Depending on the Kindness of Strangers?
Trust Relationships in Ambient Societies

Jens Riegelsberger, M. Angela Sasse, John D. McCarthy
Department of Computer Science, University College London
Gower Street, London WC1E 6BT, United Kingdom
{jriegels, a.sasse, j.mccarthy}@cs.ucl.ac.uk

ABSTRACT

The vision of the ambient society relies on the constant exchange of personal information among multiple devices, individuals and organisations. The high number of micro-trust decisions required in such a scenario calls for automated trust management. In this paper we discuss a set of contextual properties of transactions that ensure trust and trustworthy action in everyday situation and suggest how they can be incorporated in trust management for ambient technologies. We identify *institutions*, *repeated interactions* and *reputation* as contextual properties that support cooperation. We then discuss the limitations and risks of assuring cooperation based on contextual properties alone. Firstly, a subjective assessment of personal properties (e.g. benevolence and integrity) also forms an important basis for trust in others. Furthermore trust based on contextual properties is hard to establish in the case of vague outcomes, and multi-dimensional risks. Finally, establishing one's own trustworthiness requires giving access to personal information. Ambient technologies must also allow individuals to remain untrusted but private.

Keywords Ambient Society, Trust, Privacy, Personal Information, Reputation

INTRODUCTION

Imagine you are in a hotel room in a foreign city working to finish an important report when your client rings and tells you are invited to his CEO's 50th birthday party. This is a great honor and opportunity to build a strong relationship with your client but you realize you don't have anything appropriate to wear and you don't have a gift to bring along. You remember that fortunately your PDA can check local retailers' stock for garments that fit the occasion and that would go well with your wardrobe at home, taking into account your preferred brands and materials and transmitting the measurements that had been taken by your hotel bathroom that morning. Your PDA checks your schedule, organizes a bidding process among potential suppliers and books an appointment with a local fashion consultant who will bring a selection of outfits. At the same time it checks your host's music and literature collection to suggest appropriate presents, but not before it has ensured that no other guest is getting the same. When you leave the hotel the present awaits you at the hotel reception – the local branch of your trusted e-tailer has delivered it while you were busy with the fashion consultant in your hotel room, and billed your credit card directly.

Such scenarios are commonly presented when the vision of ambient technologies is discussed. This vision builds on three core technological advances: embedding intelligent devices in everyday objects (ubiquitous computing), wireless networking and communications, and natural interfaces (e.g. speech or gesture input). To realize the promises of these technologies, personal information that previously was not captured, disseminated or collated has to be constantly exchanged. In ambient societies there will be no 'practical obscurity'. There are no technological hurdles that make it too costly to collate the fragmented traces of an individual's actions. Fundamentally, the vision of ambient societies relies on trading access to personal information for convenience.

To be accepted ambient technologies need to allow users to retain control over how their personal information is used. However, the sensitivity of personal information cannot be classified *a priori* as it depends on the information receiver, context, costs and benefits. These factors and others have to be assessed for every transaction [1]. You will be more willing to give access to you measurements for increased speed and convenience when shopping under time pressure for an important event. Decisions concerning whether to release personal information and with what degree of identifiability also strongly depend on the trust we have in the information receiver [1]. Most would trust their tailor with their measurements and information about their past purchases, but how about an unknown retailer in a foreign city? The constant interactions among ambient systems would require users to make a myriad of such micro-trust decisions, rendering this approach unusable. Thus, the full potential of ambient technologies can only be realized if these trade-off decisions can be delegated to an intelligent trust management system. Agents in ambient societies need to be trusted to reach correct trust assessments on our behalf.

In this paper we describe several ways in which trust and trustworthy action is assured in everyday transactions. After identifying these social techniques, we suggest ways in which they could be incorporated into trust management for ambient technologies. We close by pointing out the limitations and risks of externalizing trust decisions to ambient technologies.

INCENTIVES FOR TRUSTWORTHY ACTION

In most situations that require trust we can identify two actors: (1) a *trustor* who acts first by taking a risk (e.g. giving information about past purchases), and (2) a *trustee* who can then behave as promised (e.g. suggest garments that fit your wardrobe) or defect (e.g. sell the contact details to third party advertisers or use the knowledge to extract maximum prices). In the broadest possible terms, defection and fulfillment will be determined by the trustee's *ability* to act as promised and – more interestingly – by his *motivation* [6]. In our scenario this translates to the question whether the fashion consultant has any knowledge of local fashion and if yes why he should give you honest advice, rather than sell you the item with the highest margin.

With regard to motivation, a prototypical trust situation is described by the economic Trust Game [3]. The trustor knows that, by trusting she gives the trustee an incentive not to fulfill. However, in most real-world situations with a Trust Game structure we observe trusting actions and fulfillment in spite of situational incentives to the contrary: Vendors deliver goods after receiving payment, banks return money, individuals do not sell their friends' phone numbers to direct marketers. This is because in many cases, trustees' actions will be governed by contextual and personal properties beyond situational incentives. Identifying and implementing the contextual properties that lead trustees to act in a trustworthy manner is a core concern for the designers of interactive systems.

Contextual Properties

Contextual properties restructure the incentives in such a way that cooperation carries a higher utility for the trustee in the face of situational incentives to defect. Your fashion consultant could steal from your hotel room, but fear of legal prosecution is likely to prevent him from doing that. Contextual properties, such as legal frameworks, are present in most everyday encounters, so we often fail to perceive their effect. In many cases we have learned to trust habitually: as long as indicators of 'situational normality' are present [12] we can be confident in an "expectation of continuity" [11]. We do not consider the trustee's incentives and the likelihood of defection. However, in unfamiliar contexts or if technology transforms the way in which we conduct transactions, as is the case with e-commerce or ambient technologies, we have no template of situational normality. Trust has to be won and assured before we can trust habitually [14]. Thus, in order to support trust and trustworthy action in ambient technologies it is important to understand how the different types of contextual properties that support trust and trustworthy action in everyday life can be implemented. We divide contextual properties into three types: institutional, temporal and social embeddedness [13].

Institutional Embeddedness

Examples of institutions that enforce trustworthy action on the part of the trustee are legal structures and crime prosecution agencies. In most societies they provide strong disincentives for theft and thus allow us to walk the streets unarmed [11]. However, such institutional punishment schemes rely on clear definitions of defection and cooperation and they need to be inexpensive compared to the risks involved in order to be viable [13]. This contextual property will probably prevent your fashion consultant from stealing from you, but what about trading your measurements or purchase history? As it is hard to quantify the cost of small infringements of privacy, relying on institutional punishment to ensure trustworthy handling of personal information in ambient societies is impractical. Not only will it be hard to pre-define appropriate uses, but also the costs of individual infringements will be too small compared to investigation and punishment.

However, institutions *can* play a role within companies by regulating the behavior of their employees. Where the trustee is a company, incentive structures (e.g. in the form of job roles) regulate the behavior of their representatives. One benefit of assuring trustworthy action in this way is that it enables potentially risky interactions with individuals about whose personal properties little is known. Trust is vested in the role and the institution that assures the appropriate behavior of its representatives [9, 10]. Your PDA can ensure that only fashion consultants that are associated with your trusted vendor at home will be considered. The problem of trust is hereby transferred from the individual level to the job role and organisation. The core task of the ambient technologies is now to unequivocally determine the consultants' identity and link with the vendor.

Sociologists observe that trust in everyday transactions is increasingly based on institutional trust and roles (e.g. franchised corner-shops) rather than in personal trust with a specific individual (e.g. in a personally known local shop keeper). This process has been termed *dis-embedding* [8]. It vastly increases the efficiency of everyday trust decisions, as we do not have to rely on observing an individual's behavior over time, but can aggregate our trust in a corporation or a brand. In the context of ambient societies, such an approach would encourage the formation of trusted networks of service providers that can be relied upon because they are known to enforce rules of trustworthy action. In our scenario, the hotel could perform such a role: Given that it is already a trusted entity as you spend a night there and as they hold your financial information, they can act as a trust aggregator for the local services of fashion or gift retailers.

Temporal Embeddedness

Temporal embeddedness refers to the expectation of future encounters. If both actors know that subsequent encounters are likely, fulfilling becomes a dominant strategy for the trustee, as he knows that defection would lead the trustee to withdraw from future interactions [2]. When looking at interactions between individuals and organisations, the organisations' interest in future business is a strong indicator of trustworthiness. In traditional retail this property can be signaled through investments in the brand, advertising, buildings or an existing relationship. This factor favors big players with many local franchises, as future interaction can be expected at many different locations. If your hotel is a branch of a global chain they risk future business in other locations by providing lower than expected service. If you stay in an independent hotel in an obscure location, they might not expect you to come back ever – whatever the quality of service they provide. Thus, if ambient technologies can signal your travel history to the hotel this can provide an incentive for them to offer a good service.

In the case of encounters with individuals that do not enact organisational roles, shared membership in small and relatively stable groups is a good indicator for the likelihood of future encounters. It is thus a strong incentive for trustworthy behavior and subsequently a good signal for trust [7]. Shared membership in social groups, such as neighborhoods, clubs, work-groups or alumni organisations could be leveraged by ambient technologies to assess the likelihood of future encounters. Thus, your PDA, when accessing information about other guests' presents can signal that you are a fellow guest.

Social Embeddedness

In addition to allowing an inference about the likelihood of future encounters, shared group membership also allows for the formation of reputations. Reputation can ensure trustworthy behavior, even if future interaction with a given actor is not expected [7]. Since only an actor with a reputation for acting trustworthy will be trusted, it becomes rational for selfish actors to acquire a good reputation by foregoing situational temptations. Reputation can thus act as a 'hostage' in the hands of a trustor [13]. A socially well-embedded trustor can threaten to tarnish the trustee's reputation if his fulfillment is below expectations. Thus, systems that support reputation formation can encourage trustworthy action in a society of rational self-interested actors. This is not so much because reputation provides detail on personal properties (such as benevolence), but because it is in the interest of the trustee to maintain this asset for future encounters.

In the context of the "birthday party" example, the user's PDA can pick vendors based on their reputation as supplied by a general reputation aggregation services (e.g. one comparable to Epinions (www.epinions.com)), and based on its reputation in the user's personal network. Clearly, this approach favors socially well-connected trustors as they can pose a more credible threat to a trustee's reputation. When interacting with individuals (e.g. checking other guests' planned presents or giving access to one's own media collection), reputation and level of embeddedness in a shared social network are the main contextual properties that encourage trustworthy action.

LIMITATIONS AND RISKS

Personal Properties

While contextual properties can ensure trustworthy action among rational self-interested actors, they do not fully explain how we reach trust decisions in everyday situations [15]. Their advantage is that they can be implemented in automatic trust management systems. However, our trust is commonly based on an assessment of the personal properties of individuals we are about to interact with. We have evolved to read signifiers of others' integrity or benevolence [15] from interpersonal cues (such as facial expression, voice, gestures) with little effort [5].

The very reason why you are invited to the dinner party is that your client can assess your personal trust-warranting properties. By exchanging small talk, by observing your behavior among others, your host can make many inferences about your integrity or benevolence. It is hard to imagine that automated trust management in ambient societies can fully replace this subjective trust assessment.

In the view of many researchers, by relying on systems that ensure trustworthy action based on purely contextual properties, we run danger of replacing personal trust with what they call *reliance* [10]. It is also well established that a higher social presence and thus higher visibility of personal properties on the part of interaction partners can call into action social norms and thus lead to higher rates of cooperation [4]. Giddens [8] holds that even in situations when cooperation is assured by contextual properties, we will vest some of our trust in the personal properties of organisational representatives. He refers to this phenomenon as *re-embedding*.

Thus in our view, ambient societies cannot rely exclusively on delegating trust decisions to technical devices that 'calculate' trust based on contextual properties. Rather they should also help us to manage our subjective trust assessment based on personal properties. At the birthday reception, ambient technologies could help us to track who we spoke and then to review this information and make sure we incorporate individuals in our social network that we considered trustworthy, friendly or interesting. Similarly, ambient technologies could combine contextual assurance and subjective assessment by only short-listing fashion consultants that have established credentials based on contextual properties, and then allowing you to make your final choice based on a short pre-recorded video or even a quick video conference chat.

Complexity of Trust

While the contextual properties introduced above suggest a clear categorization and thus easy automation of micro-trust decisions, it should be kept in mind that trust is a very complex phenomenon. While economic models commonly refer to defection and cooperation in binary terms and isolate specific risks (e.g. monetary loss), in real world transactions we often observe gradual forms of defection and multi-dimensional risks [15]. Your fashion consultant could rob you, give you poor advice, overcharge you, or sell on your personal information.

Privacy and Trustworthiness

Establishing one's own trustworthiness often requires giving access to personal information: Reputation relies on a record of past actions; identity has to be bound to actions to allow a credible threat of punishment. However, we enact different aspects of our personalities in different social networks. This *impression management* is an important factor for organizing our everyday lives (and keeping our sanity) [9]. Therefore any system that aims to support trust-decision among individuals needs to facilitate it. A single identity with an global reputation rating is not a feasible solution. Rather individuals need to be allowed to decide how much personal information they want to reveal to earn a certain amount of trust.

If – after some tiresome hours of professional socializing at the CEO's party – you want to wind down in a low-key bar, you might want to make none of your credentials of social embeddedness available to your surroundings, just chat to random punters at the bar without knowing anything about them and pay your drinks with cash, not leaving a trace.

CONCLUSIONS

In this paper we discussed how trust is typically based on a combination of contextual and personal properties. Trust decisions based on contextual properties can be delegated to ambient systems, e.g. in the form of reputation tracking, certification and corporate associations, or calculating social embeddedness. However, human trust is also vested in personal properties such as integrity or benevolence that we can quickly assess in the form of interpersonal cues. Ambient technologies cannot replace this trust-assessment, but should help us in managing it. Another problem with automating trust decisions is that most transactions have multi-dimensional risks and gradual fulfillment. The cost of explicating these factors will often be high relative to the risks involved in a single transaction. Finally, most approaches to establishing trustworthiness rely on giving access to personal information (e.g. behavior in previous transactions). In the context of ambient societies, users must be given the choice of how much information they want to reveal when establishing their trustworthiness. In many cases they might prefer to be untrusted but private.

ACKNOWLEDGEMENTS

We gratefully acknowledge the help of Richard Boardman (Imperial College London). Jens Riegelsberger is funded by a British Telecom studentship (WEB164414/CT501045).

REFERENCES

1. Adams, A. and Sasse, M. A. Privacy in Multimedia Communications: Protecting Users, Not Just Data. 49-64. 2001. Lille. *Proceedings of HCI2001*. 2001.
2. Axelrod, R., More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, vol. 24, no. 3, pp. 379-403, 1980.
3. Berg, J., Dickhaut, J., and McKabe, K., Trust, Reciprocity, and Social History, *Games and Economic Behaviour* , vol. 10 pp. 122-142, 2003.
4. Bos, N., Olson, J. S., Olson, G. M., Wright, Z., and Gergle, D. Rich media helps trust development. CHI2002 Conference Proceedings. 2002.
5. Cosmides, L. and Tooby, J., *The adapted mind: evolutionary psychology and the generation of culture* Oxford: Oxford University Press, 1992.
6. Deutsch, M., Trust and suspicion, *Journal of Conflict Resolution*, vol. 2, no. 3, pp. 265-279, 1958.
7. Fehr, E. and Fischbacher, U., The Nature of Human Altruism, *Nature*, no. 425, pp. 785-791, 2003.
8. Giddens, A., *The consequences of modernity* Stanford: Stanford University Press, 1990.
9. Goffman, E., *The Presentation of Self in Everyday Life.* Garden City: Doubleday, 1959.
10. Lahno, B., Institutional Trust: A Less Demanding Form of Trust? *Revista Latinoamericana de Estudios Avanzados (RELEA)*, 2002.
11. Luhmann, N., *Trust and Power* Cichester: Wiley, 1979.
12. McKnight, D. H. & Chervany, N. L., What is Trust? A Conceptual Analysis and An Interdisciplinary Model. 827-833. Proceedings of AIS 2000. 2000.
13. Raub, W. and Weesie, J. *The Management of Durable Relations*. 2000. Amsterdam: Thela Thesis.
14. Riegelsberger, J. & Sasse, M. A., Designing E-Commerce Applications for Consumer Trust, in Petrovic, O., Ksela, M., and Fallenboeck, M. (eds.) *Trust in the Network Economy* Wien: Springer, 2003, pp. 97-110.
15. Riegelsberger, J., Sasse, M. A., and McCarthy, J., The Researcher's Dilemma: Evaluating Trust in Computer-Mediated Communication, *International Journal of Human Computer Studies*, vol. 58, no. 6, pp. 759-781, 2003.