Trustworthiness as an Asset

(Extended abstract)

Pekka Nikander

pekka.nikander@hiit.fi

Abstract: People posses an innate capability to instinctively evaluate the trustworthiness of other people in their daily lives, and a hardwired tendency to react strongly if they feel betrayed. However, these capabilities do not work in the digital domain, for both obvious and not-so-obvious reasons. Therefore, something else is needed. In this paper we explore the prequisites and difficulties that are encountered when attempting to establish a bases for trust and trustworthy behaviour in the Internet, concentrating on the asset like nature of trustworthiness.

1 Introduction

A couple of years ago [NK01] we argued that there is a need to express trust relationships in a digital form. As an example, we showed how to represent human trust relationships with authorization certificates. We envisioned that explicit large scale expression of trust relationships would lead to a situation where people would have an incentive to engage in more trustworthy behaviour than they otherwise would. While we still support these hypothesis, we have learned quite a lot since then, especially about how complex the phenomenon called trust is, after all.

Trust seems to be built into humans in a fairly deep level. Some evolution biologist, like Steven Pinker [Pin99], argue that parts of the behaviour patterns around trust are explicitly wired to bypass rational thinking. While this may look odd at the first sight, it makes sense from the evolution's point of view. For instance, the threat of quick retaliation seems to be crucial for trust formation. Since people are more likely to survive in a society where most people behave in a trustworthy manner, it makes sense to hardwire humans to enrage as a response to betrayal. That is, if everybody believes that rage upon betrayal is hardwired, betrayal becomes less appealing even in a situation where such rage would not make sense from the rational point of view.

Properties like that make trust, as a human phenomenon, especially hard to understand and quantify. In a way, we must go beyond the usual rational bases of argument, and really understand trust as an instinct like innate phenomenon. This poses a difficulty for the digital expression of trust, too, since we must carefully understand the rational, economic, and biologically hardwired dimensions of trust.

In this paper we explore some aspects of trust, and make a stronger case for the idea that

explicitly expressing trust related information in the digital world might indeed be beneficial for both the individuals and the society. The rest of this paper is organized as follows. Firstly, in Section 2 we give one definition for trust, taking aspects from evolution biology, sociology, economics, and computer security. Next, in Section 3, we concentrate on the economics point of view, presenting an economic model which illustrates our thinking so far. In Section 4, we discuss issues related to representing trust in the digital form, based on the economic model presented in Section 3.

2   Trust defined

In general, trust is understood to designate a state of mind where a trusting party lacks some knowledge and yet willingly takes the risk of a trustee harming him or her, with the expectation that the trustee will not utilize the power after all [RSBC98][FJH00]. This behaviour is largely based on needs and emotions, and less on rational thoughts. In evolutional sense, the possibility of retaliation upon betrayal seems to be always present when we deal with trust. There is always a tension: in a high trust society, a person that decides to betray wheneven beneficial to him or her gains material benefits. In the terms of evolution, material benefits mean more offspring; thus, if untrustworthy behaviour is not punished, it will proliferate. On the other hand, it looks like that trustworthy behaviour in general is or at least has been beneficial to the species, leading to a situation where many people have an innate need to behave in a trustworthy manner even when the trusting party does not have a real possibility to retaliate Additionally, evolution has formed other behaviour patterns that help to maintain the balance, rage and gossip being the maybe most important ones.

Our evolutionally developed ability to differentiate trustworthy behaviour from untrustworthy one does not apply to digitally conveyed communciations. We do not necessarily perceive new people behind the network as our tribal peers but as strangers. Rage doesn't pose a real threat of physical injury. Thus, our ability to gossip and otherwise propagate our experiences becomes more important than ever before [NK01].

From the economic point of view, one of the major consequences of trust is that the average transaction costs are lowered. On a higher level, it has been convincingly shown that there exists a positive correlation between the generic level of trust within a society and the economic prosperity of the society [Fuk96]. In general, people are ready to pay a fairly high fee to punish others that they perceive having acted in an untrustworthy way, even in a case where they have not personally been hurt. Keeping in mind our earlier discussion about the evolutional bases of trust, this makes sense. Punishing untrustworthy behaviour early helps to weed out freeloaders.

In the computer security literature, the term trust appears to designate two fairly different propositions. Firstly, in most of the literature and especially when talking about Trusted Computing Base (TCB) or otherwise trusted components, such as a Trusted Third Party, there is often an implicit assumption that the trusted parties are those that necessarily must be trusted. On the other hand, there are a number of studies that concentrate on the conditions under which a party can be trusted. That is, the trusting party decides to trust

the trusted party, i.e., decided to take a risk of being harmed, under the assumption that it will not be harmed after all. In this paper we call these two forms of trust as necessary trust and voluntary trust.

3   Trust in Economics

Our goal in this paper is to argue that establishing a method for explicitly expressing trust in a digital form is beneficial for both individuals and the society at large. To work towards that goal, we first present an economic metamodel for trust formation, instilling much of the discussion above. After that, we turn our attention to the asset-like nature of trustworthiness.
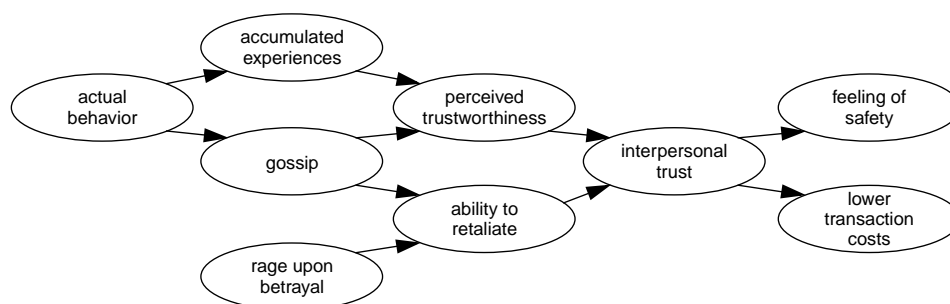

Figure 1: A model of trust formation

Fig. 1 summarises our insight on trust formation and its consequences. Reading it from right to left, the higher the overall interpersonal trust level within a society, the safer people feel and the lower the transactions costs are on the average [Fuk96]. Given any single situation, on the other hand, the level of interpersonal trust depends on the perceived trustworthiness of the parties and their perceived ability to retaliate if they are betrayed after all. For example, the threat of gaining a bad reputation acts as a fairly strong incentive for trustworthy behaviour, discouraging treason.

It looks like perceived trustworthiness is mainly based on two sources of information. First hand past experiences about the other party are very strong. The other source of information seems to be indirect, gossip in our model. And gossip is what people naturally do, they exchange information that allows them to enhance their understanding of the fellow people.

The threat of rage seems to play a diminishing role in our contemporary society. Seldom if ever we are betrayed in a face-to-face situation. (This can be also taken as an indication of the strong evolutionary powers managing our face-to-face behaviour.) In a way, outdirect rage has been mostly replaced by spreading gossips and other forms of indirect revenge.

3.1 Trustworthiness as an Asset

We now turn our attention towards the economic incentives that encourage trustworthy behaviour. As we have seen, even though people seem to have a natural tendency to behave in a more-or-less trustworthy way with people that they consider their tribal peer group, that behaviour does not necessarily extend to other people and other situations. In particular, there are many cultures where it is completely acceptable to try to gain as much advantage as possible when dealing with strangers [Fuk96].

We look at three aspects of trustworthiness here: plain monetary credibility, the implications for sellers of goods, and individual's position as an opinion former.

Credibility. When we apply for a loan at a bank or other financial institution, our credibility is assessed. The assessment does not only concern our ability to pay back the loan and the associated risk, but it also attempts to assess our tendency to willingly default or otherwise avoid our liabilities. The assessment results set both the interest premium we pay and the maximum amount of money the bank is willing to loan. Thus, more trustworthy people are able to borrow more money at a lower price, allowing them to take the inflation advantage by being close to the source of money formation [vM80].

Acting as a seller. Today, almost all commercial transactions contain an information asymmetry in one form or another. Usually the seller has more information about the goods he or she is selling and is able to better asses their real value than the buyer. Consequently, he or she may try to take advantage of the asymmetry, and sell bad goods for the price of good ones. To counteract this, a rational buyer needs to calculate a risk premium, refusing to pay the otherwise full price for a perfect looking good. Therefore, this information asymmetry may lead to a situation where no transactions are possible at all.

One way to resolve to situation is to convince the buyer about the seller's trustworthiness. In many cases such a practice is beneficial to both parties. The seller is able to get a higher price than what he would otherwise be getting, and the buyer does not need to fear getting lemons. Thus, trustworthiness works as a capital-like instrument, allowing the seller to extract higher price on the goods that he or she is selling.

Acting as a opinion former. When dealing with gossip, people have to evaluate the value of information they are receiving. That is, they need to assess whether the source of information is likely to be trustworthy or not, and form beliefs accordingly. In general, information received from a more trustworthy person is regarded more valuable than the same information received from a less trustworthy person. Trustworthy behaviour leads to more trust and higher reputation, leading to a positive value formation cycle.

4 Codifying Trust and Experience

So far we have argued that since rage and the threat of retaliation do not matter so much in the digital world as they do in the real world, gossip becomes more valuable than ever. However, since the value of gossip is based on the trustworthiness of its source, the problem becomes recursive. That is, in order to evaluate the trustworthiness of a piece of information about a person it is necessary to evaluate the trustworthiness of the source of

the information, i.e., trustworthiness of another person or a chain of persons.

In this section we briefly explore the possibility of codifying trust relationships and our perceptions of other people's trustworthiness. By making our personal evaluations known to other people we at the same time increase the amount of information available for trust evaluation and potentially increase our own trustworthiness as an opinion former.

Authorization certificates. Trust Management [BFL96] is a term usually used to denote decentralized access control conveyed with authorization certificates. Authorization certificates, in turn, are usually understood as a class of certificates that deal directly with public keys and access rights, with no human readable names. The basic idea is to use a certificate to denote authorization or delegation: an (alleged) right owner authorizes another party to perform an operation, i.e., to exercise the right. In the certificate, both parties are denoted by their public keys.

In a way, authorization certificates can be considered to be expressions of trust. The issuer of a certificate states that it trusts the subject with regard to the delegated rights and validity constraints. In other words, the issues believes that the subject is trustworthy enough to be allowed to exercise the right.

As we initially argued in [LN98], authorization certificates can be considered to express trust in general sense, not just in relation to certain rights or privileges. All depends on the context. The certificates make sense only in context, where the context defines how to interpret the public keys denoting principals (issuer and subject), and how to interpret the rights and validity constraints. In practical terms, if an evaluating party has no information about a certificate's issuer, it may as well consider the certificate nonexistent, since it is possible that the subject has just generated a key pair and signed the certificate itself.

Now, while the basic scheme is easy enough, trying to use authorization certificates for expressing generic trust relationships and opinions about trustworthiness brings forth a number of hard problems.

Differentiating reasons for trust. Even though trust may have been a more or less binary phenomenon in the original tribal human societies, it is definitely not that today. That is, in the tribal life people probably perceived other people either as trustworthy or not, and did not consider the possibility that someone might be more trustworthy in one sense and less in some other sense. This allowed people to share information easily within their tribal group.

Whenever we trust somebody in the modern society, and want to make that knowledge available, we have to be very careful in defining why and how we trust that person or other entity. Is the trust based solely on our own voluntary discretion, based on past experiences with that person, or do we rely on recommendations by others, what are our perceived chances for retaliation, etc. More importantly, we have to define the domain of activity and the peer group within which we express the trust.

The context in which trust is expressed defines the exact nature and value of such an expression. They cannot be differentiated. Furthermore, the context must be considered in each step when evaluating a chain of expressions, i.e., a chain of certificates.

Changing identities. In an all digital system, the possibility of easily creating new identities and completely dismissing old ones creates a problem that doesn't appear in real life. That is, it makes sense to try to build a high reputation, only to use this high reputation to

make a big juicy fraud and disappear. The reason for this possibility is the missing link between the digital identity and the real person.

Liability. Given the complexities of the modern society, we cannot and should not ignore the issue of liability. Only by backing the expression with the legal system can we ever hope to create a system that scales large enough. The current digital signature laws in various countries may be considered as an initial step to the right direction. However, they alone are not enough. For any automatic or semi-automatic trust evaluation system to work, people must be made liable for the effects of any misleading information they knowingly feed to the system.

Complexities combined. Given the considerations above, we start to realize the complexity of the situation. Whenever authorization certificates are used outside a well defined, simple domain, the interplay of psychological, sociological, emotional, technical, and legal issues begins. For any wide scale system to succeed, all of these aspects must be understood and balanced. That is not an easy task, and it will definitely take years and lots of experimenting before such a balance is found.

Acknowledgements

References

[BFL96]   M. Blaze, J. Feigenbaum, and J. Lacy. Decentralized Trust Management. In *IEEE Symposium on Security and Privacy*, May 1996.

[FJH00]   B. Friedman, P. H. Kahn Jr., and D. C. Howe. Trust Online. *Communications of the ACM*, 43(12):34–40, December 2000.

[Fuk96]   Francis Fukuyama. *Trust - The Social Virtues and The Creation of Prosperity*. Free Press Paperbacks, NY, 1996.

[LN98]   I. Lehti and P. Nikander. Certifying Trust. In *Practice and Theory in Public Key Cryptography (PKC) '98*, February 1998.

[NK01]   Pekka Nikander and Kristiina Karvonen. Users and Trust in Cyberspace. In Bruce Christianson, James Malcolm, Bruno Crispo, and Michael Roe, editors, *Security Protocols, 8th International Workshop*, number 2133 in LNCS, pages 24–35. Springer, 2001.

[Pin99]   Steven Pinker. *How the Mind Works*. Penguin Books, 1999.

[RSBC98] D.M. Rousseau, S.B. Sitkin, R.S. Burt, and C. Camerer. Not So Different After All: A Cross-discipline View of Trust. *Academy of Management Review*, 23(4):393–404, July 1998.

[vM80]   Ludwig von Mises. *The Theory of Money and Credit*. Liberty Fund, Inc., 1980.