contexts. Major determinant of these model is a decentralized reputation management model to locate the rightful witnesses in MAS in order to evaluate the trustworthiness of an Service Provider (SP) which is willing to communicate This model of reputation management exploits two information components. The first one contains the agent's local belief built as a result of its direct interaction with other agents. The second one includes the testimonies of third-parties that can be beneficial in the absence of local ratings. In this model, in order to estimate the total belief regarding the trustworthiness of a particular agent, the requesting agent combines a local belief in conjunction with third-party testimonies to achieve a more accurate evaluation. Furthermore, Yu and Singh propose a novel trust network which intends to locate the most appropriate witnesses in a multiagent system. In this model, each agent is surrounded by a number of acquaintances among whose subsets can be neighbours. When a requesting agent wants to evaluate the trustworthiness of a particular agent, it will send a query to the neighbours of that agent asking for their perception regarding the target agent. Unless the neighbours have not had any direct experiences with that agent they respond by their testimonies; otherwise, they will reply by returning a series of referrals. The number of referrals is limited by the branching factor and depthLimit parameters [6] so as to limit the effort expended in pursuing referrals. This process successfully terminates if an adequate number of ratings are received and it encounters failures when the depthLimit is reached and neither ratings nor referrals are gathered [45]. Note that each individual agent maintains a two-dimensional model of each acquaintance. The first dimension indicates their ability to act in a trustworthy manner, which is called expertise and the other one signifies their sociability in referring to suitable trustworthy agents. Depending on their competency in fulfilling either of the above-mentioned qualities, acquaintance models are modified to reflect their actual performance to be used in future interactions. The other major concern of this model is dealing with deceptive agents who deliberately disseminate misinformation through network for their self-interest. The proposed model considers three types of deceptions [6]: complementary, exaggerative positive and exaggerative negative. This classification is based on the behavioral model of the participants in giving ratings. For instance, if agents intentionally give controversial ratings, they may be detected as malicious agents with complementary model of deception. Such agents will lose credibility in the update phase. Similarly, an agent with exaggerative positive tendency acts rather untruthfully in the system. To clarify, even if it is not fully satisfied with the performance of a particular agent, it provides a higher rating than it actually experienced. The possible motivation for this behavior could be receiving of a commission from the other agent. Consequently, the credibility of this agent is reduced in proportion with its dishonesty. Moreover, depending on the system's circumstances, this model defines an exaggeration coefficient, which determines how much agents could lie before they are considered as being exaggerative and not a complementary deceptive agent. Note that after the actual interaction with the recommended target agent, the requesting agent re-calculates the weight of the witnesses

and updates their credibility degree for subsequent reputation prediction processes. Finally, in order to tackle with the uncertainty factors inherent in open MAS, this reputation management model benefits from the Dempster-Shafer theory of evidence [46] as an underlying computational framework. According to this theory, lack of belief does not necessarily imply disbelief in the system. Thus, instead of assuming total disbelief as initial value for newcomers, it is replaced by a state of uncertainty. In other words, with the help of the theory of evidence, Yu and Singh's model is able to differentiate between having a bad reputation and no reputation at all [45], [17]. Moreover, to predict total belief it utilizes Dempster's rule of combination [3] as an aggregation method which combines evidences to compute new a belief value. In addition, this model describes a variant of the Weighted Majority Algorithm (WMA) [47] in order to fine tune the weight of advisers for the purpose of deception detection after actual successful or unsuccessful interactions.

Finally, Yu and Singh have attempted to design a reputation model compatible with the inherent dynamicity in open MAS. For example in their approach, individuals can dynamically choose their neighbours from their current acquaintances. In addition, when the majority of agents exhibit volatile and changing behavior, it can adaptively adjust the exaggerative coefficient to a higher value to swiftly filter out deceitful agents from the system.

2.12. Sen and Sajja

In Sen and Sajja's [48] reputation model, both types of direct experiences are considered: direct interaction and observed interaction. In the scenario where this model is used, observations are noisy, i.e., the observations may differ somewhat from the actual performance. Only direct interaction gives an exact perception of the reality. Reinforcement learning is the chosen mechanism to update the reputation value. Due to the noise underlying observations, the rule used to update the reputation value when there is a new direct interaction has a greater effect than the rule used to update the value when there is a new observation. The reputation value ranges from 0 to 1. A value greater than 0.5 represents a good performer and a value less than 0.5 represents a bad performer. Agents can query other agents about the performance of a given partner. The answer is always a boolean value that says if the partner is good or not. In this model, liars are assumed to lie consistently, that means that every time they are queried, they return a good value for a bad target agent and vice versa. To decide, from the point of view of witness information, if a partner is good or not, the model uses the number of positive and negative answers received from witnesses. Knowing the number of witnesses and how many of them are liars, the model provides a mechanism to calculate how many agents should be queried to be sure that the likelihood of selecting a good partner has at least a certain value. The subset of agents to be queried is selected randomly from the set of possible witnesses although the authors claim it is easy to add a smarter selection process based on a trust mechanism.

Because the objective of this work was to study how agents use word-of-mouth reputations to select one out of several partners, agents only use witness information to take decisions. Direct experiences are only used as pieces of information to be communicated to the others. Therefore, no indication is given by the authors about how to combine direct experiences with witness information to obtain a final reputation value.

2.13. Afras

The main characteristic of this model [49] is the use of fuzzy sets to represent reputation values. Once a new fuzzy set that shows the degree of satisfaction of the latest interaction with a given partner is calculated, the old reputation value and the new satisfaction value are aggregated using a weighted aggregation. The weights of this aggregation are calculated from a single value that they call remembrance or memory. This factor allows the agent to give more importance to the latest interaction or to the old reputation value. The remembrance factor is modeled as a function of the similarity between (1) the previous reputation and the satisfaction of the last interaction and (2) the previous remembrance value. If the satisfaction of the last interaction and the reputation assigned to the partner are similar, the relevance of past experiences is increased. If the satisfaction of the last interaction and the reputation value are different, then it is the relevance of the last experience what is increased. The notion of reliability of the reputation value is modeled through the fuzzy sets themselves. A wide fuzzy set for a reputation value represents a high degree of uncertainty over that value while a narrow fuzzy set implies a reliable value. Recommendations from other agents are aggregated directly with the direct experiences. The weight given to each factor (old reputation value and new opinion) is dependent on the reputation that the recommender has. Recommendations coming from a recommender with a high reputation has the same degree of reliability as a direct experience. However, opinions from an agent with bad reputation are not taken into account. To calculate the reputation of recommenders, the agent compares the recommendation with the real behavior of the recommended agent after the interaction and increases or decreases the reputation of the recommender accordingly.

2.14. Carter et al.

The main idea behind the reputation model presented by Carter et al. [50] is that 'the reputation of an agent is based on the degree of fulfillment of roles ascribed to it by

the society'. If the society judges that they have met their roles, they are rewarded with a positive reputation, otherwise they are punished with a negative reputation.

„Each society has its own set of roles. As such, the reputation ascribed as a result of these roles only makes sense in the context of that particular society" [50]. According to this, it is impossible to universalize the calculation of reputation. The authors formalize the set of roles within an information-sharing society and propose methods to calculate the degree of satisfaction with each of these roles. An information-sharing society is a society of agents that attempt to exchange relevant information with each other in the hope of satisfying a user's request. They identify five roles:

- Social information provider: 'Users of the society should regularly contribute new knowledge about their friends to the society'. This role exemplifies the degree of connectivity of an agent with its community. Each particular recommendation made by a user has a weight associated to it. This weight indicates the strength of the recommendation and is the product of a time decay factor and the reputation of the recommender. The degree to which the social information provider role is satisfied by a given user is calculated as the summation of all these weights, mapped in the interval [0,1].
- Interactivity role: 'Users are expected to regularly use the system'. Without this participation the system becomes useless. The degree of satisfaction for this role is calculated as the number of user operations during a certain period of time divided by the total number of operations performed by all the users in the system during the same period.
- Content provider: 'Users should provide the society with knowledge objects that reflect their own areas of expertise'. The degree of satisfaction is reflected by the quality of the information agents that belong to that user. The quality of an agent is measured considering how close is the subject of that information agent tothe user's interest. The idea is that users that create information agents related to their areas of expertise will produce higher quality content related to their interest than those who do not. Administrative feedback role: 'Users are expected to provide feedback information on the quality of the system. These qualities include easy-of-use, speed, stability, and quality of information'. Users are said to satisfy this role by providing such information.
- Longevity role: 'Users should be encouraged to maintain a high reputation to promote the longevity of the system'. The degree of satisfaction of this role is measured taking into account the average reputation of the user.

Given that, the user's overall reputation is calculated as a weighted aggregation of the degree of fulfillment of each role. The weights are entirely dependent on the specific society. The reputation value for each agent is calculated by a centralized mechanism that monitors the system. Therefore, the reputation value of each user is a global measure shared by all the observers.

2.15. Castelfranchi and Falcone

The second formal model, suggested in [24], is more cognitively oriented. The model brings in the role of belief early on, and makes the basic assumption that only an agent with goals and beliefs can trust. The model considers trust to be a "cluster" mental state. Trust is thus compositional, and is made up of some basic ingredient beliefs. The degree of trust is based on the "strength" of its component beliefs. One of the interesting suggestions put forward by the model is that trust is the mental counterpart of delegation. The model is thus delegation-driven. Delegation is an action; a set of beliefs contributes to the action of delegation, and once an action is delegated, this cluster of beliefs makes up trust. Castelfranchi points out that the decision to delegate has no degrees, it is an either/or decision. However, beliefs have degrees. So trust has degrees as well. Essentially, the action of delegation arises when cumulative degrees of belief reach a threshold. This is quite similar to the idea of the cooperation threshold suggested by Marsh.

Castelfranchi breaks up trust into the following beliefs:

- Competence belief: x should believe that y can do action α
- Disposition belief: x should believe that y is willing to do α

Dependence belief: x believes it has to rely on y (strong dependence) or x believes it is good to rely on y (weak dependence)

There are other beliefs that contribute to the decision, which are related, but not entirely independent of the above beliefs:

- Fulfillment belief: x believes that goal g will be achieved (thanks to y in this case)
- Willingness belief: x has to believe that y has decided and intends to do action α
- Persistence belief: x believes that y is stable in his intentions, and will persist with α
- Self-confidence belief: x believes that y knows that y can do α.

In this case, there are some obvious limitation, One of the problems with the formulation is the broad scope of the competence belief. For instance, it is not clear why a "fulfillment belief" is needed, given the "competence" and "disposition" beliefs. If the "fulfillment

belief" refers to the nature of the task (whether it can be done or not), then the "competence" belief has to be defined more narrowly. A broad definition of competence covers fulfillment as well. The way the beliefs are defined currently, the combination of "competence" and "disposition" exhausts "fulfillment". There are some other redundancies as well.

Another problem is the definition of trust, as being strictly relative to a goal. This is not entirely true. In the case of Open Trust (mentioned earlier), an agent trusts another for an open set of tasks. No goals are specified in advance in such a case. The model also depends on the modeling of the trustee's (y's) mental state to get to a trust metric. All the beliefs x has relate to y's mental state. How x arrives at a belief about y is not specified. The role of the environment is considered only marginally, as facilitating or negatively affecting the execution of the action. Castelfranchi (forthcoming) considers the role of emotions in trust, but does not incorporate it into the model. Besides these, the model suffers from most of the problems pointed out in the Marsh model, including the large focus on the internal state of the trustor. The model also ignores communication, which is a crucial component of any trusting decision.

The trust model proposed by Castelfranchi and Falcone [51] is a clear example of a cognitive trust model. The basis of their model is the strong relation between trust and delegation. They claim that 'trust is the mental background of delegation'. In other words, the decision that takes an agent x to delegate a task to agent y is based on a specific set of beliefs and goals and this mental state is what we call 'trust'. Therefore, 'only an agent with goals and beliefs can trust'. To build a mental state of trust, the basic beliefs that an agent needs are:

- **Competence belief:** the agent should believe that y can actually do the task.
- **Dependence belief:** the agent believes that y is necessary to perform the task or that it is better to rely on y to do it.
- **Disposition belief:** not only is necessary that y could do the task, but that it will actually do the task. In case of an intentional agent, the disposition belief must be articulated in and supported by two more beliefs:
    - **Willingness belief:** the agent believes that y has decided and intends to do α (where α is the action that allows the goal g).
    - **Persistence belief:** the agent believes that y is stable in its intentions of doing α.

The first two beliefs compound what they call the core trust and together with the disposition belief, the reliance. Supported and implied by the previous beliefs, another belief arises:

- **Fulfillment belief:** if the agent "trust in y for g", the agent decides: (i) not renouncing to goal g, (ii) not personally bringing it about, (iii) not searching for alternatives to y, and (iv) to pursue g through y.

To summarize, trust (by Castelfranchi and Falcone) is a set of mental attitudes characterizing the "delegating" agent's mind (x) which prefers another agent (y) doing the action. y is a cognitive agent, so x believes that y intends to do the action and y will persist in this.

2.16. TRAVOS

The TRAVOS (Trust and Reputation model for Agent-based Virtual Organizations) system is developed to ensure highquality interaction between the participants of a large open system [51]. It exploits two information sources to assess the trustworthiness of the participants: Direct Interaction and Witness Observation. To derive trust, this model relies greatly on its direct experiences and refuses to combine others' opinions unless they are really required. For this purpose, it provides a confidence metric to determine whether the personal experiences are sufficient to make an acceptable judgment with respect to a particular SP or not. If not, it disseminates queries to obtain additional observations from other witnesses who claim to have had previous interaction with that certain SP. Specifically, this Trust and Reputation model utilizes a single rating system such that the outcomes of the interactions are summarized in a single variable which indicates an overall performance. Here, witnesses share the history of their interactions in a tuple which contains the frequency of successful and unsuccessful interaction results.

Moreover, in order to deal with inaccurate reputation providers, TRAVOS takes advantage of an exogenous approach presented in [7], [52]. According to this approach, instead of calculating the reliability of the provided recommendation based on its deviation from mainstream opinions, it calculates the probability that a particular correspondent provides accurate reports given its past opinions and proportionally adjusts the influence of its current observations afterwards. To clarify, as a first step, TRAVOS considers the actual results of all previous interactions with collection of SPs in which the agent provided similar observations. Then, by means of comparing the variables of their beta distributions it is able to measure the degree of accuracy of that certain agent. Then, by means of comparing their corresponding expected values, TRAVOS is able to conclude the honesty and accuracy of a rater's current observation [51], [4]. In the second step, this Trust and Reputation system attempts to decrease the effect of unreliable opinions on a final

computed reputation value. An untruthful agent could considerably affect the reputation of the queried SP by providing a huge number of unfair ratings. This problem arises because of its method of reputation combination, which is based on a simple summation of all the provided opinions. To rectify this, TRAVOS adopts techniques to reduce the amount of ratings unless the accuracy degree of the opinion provider is very high. TRAVOS is a probabilistic trust model which uses beta distribution probability functions to calculate the likelihood of certain SPs fulfilling agreed obligations given its past personal experiences and reputation information. Even though the performance of TRAVOS is validated in a decentralized online marketplace with pre-determined agent populations, it can extend easily to large scale open systems [4]. Yet, it lacks the ability to address the bootstrapping problem as well as the dynamicity in participants' behavior which may change their attitude overtime [2]. Besides, this model assumes that reputation information is accessible upon demand and does not present any approach for locating witnesses. Using probability theory, TRAVOS provides a novel approach for detecting and filtering malevolent witnesses. It adjusts the effect of provided opinions on the trustworthiness measurement; corresponding to the accuracy degree of their reporters. However, it does not provide any reliability measure to assess the degree of confidence of a truster agent in achieving the expected performance from the trustee. Furthermore, since TRAVOS is based on a single rating system such that the reputation is shared in the form of frequency of successful and unsuccessful interaction results, it is incapable of providing suitable recommendations in an environment with competitive SPs offering variety of services in different contexts. It is noteworthy to mention that, this Trust and Reputation system is mostly comparable with BRS in the context of handling inaccurate reports [52], [53]. Nevertheless, BRS is based on an endogenous approach which presumes that the majority of reputation sources provide an accurate opinion thus discards any opinions that deviate considerably from the average [51].

2.17. LIAR model

L.I.A.R. [54] ("Liar Identification for Agent Reputation") is a model for the implementation of a social control of agent interactions. The behaviour that is under control is the communicative behaviour of agents. L.I.A.R. provides a formalism to express the communicative rules that should be respected in the system. Then, the L.I.A.R. model gives tools and models to build agents that can reason about the other agents' interactions, detect if they violate the rules and maintain a reputation model of other agents. Since each agent has only a partial view of the system, it cannot control alone the behaviour of all the agents of the system. Therefore, social order will be achieved if several agents using the L.I.A.R. model are deployed in the MAS, the ideal situation being that every agent use it.

The L.I.A.R. reputation model has been designed to allow agents to share easily their point of view in order to improve the efficiency of the social control. The main assumption of the L.I.A.R. model is that the communicative rules are homogeneous and known by every agent. At least, they must be known by the agents participating in the social control. However, agents that are not aware of these rules can still be deployed in the system, but they may be considered as harmful agents if they do not behave as required. In order to allow an agent to evaluate some perceived interactions, the L.I.A.R. model defines a few formalisms and processes described in this section. First, we detail the formalism used to represent an interaction by a social commitment. Then, the representation of social norms is described, as well as their transcription into social policies. The last part of this section explains how all these formalisms can be used in a process to detect the violations of the social norms.

There exists different approaches to enable agents to represent and reason about the interactions of their peers. Two main approaches (Social commitments) to the representation of interactions are the cognitive approach and the social approach. The cognitive approach [55], [56], [57] consists in representing a message by a speech act. The semantics of a speech act is defined subjectively, by referring to the mental states of the sender and receiver of the message. The social approach [58], [59], [60], [61], [62] proposes to represent the occurrence of a message by a social commitment. In this case, there is no reference to agents' mental state. A social commitment represents the fact that a message has been sent and that its sender is publicly committed on the message content.

The L.I.A.R. model uses this social approach to represent interactions. This choice is motivated by the fact that we need a representation formalism that is not intrusive to the internal implementation or mental states of the agents. Interactions should be represented as an external point of view. Moreover, L.I.A.R. only requires that the utterance of the messages are recorded and, in this model, there is no need to reason on the semantics of a message. Thus, we do not make any hypothesis about the language used by the agents to communicate. However, we consider that the agents are able to map speech acts from the language they use into social commitments. Such mappings are proposed by Fornara and Colombetti [60] or Singh [59].

Social norms define the rules that must be respected by the agents during their interactions. Besides, we introduce the concept of social policy to represent the situation of a given agent, about a given social commitment and a given social norm. The use of social policies makes it possible to express norms about the state of social commitments. For instance, we can define a norm

that prohibits social commitments to be in the violated state. But if such a norm does not exist, violated social commitments would be accepted in the system and L.I.A.R. will not consider them as malicious or unauthorised behaviour. This increases the generality of the L.I.A.R. model for the definition of norms. This section presents the L.I.A.R. model for social norms and social policies.

The goal of the reputation model of L.I.A.R. is to provide an estimation over time of the compliance of other agents' behaviour with respect to the social norms. Basically, the reputation model has two 14roles: first, it uses as inputs the results of the L.I.A.R. modules presented in the previous section – social policies – to compute reputations assigned to other agents and, second, it enables agents to reason and make decisions based on these reputations. Based McKnight and Chervany's [63] distinction of trust beliefs, trust intentions and trust behaviours, we define the term "reputation" to refer to an agent's beliefs about the trustworthiness of another agent and "trust" as the act of taking a decision to trust. In short, reputation levels are the beliefs on which an agent makes its decision to trust. In the first subsection, the core concepts of the L.I.A.R. reputation model are defined. Then, the processes related to reputation (initialisation, punishment, reasoning, decision and propagation) are described.

2.18. ReGreT

ReGreT [64] [65] is a modular trust and reputation system oriented to complex small/midsize e-commerce environments where social relations among individuals play an important role. The system takes into account three different sources of information: direct experiences, information from third party agents and social structures. The system maintains three knowledge bases. The outcomes data base (ODB) to store previous contracts and their result; the information data base (IDB), that is used as a container for the information received from other partners and finally the sociograms data base (SDB) to store the graphs (sociograms) that define the agent social view of the world. These data bases feed the different modules of the system. The direct trust module deals with direct experiences and how these experiences can contribute to the trust on third party agents. Together with the reputation model they are the basis to calculate trust. The reputation model is divided in three specialized types of reputation depending on the information source that is used to calculate them:

- Witness reputation. If the reputation is calculated from the information coming from witnesses.
- Neighborhood reputation. If the reputation is calculated using the information extracted from the social relations between partners
- System reputation. It the reputation value is based on roles and general properties.

The system incorporates a credibility module that allows the agent to measure the reliability of witnesses and their information. This module is extensively used in the calculation of witness reputation. All these modules work together to offer a complete trust model based on direct knowledge and reputation. However, the modular approach in the design of the system allows the agent to decide which parts it wants to use. For instance, the agent can decide not to use neighborhood reputation to calculate a reputation value or rely only on direct trust to calculate the trust on an agent without using the reputation module. Another advantage of this modular approach is the adaptability that the system has to different degrees of knowledge. The system is operative even when the agent is a newcomer and it has an important lack of information. As long as the agent increases its knowledge about the other members of the community and its knowledge on the social relations between them, the system starts using other modules to improve the accuracy of the trust and reputation values. This allows the system to be used in a wide range of scenarios, from the most simple to the most complex. If the information is available, the system will use it.

In the ReGreT system, each trust and reputation value has an associated reliability measure. This measure tells the agent how confident the system is on that value according to how it has been calculated. Thanks to this measure, the agent can decide, for example, if it is sensible or not to use the trust and reputation values as part of the decision making mechanism. The last element in the ReGreT system is the ontological structure. The authors consider that trust and reputation are not single and abstract concepts but rather multi-facet concepts. The ontological structure provides the necessary information to combine reputation and trust values linked to simple aspects in order to calculate values associated to more complex attributes. For example, the reputation of being a good flying company summarizes the reputation of having good planes, the reputation of never losing luggage and the reputation of serving good food. In turn, the reputation of having good planes is a summary of the reputation of having a good maintenance service and the reputation of frequently renewing the fleet. Each individual can have a different ontological structure to combine trust and reputation values and a different way to weigh the importance of these values when they are combined [41].

REGRET [66], [67] is a decentralized trust and reputation model designed for complex e-commerce environments where various types of agents with different social relationships play important roles. With the help of a social structure called sociogram, it is able to model the social relationships such as cooperation, competition and

trade in a graph where the nodes represent the participants and the edges denote the nature of their relationship.

This Trust and Reputation system is based on a three-dimensional reputation model:

- individual dimension or subjective reputation which calculates trust based on the direct impressions of an agent received from Service provider (SP) and prioritizes its direct experiences according to their recency;
- social dimension which is designed to estimate the trustworthiness of SP in case the direct experiences are insufficient or the agent has newly joined the environment. This dimension is itself divided into three specialized types of reputation depending on the information sources. First, witness reputation which calculates reputation based on the information coming from the witnesses adjacent to this agent. Here, adjacency is defined as an indication that some form of relationship between two agents exists. Second, neighbourhood reputation that measures the reputation of individuals who are neighbours with the agent being evaluated by considering their social relationships and third, system reputation which assesses the trustworthiness of SP based on the general role that it plays in the sociogram. In order for REGRET to be able to calculate social reputation, it must first identify appropriate witnesses in the e-commerce environment. For doing so, it applies graph theory techniques to the sociogram to locate the most appropriate witnesses and examines their social relationships with the agent being evaluated. Furthermore, by presenting the social relationship in the form of fuzzy rules, REGRET is able to determine the honesty and credibility of the reported observations thus assigning suitable weights to them. For instance, it may declare that (IF the competition relation of witness A with the target agent is very high, THEN its recommended reputation value should be very bad).
- the third reputation dimension of REGRET is the ontological dimension, which adds the possibility of combining different aspects of reputation to calculate a complex one [2].

Note that in the last two dimensions, the agent recorded impressions are linked to single behavioral aspects and do not provide general ratings. However, with the help of the ontological structure, each agent is capable of determining the overall reputation of a particular SP by assigning the appropriate influence degree to each aspect tailored to its demand. In addition to the reputation value, REGRET comes with a reliability measurement which reflects the confidence level of the produced reputation value. Similar to SPORAS [68] and FIRE [3], reliability measurement is calculated from a combination of two factors: the number of available impressions and the variability of the impression values. In order to boost the accuracy of the reliability measure, REGRET defines the intimacy level of interaction which indicates the maximum number of impressions required for a close relationship. As the number of impressions grows, the reliability degree increases until it reaches a certain intimate value. Afterwards, reliability is not affected by the increment of the intimate parameter. It is important to mention that the value of the intimate parameter is dynamically adjustable depending on the interaction frequency of individuals as well as the quality of impressions [10].

The REGRET Trust and Reputation system takes advantage of a variety of information components to predict the trustworthiness of target SPs almost in any situation. Distinctively, in order to make more accurate judgements, it provides the neighbourhood and system reputation components in addition to the direct interaction and witness reputation components. Using social relationships, it enables newcomers to take part in the community's activities; thus provides the possibility for them to increase their knowledge and improve their social status persistently [17]. Moreover, due to the dynamic characteristic of an open environment, the population of participants varies from time to time. Besides, the agent's behavior and performances oscillate, being influenced by unexpected changes in such environments. REGRET is incapable to extensively deal with the dynamicity of an open MAS thus cannot perform effectively under all circumstances of such an environment [2]. As aforementioned, the distinguishing feature of REGRET is its use of social relationships between participants in modeling trust. With the help of the defined social relations, source agents are able to identify suitable witnesses and provide appropriate recommendations with regards to a target agent. Furthermore, REGRET proposes a mechanism to handle the ballot box stuffing and correlated evidence problem where set of witnesses express their opinions based on the same experiences. To do this, it groups the potential witnesses and considers each of them as individual sources of information and then uses a heuristics to select the best representative in the group to send the query to. However, REGRET assumes that each agent owns pre-defined sociograms which display social relationships [77] and does not address how to locate witnesses in these social structures. Subsequently, in order to ascertain the quality of the provided recommendations, any Trust and Reputation systems should develop techniques to detect deceptive and unreliable agents and following that underrate their reputation values or ignore them, accordingly. For this purpose, REGRET mainly relies on social relations and states them via fuzzy rules. Through these rules, it validates the obtained recommendations and determines their influence degree in the reputation aggregation

method. It is noteworthy to mention that REGRET examines the truthfulness of information in general and does not differentiate between dishonest third-parties and incompetent but honest ones.

2.19. PeerTrust

PeerTrust [1], [70], [5] is a coherent dynamic trust model with unique characteristics tailored for peer-to-peer ecommerce communities. For advanced assessment and quantification of peer's trust value in constantly evolving environments, this model customises a variety of common factors:

- feedback which is a judgment of other peers regarding target peer;
- feedback scope such as the amount of transactions the peer experienced with others;
- credibility factor for evaluating the honesty of feedback sources
- transaction context factor such as time and size of transactions which could act as defense mechanism against delicate fraudulent activities; and
- community context factor that addresses the feedback incentive problem.

This model proposes an innovative composite trust metric that incorporates the described parameters to enhance accuracy and reliability of predicted trustworthiness. One of common way for malicious participants to undetectably continue sabotaging in the system is maintaining their general trust value at a certain level by increasing the transaction volume which hides the effect of their frequent frauds. To alleviate the effect of those malicious attacks resulted from increase in transaction volumes it combines the first two parameters such that instead of simply aggregating generic feedback values, it equips witnessed-peers with the ability to disseminate their degree of satisfaction by calculating the average amount of successful outcomes that they experienced. Besides, to ensure the quality of the reputation information, peers are equipped with credibility measures to calculate the credible amount of satisfaction. In doing so, PeerTrust defines the personalised similarity measures [70], [5] which compute feedback similarity rate between the evaluating peer and opinion providers over a common set of peers with whom they have had previous interaction. Since trustworthy peers consistently act honestly as a role of feedback provider and do not become affected by malicious intentions such as jealousy and negative competitive attitude, in addition, this model also advocates that the trust metric can be alternatively served as a credibility measure under certain circumstances. Evidently, one of the significant parameters which is widely neglected in Trust and Reputation systems is transaction context. More explicitly, PeerTrust emphasizes that the aggregation of feedback which are only based on the credibility of their correspondents cannot efficiently reflect the trustworthiness of the agents. Thus, it incorporates various aspects of transaction such as its size, time and category under Transaction Context factors to model participants' intentions and potential fraudulent activities in the trustworthiness measurement. Furthermore, it is widely agreed that feedback are one of the foundations of Trust and Reputation systems such that these systems cannot perform effectively unless they have access to a sufficient amount of feedback [71]. Therefore, to stimulate participants' cooperation, PeerTrust embeds a reward function, called the community context factor, into the trust metric to encourage peers to persistently provide votes about others' performance. The dynamic and distributed nature of peer-to-peer systems necessitates an optimized and adaptive design of the peer location approach. To operationalize this goal, this model provides each peer with a trust manager and a data locator engine which are responsible for feedback submission and retrieval aside from trust evaluation over the underlying network.

2.20. Bayesian Reputation System - BRS

Jøsang et al. [53], [72] have proposed the flexible and adaptive Bayesian Reputation System (BRS) which supports both binomial and multinomial rating models to allow rating provision happen in different levels of precision wellsuited for open dynamic environment. Theoretically, multinomial BRS is based on computing reputation scores by statistically updating the Dirichlet Probability Density Function [73]. More explicitly, in this context, agents are allowed to rate other peers within any level from a set of predefined ratings levels. In contrast, in binomial BRS which is based on Beta distribution, the agents can only provide binary ratings for the others. That is, in multinomial BRS the reputation scores do not solely reflect the general quality of service; but are also able to distinguish between the case of polarized ratings and the case of average ratings [74]. Evidently, such differences are not noticeable in binomial ratings, resulting in uncertainty and low confidence rate in aggregated reputation score and also might prohibit the reputation scores to converge to specific values [75]. Furthermore, multinomial BRS allows the input ratings to be provided based on both discrete and continuous measures to reflect a rater's opinion more accurately when required. To operationalize this goal, it exploits the fuzzy set membership functions to transform continuous ratings into discrete ones in order to provide compatible inputs for BRS [74]. Both systems use the same principle to compute the expected reputation scores, namely by combining previous interaction records with new ratings. Moreover, BRS appears to be promising method to foster trust amongst strangers in an online environment. It takes an

innovative approach which enables trustee agents to evaluate the sincerity of the ratings provided by recommendation agents outside of its control. As such, it uses the endogenous discounting method to exclude such advisers whose probability distribution of ratings significantly deviate from the overall reputation scores of the target agent [52]. That is, it dynamically determines upper and lower bound thresholds in order to adjust the iterated filtering algorithm's sensitivity tailored to different environmental circumstances. For instance, if the majority of participants act deceitfully in the environment, the lower bound would be set to a higher value so as to increase the sensitivity of the BRS which can lead to the exclusion of more unfair raters. Besides, in order to deal with dynamicity in the participant's behavior, BRS provides a longevity factor which determines the expiry time of the old ratings and gives greater weight to more recent ones. As such, it defines a recursive updating algorithm based on the longevity factor to update the participants' reputation scores in certain time intervals. It is noteworthy to mention that, this recursive algorithm also provides a measure to calculate convergence values for the reputation scores [72]. BRS presents a set of rich features which differentiate it from some existing Trust and Reputation systems in certain ways. In particular, it proposes a novel approach to rectify the bootstrapping problem of the newly joined agent. That is, this reputation system dynamically assigns a base rate reputation score to newcomers upon arrival. It provides a method to track the average reputation scores of the whole community so as to settle the newcomers into a conservative state. Notably, such base rate could have be biased towards either positive or negative reputation scores depending on the overall participants' trustworthiness attitudes and the quality of the market at the time. [72], [74].

Furthermore, BRS takes a step towards tackling the inherent dynamicity of an open marketplace. In a dynamic environment, it is impossible to predict all the forthcoming incidents in advance. Thus, any Trust and Reputation system should be equipped with techniques to deal with unanticipated events such as changes in participant populations and attitudes. Moreover, since in open MAS it is quite probable that some information sources would not be temporarily available; conditions should be created for any participants to be able to evaluate the performance of candidate SPs ubiquitously at any time. In addition, in order to operate effectively under any circumstances, Trust and Reputation systems might provide mechanisms to monitor the behavior and relationships of all participants, including the SPs and witnesses, and thus learn and update respective information correspondingly. Unlike other available Trust and Reputation models which mainly concentrate on modeling the adviser's behaviour, BRS models the behavioral pattern of buyer and seller as well. In particular, BRS provides sellers with the ability to adaptively change their behavior to increase their benefits while maintaining a satisfactory level of honesty. For instance, based on a set of heuristics, if a certain seller agent does not succeed in conducting any business for certain period of time, it will automatically decrease the selling price while increasing its level of honesty. On the other hand, it defines the risk attitude parameter for buyer agents which affects the purchasing pattern of the buyers. That is, if the buyer makes a large loss in previous interactions, it intelligently increases the risk-aversion parameter for next rounds of transactions [53].

Furthermore BRS provides a robust protection mechanism against both positive and negative unfair ratings. As such, to diminish the risk of malicious advisers who attempt to manipulate the reputation system for their own benefits, it provides statistical iterated filtering techniques based on beta distribution to dynamically expel such advisers with unsatisfactory rating levels [53].

Finally, by the means of supporting the continuous ratings input, it not only enhances applicability and flexibility of BRS in dealing with the continuous nature of some observations; it also increases the reliability and confidence degree of the provided ratings.

3. DISCUSSION

3.1 Trust and Reputation Models Steps

As we have seen that the main target followed by every trust and reputation model is, to identify those peers who are most reliable supplying a certain service or more trustworthy carrying out a certain task. Selection of these peers differs from one model to other but, for instance, in most of them we can observe more or less the same generic steps [76], as shown in Fig. 1. First of all, an entity checks its previous experiences with a given peer in order to form what is usually called direct trust.

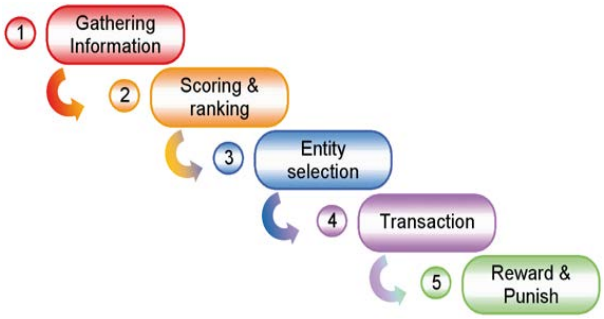

Figure 2 Trust and reputation models steps [77]

This direct trust can be assessed using complex expressions which usually take into account the number of previous transactions, the importance given to each transaction, the satisfaction obtained in each one, the time when it was performed, etc. Additionally the indirect

experiences (or experiences of other peers) are taken into account as well, obtaining what is commonly known as the reputation of a peer.

As we have already seen, trust and reputation management in P2P networks provides several benefits to electronic interactions between users, like a minimum guarantee of benevolent behavior of another interacting peer [78]. Nevertheless, this kind of systems also have several common issues and challenges that need to be addressed when developing such mechanisms. Next we are going to discuss some of them.

3.2. Potential (real) problem scenarios

Here is one example of a real trust and reputation system [7] where most of the concepts explained here are employed. For instance, eBay auction market has a feedback scheme where every buyer and seller rates each other after a transaction between them is carried out. These feedbacks are centrally aggregated in order to get a reputation value for each role. Some studies [78], [79] reveal that buyers provideratings about sellers 51.7% of the time, and sellers provide ratings about buyers 60.6% of the time. Of all ratings provided, less than 1% is negative, less than 0.5% is neutral and about 99% is positive. It was also found that there is a high correlationbetween buyer and seller ratings, suggesting that there is a degree of reciprocationof positive ratings and retaliation of negative ratings. Another distributed system modeling reputation is PageRank [80], the algorithm which the search engine of Google is based on. It represents a way of ranking the search results based on a page's reputation, which is mainly obtained by the number of links pointing to it, since the higher is the number of incoming links, the bettercontent that page is supposed to have. PageRank applies the principle of trust transitivity to the extreme since rank values can flow through looped or arbitrarily long hyperlink chains. Amazon, BizRate or Advogato are other examples of systems where a trust and or reputation scheme is applied in many different environments.

3.3. Conclusive Remarks

This paper presents an overview of Trust and Reputation models and systems that could be base for framework for classifying and comparing Trust and Reputation systems and provided an overview of some prominent current Trust and Reputation systems according to this framework pointing to ways to choose one over another for particular applications. The dimensions of framework could be helpful to system-developers to choose or build their desired Trust and Reputation system with appropriate features according to their requirements or to build some kind of Trust and Reputation tools (or even suggestion for this kind of tool) that can help in determinig Trust and Reputation factor in near future research. Understandably, there is no single solution appropriates for all kinds of tools (or appliacations) and environments. The Paper presents an attempt to provide the means to find the most appropriate path to examine the applicability and usefulness of the current Trust and Reputation systems across different application domains. Paper summarizes the most common features of trust and reputation systems and described how the existing systems support these features. Although there has been a significant number of works in Trust and Reputation systems, there are still some open fields that need further explorations. Specifically, several work has been done on reliability and honesty assessment which proposed innovative solutions in dealing with spurious feedback in uncertain environments. However, some critical aspects of this feature are not fully supported in current trust and reputation systems. To name a few, addressing discrimination detection, Trend & volatility detection in service provider behaviors, ballot box stuffing and distinguishing between malicious and victim participants is not yet addressed thoroughly [81][82][83]. In addition, it remains a challenge to build an informative rating system which supports the context diversity checking feature by providing context and criteria similarity rate to considerably improve the quality of judgements and recommendations.

REFERENCES

[1] S. Ruohomaa, L. Kutvonen, and E. Koutrouli, Reputation management survey, in Proceedings of the 2nd International Conference on Availability, Reliability and Security. Vienna, Austria. IEEE Computer Society, 2007, pp. 103-111.

[2] T. D. Huynh, Trust and Reputation in Open Multi-Agent Systems, PhD thesis, University of Southampton, 2006.

[3] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, An integrated trust and reputation model for open multiagent systems, Journal of Autonomous Agents and Multi-Agent Systems, vol. 13, no. 2, pp. 119-154, 2006.

[4] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, TRAVOS: Trust and reputation in the context of Inaccurate information sources, Autonomous Agents and Multi-Agent Systems, vol. 12, no. 2, pp. 183-198, 2006.

[5] L. Xiong and L. Liu, PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities, IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 7, pp. 843-857, 2004.

[6] B. Yu and M. P. Singh, Detecting deception in reputation management, in Proceedings of the 2nd International Joint Conference on Autonomous Agents & Multiagent Systems, Melbourne, Australia, ACM, 2003, pp. 73-80.

[7] A. Jøsang, R. Ismail, and C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems, vol. 43, no. 2, pp. 618-644, 2007.

[8] E. Chang, T. S. Dillon, and F. Hussain, Trust and reputation for Service Oriented Environments-Technologies for Building Business Intelligence and Consumer Confidence. Location: US: John Wiley & Sons, 2006

[9] D. Artz and Y. Gil, A survey of trust in computer science and the semantic web, Journal of Web Semantics, vol. 5, no. 2, pp. 58-71, 2007

[10] S. Ramchurn, D. Huynh, and N. R. Jennings, Trust in multi-agent systems, The Knowledge Engineering Review, vol. 19, no. 1, pp. 1-25, 2004

[11] Barber, K. S. and J. Kim: 2001, 'Belief Revision Process based on Trust: Simulation Experiments'. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. pp. 1—12.

[12] Montaner, M., B. Lopez, and J. de la Rosa: 2002, 'Developing turst in recommender agents'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 304—305.

[13] Dellarocas, C.: 2003, 'The digitalization of Word-Of-Mouth: Promise and Challenges of Online Reputation Mechanisms'. Management Science.

[14] Grandison, T. and M. Sloman: 2000, 'A survey of trust in Internet application, IEEE, Communications Surveys, Fourth Quarter, 2000'.

[15] Mui, L., A. Halberstadt, and M. Mohtashemi: 2002, 'Notions of Reputation in Multi-Agent Systems: A Review'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 280—287.

[16] McKnight, D. H. and N. L. Chervany: 2002, 'Notions of Reputation in Multi-Agent Systems: A Review'. In: Proceedings of the 34th Hawaii International Conference on System Sciences.

[17] Sabater, J., & Sierra, C. (2005). Review on Computational Trust and Reputation Models.Artificial Intelligence Review, 24(1), 33-60. Kluwer Academic Publishers. Retrieved from http://www.springerlink.com/index/10.1007/s10462-004-0041-5

[18] Gambetta D. (1990). Can we trust Trust? In Gambetta, D., Ed., Trust: Making and Making and Breaking Cooperative Relations. Oxford: Basil Blackwell

[19] Dunn. J. (1990). Trust and Political Agency. In Gambetta, D., Ed., Trust: Making and Making and Breaking Cooperative Relations. Oxford: Basil Blackwell

[20] Marsh, S. (1994). Formalising Trust as a computational concept, Ph.D. Thesis, Department of Computing science and Mathematics, Stirling: University of Stirling

[21] Luhmann, N. (1979). Trust and Power. Chichester: Wiley.

[22] Goldbeck, J. A definition of trust for computing with social networks, Jennifer Golbeck, MINDSWAP, University of Maryland, College Park, November 2011. Available at www.mindswap.org/papers/TrustDef.doc

[23] Abdul-Rahman, A. and S. Hailes: 2000, 'Supporting Trust in Virtual Communities'. In: Proceedings of the Hawaii's International Conference on Systems Sciences, Maui, Hawaii.

[24] Castelfranchi C. & Falcone R. The Dynamics of Trust: from beliefs to action. In the Proceedings of the Autonomous Agents workshop on deception, fraud and trust in agent societies. Seattle, 1999.

[25] Lagenspetz O. (1992). Legitimacy and Trust. Philosophical Investigations, 15:1

[26] Dasgupta, P., (1990). Trust as a Commodity. In Gambetta, D., Ed. Trust: Making and Making and Breaking Cooperative Relations. Oxford: Basil Blackwell

[27] eBay: 2002, 'eBay'. http://www.eBay.com

[28] Amazon: 2002, 'Amazon Auctions'. http://auctions.amazon.com

[29] Pik.ba: 2012, pik. http://www.pik.ba

[30] Zacharia, G.: 1999, 'Collaborative Reputation Mechanisms for Online Communities'. Master's thesis, Massachusetts Institute of Technology.

[31] Glickman, M. E.: 1999, 'Parameter estimation in large dynamic paired comparison experiments'. Applied Statistics (48), 377—394.

[32] Schillo, M., P. Funk, and M. Rovatsos: 2000, 'Using Trust for Detecting Deceitful Agents in Artificial Societites'. Applied Artificial Intelligence (Special Issue on Trust, Deception and Fraud in Agent Societies).

[33] T. D. Huynh, N. R. Jennings, and N. Shadbolt, Developing an integrated trust and reputation model for open multi-agent systems, in Proceedings of the 7th International Workshop on Trust in Agent Societies, New York, 2004, pp. 62–77

[34] Mario Gómez, Javier Carbó, Clara Benac Earle. An Anticipatory Trust Model for Open Distributed Systems. In Proceedings of SAB ABiALS'2006. pp.307-324.

[35] Alfarez Abdul-Rahman. The PGP Trust Model. EDI-Forum, April 1997. Available at http://www.cs.ucl.ac.uk/staff/F.AbdulRahman/docs

[36] M. Clifford, C. Lavine, and M. Bishop, "The Solar Trust Model: Authentication Without Limitation,"Proceedings of the 14th Annual Computer Security Applications Conference" pp. 300–307 (Dec. 1998)

[37] Young, A. Cicovic, N.K. and Chadwick, D. "Trust models in ICE-TEL". Proceedings. 1997 Symposium on Network and Distributed System Security. (pp. 122-133), Feb. 1997

[38] Pearl, J.: 1988, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann

[39] Esfandiari, B. and S. Chandrasekharan: 2001, 'On How Agents Make friends: Mechanisms for Trust Acquisition'. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. pp. 27—34

[40] Lashkari, Y., M. Metral, and P. Maes: 1994, 'Collaborative Interface Agents'. In: Proceedings of the Twelfth National Conference on Artificial Intelligence,AAAIPress

[41] Sabater, J. and C. Sierra: 2001, 'REGRET: A reputation model for gregarious societies'. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. pp. 61—69.

[42] Sabater, J. and C. Sierra: 2002, 'Reputation and Social Network Analysis in Multi-Agent Systems'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 475—482.

[43] Yu, B. and M. P. Singh: 2001, 'Towards a Probabilistic Model of Distributed Reputation Management'. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. pp. 125—137.

[44] Yu, B. and M. P. Singh: 2002a, 'Distributed Reputation Management for Electronic Commerce'. Computational Intelligence 18(4), 535—549.

[45] Yu, B. and M. P. Singh: 2002b, 'An Evidential Model of Distributed Reputation Management'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 294—301.

[46] J. Gordon and E. H. Shortliffe, The Dempster–Shafer theory of evidence, in Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project, in Buchanan and E. H. Shortliffe, eds. Addison Wesley, Reading, MA,1984, pp. 272-292

[47] N. Littlestone and M. K. Warmuth, The weighted majority algorithm, Information and Computation, vol. 108, no. 2, pp. 212-261, 1994.

[48] Sen, S. and N. Sajja: 2002, 'Robustness of Reputation-based Trust: Booblean Case'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 288—293

[49] Carbo, J., J. Molina, and J. Davila: 2002, 'Comparing predictions of SPORAS vs. a Fuzzy Reputation Agent System'. In: 3rd International Conference on Fuzzy Sets and Fuzzy Systems, Interlaken. pp. 147—153

[50] Carter, J., E. Bitting, and A. Ghorbani: 2002, 'Reputation Formalization for an Information-Sharing Multi-Agent Sytem'. Computational Intelligence **18**(2), 515—534.

[51] Castelfranchi, C. and R. Falcone: 1998, 'Principles of Trust for MAS: Cognitive Anatomy, Social

Importance, and Quantification'. In: Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98),Paris,France. pp. 72—79

[52] J. Patel, W. L. Teacy, N. R. Jennings, and M. Luck, A probabilistic trust model for handling inaccurate reputation sources, in Proceedings of 3rd International Conference on Trust Management, Rocquencourt, France, 2005, pp. 193-209.

[53] A. Whitby, A. Jøsang, and J. Indulska, Filtering out unfair ratings in Bayesian reputation systems, in Proceedings of the 7th International Workshop on Trust in Agent Societies, New York, USA, 2004, pp 19-23.

[54] A. Jøsang, and R. Ismail, The beta reputation system, in Proceedings of the 15th Bled Conference on electronic Commerce, Bled, Slovenia. 2002

[55] G. Muller and L. Vercouter, Decentralized Monitoring of Agent Communication with a Reputation Model, Trusting Agents for trusting Electronic Societies, LNCS 3577, 2005.

[56] P. Cohen and H. Levesque. Communicative actions for artificial agents. In Proceedings of the International Conference on Multi Agent Systems (ICMAS'95), pages 65–72, Cambridge, MA, United States of America, 1995. MIT Press.

[57] Y. Labrou and T. Finin. A semantics approach for kqml - a general purpose communication language for software agents. In Proceedings of the Conference on "Information and Knowledge Management" (CIKM'94), pages 447–455, Gaithersburg, MD, United States of America, November 1994. ACM Press, New York, NY, United States of America.

[58] FIPA. FIPA communicative act library specification. Technical Report SC00037J, Fundation For Intelligent Physical Agents (FIPA), December 2002. Standard Status

[59] M. P. Singh. Social and psychological commitments in multi-agent systems. In Proceedings of the AAAI Fall Symposium on Knowledge and Action at Social and Organizational Levels (longer version), pages 104–106, Monterey, CA, United States of America, November 1991

[60] M. P. Singh. A social semantics for agent communication languages. In F. Dignum and M. Greaves, editors, Proceedings of the Workshop on "Agent Communication Languages" at the International Joint Conference on Artificial Intelligence (IJCAI'99), pages 31–45, Heidelberg, Germany, 2000. Springer-Verlag.

[61] N. Fornara and M. Colombetti. Defining interaction protocols using a commitment-based agent communication language. In Proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS'03), pages 520–527, Melbourne, Australia, July 2003. ACM Press, New York, NY, United States of America.

[62] J. Bentahar, B. Moulin, and B. Chaib-Draa. Towards a formal framework for conversational agents. In M.-P. Huget and F. Dignum, editors, Proceedings of the Workshop on "Agent Communication Languages and Conversation Policies" at Autonomous Agents and Multi-Agent Systems (AAMAS'03), Melbourne, Australia, July 2003.

[63] P. Pasquier, R. A. Flores, and B. Chaib-draa. Modelling flexible social commitments and their enforcement. In Proceedings of Engineering Societies in the Agents' World (ESAW'04), volume 3451 of Lecture Notes in Computer Science, pages 139–151, Toulouse, France, October 2004.

[64] D.H. McKnight and N.L. Chervany. Trust and distrust definitions: One bite at a time. In Proceedings of the Workshop on "Deception, Fraud and Trust in Agent Societies" at Autonomous Agents and Multi-Agent Systems (AAMAS'01), volume 2246 of Lecture Notes In Computer Science, pages 27–54, Montreal, Canada, May 2001. Springler-Verlag, London, United Kingdom

[65] Sabater, J. and C. Sierra: 2001, 'REGRET: A reputation model for gregarious societies'. In: Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies, Montreal, Canada. pp. 61—69.

[66] Sabater, J. and C. Sierra: 2002, 'Reputation and Social Network Analysis in Multi-Agent Systems'. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS-02), Bologna, Italy. pp. 475—482.

[67] J. Sabater, Evaluating the regret system, Applied Artificial Intelligence, vol. 18, no. 9-10, pp. 797-813, 2004.

[68] J. Sabater, and C. Sierra, Social ReGret, A reputation model based on social relations. SIGecom Exchanges, vol. 3, no. 1, pp. 44-56, 2002

[69] G. Zacharia and P. Maes, Trust management through reputation mechanisms. Applied Artificial Intelligence, vol. 14, no. 9, pp. 881-908, 2000.

[70] J. Sabater, and C. Sierra, Social ReGret, A reputation model based on social relations. SIGecom Exchanges, vol. 3, no. 1, pp. 44-56, 2002

[71] L. Xiong and L. Liu, A reputation-based trust model for peer-to-peer ecommerce communities, in Proceedings of IEEE Conference on E-Commerce, San Diego,CA,USA, 2003, pp 228-229

[72] R. Jurca and B. Faltings, An incentive compatible reputation mechanism, In Proceedings of the IEEE Conference on E-Commerce, Melbourne, Australia 2003, pp.1026-1027.

[73] A. Jøsang and W. Quattrociocchi, Advanced features in Bayesian reputation systems, in Trust, Privacy and Security in Digital Business, vol. 5695, Heidelberg: Springer, 2009, pp. 105-114.

[74] A. Jøsang and J. Haller, Dirichlet reputation systems, in Proceedings of the International Conference on Availability, Reliability and Security, Vienna, Austria, 2007

[75] A. Jøsang, X. Luo, X. Chen, Continuous ratings in discrete bayesian reputation systems. In Proceedings of the Joint iTrust and PST Conferences on Privacy, Trust Management and Security (IFIPTM 2008), Trondheim, 2008

[76] Y. Wang and M. P. Singh,: Formal trust model for multiagent systems, in Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad,India, 2007, pp. 1551-1556

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

[78] Félix Gómez Mármol, Gregorio Martínez Pérez, "Trust and reputation models comparison", Internet Research, Vol. 21 Issue: 2, pp.138 – 153

[79] Felix Gomez Marmol and Gregorio Martinez Perez, "State of the Art in Trust and Reputation Models in P2P networks," in Handbook of Peer-to-Peer Networking. USA: Springer, 2010, pp. 761-784.

[80] N. Sundaresan, Online trust and reputation systems, in: Proceedings of the 8th ACM conference on Electronic Commerce, 2007, pp. 366–367.

[81] P. Resnick, R. Zeckhauser, Trust among strangers in Internet transactions: Empirical analysisof eBay's reputation system, in: The Economics of the Internet and E-Commerce, Vol. 11,2002, pp. 127–157

[82] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web (1998)

[83] Z. Noorian, M. Ulieru, The State of the Art in Trust and Reputation Systems: A Framework for Comparison, Journal of Theoretical and Applied Electronic Commerce Research ISSN 0718–1876 Electronic Version vol. 5, Issue: 2, August 2010, pp. 97-117