COVER SHEET

A Survey of Trust and Reputation Systems for Online Service Provision

Audun Jøsang [a,1,4] Roslan Ismail [b,2] Colin Boyd [a,3]

[a] *Information Security Research Centre*
Queensland University of Technology
Brisbane, Australia

[b] *College of Information Technology*
Universiti Tenaga Nasional (UNITEN)
Malaysia

Abstract

Trust and reputation systems represent a significant trend in decision support for Internet mediated service provision. The basic idea is to let parties rate each other, for example after the completion of a transaction, and use the aggregated ratings about a given party to derive a trust or reputation score, which can assist other parties in deciding whether or not to transact with that party in the future. A natural side effect is that it also provides an incentive for good behaviour, and therefore tends to have a positive effect on market quality. Reputation systems can be called collaborative sanctioning systems to reflect their collaborative nature, and are related to collaborative filtering systems. Reputation systems are already being used in successful commercial online applications. There is also a rapidly growing literature around trust and reputation systems, but unfortunately this activity is not very coherent. The purpose of this article is to give an overview of existing and proposed systems that can be used to derive measures of trust and reputation for Internet transactions, to analyse the current trends and developments in this area, and to propose a research agenda for trust and reputation systems.

Key words: Trust, reputation, transitivity, collaboration, e-commerce, security, decision

[1] Corresponding author. QUT, Brisbane Qld 4001, Australia. Email: a.josang@qut.edu.au
[2] Email: roslan@uniten.edu.my
[3] Email: c.boyd@qut.edu.au
[4] The work reported in this paper has been funded in part by the Co-operative Research Centre for Enterprise Distributed Systems Technology (DSTC) through the Australian Federal Government's CRC Programme (Department of Education, Science, & Training).

Preprint of article published in Decision Support Systems, 43(2) 2007, p.618-644

1 Introduction

Online service provision commonly takes place between parties who have never transacted with each other before, in an environment where the service consumer often has insufficient information about the service provider, and about the goods and services offered. This forces the consumer to accept the "risk of prior performance", i.e. to pay for services and goods before receiving them, which can leave him in a vulnerable position. The consumer generally has no opportunity to see and try products, i.e. to "squeeze the oranges", before he buys. The service provider, on the other hand, knows exactly what he gets, as long as he is paid in money. The inefficiencies resulting from this information asymmetry can be mitigated through trust and reputation. The idea is that even if the consumer can not try the product or service in advance, he can be confident that it will be what he expects as long as he trusts the seller. A trusted seller therefore has a significant advantage in case the product quality can not be verified in advance.

This example shows that trust plays a crucial role in computer mediated transactions and processes. However, it is often hard to assess the trustworthiness of remote entities, because computerised communication media are increasingly removing us from familiar styles of interaction. Physical encounter and traditional forms of communication allow people to assess a much wider range of cues related to trustworthiness than is currently possible through computer mediated communication. The time and investment it takes to establish a traditional brick-and-mortar street presence provides some assurance that those who do it are serious players. This stands in sharp contrast to the relative simplicity and low cost of establishing a good looking Internet presence which gives little evidence about the solidity of the organisation behind it. The difficulty of collecting evidence about unknown transaction partners makes it hard to distinguish between high and low quality service providers on the Internet. As a result, the topic of trust in open computer networks is receiving considerable attention in the academic community and e-commerce industry.

There is a rapidly growing literature on the theory and applications of trust and reputation systems, and the main purpose of this document is to provide a survey of the developments in this area. An earlier brief survey of reputation systems has been published by Mui *et al.* (2002) [48]. Overviews of agent transaction systems are also relevant because they often relate to reputation systems [24,42,37]. There is considerable confusion around the terminology used to describe these systems, and we will try to describe proposals and developments using a consistent terminology in this study. There also seems to be a lack of coherence in this area, as indicated by the fact that authors often propose new systems from scratch, without trying to extend and enhance previous proposals.

Sec.2 attempts to define the concepts of trust and reputation, and proposes an

agenda for research into trust and reputation systems. Sec.3 describes why trust and reputation systems should be regarded as security mechanisms. Sec.4 describes the relationship between collaborative filtering systems and reputation systems, where the latter can also be defined in terms of collaborative sanctioning systems. In Sec.5 we describe different trust classes, of which *provision trust* is a class of trust that refers to service provision. Sec.6 describes four categories for reputation and trust semantics that can be used in trust and reputation systems, Sec.7 describes centralised and distributed reputation system architectures, and Sec.8 describes some reputation computation methods, i.e. how ratings are to be computed to derive reputation scores. Sec.9 provides and overview of reputation systems in commercial and live applications. Sec.10 describes the main problems in reputation systems, and provides an overview of literature that proposes solutions to these problems. The study is rounded off with a discussion in Sec.11.

2 Background for Trust and Reputation Systems

2.1 The Notion of Trust

Manifestations of trust are easy to recognise because we experience and rely on it every day, but at the same time trust is quite challenging to define because it manifests itself in many different forms. The literature on trust can also be quite confusing because the term is being used with a variety of meanings [46]. Two common definitions of trust which we will call *reliability trust* and *decision trust* respectively will be used in this study.

As the name suggest, reliability trust can be interpreted as the reliability of something or somebody, and the definition by Gambetta (1988) [21] provides an example of how this can be formulated:

Definition 1 (Reliability Trust) *Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends.*

This definition includes the concept of *dependence* on the trusted party, and the *reliability* (probability) of the trusted party, as seen by the trusting party.

However, trust can be more complex than Gambetta's definition indicates. For example, Falcone & Castelfranchi (2001) [18] recognise that having high (reliability) trust in a person in general is not necessarily enough to decide to enter into a situation of dependence on that person. In [18] they write: *"For example it is possible that the value of the damage per se (in case of failure) is too high to choose a given decision branch, and this independently either from the probability of the failure*

(even if it is very low) or from the possible payoff (even if it is very high). In other words, that danger might seem to the agent an intolerable risk." In order to capture this broad concept of trust, the following definition inspired by McKnight & Chervany (1996) [46] can be used.

Definition 2 (Decision Trust) *Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible.*

The relative vagueness of this definition is useful because it makes it the more general. It explicitly and implicitly includes aspects of a broad notion of trust which are *dependence* on the trusted entity or party, the *reliability* of the trusted entity or party, *utility* in the sense that positive utility will result from a positive outcome, and negative utility will result from a negative outcome, and finally a certain *risk attitude* in the sense that the trusting party is willing to accept the situational risk resulting from the previous elements. Risk emerges, for example, when the value at stake in a transaction is high, and the probability of failure is non-negligible (i.e. reliability < 1). Contextual aspects, such law enforcement, insurance and other remedies in case something goes wrong, are only implicitly included in the definition of trust above, but should nevertheless be considered to be part of trust.

There are only a few computational trust models that explicitly take risk into account [22]. Studies that combine risk and trust include Manchala (1998) [44] and Jøsang & Lo Presti (2004) [32]. Manchala explicitly avoids expressing measures of trust directly, and instead develops a model around other elements such as transaction values and the transaction history of the trusted party. Jøsang & Lo Presti distinguish between reliability trust and decision trust, and develops a mathematical model for decision trust based on more finely grained primitives, such as agent reliability, utility values, and the risk attitude of the trusting agent.

The difficulty of capturing the notion of trust in formal models in a meaningful way has led some economists to reject it as a computational concept. The strongest expression for this view has been given by Williamson (1993) [67] who argues that the notion of trust should be avoided when modelling economic interactions, because it adds nothing new, and that well known notions such as reliability, utility and risk are adequate and sufficient for that purpose. According to Williamson, the only type of trust that can be meaningful for describing interactions is personal trust. He argues that personal trust applies to emotional and personal interactions such as love relationships where mutual performance is not always monitored and where failures are forgiven rather than sanctioned. In that sense, traditional computational models would be inadequate e.g. because of insufficient data and inadequate sanctioning, but also because it would be detrimental to the relationships if the involved parties were to take a computational approach. Non-computation models for trust can be meaningful for studying such relationships according to Williamson, but developing such models should be done within the domains of sociology and psychology,

rather than in economy.

2.2 Reputation and Trust

The concept of reputation is closely linked to that of trustworthiness, but it is evident that there is a clear and important difference. For the purpose of this study, we will define reputation according to the Concise Oxford dictionary.

Definition 3 (Reputation) *Reputation is what is generally said or believed about a person's or thing's character or standing.*

This definition corresponds well with the view of social network researchers [19,45] that reputation is a quantity derived from the underlying social network which is globally visible to all members of the network. The difference between trust and reputation can be illustrated by the following perfectly normal and plausible statements:

(1) *"I trust you because of your good reputation."*
(2) *"I trust you despite your bad reputation."*

Assuming that the two sentences relate to identical transactions, statement (1) reflects that the relying party is aware of the trustee's reputation, and bases his trust on that. Statement (2) reflects that the relying party has some private knowledge about the trustee, e.g. through direct experience or intimate relationship, and that these factors overrule any reputation that a person might have. This observation reflects that trust ultimately is a personal and subjective phenomenon that that is based on various factors or evidence, and that some of those carry more weight than others. Personal experience typically carries more weight than second hand trust referrals or reputation, but in the absence of personal experience, trust often has to be based on referrals from others.

Reputation can be considered as a collective measure of trustworthiness (in the sense of reliability) based on the referrals or ratings from members in a community. An individual's subjective trust can be derived from a combination of received referrals and personal experience. In order to avoid dependence and loops it is required that referrals be based on first hand experience only, and not on other referrals. As a consequence, an individual should only give subjective trust referral when it is based on first hand evidence or when second hand input has been removed from its derivation base [33]. It is possible to abandon this principle for example when the weight of the trust referral is normalised or divided by the total number of referrals given by a single entity, and the latter principle is applied in Google's PageRank algorithm [52] described in more detail in Sec.9.5 below.

Reputation can relate to a group or to an individual. A group's reputation can for example be modelled as the average of all its members' individual reputations, or as the average of how the group is perceived as a whole by external parties. Tadelis' (2001) [66] study shows that an individual belonging to to a given group will inherit an *a priori* reputation based on that group's reputation. If the group is reputable all its individual members will *a priori* be perceived as reputable and vice versa.

2.3 A Research Agenda for Trust and Reputation Systems

There are two fundamental differences between traditional and online environments regarding how trust and reputation are, and can be, used.

Firstly, as already mentioned, the traditional cues of trust and reputation that we are used to observe and depend on in the physical world are missing in online environments, so that electronic substitutes are needed. Secondly, communicating and sharing information related to trust and reputation is relatively difficult, and normally constrained to local communities in the physical world, whereas IT systems combined with the Internet can be leveraged to design extremely efficient systems for exchanging and collecting such information on a global scale.

Motivated by this basic observation, the purposes of research in trust and reputation systems should be to:

a. Find adequate online substitutes for the traditional cues to trust and reputation that we are used to in the physical world, and identify new information elements (specific to a particular online application) which are suitable for deriving measures of trust and reputation.
b. Take advantage of IT and the Internet to create efficient systems for collecting that information, and for deriving measures of trust and reputation, in order to support decision making and to improve the quality of online markets.

These simple principles invite rigorous research in order to answer some fundamental questions: What information elements are most suitable for deriving measures of trust and reputation in a given application? How can these information elements be captured and collected? What are the best principles for designing such systems from a theoretic and from a usability point of view? Can they be made resistant to attacks of manipulation by strategic agents? How should users include the information provided by such systems into their decision process? What role can these systems play in the business model of commercial companies? Do these systems truly improve the quality of online trade and interactions? These are important questions that need good answers in order to determine the potential for trust and reputation systems in online environments.

According to Resnick *et al.* [56], reputation systems must have the following three

properties to operate at all:

1. Entities must be long lived, so that with every interaction there is always an expectation of future interactions.
2. Ratings about current interactions are captured and distributed.
3. Ratings about past interactions must guide decisions about current interactions.

The longevity of agents means, for example, that it should be impossible or difficult for an agent to change identity or pseudonym for the purpose of erasing the connection to its past behaviour. The second property depends on the protocol with which ratings are provided, and this is usually not a problem for centralised systems, but represents a major challenge for distributed systems. The second property also depends on the willingness of participants to provide ratings, for which there must be some form of incentive. The third property depends on the usability of reputation system, and how people and systems respond to it, and this is reflected in the commercial and live reputation systems described in Sec.9, but only to a small extent in the theoretic proposals described in Sec.8 and Sec.10.

The basic principles of reputation systems are relatively easy to describe (see Sec.7 and Sec.8). However, because the notion of trust itself is vague, what constitutes a *trust system* is difficult to describe concisely. A method for deriving trust from a transitive trust path is an element which is normally found in trust systems. The idea behind trust transitivity is that when Alice trusts Bob, and Bob trusts Claire, and Bob refers Claire to Alice, then Alice can derive a measure of trust in Claire based on Bob's referral combined with her trust in Bob. This is illustrated in fig.1 below.

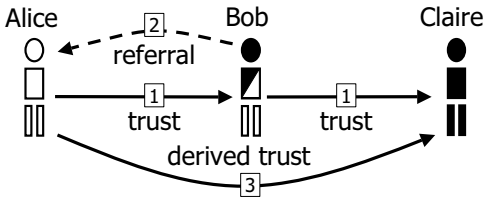


Fig. 1. Trust transitivity principle

The type of trust considered in this example is obviously reliability trust (and not decision trust). In addition there are semantic constraints for the transitive trust derivation to be valid, i.e. that Alice must trust Bob to recommend Claire for a particular purpose, and that Bob must trust Claire for that same purpose [33].

The main differences between trust and reputation systems can be described as follows: Trust systems produce a score that reflects the relying party's subjective view of an entity's trustworthiness, whereas reputation systems produce an entity's (public) reputation score as seen by the whole community. Secondly, transitivity is an explicit component in trust systems, whereas reputation systems usually only take transitivity implicitly into account. Finally, trust systems usually take subjective

and general measures of (reliability) trust as input, whereas information or ratings about specific (and objective) events, such as transactions, are used as input in reputation systems.

There can of course be trust systems that incorporate elements of reputation systems and vice versa, so that it is not always clear how a given systems should be classified. The descriptions of the various trust and reputation systems below must therefor be seen in light of this.

3 Security and Trust

3.1 Trust and Reputation Systems as Soft Security Mechanisms

In a general sense, the purpose of security mechanisms is to provide protection against malicious parties. In this sense there is a whole range of security challenges that are not met by traditional approaches. Traditional security mechanisms will typically protect resources from malicious users, by restricting access to only authorised users. However, in many situations we have to protect ourselves from those who offer resources so that the problem in fact is reversed. Information providers can for example act deceitfully by providing false or misleading information, and traditional security mechanisms are unable to protect against this type of threat. Trust and reputation systems on the other hand can provide protection against such threats. The difference between these two approaches to security was first described by Rasmussen & Jansson (1996) [53] who used the term *hard security* for traditional mechanisms like authentication and access control, and *soft security* for what they called social control mechanisms in general, of which trust and reputation systems are examples.

3.2 Computer Security and Trust

Security mechanisms protect systems and data from being adversely affected by malicious and non-authorised parties. The effect of this is that those systems and data can be considered more reliable, and thus more trustworthy. The concepts of Trusted Systems and Trusted Computing Base have been used in the IT security jargon (see e.g. Abrams 1995 [3]), but the concept of security assurance level is more standardised as a measure of security [5] . The assurance level can be interpreted as a system's strength to resist malicious attacks, and some organisations require

[5] See e.g. the UK CESG at http://www.cesg.gov.uk/ or the Common Criteria Project at http://www.commoncriteriaportal.org/

systems with high assurance levels for high risk or highly sensitive applications. In an informal sense, the assurance level expresses a level of public (reliability) trustworthiness of given system. However, it is evident that additional information, such as warnings about newly discovered security flaws, can carry more weight than the assurance level when people form their own subjective trust in the system.

3.3 Communication Security and Trust

Communication security includes encryption of the communication channel and cryptographic authentication of identities. Authentication provides so-called *identity trust*, i.e. a measure of the correctness of a claimed identity over a communication channel. The term *"trust provider"* is sometimes used in the industry to describe CAs [6] and other authentication service providers with the role of providing the necessary mechanisms and services for verifying and managing identities. The type of trust that CAs and identity management systems provide is simply identity trust. In case of chained identity certificates, the derivation of identity trust is based on trust transitivity, so in that sense these systems can be called identity trust systems.

However, users are also interested in knowing the reliability of authenticated parties, or the quality of goods and services they provide. This latter type of trust will be called *provision trust* in this study, and only trust and reputation systems (i.e. *soft* security mechanisms) are useful tools for deriving provision trust.

It can be observed that identity trust is a condition for trusting a party behind the identity with anything more than a baseline or default provision trust that applies to all parties in a community. This does not mean that the real world identity of the principal must be known. An anonymous party, who can be recognised from interaction to interaction, can also be trusted for the purpose of providing services.

4 Collaborative filtering and Collaborative Sanctioning

Collaborative filtering systems (CF) have similarities with reputation systems in that both collect ratings from members in a community. However they also have fundamental differences. The assumptions behind CF systems is that different people have different tastes, and rate things differently according to subjective taste. If two users rate a set of items similarly, they share similar tastes, and are called *neighbours* in the jargon. This information can be used to recommend items that one participant likes, to his or her neighbours, and implementations of this technique

[6] Certification Authority

are often called *recommender systems*. This must not be confused with reputation systems which are based on the seemingly opposite assumption, namely that all members in a community should judge the performance of a transaction partner or the quality of a product or service consistently. In this sense the term *"collaborative sanctioning"* (CS) [49] has been used to describe reputation systems, because the purpose is to sanction poor service providers, with the aim of giving an incentive for them to provide quality services.

CF takes ratings subject to taste as input, whereas reputation systems take ratings assumed insensitive to taste as input. People will for example judge data files containing film and music differently depending on their taste, but all users will judge files containing viruses to be bad. CF systems can be used to select the preferred files in the former case, and reputation systems can be used to avoid the bad files in the latter case. There will of course be cases where CF systems identify items that are invariant to taste, which simply indicates low usefulness of that result for recommendation purposes. Inversely, there will be cases where ratings that are subject to personal taste are being fed into reputation systems. The latter can cause problems, because most reputation systems will be unable to distinguish between variations in service provider performance, and variations in the observer's taste, potentially leading to unreliable and misleading reputation scores.

Another important point is that CF systems and reputation systems assume an optimistic and a pessimistic world view respectively. To be specific CF systems assume all participants to be trustworthy and sincere, i.e. to their job as best they can and to always report their genuine opinion. Reputation systems, on the other hand, assume that some participants will try to misrepresent the quality of services in order to make more profit, and to lie or provide misleading ratings in order to achieve some specific goal. It can be very useful to combine CF and reputation systems, and Amazon.com described in Sec.9.3.3 does this to a certain extent. Theoretic schemes include Damiani *et al.*'s (2002) proposal to separate between provider reputation and resource reputation in P2P networks [13].

5 Trust Classes

In order to be more specific about trust semantics, we will distinguish between a set of different trust classes according to Grandison & Sloman's classification (2000) [22]. This is illustrated in fig.2 below [7] . The highlighting of provision trust in fig.2 is done to illustrate that it is the focus of the trust and reputation systems described in this study.

[7] Grandison & Sloman use the terms *service provision trust*, *resource access trust*, *delegation trust*, *certification trust*, and *infrastructure trust*.

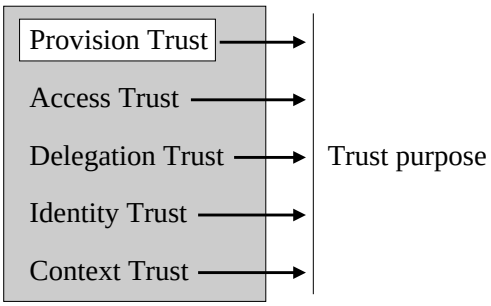


Fig. 2. Trust Classes (Grandison & Sloman 2000)

- **Provision trust** describes the relying party's trust in a service or resource provider. It is relevant when the relying party is a user seeking protection from malicious or unreliable service providers. The Liberty Alliance Project [8] uses the term "business trust" [40] to describes mutual trust between companies emerging from contract agreements that regulate interactions between them, and this can be interpreted as provision trust. For example when a contract specifies quality requirements for the delivery of services, then this business trust would be provision trust in our terminology.
- **Access trust** describes trust in principals for the purpose of accessing resources owned by or under the responsibility of the relying party. This relates to the access control paradigm which is a central element in computer security. A good overview of access trust systems can be found in Grandison & Sloman (2000) [22].
- **Delegation trust** describes trust in an agent (the delegate) that acts and makes decision on behalf of the relying party. Grandison & Sloman point out that acting on one's behalf can be considered to a special form of service provision.
- **Identity trust** [9] describes the belief that an agent identity is as claimed. Trust systems that derive identity trust are typically authentication schemes such as X.509 and PGP [74]. Identity trust systems have been discussed mostly in the information security community, and a brief overview and analysis can be found in Reiter & Stubblebine (1997) [54].
- **Context trust** [10] describes the extent to which the relying party believes that the necessary systems and institutions are in place in order to support the transaction and provide a safety net in case something should go wrong. Factors for this type of trust can for example be critical infrastructures, insurance, legal system, law enforcement and stability of society in general.

Trust purpose is an overarching concept that that can be used to express any operational instantiation of the trust classes mentioned above. In other words, it defines the specific scope of a give trust relationship. A particular trust purpose can for example be *"to be a good car mechanic"*, which can be grouped under the provision

[8] http://www.projectliberty.org/
[9] Called "authentication trust" in Liberty Alliance (2003) [40]
[10] Called "system trust" in McKnight & Chervany (1996) [46]

trust class.

Conceptually, identity trust and provision trust can be seen as two layers on top of each other, where provision trust normally can not exist without identity trust. In the absence of identity trust, it is only possible to have a baseline provision trust in an agent or entity.

6 Categories of Trust Semantics

The semantic characteristics of ratings, reputation scores and trust measures are important in order for participants to be able to interpret those measures. The semantics of measures can be described in terms of a *specificity-generality* dimension and a *subjectivity-objectivity* dimension as illustrated in Table 1 below.

A specific measure means that it relates to a specific trust aspect such as the ability to deliver on time, whereas a general measure is supposed to represent an average of all aspects.

A subjective measure means that an agent provides a rating based on subjective judgement whereas an objective measure means that the rating has been determined by objectively assessing the trusted party against formal criteria.

Table 1
Classification of trust and reputation measures.

	Specific, vector based	**General, synthesised**
Subjective	Survey questionnaires	eBay, voting
Objective	Product tests	Synthesised general score from product tests, D&B rating

- **Subjective and specific** measures are for example used in survey questionnaires where people are asked to express their opinion over a range of specific issues. A typical question could for example be: *"How do you see election candidate X's ability to handle the economy?"* and the possible answers could be on a scale 1 - 5 which could be assumed equivalent to *"Disastrous"*, *"Bad"*, *"Average"*, *"Good"* and *"Excellent"*. Similar questions could be applied to foreign policy, national security, education and health, so that a person's answer forms a subjective vector of his or her trust in candidate X.
- **Subjective and general** measures are for example used on eBay's reputation system which is described in detail in Sec.9.1. An inherent problem with this type of measure is that it often fails to assign credit or blame to the right aspect or even the right party. For example, if a shipment of an item bought on eBay arrives late or is broken, the buyer might give the seller a negative rating, whereas the post office might have caused the problem.

A general problem with all subjective measures is that it is difficult to protect against unfair ratings. Another potential problem is that the act of referring negative general and subjective trust in an entity can lead to accusations of slander. This is not so much a problem in reputation systems because the act of rating a particular transaction negatively is less sensitive than it is to refer negative trust in an entity in general.

- **Objective and specific** measures are for example used in technical product tests where the performance or the quality of the product can be objectively measured. Washing machines can for example be tested according to energy consumption, noise, washing program features etc. Another example is to rate the fitness of commercial companies based on specific financial measures, such as earning, profit, investment, R&D expenditure etc. An advantage with objective measures is that the correctness of ratings can be verified by others, or automatically generated based on automated monitoring of events.
- **Objective and general** measures can for example be computed based on a vector of objective and specific measures. In product tests, where a range of specific characteristics are tested, it is common to derive a general score which can be a weighted average of the score of each characteristic. Dunn & Bradstreet's business credit rating is an example of a measure that is derived from a vector of objectively measurable company performance parameters.

7 Reputation Network Architectures

The technical principles for building reputation systems are described in this and the following section. The network architecture determines how ratings and reputation scores are communicated between participants in a reputation systems. The two main types are centralised and distributed architectures.

7.1 Centralised Reputation Systems

In centralised reputation systems, information about the performance of a given participant is collected as ratings from other members in the community who have had direct experience with that participant. The central authority (reputation centre) that collects all the ratings typically derives a reputation score for every participant, and makes all scores publicly available. Participants can then use each other's scores, for example, when deciding whether or not to transact with a particular party. The idea is that transactions with reputable participants are likely to result in more favourable outcomes than transactions with disreputable participants.

Fig.3 below shows a typical centralised reputation framework, where A and B denote transaction partners with a history of transactions in the past, and who consider

transacting with each other in the present.

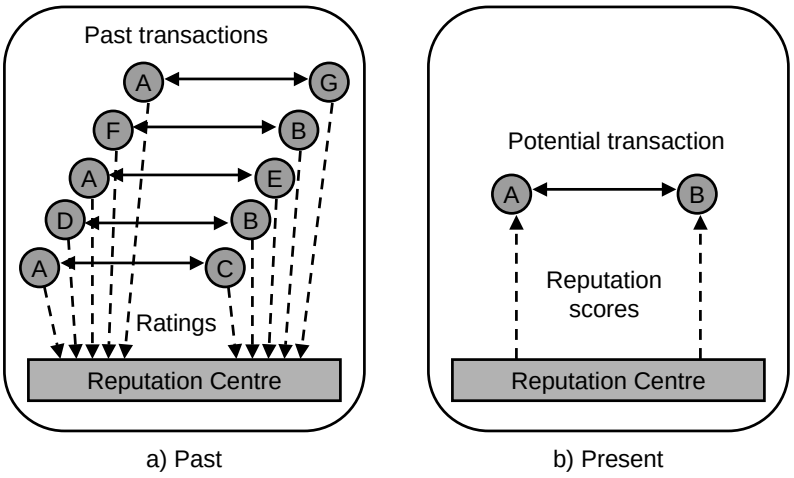


Fig. 3. General framework for a centralised reputation system

After each transaction, the agents provide ratings about each other's performance in the transaction. The reputation centre collects ratings from all the agents, and continuously updates each agent's reputation score as a function of the received ratings. Updated reputation scores are provided online for all the agents to see, and can be used by the agents to decide whether or not to transact with a particular agent.

The two fundamental aspects of centralised reputation systems are:

1. *Centralised communication protocols* that allow participants to provide ratings about transaction partners to the central authority, as well as to obtain reputation scores of potential transaction partners from the central authority.
2. *A reputation computation engine* used by the central authority to derive reputation scores for each participant, based on received ratings, and possibly also on other information. This is described in Sec.8 below.

7.2 Distributed Reputation Systems

There are environments where a distributed reputation system, i.e. without any centralised functions, is better suited than a centralised system. In a distributed system there is no central location for submitting ratings or obtaining reputation scores of others. Instead, there can be distributed stores where ratings can be submitted, or each participant simply records the opinion about each experience with other parties, and provides this information on request from relying parties. A relying party, who considers transacting with a given target party, must find the distributed stores, or try to obtain ratings from as many community members as possible who have had direct experience with that target party. This is illustrated in fig.4 below.

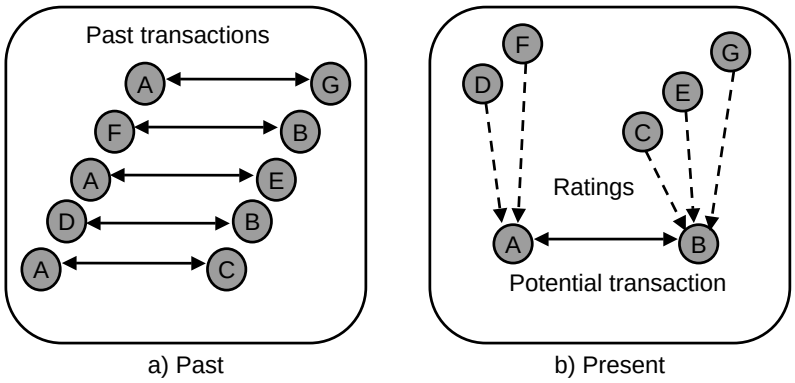


Fig. 4. General framework for a distributed reputation system

The relying party computes the reputation score based on the received ratings. In case the relying party has had direct experience with the target party, the experience from that encounter can be taken into account as private information, possibly carrying a higher weight than the received ratings.

The two fundamental aspects of distributed reputation systems are:

1. *A distributed communication protocol* that allows participants to obtain ratings from other members in the community.
2. *A reputation computation method* used by each individual agent to derive reputation scores of target parties based on received ratings, and possibly on other information. This is described in Sec.8 below.

Peer-to-Peer (P2P) networks represent a environment well suited for distributed reputation management. In P2P networks, every node plays the role of both client and server, and is therefore sometimes called a *servent*. This allows the users to overcome their passive role typical of web navigation, and to engage in an active role by providing their own resources. There are two phases in the use of P2P networks. The first is the *search* phase, which consists of locating the servent where the requested resource resides. In some P2P networks, the search phase can rely on centralised functions. One such example is Napster [11] which has a resource directory server. In pure P2P networks like Gnutella [12] and Freenet [13], also the search phase is distributed. Intermediate architectures also exist, e.g. the FastTrack architecture which is used in P2P networks like KaZaA [14], grokster [15] and iMesh [16]. In FastTrack based P2P networks, there are nodes and supernodes, where the latter keep tracks of other nodes and supernodes that are logged onto the network, and thus act as directory servers during the search phase.

[11] http://www.napster.com/
[12] http://www.gnutella.com
[13] http://www.zeropaid.com/freenet
[14] http://www.kazaa.com
[15] http://www.grokster.com/
[16] http://imesh.com

After the search phase, where the requested resource has been located, comes the *download phase*, which consists of transferring the resource from the exporting to the requesting servent.

P2P networks introduce a range of security threats, as they can be used to spread malicious software, such as viruses and Trojan horses, and easily bypass firewalls. There is also evidence that P2P networks suffer from free riding [4]. Reputation systems are well suited to fight these problems, e.g. by sharing information about rogue, unreliable or selfish participants. P2P networks are controversial because they have been used to distribute copyrighted material such as MP3 music files, and it has been claimed that content poisoning [17] has been used by the music industry to fight this problem. We do not defend using P2P networks for illegal file sharing, but it is obvious that reputation systems could be used by distributors of illegal copyrighted material to protect themselves from poisoning.

Many authors have proposed reputation systems for P2P networks [2,12,13,17,23,35,39]. The purpose of a reputation system in P2P networks is:

1. To determine which servents are most reliable at offering the best quality resources, and
2. To determine which servents provide the most reliable information with regard to (1).

In a distributed environment, each participant is responsible for collecting and combining ratings from other participants. Because of the distributed environment, it is often impossible or too costly to obtain ratings resulting from all interactions with a given agent. Instead the reputation score is based on a subset of ratings, usually from the relying party's "neighbourhood".

8 Reputation Computation Engines

Seen from the relying party's point of view, trust and reputation scores can be computed based on own experience, on second hand referrals, or on a combination of both. In the jargon of economic theory, the term *private information* is used to describe first hand information resulting from own experience, and *public information* is used to describe publicly available second hand information, i.e. information that can be obtained from third parties.

Reputation systems are typically based on public information in order to reflect the community's opinion in general, which is in line with Def.3 of reputation. A party

[17] Poisoning music file sharing networks consists of distributing files with legitimate titles - and put inside them silence or random noise.

who relies on the reputation score of some remote party, is in fact trusting that party through *trust transitivity* [33].

Some systems take both public and private information as input. Private information, e.g. resulting from personal experience, is normally considered more reliable than public information, such as ratings from third parties.

This section describes various principles for computing reputation and trust measures. Some of the principles are used in commercial applications, whereas others have been proposed by the academic community.

8.1 Simple Summation or Average of Ratings

The simplest form of computing reputation scores is simply to sum the number of positive ratings and negative ratings separately, and to keep a total score as the positive score minus the negative score. This is the principle used in eBay's reputation forum which is described in detail in [55]. The advantage is that anyone can understand the principle behind the reputation score, the disadvantage that it is primitive and therefore gives a poor picture participants' reputation score although this is also due to the way rating is provided, see Sec.10.1 and Sec.10.2.

A slightly more advanced scheme proposed in e.g. [63] is to compute the reputation score as the average of all ratings, and this principle is used in the reputation systems of numerous commercial web sites, such as Epinions, and Amazon described in Sec.9.

Advanced models in this category compute a weighted average of all the ratings, where the rating weight can be determined by factors such as rater trustworthiness/reputation, age of the rating, distance between rating and current score etc.

8.2 Bayesian Systems

Bayesian systems take binary ratings as input (i.e. positive or negative), and are based on computing reputation scores by statistical updating of beta probability density functions (PDF). The *a posteriori* (i.e. the updated) reputation score is computed by combining the *a priori* (i.e. previous) reputation score with the new rating [31,49–51,68]. The reputation score can be represented in the form of the beta PDF parameter tuple $(\alpha,\ \beta)$ (where α and β represent the amount of positive and negative ratings respectively), or in the form of the probability expectation value of the beta PDF, and optionally accompanied with the variance or a confidence parameter. The advantage of Bayesian systems is that they provide a theoretically sound basis for computing reputation scores, and the only disadvantage that it might be

too complex for average persons to understand.

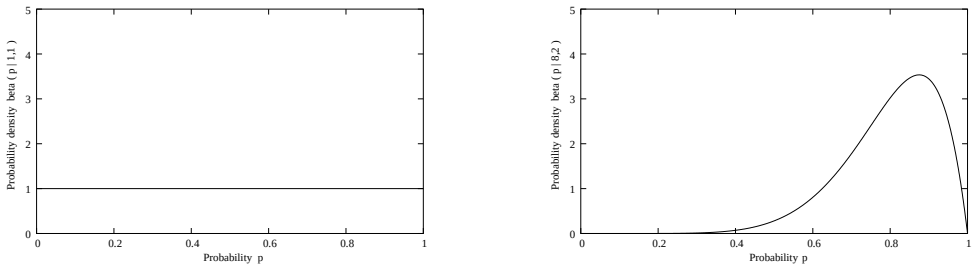
The beta-family of distributions is a continuous family of distribution functions indexed by the two parameters α and β. The beta PDF denoted by beta$(p \mid \alpha, \beta)$ can be expressed using the gamma function Γ as:

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

with the restriction that the probability variable $p \neq 0$ if $\alpha < 1$, and $p \neq 1$ if $\beta < 1$. The probability expectation value of the beta distribution is given by:

$$\mathrm{E}(p) = \alpha/(\alpha+\beta). \qquad (2)$$

When nothing is known, the *a priori* distribution is the uniform beta PDF with $\alpha = 1$ and $\beta = 1$ illustrated in fig.5.a. Then, after observing r positive and s negative outcomes, the *a posteriori* distribution is the beta PDF with $\alpha = r + 1$ and $\beta = s + 1$. For example, the beta PDF after observing 7 positive and 1 negative outcomes is illustrated in fig.5.b.



(a) Uniform PDF beta$(p \mid 1, 1)$ (b) PDF beta$(p \mid 8, 2)$

Fig. 5. Example beta probability density functions

A PDF of this type expresses the uncertain probability that future interactions will be positive. The most natural is to define the reputation score as a function of the expectation value. The probability expectation value of fig.5.b according to Eq.(2) is $\mathrm{E}(p) = 0.8$. This can be interpreted as saying that the relative frequency of a positive outcome in the future is somewhat uncertain, and that the most likely value is 0.8.

8.3 Discrete Trust Models

Humans are often better able to rate performance in the form of discrete verbal statements, than continuous measures. This is also valid for determining trust measures, and some authors, including [1,8,9,44], have proposed discrete trust models.

For example, in the model of Abdul-Rahman & Hailes (2000) [1] trustworthiness of an agent x can be referred as *Very Trustworthy*, *Trustworthy*, *Untrustworthy* and *Very Untrustworthy*. The relying party can then apply his or her own perception about the trustworthiness of the referring agent before taking the referral into account. Look-up tables, with entries for referred trust and referring party downgrade/upgrade, are used to determine derived trust in x. Whenever the relying party has had personal experience with x, this can be used to determine referring party trustworthiness. The assumption is that personal experience reflects x's real trustworthiness and that referrals about x that differ from the personal experience will indicate whether the referring party underrates or overrates. Referrals from a referring party who is found to overrate will be downgraded, and vice versa.

The disadvantage of discrete measures is that they do not easily lend themselves to sound computational principles. Instead, heuristic mechanisms like look-up tables must be used.

8.4 Belief Models

Belief theory is a framework related to probability theory, but where the sum of probabilities over all possible outcomes not necessarily add up to 1, and the remaining probability is interpreted as uncertainty.

Jøsang (1999,2001) [28,29] has proposed a belief/trust metric called *opinion* denoted by $\omega_x^A = (b, d, u, a)$, which expresses the relying party A's belief in the truth of statement x. Here b, d, and u represent belief, disbelief and uncertainty respectively where $b, d, u \in [0, 1]$ and $b + d + u = 1$. The parameter $a \in [0, 1]$, which is called the relative atomicity, represents the base rate probability in the absence of evidence, and is used for computing an opinion's probability expectation value $E(\omega_x^A) = b + au$, meaning that a determines how uncertainty shall contribute to $E(\omega_x^A)$. When the statement x for example says *"David is honest and reliable"*, then the opinion can be interpreted as reliability trust in David. As an example, let us assume that Alice needs to get her car serviced, and that she asks Bob to recommend a good car mechanic. When Bob recommends David, Alice would like to get a second opinion, so she asks Claire for her opinion about David. This situation is illustrated in fig. 6 below.

When trust and trust referrals are expressed as opinions, each transitive trust path Alice→Bob→David, and Alice→Claire→David can be computed with the *discounting operator*, where the idea is that the referrals from Bob and Claire are discounted as a function Alice's trust in Bob and Claire respectively. Finally the two paths can be combined using the *consensus operator*. These two operators form part of *Subjective Logic* [29], and semantic constraints must be satisfied in order for the transitive trust derivation to be meaningful [33]. Opinions can be uniquely

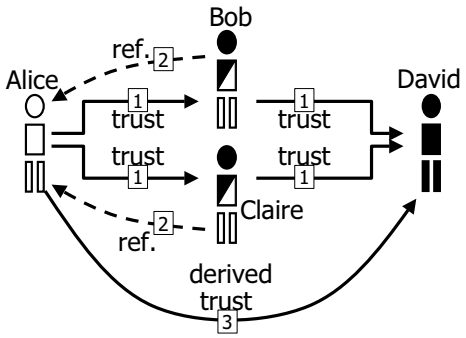


Fig. 6. Deriving trust from parallel transitive chains

mapped to beta PDFs, and this sense the consensus operator is equivalent to the Bayesian updating described in Sec.8.2. This model is thus both belief-based and Bayesian.

Yu & Singh (2002) [70] have proposed to use belief theory to represent reputation scores. In their scheme, two possible outcomes are assumed, namely that an agent A is trustworthy (T_A) or not trustworthy ($\neg\, T_A$), and separate beliefs are being kept about whether A is trustworthy or not, denoted by $m(T_A)$ and $m(\neg\, T_A)$ respectively. The reputation score Γ of an agent A is then defined as:

$$\Gamma(A) = m(T_A) - m(\neg\, T_A)\ , \qquad \begin{array}{l} \text{where } m(T_A), m(\neg\, T_A) \in [0,1] \\ \text{and } \Gamma(A) \in [-1,1]\ . \end{array} \tag{3}$$

Without going into detail, the ratings provided by individual agents are belief measures determined as a function of A's past history of behaviours with individual agents as trustworthy or not trustworthy, using predefined threshold values for what constitutes trustworthy and untrustworthy behaviour. These belief measures are then combined using Dempster's rule [18], and the resulting beliefs are fed into Eq.3 to compute the reputation score. Ratings are considered valid if they result from a transitive trust chain of length less or equal to a predefined limit.

8.5 Fuzzy Models

Trust and reputation can be represented as linguistically fuzzy concepts, where membership functions describe to what degree an agent can be described as e.g. trustworthy or not trustworthy. Fuzzy logic provides rules for reasoning with fuzzy measures of this type. The scheme proposed by Manchala (1988) [44] described in Sec.2 as well as the REGRET reputation system proposed by Sabater & Sierra (2001,2002) [59–61] fall in this category. In Sabater & Sierra's scheme, what they call *individual reputation* is derived from private information about a given agent,

[18] Dempster's rule is the classical operator for combining evidence from different sources.

what they call *social reputation* is derived from public information about an agent, and what they call *context dependent reputation* is derived from contextual information.

8.6 Flow Models

Systems that compute trust or reputation by transitive iteration through looped or arbitrarily long chains can be called flow models.

Some flow models assume a constant trust/reputation weight for the whole community, and this weight can be distributed among the members of the community. Participants can only increase their trust/reputation at the cost of others. Google's PageRank [52] described in Sec.9.5, the Appleseed algorithm [73] and Advogato's reputation scheme [38] described in Sec.9.2 belong to this category. In general, a participant's reputation increases as a function of incoming flow, and decreases as a function of outgoing flow. In the case of Google, many hyperlinks to a web page contributes to increased PageRank whereas many hyperlinks from a web page contributes to decreased PageRank for that web page.

Flow models do not always require the sum of the reputation/trust scores to be constant. One such example is the EigenTrust model [35] which computes agent trust scores in P2P networks through repeated and iterative multiplication and aggregation of trust scores along transitive chains until the trust scores for all agent members of the P2P community converge to stable values.

9 Commercial and Live Reputation Systems

This section describes the most well known applications of online reputation systems. All analysed systems have a centralised network architecture. The computation is mostly based on the summation or average of ratings, but two systems use the flow model.

9.1 eBay's Feedback Forum

eBay [19] is a popular auction site that allows sellers to list items for sale, and buyers to bid for those items. The so-called Feedback Forum on eBay gives buyer and seller the opportunity to rate each other (provide *feedback* in the eBay jargon) as positive, negative, or neutral (i.e. 1, -1, 0) after completion of a transaction. Buyers

[19] http://ebay.com/

and sellers also have the possibility to leave comments like *"Smooth transaction, thank you!"* which are typical in positive case or *"Buyers beware!"* in the rare negative case. The Feedback Forum is a centralised reputation system, where eBay collects all the ratings and computes the scores. The running total reputation score of each participant is the sum of positive ratings (from unique users) minus the sum of negative ratings (from unique users). In order to provide information about a participant's more recent behaviour, the total of positive, negative and neutral ratings for the three different time windows i) past six months, ii) past month, and iii) past 7 days are also displayed.

There are many empirical studies of eBay's reputation system, see Resnick *et al.* (2002) [57] for an overview. In general the observed ratings on eBay are surprisingly positive. Buyers provide ratings about sellers 51.7% of the time, and sellers provide ratings about buyers 60.6% of the time [55]. 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 correlation between buyer and seller ratings, suggesting that there is a degree of reciprocation of positive ratings and retaliation of negative ratings. This is problematic if obtaining honest and fair ratings is a goal, and a possible remedy could be to not let sellers rate buyers.

The problem of ballot stuffing, i.e. that ratings can be repeated many times, e.g. to unfairly boost somebody's reputation score, seems to be a minor problem on eBay because participants are only allowed to rate each other after the completion of a transaction, which is monitored by eBay. It is of course possible to create fake transactions, but because eBay charges a fee for listing items, there is a cost associated with this practice. However, unfair ratings for genuine transactions can not be avoided.

The eBay reputation system is very primitive and can be quite misleading. With so few negative ratings, a participant with 100 positive and 10 negative ratings should intuitively appear much less reputable than a participant with 90 positive and no negatives, but on eBay they would have the same total reputation score. Despite its drawbacks and primitive nature, the eBay reputation system seems to have a strong positive impact on eBay as a marketplace. Any system that facilitates interaction between humans depend on how they respond to it, and people appear to respond well to the eBay system and its reputation component.

9.2 Expert Sites

Expert sites have a pool of individuals that are willing to answer questions in their areas of expertise, and the reputation systems on those sites are there to rate the experts. Depending on the quality of a reply, the person who asked the question can rate the expert on various aspects of the reply such as clarity and timeliness.

AllExperts [20] provides a free expert service for the public on the Internet with a business model based on advertising. The reputation system on AllExperts uses the aspects: *Knowledgeable*, *Clarity of Response*, *Timeliness* and *Politeness* where ratings can be given in the in the interval $[1, 10]$. The score in each aspect is simply the numerical average of ratings received. The number of questions an expert has received is also displayed in addition to a *General Prestige* score that is simply the sum of all average ratings an expert has received. Most experts receive ratings close to 10 on all aspects, so the General Prestige is usually close to $10\times$the number of questions received. It is also possible to view charts of ratings over periods from 2 months to 1 year.

AskMe [21] is an expert site for a closed user group of companies and their employees, and the business model is based on charging a fee for participating in the AskMe network. Ask Me does not publicly provide any details of how the system works.

Advogato [22] is a community of open-source programmers. Members rank each other according to how skilled they perceive each other to be, using Advogato's trust scheme [23], which in essence is a centralised reputation system based on a flow model. The reputation engine of Advogato computes the reputation flow through a network where members constitute the nodes and the edges constitute referrals between nodes. Each member node is assigned a capacity between 800 and 1 depending on the distance from the source node that is owned by Raph Levien who is the creator of Advogato. The source node has a capacity of 800 and the further away from the source node, the smaller the capacity. Members can refer each other with the status of *Apprentice* (lowest), *Journeyer* (medium)or *Master* (highest). A separate flow graph is computed for each type of referral. A member will get the highest status for which there is a positive flow to his or her node. For example if the flow graph of Master referrals and the flow graph of Apprentice referrals both reach member x then that member will have Master status, but if only the flow graph of Apprentice referrals reaches member x then that member will have Apprentice status. The Advogato reputation systems does not have any direct purpose other than to boost the ego of members, and to be a stimulant for social and professional networking within the Advogato community.

9.3 Product Review Sites

Product review sites have a pool of individual reviewers who provide information for consumers for the purpose of making better purchase decisions. The reputation

[20] http://www.allexperts.com/
[21] http://www.askmecorp.com/
[22] http://www.advogato.org/
[23] http://www.advogato.org/trust-metric.html

systems on those sites apply to products as well as to the reviewers themselves.

9.3.1 Epinions

Epinions [24] founded in 1999 is a product and shop review site with a business model mainly based on so-called cost-per-click online marketing, which means that Epinions charges product manufacturers and online shops by the number of clicks consumers generate as a result of reading about their products on Epinions' web site. Epinions also provides product reviews and ratings to other web sites for a fee.

Epinions has a pool of members who write product and shop reviews. Anybody from the public can become a member simply by signing up. The product and shop reviews written by members consist of prose text and quantitative ratings from 1 to 5 stars for a set of aspects such as *Ease of Use*, *Battery Life* etc. in case of products, and *Ease of Ordering*, *Customer Service*, *On-Time Delivery* and *Selection* in case of shops. Other members can rate reviews as *Not Helpful*, *Somewhat Helpful*, *Helpful*, and *Very Helpful*, and thereby contribute to determining how prominently the review will be placed, as well as to giving the reviewer a higher status. A member can obtain the status *Advisor*, *Top Reviewer* or *Category Lead* (highest) as a function of the accumulated ratings on all his or her reviews over a period. It takes considerable reviewing effort to obtain a status above member, and most members don't have any status.

Category Leads are selected at the discretion of Epinions staff each quarter based on nominations from members. Top Reviewers are automatically selected every month based on how well their reviews are rated, as well as on the Epinions Web of Trust (see below), where a member can *Trust* or *Block* another member. Advisors are selected in the same way as Top Reviewers, but with a lower threshold for review ratings. Epinions does not publish the exact thresholds for becoming Top Reviewer or Advisor, in order to discourage members from trying to manipulate the selection process.

The Epinions Web of Trust is a simple scheme, whereby members can decide to either *trust* or *block* another member. A member's list of trusted members represents that member's personal Web of Trust. As already mentioned, the Web of Trust influences the automated selection of Top Reviewers and Advisors. The number of members (and their status) who trust a given member, will contribute to that member getting a higher status. The number of members (and their status) who block another member will have a negative impact on that member getting a higher status.

Epinions has an incentive systems for reviewers called the *Income Share Program*, whereby members can earn money. Income Share is automatically deter-

[24] http://www.epinions.com/

mined based on general use of reviews by consumers. Reviewers can potentially earn as much for helping someone make a buying decision with a positive review, as for helping someone avoid a purchase with a negative review. This is important in order not to give an incentive to write biased reviews just for profit. As stated on the Epinions FAQ pages: *"Epinions wants you to be brutally honest in your reviews, even if it means saying negative things."* The Income Share pool is a portion of Epinions' income. The pool is split among all members based on the utility of their reviews. Authors of more useful reviews earn more than authors of less useful reviews.

The Income Share formula is not specified in detail in order to discourage attempts to defraud the system. Highly rated reviews will generate more revenue than poorly rated reviews, because the former are more prominently placed so that they are more likely to be read and used by others. Category Leads will normally earn more than Top Reviewers who in turn will normally earn more than Advisors, because their reviews per definition are rated and listed in that order.

Providing high quality reviews is Epinions core value proposition to consumers, and the reputation system is instrumental in achieving that. The reputation system can be characterised as highly sophisticated because of the revenue based incentive mechanism. Where other reputation systems on the Internet only provide immaterial incentives like status or karma, the Epinions system can provide hard cash.

9.3.2 BizRate

BizRate runs a *Customer Certified Merchant* scheme whereby consumers who buy at a BizRate listed store are asked to rate site navigation, selection, prices, shopping options and how satisfied they were with the shopping experience. Consumers participating in this scheme become registered BizRate members. A *Customer Certificate* is granted to a merchant if a sufficient number of surveys over a give period are positive, and this allows the merchant to display the BizRate Customer Certified seal of approval on it's web site. As an incentive to fill out survey forms BizRate, members get discounts at listed stores. This scheme does not capture the frustrated customers who give up before they reach the check, and therefore tends to provide a positive bias of web stores. Thus is understandable from a business perspective, because it provides an incentive for stores to participate in the Customer Certificate scheme.

BizRate also runs a product review service similar to Epinions, but which uses a much simpler reputation system. Members can write product reviews on BizRate, and anybody can become a member simply by signing up. Users, including non-members, who browse BizRate for product reviews can vote on reviews as being *helpful*, *not helpful* or *off topic*, and the reputation systems stops there. Reviews are ordered according to the ratio of helpful over total votes, where the reviews with the

highest ratios are listed first. It is also possible to have the reviews sorted by rating, so that the best reviews are listed first. Reviewers do not get any status and they can not earn money by writing reviews for BizRate. There is thus less incentive for writing reviews on BizRate than there is on Epinions, but it is uncertain how this influences the quality of the reviews. The fact that anybody can sign up to become a member and write reviews and that anybody including non members can vote on the helpfulness of reviews makes this reputation scheme highly vulnerable to attack. A simple attack could consist of writing many positive reviews for a product and ballot stuff them so that they get presented first and result in a high average score for that product.

9.3.3 Amazon

Amazon [25] is mainly an online bookstore that allows members to write book reviews. Amazon's reputation scheme is quite similar to the one BizRate uses. Anybody can become a member simply by signing up. Reviews consist of prose text and a rating in the range 1 to 5 stars. The average of all ratings gives a book its average rating. Users, including non-members, can vote on reviews as being *helpful* or *not helpful*. The numbers of helpful as well as the total number of votes are displayed with each review. The order in which the reviews are listed can be chosen by the user according to criteria such as "newest first", "most helpful first" or "highest rating first".

As a function of the number of helpful votes each reviewer has received, as well as other parameters not publicly revealed, Amazon determines each reviewer's rank, and those reviewers who are among the 1000 highest get assigned the status of Top 1000, Top 500, Top 100, Top 50, Top 10 or #1 Reviewer. Amazon has a system of *Favourite People*, where each member can choose other members as favourite reviewers, and the number of other members who has a specific reviewer listed as favourite person also influences that reviewer's rank. Apart from giving some members status as top reviewers, Amazon does not give any financial incentives. However there are obviously other financial incentives external to Amazon that can play an important role. It is for example easy to imagine why publishers would want to pay people to write good reviews for their books on Amazon.

There are many reports of attacks on the Amazon review scheme where various types of ballot stuffing has artificially elevated reviewers to top reviewer, or various types of "bad mouthing" has dethroned top reviewers. This is not surprising due to the fact that users can vote without becoming a member. For example the Amazon #1 Reviewer usually is somebody who posts more reviews than any living person could possibly do if it would require that person to read each book, thus indicating that the combined effort of a group of people, presented as a single person's work,

[25] http://www.amazon.com/

is needed to get to the top. Also, reviewers who have reached the Top 100 rank have reported a sudden increase in negative votes which reflects that there is a cat fight taking place in order to get into the ranks of top reviewers. In order to reduce the problem, Amazon only allows one vote per registered cookie for any given review. However deleting that cookie or switching to another computer will allow the same user to vote on the same review again. There will always be new types of attacks, and Amazon needs to be vigilant and respond to new types of attacks as they emerge. However, due to the vulnerability of the review scheme it can not be described as a robust scheme.

9.4 Discussion Fora

9.4.1 Slashdot

Slashdot [26] was started in 1997 as a *"news for nerds"* message board. More precisely it is a forum for posting articles and comments to articles. In the early days when the community was small, the signal to noise ratio was very high. As is the case with all mailing lists and discussion fora where the number of members grow rapidly, spam and low quality postings emerged to become a major problem, and this forced Slashdot to introduce moderation. To start with there was a team of 25 moderators which after a while grew to 400 moderators to keep pace with the growing number of users and the amount of spam that followed. In order to create a more democratic and healthy moderation scheme, automated moderator selection was introduced, and the Slashdot reputation system forms an integral part of that as explained below. The moderation scheme actually consists of two moderation layers where M1 is for moderating comments to articles, and M2 is for moderating M1 moderators.

The articles posted on Slashdot are selected at the discretion of the Slashdot staff based on submissions from the Slashdot community. Once an article has been posted, anyone can give comments to that article.

Users of Slashdot can be *Logged In Users* or just anonymous persons browsing the web. Anybody can become a Logged In User simply by signing up. Reading articles and comments as well as writing comments to articles can be done anonymously. Because anybody can write comments, they need to be moderated, and only Logged In Users are eligible to become moderators.

Regularly (typically every 30 minutes), Slashdot automatically selects a group of M1 moderators among long time regular Logged In Users, and gives each moderator 3 days to spend a given number of (typically 5) moderation points. Each moderation point can be spent moderating 1 comment by giving it a rating selected

[26] http://slashdot.org/

from a list of negative (*offtopic*, *flamebait*, *troll*, *redundant*, *overrated*) or positive (*insightful*, *interesting*, *informative*, *funny*, *underrated*) adjectives. An integer score in the range [-1, 5] is maintained for each comment. The initial score is normally 1 but can also be influenced by the comment provider's Karma as explained below. A moderator rating a comment positively causes a 1 point increase in the comment's score, and a moderator rating a comment negatively causes a 1 point decrease in the comment's score, but within the range [-1, 5].

Each Logged In User maintains a *Karma* which can take one of the discrete values *Terrible*, *Bad*, *Neutral*, *Positive*, *Good* and *Excellent*. New Logged In Users start with neutral Karma. Positive moderation of a user's comments contributes to higher Karma whereas a negative moderation of a user's comments contributes to lower Karma of that user. Comments by users with very high Karma will get initial score 2 whereas comments by users with very low Karma will get initial score 0 or even -1. High Karma users will get more moderation points and low Karma users will get less moderation points to spend when they are selected as moderators.

The purpose of the comment score is to be able to filter the good comments from the bad and to allow users to set thresholds when reading articles and postings on Slashdot. A user who only wants to read the best comments can set the threshold to 5 whereas a user who wants to read everything can set the threshold to -1.

To address the issue of unfair moderations, Slashdot has introduced a metamoderation layer called M2, (the moderation layer described above is called M1) with the purpose of moderating the M1 moderators. Any longstanding Logged In user can metamoderate several times per day if he or she so wishes. A user who wants to metamoderate will be asked to moderate the M1 ratings on 10 randomly selected comment postings. The metamoderator decides if a moderator's rating was fair, unfair, or neither. This moderation affects the Karma of the M1 moderators which in turn influences their eligibility to become M1 moderators in the future.

The Slashdot reputation system recognises that a moderator's taste can influence how he or she rates a comment. Having one set of positive ratings and one set of negative ratings, each with different types of taste dependent rating choices, is aimed at solving that problem. The idea is that moderators with different taste can give different ratings (e.g. insightful or funny) to a comment that has merit, but every rating will still be uniformly positive. Similarly, moderators with different taste can give different ratings (e.g. offtopic or overrated) to a comment without merit, but every rating will still be uniformly negative. Slashdot staff are also able to spend arbitrary amounts of moderation points making these people omnipotent and thereby able to manually stabilise the system in case Slashdot would be attacked by extreme volumes of spam and unfair ratings.

The Slashdot reputation system directs and stimulates the massive collaborative effort of moderating thousands of postings every day. The system is constantly

being tuned and modified and can be described as an ongoing experiment in search for the best practical way to promote quality postings, discourage noise and to make Slashdot as readable and useful as possible for a large community.

9.4.2 Kuro5in

Kuro5hin [27] is a web site for discussion of technology and culture started in 1999. It allows members to post articles and comments similarly to Slashdot. The reputation system on Kuro5hin is called *Mojo*. It underwent major changes in October 2003 because it was unable to effectively counter noise postings from throwaway accounts, and because attackers rated down comments of targeted members in order to make them lose their reputation scores. Some of the changes introduced in Mojo to solve these problems include to only let a comment's score influence a user's Mojo (i.e. reputation score) when there are at least six ratings contributing to it, and to only let one rating count from any single IP address.

It is possible that the problems experienced by Kuro5hin could have been avoided had they used Slashdot's principle of only allowing longstanding members to moderate because throwaway accounts would have been less effective as an attack tool.

9.5 Google's Web Page Ranking System

The early web search engines such as Altavista simply presented every web page that matched the key words entered by the user, which often resulted in too many and irrelevant pages being listed in the search results. Altavista's proposal for handling this problem was to offer advanced ways to combine keywords based on binary logic. This was too complex for users and therefore did not represent a good solution.

PageRank proposed by Page *et al.* (1998) [52] represents a way of ranking the best search results based on a page's reputation. Roughly speaking, PageRank ranks a page according to how many other pages are pointing at it. This can be described as a reputation system, because the collection of hyperlinks to a given page can be seen as public information that can be combined to derive a reputation score. A single hyperlink to a given web page can be seen as a positive rating of that web page. Google's search engine [28] is based on the PageRank algorithm and the rapidly rising popularity of Google at the cost of Altavista was obviously caused by the superior search results that the PageRank algorithm delivered.

The definition of PageRank from Page *et al.* (1998) [52] is given below:

[27] http://www.kuro5hin.org/
[28] http://www.google.com/

Definition 4 *Let P be a set of hyperlinked web pages and let u and v denote web pages in P. Let $N^-(u)$ denote the set of web pages pointing to u and let $N^+(v)$ denote the set of web pages that v points to. Let E be some vector over P corresponding to a source of rank. Then, the PageRank of a web page u is:*

$$R(u) = cE(u) + c \sum_{v \in N^-(u)} \frac{R(v)}{|N^+(v)|} \tag{4}$$

where c is chosen such that $\sum_{u \in P} R(u) = 1$.

In [52] it is recommended that E be chosen such that $\sum_{u \in P} E(u) = 0.15$. The first term $cE(u)$ in Eq.(4) gives rank value based on initial rank. The second term $c \sum_{v \in N^-(u)} \frac{R(v)}{|N^+(v)|}$ gives rank value as a function of hyperlinks pointing at u.

According to Def.4 above $R \in [0, 1]$, but the PageRank values that Google provides to the public are scaled to the range [0,10] in increments of 0.25. We will denote the public PageRank of a page u as $PR(u)$. This public PageRank measure can be viewed for any web page using Google's toolbar which is a plug-in to the MS Explorer browser. Although Google do not specify exactly how the public PageRank is computed, the source rank vector E can be defined over the root web pages of all domains weighted by the cost of buying each domain name. Assuming that the only way to improve a page's PageRank is to buy domain names, Clausen (2004) [11] shows that there is a lower bound to the cost of obtaining an arbitrarily good *PR*.

Without specifying many details, Google state that the PageRank algorithm they are using also take other elements into account, with the purpose of making it difficult or expensive to deliberately influence PageRank.

In order to provide a semantic interpretation of a PageRank value, a hyperlink can be seen as a positive referral of the page it points to. Negative referrals do not exist in PageRank so that it is impossible to blacklist web pages with the PageRank algorithm of Eq.(4) alone. Before Google with it's PageRank algorithm entered the search engine market, some webmasters would promote web sites in a spam-like fashion by filling web pages with large amounts of commonly used search key words as invisible text or as metadata in order for the page to have a high probability of being picked up by a search engine no matter what the user searched for. Although this still can occur, PageRank seems to have reduced that problem because a high *PR* is also needed in addition to matching key words in order for a page to be presented to the user.

PageRank applies the principle of trust transitivity to the extreme because rank values can flow through looped or arbitrarily long hyperlink chains. Some theoretic models including [35,38,73] do also allow looped and/or infinite transitivity.

Many suppliers and subcontractors have established a web presence in order to get a broader and more global exposure to potential contract partners. However as described in Sec.1 the problem of information asymmetry and uncertainty about supplier reliability can make it risky to establish supply chain and subcontract agreements online. Reputation systems have the potential to alleviate this problem by providing the basis for making more informed decisions and commitments about suppliers and subcontractors.

Open Ratings [29] is a company that sells *Past Performance* reports about supply chain subcontractors based on ratings provided by past contract partners. Ratings are provided on a 1-100 scale on the following 9 aspects: *Reliability*, *Cost*, *Order Accuracy*, *Delivery/Timeliness*, *Quality*, *Business Relations*, *Personnel*, *Customer Support* and *Responsiveness* and a suppliers score is computed as a function of recently received ratings. The reports also contain the number and business categories of contract partners that provided the ratings.

9.7 Scientometrics

Scientometrics [25] is the study of measuring research output and impacts thereof based on the scientific literature. Scientific papers cite each other, and each citation can be seen as a referral of other scientific papers, their authors and the journals where the papers are published. The basic principle for ranking scientific papers is to simply count the number of times each scientific paper has been cited by another paper, and rank them accordingly. Journals can be ranked in a similar fashion by summing up citations of all articles published in each journal and rank the journals accordingly. Similarly to Google's PageRank algorithm, only positive referrals are possible with cross citations. This means that papers that, for example, are known to be plagiarisms or to contain falsified results can not easily be sanctioned with scientometric methods.

As pointed out by Makino (1998) [43], even though scientometrics normally provide reasonable indicators of quality and reputation, it can sometimes give misleading results.

There is an obvious similarity between hyperlinked web pages and literature cross references, and it would be interesting to apply the concepts of PageRank to scientific cross citations in order to derive a new way of ranking authors and journals. We do not know of any attempt in this direction.

[29] http://openratings.com/

10 Problems and Proposed Solutions

Numerous problems exist in all practical and academic reputation systems. This section describes problems that have been identified and some proposed solutions.

10.1 Low Incentive for Providing Rating

Ratings are typically provided after a transaction has taken place, and the transaction partners usually have no direct incentive for providing rating about the other party. For example when the service provider's capacity is limited, participants may not want to share the resource with others and therefore do not want to give referrals about it. Another example is when buyers withhold negative ratings because they are "nice" or because they fear retaliation from the seller. Even without any of these specific motives, a rater does not benefit directly from providing a rating. It serves the community to provide ratings and the potential for free-riding (i.e. letting the others provide the ratings) therefore exists.

Despite this fact many do provide ratings. In their study, Resnick & Zeckhauser (2002) [55] found that 60.7% of the buyers and 51.7% of the sellers on eBay provided ratings about each other. Possible explanations for these relatively high values can for example be that providing reciprocal ratings simply is an expression of politeness. However lack of incentives for providing ratings is a general problem that needs special attention and that might require specific incentive mechanisms.

Miller *et al.* (2003) [47] have proposed a scheme for eliciting honest feedback based on financial rewards. Jurca & Faltings (2003) [34] have proposed a similar incentive scheme for providing truthful ratings based on payments.

10.2 Bias Toward Positive Rating

There is often a positive bias when ratings are provided. In Resnick & Zeckhauser (2002) [55], it was found that only 0.6% of all the ratings provided by buyers and only 1.6% of all the ratings provided by sellers were negative, which seems too low to reflect reality. Possible explanations for the positive rating bias are that a positive ratings simply represents an exchange of courtesies (Resnick & Zeckhauser 2002), that positive ratings are given in the hope of getting a positive rating in return (Chen & Singh 2001)[10] or alternatively that negative ratings are avoided because of fear of retaliation from the other party (Resnick & Zeckhauser 2002). After all, nobody is likely to be offended by an unfairly positive rating, but badmouthing and unfairly negative ratings certainly have the potential of provoking retaliation or even a law suit.

An obvious method for avoiding positive bias can consist of providing anonymous reviews. A cryptographic scheme for anonymous ratings is proposed by Ismail *et al.* (2003) [27].

10.3 Unfair Ratings

Finding ways to avoid or reduce the influence of unfairly positive or unfairly negative ratings is a fundamental problem in reputation systems where ratings from others are taken into account. This is because the relying party can not control the sincerity of the ratings when they are provided on a subjective basis. Authors proposing methods to counter this problem include [2,5,6,10,12–14,47,64,58,68,69,71]. The methods of avoiding bias from unfair ratings can broadly be grouped into two categories described below.

10.3.1 Endogenous Discounting of Unfair Ratings

This category covers methods that exclude or give low weight to presumed unfair ratings based on analysing and comparing the rating values themselves. The assumption is that unfair ratings can be recognised by their statistical properties.

Dellarocas (2000) [14] and Withby *et al.* [68] have proposed two different schemes for detecting and excluding ratings that are likely to be unfair when judged by statistical analysis. Chen & Singh (2001) [10] have proposed a scheme that uses elements from collaborative filtering for grouping raters according to the ratings they give to the same objects.

10.3.2 Exogenous Discounting of Unfair Ratings

This category covers methods where the externally determined reputation of the rater is used to determine the weight given to ratings. The assumption is that raters with low reputation are likely to give unfair ratings and vice versa.

Private information e.g. resulting from personal experience is normally considered more reliable than public information such as ratings from third parties. If the relying party has private information, then this information can be compared to public information in order to give an indication of the reliability of the public information.

Buchegger & Le Boudec (2003) [6] have proposed a scheme based on a Bayesian reputation engine and a deviation test that is used to classify raters as trustworthy and not trustworthy. Cornelli *at al.* (2002) [12] have described a reputation scheme

to be used on top of the Gnutella [30] P2P network. Ekström & Björnson (2002) [16] have proposed a scheme and built a prototype called TrustBilder for rating subcontractors in the Architecture Engineering Construction (AEC) industry. Yu & Singh (2003) [71] have proposed to use a variant of the Weighted Majority Algorithm [41] to determine the weights given to each rater.

10.4 Change of Identities

Reputation systems are based on the assumption that identities and pseudonyms are long lived, allowing ratings about a particular party from the past to be related to the same party in the future. In case a party has suffered significant loss of reputation it might be in his interest to change identity or pseudonym in order to cut with the past and start from fresh. However, this practice is not in the general interest of the community [20] and should be prevented or discouraged. Authors proposing methods to counter this practice include Zacharia, Moukas & Maes (1999) [72]. Their reputation scheme, which we call the ZMM-scheme, was used in the 1996-1999 MIT based Kasbah multi-agent C2C transaction system. Upon completion of a transaction, both parties were able to rate how well the other party behaved. The Kasbah agents used the resulting reputation score when negotiating future transactions. A main goal in the design of the ZMM-scheme was to discourage users from changing identities, and the ZMM scheme was deliberately designed to penalise newcomers. This approach has the disadvantage that it can be difficult to distinguish between good and bad newcomers.

10.5 Quality Variations Over Time

Economic theory indicates that there is a balance between the cost of establishing a good reputation and the financial benefit of having a good reputation, leading to an equilibrium [36,62]. Variations in the quality of services or goods can be a result of deliberate management decisions or uncontrolled factors, and whatever the cause, the changes in quality will necessarily lead to variations in reputation. Although a theoretic equilibrium exists, there will always be fluctuations, and it is possible to characterise the conditions under which oscillations can be avoided [65] or converge towards the equilibrium [26]. In particular, discounting of the past is shown to be a condition for convergence towards an equilibrium [26]. Discounting of the past can be implemented in various ways, and authors use different names to describe what is basically the same thing. Past ratings can be discounted by a *forgetting factor* [31], *aging factor* [7] or *fading factor* [6]. Inversely a *longevity factor*[30] can be used to determine a rating's time to live. Yet another way to

[30] http://www.gnutella.com

describe it is by *reinforcement learning* [64]. The discounting of the past can be a function of time or of the frequency of transactions, or a combination of both [6].

10.6 Discrimination

Discriminatory behaviour can occur both when providing services and when providing ratings. A seller can for example provide good quality to all buyers except one single buyer. Ratings about that particular seller will indicate that he is trustworthy expect for the ratings from the buyer victim. The filtering techniques described in Sec.10.3.1 will give false positives, i.e. judge the buyer victim to be unfair in such situations. Only systems that are able to recognise the buyer victim as trustworthy, and thereby give weight to his ratings, would be able to handle this situation well. Some of the techniques described in Sec.10.3.2 would theoretically be able to protect against this type of discrimination, but no simulations have been done to prove this.

Discrimination can also take the form of a single rater giving fair ratings except when dealing with a specific partner. The filtering techniques described in Sec.10.3.1 and Sec.10.3.1 are designed to handle this type of discrimination.

10.7 Ballot Box Stuffing

Ballot stuffing means that more than the legitimate number of ratings are provided. This problem is closely related to unfair ratings because ballot stuffing usually consists of too many unfair ratings. In traditional voting schemes, such as political elections, ballot stuffing means that too many votes are cast in favour of a candidate, but in online reputation systems, ballot stuffing can also happen with negative votes. This is a common problem in many online reputation systems described in Sec.9 and they usually have poor protection against it. Among the commercial and live reputation systems, eBay's Feedback forum seems to provide adequate protection against ballot stuffing, because ratings can only be provided after transactions completion. Because eBay charges a fee for each transaction ballot stuffing would be expensive. Epinions' and Slashdot's reputation system also provides some degree of protection because only registered members can vote in a controlled way on the merit of reviews and comments.

11 Discussion and Conclusion

The purpose of this work has been to describe and analyse the state of the art in trust and reputation systems. Dingledine *al.* [15] have proposed the following set

of basic criteria for judging the quality and soundness of reputation computation engines.

1. *Accuracy for long-term performance*. The system must reflect the confidence of a given score. It must also have the capability to distinguish between a new entity of unknown quality and an entity with poor long-term performance.
2. *Weighting toward current behaviour*. The system must recognise and reflect recent trends in entity performance. For example, an entity that has behaved well for a long time but suddenly goes downhill should be quickly recognised as untrustworthy.
3. *Robustness against attacks*. The system should resist attempts of entities to manipulate reputation scores.
4. *Smoothness*. Adding any single rating should not influence the score significantly.

The criteria (1), (2) and (4) are easily satisfied by most reputation engines except for the most primitive such as taking a rating score as the sum of positive minus negative ratings such as in eBays feedback forum. Criterion (3) on the other hand will probably never be solved completely because there will always be new and unforeseen attacks for which solutions will have to be found.

The problems of unfair ratings and ballot stuffing are probably the hardest to solve in any reputation system that is based on subjective ratings from participants, and large number of researchers are working on this in the academic community. Instead of having one solution that works well in all situations there will be multiple techniques with advantages, disadvantages and trade-offs. Lack of incentives to provide ratings is also a fundamental problem, because there often is no rational reason for providing feedback. Among the commercial and online reputation systems that take ratings from users into account, financial incentives are only provided by Epinions (hard cash) and BizRate (price discount), all the other web sites only provide immaterial incentives in the form of status or rank.

Given that reputation systems used in commercial and online applications have serious vulnerabilities, it is obvious that the reliability of these systems sometimes is questionable. Assuming that reputation systems give unreliable scores, why then are they used? A possible answer to this question is that in many situations the reputation systems do not need to be robust because their value lies elsewhere. Resnick & Zeckhauser (2002) [55] consider two explanations in relation to eBays reputation system: (a) Even though a reputation system is not robust it might serve its purpose of providing an incentive for good behaviour if the participants think it works, and (b) even though the system might not work well in the statistical normative sense, it may function successfully if it swiftly reacts against bad behaviour (called *"stoning"*) and if it imposes costs for a participant to get established (called *"label initiation dues"*).

Given that some online reputation systems are far from being robust, it is obvious that the organisations that run them have a business model that is relatively insensitive to their robustness. It might be that the reputation system serves as a kind of social network to attract more people to a web site, and if that is the case, then having simple rules for participating is more important than having strict rules for controlling participants' behaviour. Any reputation system with user participation will depend on how people respond to it, and must therefore be designed with that in mind. Another explanation is that, from a business perspective, having a reputation system that is not robust can be desirable if it generally gives a positive bias. After all, commercial web stores are in the business of selling, and positively biased ratings are more likely to promote sales than negative ratings.

Whenever the robustness of a reputation system is crucial, the organisation that runs it should take measures to protect the stability of the system and robustness against attacks. This can for example be by including routine manual control as part of the scheme, such as in Epinions' case when selecting Category Lead reviewers, or in Slashdot's case where Slashdot staff are omnipotent moderators. Exceptional manual control will probably always be needed, should the system come under heavy attack. Another important element is to keep the exact details of the computation algorithm and how the system is implemented confidential (called *"security by obscurity"*), such as in the case of Epinions, Slashdot and Google. Ratings are usually based on subjective judgement, which opens up the Pandora's box of unfair ratings, but if ratings can be based on objective criteria it would be much simpler to achieve high robustness.

The trust and reputation schemes presented in this study cover a wide range of application and are based on many different types mechanisms, and there is no single solution that will be suitable in all contexts and applications. When designing or implementing new systems, it is necessary to consider the constraints and the type of information that can be used as input ratings.

The rich literature growing around trust and reputation systems for Internet transactions, as well as the implementation of reputation systems in successful commercial application, give a strong indication that this is an important technology. The commercial and live implementation seems to have settled around relatively simple schemes, whereas a multitude of different systems with advanced features are being proposed by the academic community. A general observation is that the proposals from the academic community so far lack coherence. The systems being proposed are usually designed from scratch, and only in very few cases are authors building on proposals by other authors. The period we are in can therefore be seen as a period of pioneers, and we hope that the near future will bring consolidation around a set of sound and well recognised principles for building trust and reputation systems, and that these will find their way into practical and commercial applications.

References

[1] A. Abdul-Rahman and S. Hailes. Supporting Trust in Virtual Communities. In *Proceedings of the Hawaii International Conference on System Sciences*, Maui, Hawaii, 4-7 January 2000 2000.

[2] K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In Henrique Paques, Ling Liu, and David Grossman, editors, *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM01)*, pages 10–317. ACM Press, 2001.

[3] M.D. Abrams. Trusted System Concepts. *Computers and Security*, 14(1):45–56, 1995.

[4] E. Adar and B.A. Huberman. Free Riding on Gnutella. *First Monday (Peer-reviewed Journal on the Internet)*, 5(10):8, Otober 2000.

[5] S. Braynov and T. Sandholm. Incentive Compatible Mechanism for Trust Revelation. In *Proceedings of the First Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, July 2002.

[6] S. Buchegger and J.-Y. Le Boudec. A Robust Reputation System for Mobile Ad-hoc Networks. Technical Report IC/2003/50, EPFL-IC-LCA, 2003.

[7] S. Buchegger and J.-Y. Le Boudec. The Effect of Rumor Spreading in Reputation Systems for Mobile Ad-hoc Networoks. In *Proceedings of the Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, March 2003.

[8] V. Cahill, B. Shand, E. Gray, et al. Using Trust for Secure Collaboration in Uncertain Environments. *Pervasive Computing*, 2(3):52–61, July-September 2003.

[9] M. Carbone, M. Nielsen, and V. Sassone. A Formal Model for Trust in Dynamic Networks. In *Proc. of International Conference on Software Engineering and Formal Methods (SEFM'03)*, Brisbane, September 2003.

[10] M. Chen and J.P. Singh. Computing and Using Reputations for Internet Ratings. In *Proceedings of the Third ACM Conference on Electronic Commerce (EC'01)*. ACM, October 2001.

[11] A. Clausen. The Cost of Attack of PageRank. In *Proceedings of The International Conference on Agents, Web Technologies and Internet Commerce (IAWTIC'2004)*, Gold Coast, July 2004.

[12] F. Cornelli et al. Choosing Reputable Servents in a P2P Network. In *Proceedings of the eleventh international conference on World Wide Web (WWW'02)*. ACM, May 2002.

[13] E. Damiani et al. A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks. In *Proceedings of the 9th ACM conference on Computer and Communications Security (CCS'02)*, pages 207–216. ACM, 2002.

[14] C. Dellarocas. Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In *ACM Conference on Electronic Commerce*, pages 150–157, 2000.

[15] R. Dingledine, M.J. Freedman, and D. Molnar. Accountability Measures for Peer-to-Peer Systems. In *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly Publishers, 2000.

[16] M. Ekstrom and H. Bjornsson. A rating system for AEC e-bidding that accounts for rater credibility. In *Proceedings of the CIB W65 Symposium*, pages 753–766, September 2002.

[17] D. Fahrenholtz and W. Lamesdorf. Transactional Security for a Distributed Reputation Management System. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web)*, volume LNCS 2455, pages 214–223. Springer, September 2002.

[18] R. Falcone and C. Castelfranchi. *Social Trust: A Cognitive Approach*, pages 55–99. Kluwer, 2001.

[19] L.C. Freeman. Centrality on Social Networks. *Social Networks*, 1:215–239, 1979.

[20] E. Friedman and P. Resnick. The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 2001.

[21] D. Gambetta. Can We Trust Trust? In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–238. Basil Blackwell. Oxford, 1990.

[22] T. Grandison and M. Sloman. A Survey of Trust in Internet Applications. *IEEE Communications Surveys and Tutorials*, 3, 2000.

[23] M. Gupta, P. Judge, and M. Ammar. A reputation system for peer-to-peer networks. In *Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video (NOSSDAV)*, 2003.

[24] R. Guttman, A. Moukas, and P. Maes. Agent-mediated Electronic Commerce: A Survey. *Knowledge Engineering Review*, 13(3), June 1998.

[25] W. Hood and C.S. Wilson. The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, 52(2):291–314, 2001.

[26] B.A. Huberman and F. Wu. The Dynamics of Reputations. *Computing in Economics and Finance*, 18, 2003.

[27] R. Ismail, C. Boyd, A. Jøsang, and S. Russel. Strong Privacy in Reputation Systems. In *Proceedings of the 4th International Workshop on Information Security Applications (WISA)*, Jeju Island, Korea, August 2003.

[28] A. Jøsang. Trust-Based Decision Making for Electronic Transactions. In L. Yngström and T. Svensson, editors, *Proceedings of the 4th Nordic Workshop on Secure Computer Systems (NORDSEC'99)*. Stockholm University, Sweden, 1999.

[29] A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.

[30] A. Jøsang, S. Hird, and E. Faccer. Simulating the Effect of Reputation Systems on e-Markets. In P. Nixon and S. Terzis, editors, *Proceedings of the First International Conference on Trust Management (iTrust)*, Crete, May 2003.

[31] A. Jøsang and R. Ismail. The Beta Reputation System. In *Proceedings of the 15th Bled Electronic Commerce Conference*, June 2002.

[32] A. Jøsang and S. Lo Presti. Analysing the Relationship Between Risk and Trust. In T. Dimitrakos, editor, *Proceedings of the Second International Conference on Trust Management (iTrust)*, Oxford, March 2004.

[33] A. Jøsang and S. Pope. Semantic Constraints for Trust Tansitivity. In S. Hartmann and M. Stumptner, editors, *Proceedings of the Asia-Pacific Conference of Conceptual Modelling (APCCM) (Volume 43 of Conferences in Research and Practice in Information Technology)*, Newcastle, Australia, February 2005.

[34] R. Jurca and B. Faltings. An Incentive Compatible Reputation Mechanism. In *Proceedings of the 6th Int. Workshop on Deception Fraud and Trust in Agent Societies (at AAMAS'03)*. ACM, 2003.

[35] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina. The EigenTrust Algorithm for Reputation Management in P2P Networks. In *Proceedings of the Twelfth International World Wide Web Conference*, Budapest, May 2003.

[36] D. Kreps and R. Wilson. Reputation and Imperfect Information. *Journal of Economic Theory*, 27(2):253–279, 1982.

[37] I. Kurbel, K.and Loutchko. A Framework for Multi-agent Electronic Marketplaces: Analysis and Classification of Existing Systems. In *Proceedings of the International ICSC Congress on Information Science Innovations (ISI'01)*, American University in Dubai, U.A.E., March 2001.

[38] R. Levien. *Attack Resistant Trust Metrics*. PhD thesis, University of California at Berkeley, 2004.

[39] C.Y. Liau et al. Efficient Distributed Reputation Scheme for Peer-to-Peer Systems. In *Proceedings of the 2nd International Human.Society@Internet Conference (HSI)*, volume LNCS 2713, pages 54–63. Springer, 2003.

[40] Liberty-Alliance.
Liberty Trust Models Guidelines. http://www.projectliberty.org/specs/liberty-trust-models-guidelines-v1.0.pdf, Draft Version 1.0-15 edition, 2003.

[41] N. Littlestone and M.K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[42] P. Maes, R. Guttman, and A. Moukas. Agents that Buy and Sell: Transforming Commerce as We Know It. *Communications of the ACM*, 42(3):81–91, 1999.

[43] J. Makino. Productivity of Research Groups. Relation between Citation Analysis and Reputation within Research Communities. *Scientometrics*, 43(1):87–93, 1988.

[44] D.W. Manchala. Trust Metrics, Models and Protocols for Electronic Commerce Transactions. In *Proceedings of the 18th International Conference on Distributed Computing Systems*, 1998.

[45] P.V. Marsden and N. Lin, editors. *Social Structure and Network Analysis*. Beverly Hills: Sage Publications, 1982.

[46] D.H. McKnight and N.L. Chervany. The Meanings of Trust. Technical Report MISRC Working Paper Series 96-04, University of Minnesota, Management Information Systems Reseach Center, 1996.

[47] N. Miller, P. Resnick, and R. Zeckhauser. *Eliciting Honest Feedback in Electronic Markets*. Working paper originally prepared for the SITE'02 workshop, availale at http://www.si.umich.edu/ presnick/papers/elicit/, February 11 2003.

[48] L. Mui, A. Halberstadt, and M. Mohtashemi. Notions of Reputation in Multi-agent Systems: A Review. In *Proceedings of the First Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, July 2002.

[49] L. Mui, M. Mohtashemi, and C. Ang. A Probabilistic Rating Framework for Pervasive Computing Environments. In *Proceedings of the MIT Student Oxygen Workshop (SOW'2001)*, 2001.

[50] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt. Ratings in Distributed Systems: A Bayesian Approach. In *Proceedings of the Workshop on Information Technologies and Systems (WITS)*, 2001.

[51] L. Mui, M. Mohtashemi, and A. Halberstadt. A Computational Model of Trust and Reputation. In *Proceedings of the 35th Hawaii International Conference on System Science (HICSS)*, 2002.

[52] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.

[53] L. Rasmusson and S. Janssen. Simulated Social Control for Secure Internet Commerce. In Catherine Meadows, editor, *Proceedings of the 1996 New Security Paradigms Workshop*. ACM, 1996.

[54] M.K. Reiter and S.G. Stubblebine. Toward acceptable metrics of authentication. In *Proceedings of the 1997 IEEE Symposium on Research in Security and Privacy*, Oakland, CA, 1997.

[55] P. Resnick and R. Zeckhauser. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In M.R. Baye, editor, *The Economics of the Internet and E-Commerce*, volume 11 of *Advances in Applied Microeconomics*. Elsevier Science, 2002.

[56] P. Resnick, R. Zeckhauser, R. Friedman, and K. Kuwabara. Reputation Systems. *Communications of the ACM*, 43(12):45–48, December 2000.

[57] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The Value of Reputation on eBay:
A Controlled Experiment. *Experimental Economics*, 9(2):79–101, 2006. Avaliable from http://www.si.umich.edu/˜presnick/papers/postcards/PostcardsFinalPrePub.pdf.

[58] T. Riggs and R. Wilensky. An Algorithm for Automated Rating of Reviewers. In *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pages 381–387, 2001.

[59] J. Sabater and C. Sierra. REGRET: A reputation model for gregarious societies. In *Proceedings of the 4th Int. Workshop on Deception, Fraud and Trust in Agent Societies, in the 5th Int. Conference on Autonomous Agents (AGENTS'01)*, pages 61–69, Montreal, Canada, 2001.

[60] J. Sabater and C. Sierra. Reputation and Social Network Analysis in Multi-Agent Systems. In *Proceedings of the First Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, July 2002.

[61] J. Sabater and C. Sierra. Social ReGreT, a reputation model based on social relations. *SIGecom Exchanges*, 3.1:44–56, 2002.

[62] A. Schiff and J. Kennes. The Value of Reputation Systems. In *Proceedings of the First Summer Workshop in Industrial Organization (SWIO)*, Auckland NZ, March 2003.

[63] J. Schneider et al. Disseminating Trust Information in Wearable Communities. In *Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing (HUC2K)*, September 2000.

[64] S. Sen and N. Sajja. Robustness of Reputation-based Trust: Boolean Case. In *Proceedings of the First Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*. ACM, July 2002.

[65] C. Shapiro. Consumer Information, Product Quality, and Seller Reputation. *The Bell Journal of Economics*, 13(1):20–35, 1982.

[66] S. Tadelis. Firm Reputation with Hidden Information. *Economic Theory*, 21(2):635–651, 2003.

[67] O.E. Williamson. Calculativeness, Trust and Economic Organization. *Journal of Law and Economics*, 36:453–486, April 1993.

[68] A. Withby, A. Jøsang, and J. Indulska. Filtering Out Unfair Ratings in Bayesian Reputation Systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005.

[69] B. Yu and M.P Singh. A Social Mechanism of Reputation Management in Electronic Communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents*, pages 154–165, 2000.

[70] B. Yu and M.P. Singh. An Evidential Model of Distributed Reputation Management. In *Proceedings of the First Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*. ACM, July 2002.

[71] B. Yu and M.P. Singh. Detecting Deception in Reputation Management. In *Proceedings of the Second Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, pages 73–80. ACM, 2003.

[72] G. Zacharia, A. Moukas, and P. Maes. Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Science*. IEEE, 1999.

[73] C.-N. Ziegler and G. Lausen. Spreading Activation Models for Trust Propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE '04)*, Taipei, March 2004.

[74] P.R. Zimmermann. *The Official PGP User's Guide*. MIT Press, 1995.

Audun Jøsang is the research leader of the Security Unit at the Distributed Systems Technology Centre in Brisbane. His research focuses on trust and reputation systems in addition to information security. Audun received his PhD from the Norwegian University of Science and Technology in 1998, and has a MSc in Information Security from Royal Holloway College, University of London, and a BSc in Telematics from the Norwegian Institute of Technology.

Roslan Ismail is a senior lecturer at the Malaysian National Tenaga University and a PhD student at the Information Security Research Centre at Queensland University of Technology. His research interests are in computer security, reputation systems, e-commerce security, forensic security, security in mobile agents and trust management in general. He has a MSc in Computer Science from the The Malaysian University of Technology, and a BSc from Pertanian University.

Colin Boyd is an Associate Professor at Queensland University of Technology and Deputy Director of the Information Security Research Centre there. His research interests are in the theory and applications of cryptography. He has authored over 100 fully refereed publications including a recent book on protocols for authentication and key establishment. Colin received the B.Sc. and Ph.D. degrees in mathematics from the University of Warwick in 1981 and 1985, respectively.