# Can We Manage Trust?[⋆]

Audun Jøsang[1] and Claudia Keser[2] and Theo Dimitrakos[3]

[1] DSTC [⋆⋆]
UQ Qld 4072, Australia.
ajosang@dstc.edu.au
[2] IBM T. J. Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598, USA
ckeser@us.ibm.com
[3] British Telecom
2A Rigel House, Adastral Park, Martlesham Heath
Ipswich, Suffolk, IP5 3RE, UK
Theo.Dimitrakos@bt.com

**Abstract.** The term trust management suggests that trust can be managed, for example by creating trust, by assessing trustworthiness, or by determining optimal decisions based on specific levels of trust. The problem to date is that trust management in online environments is a diverse and ill defined discipline. In fact, the term trust management is being used with very different meanings in different contexts. This paper examines various approaches related to online activities where trust is relevant and where there is potential for trust management. In some cases, trust management has been defined with specific meanings. In other cases, there are well established disciplines with different names that could also be called trust management. Despite the confusion in terminology, trust management, as a general approach, represents a promising development for making online transactions more dependable, and in the long term for increasing the social capital of online communities.

## 1 Introduction

How can we manage trust? This question is increasingly catching the attention of stakeholders in the Internet and online services industry as concerns regarding privacy, online payment security or reliability of service providers and vendors seems to negatively affect the growth of the Internet.

Lack of trust is like sand in the social machinery, and represents a real obstacle for the uptake of online services, for example for entertainment, for building personal relationships, for conducting business and for interacting with governments. Lack of trust also makes us waste time and resources on protecting ourselves against possible harm, and thereby creates significant overhead in economic transactions.

[⋆] Appears in the Proceedings of the 3rd International Conference on Trust Management, (iTrust), Paris, May 2005.

[⋆⋆] 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, and Training).

Positive trust on the other hand is a like catalyst for human cooperation. It enables people to interact spontaneously and helps the economy to operate smoothly. A high level of well placed general trust is therefore very desirable for the prosperity of a community.

However, the ability to distrust, when warranted, enables us to avoid harm when confronted with unreliable systems or dishonest people and organisations. Similarly, trust, when unwarranted, results in exposure to risk and hazards. Trust is like a compass for guiding us safely through a world of uncertainty, risk and moral hazards.

The important role that trust plays for online interaction has resulted in the emergence of trust management as a new research discipline in the intersection between sociology, commerce, law and computer science. It focuses on understanding and facilitating people's trust in each other, in the network infrastructure and in the environment within which online interactions are embedded. An important goal of trust management, is to stimulate people's and organisations' acceptance of online environments as safe places for interacting and doing business.

This paper takes a closer look at current approaches to trust management, with the goal of assessing their potential for stimulating online activities and increasing the quality of online communities. In many cases, trust management simply represents a specific approach to activities that are often already well established under different names. The unique aspect of trust management is that it specifically focuses on assessing and increasing the dependability of systems and players in the online environment, and thereby on stimulating the growth and prosperity of online communities.

## 2 Understanding Trust Management

Trust is a directional relationship between two parties that can be called *trustor* and *trustee*. One must assume the trustor to be a "thinking entity" in some form, whereas the trustee can be anything from a person or physical entity, to abstract notions such as software or a cryptographic key [20].

A trust relationship has a *scope*, meaning that it applies to a specific purpose or domain of action, such as "being authentic" in the case of a an agent's trust in a cryptographic key. Mutual trust is when both parties trust each other for the same purpose, but this is obviously only possible when both parties are thinking entities. Trust influences the trustor's attitudes and actions, but can also have effects on the trustee and other elements in the environment, for example, by stimulating reciprocal trust [12].

The literature uses the term trust with a variety of meanings [30]. A distinction between *context independent trust*, (which we nickname *"reliability trust"*), and *context dependent trust* (which we nickname *"decision trust"*), can often be recognised in the literature, although usually not explicitly expressed in those terms.

As the name suggest, reliability trust can be interpreted as the reliability of something or somebody independently of the context, and the definition by Gambetta (1988) [15] 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.*

In Def.1, trust is primarily defined as the trustor's estimate of the trustee's reliability (in the sense of probability), and includes the concept of *dependence* on the trustee. It can be noted that this definition does not take the context into account.

However, trust can be more complex than Gambetta's definition indicates. For example, Falcone & Castelfranchi (2001) [13] 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 [13] 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."*

For example, consider a person who distrusts an old rope for climbing from the third floor of a house during a fire exercise. Imagine now that the same person is trapped in a real fire in the same house, and that the only escape is to climb from the third floor window with the same old rope. In a real fire, most people would trust the rope. Although the *reliability trust* in the rope is the same in both situations, the *decision trust* changes as a function of the utility values associated with the possible courses of action.

The following definition captures the concept of trust seen within a context.

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

In Def.2, trust is primarily defined as the willingness to rely on a given object, and specifically includes the notions of *dependence* on the trustee, and its *reliability*. In addition, Def.2 implicitly also covers contextual elements such as *utility* (of possible outcomes), *environmental factors* (law enforcement, contracts, security mechanisms etc.) and *risk attitude* (risk taking, risk averse, etc.).

Both reliability trust and decision trust are based on positive belief about the object that the trustor depends on for his welfare. People hold this type of beliefs about almost everything around them, and would in fact be unable to live without it, as commented by Luhmann (1979):

> *"But a complete absence of trust would prevent him even from getting up in the morning. He would be prey to a vague sense of dread, to paralysing fears. He would not even be capable of formulating distrust and making that a basis for precautionary measures, since this would presuppose trust in other directions. Anything and everything would be possible. Such abrupt confrontation with the complexity of the world at its most extreme is beyond human endurance."* [29].

Trying to define trust management around this broad view of trust would create an extremely general, and effectively useless notion, because it would have to deal with our beliefs about almost everything around us. Simplifying assumptions about the stability of the world and our existential situation are therefore necessary.

Fortunately, we are concerned only with the online environment, such as the Internet, and this allows us to reduce the scope of trust management to a level where it can be useful [16, 23]. Because trust is an asymmetric relationship between a trustor and a trustee, trust management can be seen from two sides. The ability to gain the trust of others is an important criterion for the success and survival of an entity, because it makes others willing to collaborate. Humans and animals therefore have a set of genetically determined and culturally acquired strategies for appearing trustworthy. The safest and most used strategy is probably to simply act in a responsible and trustworthy manner. However, it is not uncommon that people attempt to give a false impression of trustworthiness for various reasons, such as for personal gain. The ability to correctly assess the trustworthiness of a target is therefore an equally important criterion for the performance, success or survival of an entity. Humans and animals have a strong ability to assess trustworthiness of other entities, based on a multitude of cognitive and affective factors. Based on this duality of trust, we define trust management for online environments as follows:

**Definition 3 (Trust Management).** *The activity of creating systems and methods that allow relying parties to make assessments and decisions regarding the dependability of potential transactions involving risk, and that also allow players and system owners to increase and correctly represent the reliability of themselves and their systems.*

Computer networks are removing us from a familiar direct style of interacting. We may now collaborate online with people or organisations we have never met, perhaps never heard of before, and that we might never meet again. Many of the traditional strategies for representing and assessing trustworthiness in the physical world can no longer be used in online environments. It can therefore be difficult to assess whether the services and information provided by a remote party are reliable and correctly represented. Organisations that engage in online service provision also face the challenge of building online systems and online customer relationships which engender trust.

There is thus a need for methodologies that enable relying parties to determine the trustworthiness of remote parties through computer mediated communication and collaboration. At the same time, trustworthy entities need methodologies that enable them to be recognised as such. In a nutshell, developing and applying these methodologies can be called trust management.

## 3 Sociological and Economic Interpretations of Trust

Online activities are directly and indirectly influenced by the social and economic environment in which they are embedded. It is therefore useful to understand the role trust plays in socio-economic activities in general. The following two subsections will look at trust from sociological and economic perspectives respectively, in order to assess the meaning and value trust management in those domains.

### 3.1 Trust and Sociology

From a sociological viewpoint, Fukuyama (1995) [14] and Putnam (1995) [36] argue that trust creates *social capital*. Social capital has been defined by Coleman (1988) [10] as *"the ability of people to work together for common purposes in groups and organizations"*.

The social capital embedded in trusting relationships is present in communities of all sizes, ranging from families to corporations and nations. Fukuyama argues that social capital fuels a society's economic performance. Based on a historical analysis he demonstrates that low trust countries tend to show a lower overall economic performance than high trust countries.

According to Fukuyama, the ability of a society to develop strong civic institutions and efficient organisations depends to a large degree on its social capital:

> *"One of the most immediate consequences of a culture with a high propensity for spontaneous sociability is the ability to form large modern corporations. The emergence of large, professionally managed corporations was driven, of course, by a host of technological and market-size factors as producers and distributors sought optimum scale efficiencies. But the development of large organisations able to exploit such efficiencies was greatly facilitated by the prior existence of a culture inclined to spontaneous social organization. It would appear to be no accident that three high-trust societies, Japan, Germany and the United States, pioneered the development of large-scale, professionally managed enterprises. Low trust societies like France, Italy and noncommunist Chinese states including Taiwan and Hong Kong, by contrast, were relatively late in moving beyond large family businesses to modern corporations."* [14], p.338.

Knack and Keefer (1997) [27] provide empirical support to Fukuyama's thesis based on general measures of trust derived from the *World Values Survey* of the National Opinion Research Center. Knack and Keefer took the percentage of respondents from each country with a high trust attitude as a measure of how trusting that country's populace was. They found a strong correlation between a trusting attitude and economic growth.

While the sociological interpretation of trust as social capital described above was aimed at explaining patterns in traditional cultures and societies, it seems quite obvious that social capital also is important in online communities, by stimulating their creation and prosperity [19, 35]. Huang *et al.* (2003) [17] show that trust has a statistically significant influence on levels of Internet penetration across countries, thereby reflecting this effect.

### 3.2 Trust and Economic Exchange

Neoclassical economic theory is based on the paradigm of self-interested rational behaviour. This postulated egoism of the economic agents leaves no room for trustworthy behaviour or trust. As a result, game theoretic economic models often predict breakdown in collaboration and inefficient economic exchange [3].

In real life however, people behave less selfishly than postulated by economic theory. In an experimental economics laboratory setting, Keser (2000) observed that people do trust, although the level of trust may depend on many factors [24]. For example, as already pointed out, the level of trust varies across countries. It may also play a role whether and how well the interacting parties know each other and whether the interaction is going to be repeated or not. Obviously, people are less inclined to trust a stranger with whom they interact for the first and probably only time than they are inclined to trust someone whom they have known from many previous interactions.

Repeated interaction with the same partner allows for reciprocal behaviour: cooperative behaviour by the others may be rewarded by cooperation while failure to cooperate may be punished by avoidance of cooperation. Mutual cooperation can be easily established and maintained if all of the players start out by playing cooperatively, hereby signalling their willingness to cooperate, and then adhere to the reciprocity principle. Axelrod (1984)[2], Selten *et al.* (1997) [41] and Keser (2000) [24] demonstrate that this kind of strategy tends to be very successful over many encounters with other players.

The reciprocity principle can play a role in a single encounters with others. This is due to *indirect reciprocity*(i.e. *"I'm doing a good deed today, hoping that it will come back to me at an other occasion"*) . However, as shown for example by Keser and van Winden (2000) [26], the efficiency level observed in single encounters is significantly lower than in repeated encounters. This implies for the Internet, as it increases the likelihood of single anonymous encounters, that we can expect reciprocity to play a lesser role, than in more traditional types of interactions. Fortunately, technological mitigation strategies, e.g. by using trust and reputation systems, can be used to change this trend.

Williamson (1993) [44] discusses the role of trust in the economic organisation. The economic approach to economic organisation is principally calculative. The awkwardness of including the notion of trust in calculative models leads him to reject trust as a computational concept. Williamson 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. Because these terms already have a clear meaning, they give us better model control than the term trust provides.

A question arises whether trust related to the environment can be dissected in the similar way that direct calculative trust in agents can. By trust in the environment, we mean trust in elements of the environment within which economic transactions are embedded. This influences economic exchange in an indirect way. According to Williamson (1993) [44], that is because the need for transaction specific safeguards varies systematically with the institutional environments within which transactions are embedded. Changes in the conditions of the environment are therefore factored in, by adjusting transaction specific controls in cost-effective ways. In effect, institutional environments that provide general purpose safeguards relieve the need for added transaction specific controls. Accordingly, transactions that are viable in contexts that provide strong safeguards may not be viable in contexts that are weak, because it is not cost effective for the parties to craft transaction specific governance in the latter circumstances.

## 4 Principles for Building Trust

Trust requires two parties, and can therefore be seen from two sides. The relying party is interested in assessing the trustworthiness of the trustee as correctly as possible. The trustee, on the other hand, is interested in painting the best possible picture of himself, which is often done through marketing. This section briefly explains the model for establishing trust in e-commerce, and situations where it can be ethical to deceive by deliberately creating false trust.

## 4.1 Building Trust in e-Commerce

The following Trust Transition Model is adapted from Cheskin (1999) [8] and extended to give a more complete description of realistic transition stages for trust. The various technologies for building trust used by e-commerce players [33] fit into this model.

A distinction can be made between *extrinsic trust factors* which roughly speaking are information elements that are communicated between parties, and *intrinsic trust factors* which roughly speaking are information elements emerging from personal experience. These take precedence at different stages in the trust transition model of Fig.1.

Cheskin found that an unaware (potential) person first has to be attracted to, or made aware of a service provider, in order for that person to be able to develop trust in that service provider. Let us assume that the principal is aware of several service providers with similar offerings regarding specified price and quality. Then in the initial phase, the extrinsic trust factors will determine the principal's choice of service provider. These extrinsic factors can be, for instance, that the service provider has a well-known brand with a good reputation, that the web site offers ease of navigation, that the service provider presents satisfactory privacy and complaint policies, and that the communication channel is encrypted.

The customer then gets first hand experience with the chosen service provider. A good experience will improve trust, whereas a bad experience normally will cause the trust level to drop below the purchase threshold, as indicated by the dotted arrow in Fig.1. There are techniques for regaining lost trust. The provider must be able to determine when something has gone wrong and give the dissatisfied customer adequate apology and compensation. Studies show that a dissatisfied customer who receives compensation can end up having stronger trust that a satisfied customer. The obvious explanation is that receiving apology and compensation is evidence of good fulfilment, which in the end becomes a very strong positive intrinsic trust factor.

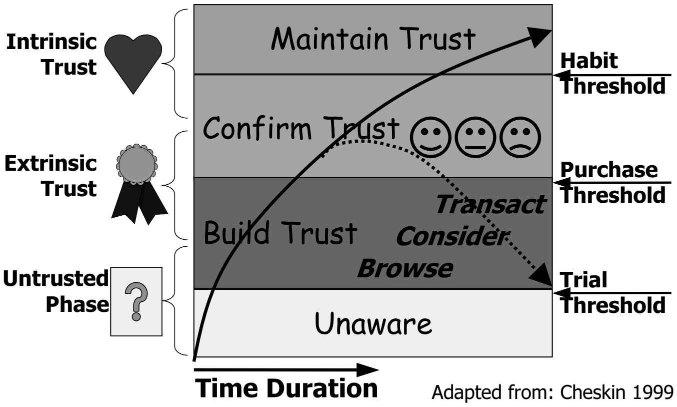

**Fig. 1.** e-Commerce trust transition model

Over time, assuming that the extrinsic aspects of the service provider have not dramatically changed, the customer's mental model of the service provider, and its perceived trustworthiness, tends to become implicit, becoming an internalised product of interacting with that service provider. The personal direct experience represents intrinsic trust factors that overtime will outweigh the extrinsic factors.

Since deep and stable trust is based on experience over time, establishing initial trust can be a major challenge to newcomers in e-commerce, particularly those who do not have well established off-line brands. Without initial trust, merchants can not build a good transaction history - and without a good transaction history, consumers may not build trust in these merchants. Pichler [34] describes how, to some extent, merchants can 'buy' trust though advertising: this evidence of financial investment implies to consumers that a firm will not engage in quick gain deception.

Economic theory indicates that there is a balance between the cost of establishing a brand with a good reputation, and the financial benefit of having a it, leading to an equilibrium [28, 40]. 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 the perceived trustworthiness. Although a theoretic equilibrium exists, there will always be fluctuations, and it is possible to characterise the conditions under which oscillations can be avoided [42] or converge towards the equilibrium [18].

### 4.2 Deception by Creating False Trust

Attempts to create false trust is normally seen as unethical. However, there are situations where most people would agree that it is morally acceptable to create false trust, and self defence is one such case.

In case a system is being attacked, counter attack against the perpetrators' system is generally illegal. Since attackers normally trust systems to provide genuine resources and services, an effective and legal defence strategy may be to let the attacked systems deceive [39]. This would waste the attacker's time and resources, and give the system owners the opportunity to organise an appropriate defence, and to collect, if needed, forensic information.

*Honeypots*, *tar pits* and *poisoning* are examples of computer deception. A honeypot is a system with no purpose, except to encourage attacks, so that data can be collected about the attacks and the attackers. A tar pit is an application, such as a protocol layer entity, that will intentionally respond slowly to requests from an attacker, with the purpose of slowing down the attacker's interaction with the target system. Poisoning is the deliberate introduction of false or incorrect data in a computer system, with purpose of making attackers believe that it is correct data.

It can also be useful to prevent the creation of trust, or even to create distrust, in order to divert attention away from an attractive but scarce resource. Agents who want to avoid sharing a scarce resource with others could deliberately destroy or hide positive evidence that otherwise would increase peoples trust in the resource, or deliberately fabricate false negative evidence, on order to induce negative trust in the resource.

# 5 Trust and Security

In a general sense, the purpose of security mechanisms is to provide protection against malicious parties. 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. In this sense there is a whole range of security challenges that are not met by traditional approaches. 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) [37], 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. This section discusses trust related traditional security activities, whereas soft security and reputation systems will be described in more detail in Sec.6

## 5.1 Trust and Computer Systems Security

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 [1]), but the concept of security assurance level is more standardised as a measure of security[4]. The assurance level can be interpreted as a system's strength to resist malicious attacks, and some organisations require 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) trust in a 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 actually form their own subjective trust in the system.

In case of multiple parties with conflicting interests, what is trusted by one party can be distrusted by another. For example, in the discussion around trusted computing (TC), there are two views on increasing security assurance regarding Digital Rights Management (DRM) by increasing monitoring of user activities. The service provider's perspective can be expressed as follows:

> *"A mildly charitable view of TC was put forward by the late Roger Needham who directed Microsoft's research in Europe: there are some applications in which you want to constrain the user's actions. For example, you want to stop people fiddling with the odometer on a car before they sell it. Similarly, if you want to do DRM on a PC then you need to treat the user as the enemy."*[5]

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

[5] Quoted from Ross Anderson's "Trusted Computing FAQ" where he cites Roger Needham [31]. http://www.cl.cam.ac.uk/ rja14/tcpa-faq.html

Although this would increase the digital content providers' trust, it would negatively affect the users' trust in those systems because it would threaten their privacy. This is an example of how measures to increase security assurance could possibly back-fire, thus making trust management controversial.

### 5.2 Trust in User Authentication and Access Control

Owners of systems and resources usually want to control who can access them. This is traditionally based on having a process for initial authorisation of identified parties, combined with operational mechanisms for authentication, and for controlling what resources those authenticated parties can access. There is thus a separation between authorisation, authentication and access control.

The first common use of the term trust management was closely linked to the combination of authorisation, authentication and access control in distributed systems, as expressed by by Blaze *et al.* (1996) [5]. The main idea behind their approach was that a system does not need to know the identities of those who are accessing its resources, only that they are trusted to do so. This type of trust management is thus about verifying access credentials without necessarily authenticating entities. Blaze *et al.* defined trust management as:

> *"a unified approach to specifying and interpreting security policies, credentials, relationships which allow direct authorisation of security-critical actions."* [5]

The traditional purpose of X.509 and PGP public-key certificates is to link identities to public keys, not to specify access privileges. This contrasts with the approach by Blaze *et al.* which consists of assigning access privileges directly to a public key in the form of an access privilege certificate. Anyone possessing the corresponding private key then automatically has the access privileges assigned to the public key.

This model was implemented by AT&T Research Laboratories in PolicyMaker [6], and later enhanced in KeyNote [4]. REFEREE [9] is a system based on PolicyMaker aimed at controlling access to Web resources.

### 5.3 The Role of Trust in Security Protocols

From a trust perspective, security protocols can be described as mechanisms for propagating trust from where it initially exists to where it is needed [43]. Any security protocol relies on a set of initial assumptions about the protocol building blocks, participants and execution environment. More specifically, these assumptions can be divided into external and internal assumptions. The *external assumptions* are not specifically included when specifying and analysing a protocol, but are nevertheless needed in order for the implementation and instantiation of the protocol to be secure. The *Internal assumptions* are explicitly expressed in the formalism of the security protocol, and form the basis for deriving the conclusion of the protocol. This separation between internal and external assumptions can clearly be discerned in the description of BAN logic (1990) [7] for security protocol verification.

> *"Since we operate at an abstract level, we do not consider errors introduced by concrete implementations of a protocol, such as deadlock, or even inappropriate use of cryptosystems. Furthermore, while we allow for the possibility of hostile intruders, there is no attempt to deal with the authentication of an untrustworthy principal, nor detect weaknesses of encryption schemes or unauthorised release of secrets. Rather, our study concentrates on the beliefs of trustworthy parties involved in the protocols and in the evolution of these beliefs as a consequence of communication."* [7]

Formalisms for security protocol verification can prove that the intended conclusion of the protocol can be derived from the initial internal assumptions and a correct instantiation of the protocol. However, because the internal assumptions only represent a subset of all the necessary assumption, these formalisms have limited scope for verifying the overall security of protocols. Nevertheless, security protocols are used in a broad range of security applications, and represent an efficient way of communicating and deriving security critical beliefs through otherwise insecure networks.

## 6 Systems for Deriving Trust and Reputation

There are two fundamental differences between traditional and online environments regarding how trust and reputation are, and can be used. Firstly, 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. In order to build good trust and reputation systems we therefore need to:

- 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.
- Take advantage of IT and the Internet to create efficient systems for collecting that information, and to derive measures of trust and reputation in order to support decision making and to improve the quality of online markets.

Reputation systems[38, 21] create word-of-mouth networks in online communities, and provide a quantitative measure of quality of service. Parties can rate each other, for example after the completion of a transaction. A reputation score can be derived from aggregated ratings about a given party, to assist other parties in deciding whether to transact with that party in the future. A natural side effect is the incentive this provides for good behaviour among participants. Examples of Web sites that use reputation systems are eBay, Epinions, Amazon, BizRate and Google.

For environments where trust is needed for any economic transaction to take place, it has been shown [25] that reputation systems can considerably increase the levels of both trust and good behaviour. This indicates that the introduction of reputation systems facilitates growth of online markets such as eBay. Reputation influences interactions between agents in two important ways. Firstly, it positively affects the trustor's reliability trust in the trustee. Secondly, it disciplines the trustee, because he knows that bad behaviour will be seen, and thereby sanctioned by the community. Trust and reputation are closely linked. Take e.g. the Concise Oxford dictionary's definition of reputation:

**Definition 4 (Reputation).** *Reputation is what is generally said or believed about a person's or thing's character or standing.*

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 statements relate to the same transaction, statement 1 reflects that the relying party bases his trust on the trustee's reputation. 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 positive or negative reputation that a person might have. Whenever trust is based on multiple factors, some will carry more weight than others. Personal experience (intrinsic factors) typically carries more weight than second hand trust referrals or reputation (extrinsic factors), but in the absence of personal experience, trust often has to be based on referrals from others.

While the basic principles of reputation systems are relatively easy to describe, what constitutes a *trust system* is less precise. Trust systems can include methods for analysing private information in addition to public ratings, and usually have a method for analysing transitive trust [21]. The idea behind trust transitivity is that when, for example, Alice trusts Bob, and Bob trusts Claire, then Alice can derive a measure of trust in Claire, as illustrated in Fig.2 below. The type of trust considered in transitivity models is obviously reliability trust, not decision trust. In addition there are semantic constraints for the transitive trust derivation to be valid, e.g. that all involved trust relationships have the same purpose [22]. PKIs[6], PGP[7] [45], and even Google's PageRank algorithm [32], are all based on trust transitivity, and can be called trust systems.

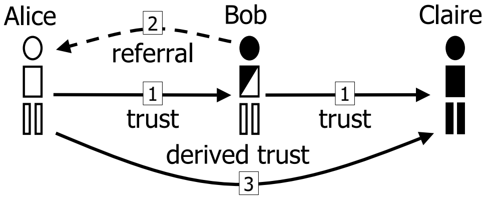

**Fig. 2.** Trust transitivity principle

[6] Public-Key Infrastructure
[7] Pretty Good Privacy

## 7 Discussion and Conclusion

All is not as it should be in the online world. People have well based concerns about the reliability of players, systems and infrastructure, and this often represents a barrier for leveraging computer systems and the Internet to make communication and organisation more efficient. The emergence of trust management as a specific activity is aimed at mitigating this problem.

Trust is a relationship between two parties, and the goal of trust management can therefore be seen from two sides: Firstly, it aims at enabling players to efficiently assess the reliability of systems, services and remote parties. Secondly, it aims at allowing parties and system owners to increase the security and reliability of their systems, and correctly represent and communicate this to other players.

There can of course be many different approaches to this. Some approaches are already well established under separate names, whereas other approaches have taken shape together with the emergence of trust management as a discipline in its own right. In order to avoid misunderstanding it is always good to be as specific as possible when using the term trust management, by providing additional information about its meaning in a given context. For example, if someone in the access control research community uses the term trust management without any further explanation when talking to a web marketing specialist, it is very likely to be misunderstood. This sort of confusion can easily be avoided by being more specific.

A current research project that combines multiple approaches to trust management is the European TrustCoM[8] initiative [11] that focuses on developing technologies that will enable multiple companies to integrate services, processes and resources, for example to provide integrated services through ad hoc virtual organisations, while at the same time allowing them to protect their respective assets. This builds on a framework for establishing trust between parties based on electronic contracts, policies and reputation systems, as well as a federated security architecture for authentication and access control. This can be seen as a very pragmatic approach to trust management.

A high level of general trust in a community is an important factor for the prosperity and economic fitness of that community. This applies to both traditional geographically bound communities as well as to online and virtual communities. Unfortunately, stakeholders in the development of online services and systems find it hard to ensure that systems are secure and that participants always behave in good faith. Finding ways to make the online environment more secure and dependable, and to increase the transparency regarding the participants' honesty and reliability, is crucial for attracting people and organisations to create and use online applications.

By leveraging the tremendous efficiency with which information can be disseminated and automatically analysed through computer systems and networks, trust management certainly has the potential of making the online environment an even safer place to interact and transact than the traditional environment currently is. Increased social capital in online communities will be the result of a successful outcome of this endeavour.

[8] http://www.eu-trustcom.com/

## References

1. M.D. Abrams. Trusted System Concepts. *Computers and Security*, 14(1):45–56, 1995.
2. Robert Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.
3. J. Berg, J. Dickhaut, and K. McCabe. Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10:122–142, 1996.
4. Matt Blaze, Joan Feigenbaum, and Angelos D. Keromytis. KeyNote: Trust Management for Public-Key Iinfrastructures. In *Proceedings of the 1998 Secure Protocols International Workshop*, Cambridge, England, 1998.
5. Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized trust management. In *Proceedings of the 1996 IEEE Conference on Security and Privacy*, Oakland, CA, 1996.
6. Matt Blaze, Joan Feigenbaum, and Martin Strauss. Compliance Checking in the PolicyMaker Trust Management System. In *Proceedings of Financial Crypto*, 1998.
7. Michael Burrows, Marín Abadi, and Roger Needham. A logic of authentication. *ACM Transactions on Computer Systems*, 8(1):18–36, 1990.
8. Cheskin Research & Studio Archetype/Sapient. *eCommerce Trust Study*. Sapient, http://www.sapient.com/cheskin/, January 1999.
9. Y.-H. Chu, J. Feigenbaum, B. LaMacchia, P. Resnick, and M. Strauss. REFEREE: Trust Management for Web Applications. *World Wide Web Journal*, 2:127–139, 1997.
10. J. S. Coleman. Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94:95–120, 1988.
11. T. Dimitrakos. Towards a Trust and Contract Management Framework for Dynamic Virtual Organisations. In *Proceedings of eAdoptions and the Knowledge Economy: eChallenges 2004*. IOS Press, November 2004.
12. R. Falcone and C. Castelfranchi. How trust enhances and spread trust. In *Proceedings of the 4th Int. Workshop on Deception Fraud and Trust in Agent Societies, in the 5th International Conference on Autonomous Agents (AGENTS'01)*, May 2001.
13. R. Falcone and C. Castelfranchi. *Social Trust: A Cognitive Approach*, pages 55–99. Kluwer, 2001.
14. Francis Fukuyama. *Trust: The Social Virtues and the Creation of Prosperity*. The Free Press, New York, 1995.
15. D. Gambetta. Can We Trust Trust? In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–238. Basil Blackwell. Oxford, 1990.
16. T. Grandison and M. Sloman. Specifying and Analysing Trust for Internet Applications. In *Proceedings of the 2nd IFIP Conference on e-Commerce, e-Business and e-Government (I3E2002)*, Lisbon, 2002.
17. H. Huang, C. Keser, J. Leland, and J. Shachat. Trust, the Internet, and the Digital Divide. *IBM Systems Journal*, 42(3):507–518, 2003.
18. B.A. Huberman and F. Wu. The Dynamics of Reputations. *Computing in Economics and Finance*, 18, 2003.
19. Sirkka L. Järvenpää and D. E. Leidner. Communication and trust in global virtual teams. *Organization Science*, 10(6):791–815, 1999.
20. A. Jøsang. The right type of trust for distributed systems. In C. Meadows, editor, *Proc. of the 1996 New Security Paradigms Workshop*. ACM, 1996.
21. A. Jøsang, R. Ismail, and C. Boyd. A Survey of Trust and Reputation Systems for Online Service Provision (to appear). *Decision Support Systems*, 2005.
22. A. Jøsang and S. Pope. Semantic Constraints for Trust Tansitivity. In *Proceedings of the Asia-Pacific Conference of Conceptual Modelling (APCCM)*, Newcastle, Australia, February 2005.

23. A. Jøsang and N. Tran. Trust Management for e-Commerce. In *VirtualBanking2000*. Virtual Conference hosted at http://virtualbanking2000.com, 2000.
24. C. Keser. Strategically Planned Behavior in Public Goods Experiments. Technical report, CIRANO Scientific Series 2000s-35, Montreal, Canada, 2000.
25. C. Keser. Experimental Games for the Design of Reputation Management Systems. *IBM Systems Journal*, 42(3):498–506, 2003.
26. C. Keser and F. van Winden. Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics*, 102:23–39, 2000.
27. S. Knack and P. Keefer. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112(4):1251–1288, 1997.
28. D. Kreps and R. Wilson. Reputation and Imperfect Information. *Journal of Economic Theory*, 27(2):253–279, 1982.
29. N. Luhmann. *Trust and Power*. Wiley, Chichester, 1979.
30. 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, URL: http://misrc.umn.edu/wpaper/, 1996.
31. R.M Needham. *Security and open source*. Presentation at the Open Source Software Workshop, 2002. http://nats-www.informatik.uni-hamburg.de/view/OSS2004/PaperCollection.
32. 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.
33. M.A. Patton and A. Jøsang. Technologies for Trust in Electronic Commerce. *Electronic Commerce Research*, 4(1-2):9–21, 2004.
34. R Pichler. *Trust and Reliance - Enforcement and Compliance: Enhancing Consumer Confidence in the Electronic Marketplace*. Stanford Law School, www.oecd.org/dsti/sti/it/secur/act/online_trust/Consumer_Confidence.pdf, May 2000.
35. A. Powell, G. Piccoli, and B. Ives. Virtual teams: a review of current literature and directions for future research. *SIGMIS Database*, 35(1):6–36, 2004.
36. R.D. Putnam. Tuning in, tuning out: The strange disappearance of social capital in America. *Political Science and Politics*, 28(4):664–683, 1995.
37. 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.
38. P. Resnick, R. Zeckhauser, R. Friedman, and K. Kuwabara. Reputation Systems. *Communications of the ACM*, 43(12):45–48, December 2000.
39. N.C. Rowe. Designing Good Deceptions in Defense of Information Systems. In *Proceedings of the Annual Computer Security Applications Conferencen (ACSAC)*, Tucson, December 2004.
40. 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.
41. R. Selten, M. Mitzkewitz, and G. Uhlich. Duopoly Strategies Programmed by Experienced Player. *Econometrica*, 65:517–555, 1997.
42. C. Shapiro. Consumer Information, Product Quality, and Seller Reputation. *The Bell Journal of Economics*, 13(1):20–35, 1982.
43. G.J. Simmons and C. Meadows. The role of trust in information integrity protocols. *Journal of Computer Security*, 3(1):71–84, 1995.
44. O.E. Williamson. Calculativeness, Trust and Economic Organization. *Journal of Law and Economics*, 36:453–486, April 1993.
45. P.R. Zimmermann. *The Official PGP User's Guide*. MIT Press, 1995.