FIDIS

Future of Identity in the Information Society

Title:	"D17.4: Trust and Identification in the Light of Virtual Persons"
Author:	WP17
Editors:	David-Olivier Jaquet-Chiffelle (VIP, Switzerland)
	Hans Buitelaar (TILT, Netherlands)
Reviewers:	James Backhouse (LSE, England)
	Sandra Steinbrecher (TU Dresden, Germany)
	Jozef Vyskoc (VaF, Slovakia)
Identifier:	D17.4
Type:	[Deliverable]
Version:	1.2
Date:	Thursday, 25 June 2009
Status:	[Final Version]
Class:	[Public]
File:	fidis_D17.4_v1.2.doc

Summary

The purpose of this deliverable is to pinpoint and explicit fundamental difficulties related to identification and trust in the digital world. The idea is to illuminate and analyze them from the perspective of the "virtual person" model, which has been a major research topic in previous FIDIS deliverables.

First the core concepts of trust, confidence and identification are defined and analyzed in traditional contexts. Then, we discuss how these concepts can be (re-)interpreted in the digital world and, more generally, in the light of virtual persons in the virtual world.

Copyright © 2004-09 by the FIDIS consortium - EC Contract No. 507512
The FIDIS NoE receives research funding from the Community's Sixth Framework Program

Copyright Notice:

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the editors and authors of the document. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

PLEASE NOTE: This document may change without notice – Updated versions of this document can be found at the FIDIS NoE website at www.FIDIS.net.

Members of the FIDIS consortium

1. Goethe University Frankfurt	Germany
2. Joint Research Centre (JRC)	Spain
3. Vrije Universiteit Brussel	Belgium
4. Unabhängiges Landeszentrum für Datenschutz (ICPP)	Germany
5. Institut Européen D'Administration Des Affaires (INSEAD)	France
6. University of Reading	United Kingdom
7. Katholieke Universiteit Leuven	Belgium
8. ***Tilburg University***[1]	Netherlands
9. Karlstads University	Sweden
10. Technische Universität Berlin	Germany
11. Technische Universität Dresden	Germany
12. Albert-Ludwig-University Freiburg	Germany
13. Masarykova universita v Brne (MU)	Czech Republic
14. VaF Bratislava	Slovakia
15. London School of Economics and Political Science (LSE)	United Kingdom
16. Budapest University of Technology and Economics (ISTRI)	Hungary
17. IBM Research GmbH	Switzerland
18. Centre Technique de la Gendarmerie Nationale (CTGN)	France
19. ***Netherlands Forensic Institute (NFI)***[2]	Netherlands
20. ***Virtual Identity and Privacy Research Center (VIP)***[3]	Switzerland
21. Europäisches Microsoft Innovations Center GmbH (EMIC)	Germany
22. Institute of Communication and Computer Systems (ICCS)	Greece
23. AXSionics AG	Switzerland
24. SIRRIX AG Security Technologies	Germany

[1] Legal name: Stichting Katholieke Universiteit Brabant
[2] Legal name: Ministerie Van Justitie
[3] Legal name: Berner Fachhochschule

Versions

Version	*Date*	*Description (Editor)*
0.1	14.01.09	• Initial template release + eBay scenario
0.2	02.02.09	• Further Scenarios added
0.3	17.02.09	• Inclusion of TILT contribution & Identification (Jaquet-Chiffelle, VIP)
0.4	24.02.09	• Inclusion of 5.2 (Buitelaar, TILT), 6.2 (Anrig, VIP), 2.0, 2.1 (Haenni, VIP), 3.4 (Benoist, VIP)
0.5	25.04.09	• Inclusion of 5.0, 5.1 (van der Wees, TILT)
0.6	04.05.09	• §3.0, §3.6, §4.1, §6.1, §7.0, §7.2 added (Haenni, Anrig, VIP)
0.7	9.5.09	• Inclusion of 2.2.2 and 6.1.2 (Benoist, VIP), minor corrections (van der Wees, TILT), consolidation (Jaquet-Chiffelle, VIP)
0.8	17.5.09	• Corrections and editing (Jaquet-Chiffelle, VIP)
0.9	21.5.09	• Executive summary, summary, conclusion, 7.3, corrections and editing in the whole document (Jaquet-Chiffelle, VIP), partial rewriting of 5.2 (Buitelaar, TILT) • Editing and changes (Haenni, Benoist, Anrig, VIP)
1.0 **(final draft for internal review)**	31.5.09	• 7.1.1 + rewriting of 5.1 (van der Wees, TILT) • 7.1.2 (Buitelaar, TILT) • Editing (Jaquet-Chiffelle, VIP)
1.1	18.6.09	• Corrections and changes taking into account feedback of the reviewers Jozef Vyskoc's, Sandra Steinbrecher and James Backhouse (all)
1.2 **Final version**	25.6.09	• Inclusion of TILT's corrections, final check and editing (Jaquet-Chiffelle, VIP)

Foreword

FIDIS partners from various disciplines have contributed as authors to this document. The following list names the main contributors for the sections of this document:

Section	*Contributor(s)*
1 (Executive Summary)	D.-O. Jaquet-Chiffelle (VIP)
2 (Introduction)	All (VIP & TILT)
3 (Case studies)	All (VIP & TILT)
4 (Core concepts)	R. Haenni, D.-O. Jaquet-Chiffelle with B. Anrig, E. Benoist (VIP)
5 (Response mechanisms)	H. Buitelaar, L. van der Wees (TILT)
6 (Trust and identification)	B. Anrig, E. Benoist with R. Haenni, D.-O. Jaquet-Chiffelle (VIP)
7 (Trust in the Virtual World)	All (VIP & TILT)
8 (Conclusion)	D.-O. Jaquet-Chiffelle (VIP)
9 (References)	All (VIP & TILT)

Table of Contents

1 Executive Summary

The purpose of this deliverable is to pinpoint and explore fundamental difficulties related to identification and trust in the digital world. The idea is to illuminate and analyze them from the perspective of the "virtual person" model, which has been a major research topic in previous FIDIS deliverables.

After an introduction that shows the importance of the topics covered in this deliverable and situate this research with respect to other related FIDIS activities, we give several case studies in order to illustrate how trust, confidence and identification often intervene in today's typical applications on the Internet, i.e., in the digital world. These case studies also show how deeply trust and identification are intertwined. We observe fundamental similarities between the challenges that Amazon, eBay, etc. are faced with in respect to trust and identification in the digital world. We also present some typical counter-measures to overcome these challenges.

In Section 4, the core concepts of trust, confidence and identification are defined and analyzed in both the traditional and the digital contexts. The key concepts are precisely defined in order to guarantee a common language and consistency for the rest of the deliverable. The in-depth analysis and synthesis of current definitions and approaches for trust gives an overview of the state-of-the-art that opens the doors for future research. This scientific analysis and synthesis is first addressed to specialists in trust-related domains, but could be interesting as well for policy makers who need an in-depth understanding of trust mechanisms. The precise definition with a thorough discussion of *identification* brings the traditional IT approach closer to what is usually adopted in forensic sciences.

Then, we discuss the importance of response mechanisms (social pressure, reputation, legal liability) as trust enablers or to regulate behaviours (to discourage entities from acting in an inappropriate way) to build trust.

Section 6 explores further the close relations between trust and identification: the identification components in trust and the trust components in identification.

Eventually, in Section 7, we discuss how the concepts of trust and identification can be modelled in the digital world and, more generally, in the virtual world in the light of virtual persons.

2 Introduction

The transition of many areas of everyone's daily life into the digital world has created tremendous new opportunities in how people communicate and interact. Whereas traditional human interaction is largely based on direct (person-to-person) contacts in the physical world, which inherently restricts its extent to a relatively small spatial perimeter, it is no longer necessary in a purely digitalized environment to restrict oneself to any such physical interactions. It is thus common today to have a large number of "online" contacts, friendships, or business relationships with people that never met in real life. In principal, there is no limitation of where these online contacts live and who they are, and there is often no way of finding it out. A prominent example of a platform that supports and popularizes this type of online relationships is the social network platform Facebook.

The downside of the digitalization and virtualization of our social life is that it creates a whole class of entirely new problems and threats. One of the key problems is the increased difficulty of identifying people in applications where identification is of crucial importance. While the traditional way of identifying a person relies on people's strong face and voice recognition ability to match the physical appearance of that person (what is seen or heard) with what is stored in memory, it is now a question of interpreting various types of "digital evidence" about that person. Even if high-definition pictures, sound, movies are used to represent the physical appearance of a person with arbitrary accuracy, they remain an accumulation of bits and bytes, which needs to be interpreted before being useful in the identification process. In general, this intermediary interpretation is an additional potential source of failure, which may confound the identification process or render it impossible.

Another key problem in many applications of the digitalized world is the question of whether trust can be granted to an unknown virtual counterpart in the same way as trust is granted to a person in the real physical world. This is a very difficult question in general, but the answer is at least partly affirmative if the virtual counterpart can be unambiguously linked to a real trustworthy person. Consider for example a digitally signed e-mail from a trustworthy friend compared to some spam e-mail received from a dubious address: what would be the degrees of trust attributed to the person behind the bits and bytes of the received messages in these two cases? And how much weight or confidence would then be attributed to the respective content of the messages? Even if the anonymity of an e-mail does not generally allow to entirely discredit its content, this example shows that the problems of identification and assigning trust in a digital world are intrinsically interwoven.

The purpose of this deliverable is to pinpoint and make explicit fundamental difficulties related to identification and trust in the digital world. The idea is to illuminate and analyze them from the perspective of the "virtual person" model, which has been a major research topic in previous FIDIS deliverables. In the remainder of this section, we try to further motivate the importance of trust and identification in situations like the ones mentioned above. We will also work out the connection to previous deliverables.

2.1 Trust

Trust has always played an important role in human societies. Many observable patterns of social interaction and corresponding relationships between individuals or groups of

individuals are intrinsically tied to various forms of trust. In a broad sense, trust is usually understood as someone's firm belief in the reliability, competence, qualification, ability, strength, integrity, truthfulness, honesty, sincerity, loyalty, etc. of someone else. It is thus a relationship of reliance between a trusting and a trusted party.

The usual subjects and objects of trust relationships are physical persons (or groups of physical persons) involved in everyday social interactions. With the emergence of information technologies and the resulting trend towards a ubiquitously interconnected information society, the range of applicability of trust-related questions needs to be enlarged more and more from personally connected local communities towards globally distributed virtual communities. Then, one of the key questions to answer is whether and to what extent trust is also a matter between so-called "virtual persons", which are hidden behind ambiguous descriptions or pseudonyms (Cofta, 2007). This problem is a particular instance of the following more fundamental question: "*How can I trust bits and bytes?*" (Gerck, 2002). Most generally, we may pose the question of the possibility of human-machine or machine-machine trust relationships (Muir, 1987; Lee, Moray, 1992).

Due to its fundamental role for social groups like organizations, communities, institutions, or even whole economies to function, trust has been a very popular area of scientific study and research in many different fields, most notably in the social sciences and its sub-branches, e.g. in sociology, psychology, economics, and political sciences. It has also been an area of interest for numerous philosophers, who strive to explore the conceptual nature, moral foundations, and the epistemology of trust and trustworthiness. More recently, physicists and system scientists interested in collective processes, dynamical complex systems, or generally in cybernetics have discovered trust as an important issue. Simultaneously, trust has been recognized in computer science to be fundamental for building up and managing public-key infrastructures, peer-to-peer networks, large-scale e-commerce applications, web-services, the semantic web, or interactive online communities. The problems of applications of that kind are intrinsically tied to the more general problem of establishing trust in IT applications and services or in new technology in general (Flowerday et al., 2006).

In the light of its widespread scientific relevance, trust should be regarded as a multi-disciplinary research topic with many different meanings and varying perceptions. We do not further expose the diversity of the possible perceptions in this deliverable, but we will later try to take over some of the most common themes and patterns found in the literature. A more comprehensive essay on trust is included in (Jaquet-Chiffelle et al., 2009).

2.2 Identification

2.2.1 In e-commerce

"***Trade*** *is the willing exchange of goods, services, or both. Trade is also called commerce.*" [4]

For instance, let us look at the market place of a city. The exchange is trivial and does not require any identification: the provider delivers his goods and the client pays for them. The transaction is synchronous (or almost) and so both do not use any explicit identification. Buying goods in such a market place seems to be an anonymous transaction. Cash cannot be

[4] <en.wikipedia.org/wiki/Trade>

tracked, the client and provider do not know each other, the transaction does not involve any identification. This first impression is nevertheless reducing the complexity of the reality. In reality, a client will not visit just any merchant, he will know where to find the best (or the cheapest) bread or the best poultry because of the reputation of the vendor, or because of previous experiences. The merchant will also not select the same meat for a regular customer as for a tourist that he will never see again. So even if the transaction seems anonymous, the client identifies the vendor (maybe with an expression such as « The first butcher in front of the church ») and the client is also identified, sometimes with his name for some regular client, sometime as « the lady who buys 3 loaves every day » or even by his belonging into a group of clients that one has never seen (and probably will never see any more) like the « tourist » group.

We have seen that even on a market identification implicitly takes place, but commerce is not restricted to people visiting a market and paying cash. In most cases, an explicit identification is used in the exchange process. Identification is used for ensuring that the exchange is fair (goods delivered and money paid) as well as for ensuring the possibility to judge later complaints concerning exchange fairness (e.g. with respect to quality of goods).

A form of identification is required as soon as the exchange of goods and money is asynchronous. For instance, if a client pays using a credit card, the delivering of the good is straightforward, but the payment will occur later. Therefore the merchant needs to be certain that he will be paid for the thing he sold. The credit card offers a sufficient identification, the seller is not interested in the name of the client per-se, he just wants to be certain that he will be paid (i.e. that the bank will pay). On modern credit cards, authentication is done using a PIN (personal identification number) code that someone can know even without being the legitimate owner of the card. In this case, the seller is satisfied, the identification succeeded, which means he is assured to be paid by the bank, even if the identity of the person was not checked (and was maybe even false).

Identification in e-commerce plays the same role as in the physical world, both partners want to be sure that the transaction will occur as expected. Because of the distributed structure of the Internet, identification in e-commerce will take very various forms. Explicit identification of a person will occur between a client and his credit card company. The company requires a copy of the identity card, of the passport or of any document proving the official identity of the person and also a financial profile of the future client. Once this has been established, the person will be granted a credit. The identity will be used to sue the client if he does not pay his bills. Once this step is taken, the credit card owner will receive a valid credit card, which means that the bank (or credit card institution) provides a new form of "identity" to the user. The merchant is not interested in any strong identification of the client, he just want to be assured by the credit card company that they will pay. On the other side, the client wants to be sure that the provider will deliver the promised goods. In this case also, he uses identification to achieve this goal. In most of the cases, the client contacts a renown website (such as Amazon[5], Fnac[6], leShop[7], ...) that he knows. Such websites have a reputation, and the user knows a lot of persons that enjoyed good experiences with this partner. So in this case, he simply needs to be assured that he is really connected with the right site. This is achieved

[5] <www.amazon.com>
[6] <www.fnac.com>
[7] <www.leshop.ch>

using HTTPS (the secure version of the standard web protocol HTTP). HTTPS[8] uses a public key infrastructure to certify the "identity" of a website. The web server uses a certificate signed by a trusted third party that contains its public key. This key is used by the client to create a secure channel between his browser and the server. Because of the public certificate, he is assured that only that web server can be on the other side of this channel. So both the site and the user have enough information about the identity of their partner to make a deal.

In the Web 2.0, e-commerce in not restricted to customer purchasing goods at renown sites. New forms of e-commerce are more and more including the clients of a website both as vendors and buyers (see eBay[9] or ricardo[10] for instance). In this case, anonymity of the partners is an important part of the business model. Clients do not want everybody to know what they buy or sell on those sites. So the transaction must occur between two actors that must remain partially anonymous to each other. In this example, a third party (the website) is responsible for the "identities" of the actors. The website verifies that the user is a legitimate one (by checking his password for instance), and it creates a new identity (the pseudonym) that has some of the features of an identity in the real world: it can have a reputation, and is responsible for paying and delivering the corresponding goods.

Some sort of identification always plays a central role both in commerce and in e-commerce, the partners use it in order to trust each other.

2.2.2 In e-government

Identification plays also a significant role in the e-government context; this role is nevertheless somehow different in comparison e-commerce. In e-government, the interlocutors know each other, the citizen/inhabitant knows the administration and the administration knows its inhabitants. Identification does not bring much more than what it is supposed to: "verify the identity of the partner" (where in e-commerce, identification is also used to create or increase trust in each-other).

E-government is a very large topic that covers areas such as tax-declaration, public key infrastructure for citizens, e-voting, or even part of e-health in some countries. Identification does not play the same role in all these applications: whereas a strong identification is the goal of establishing a PKI infrastructure, it is a prerequisite for establishing tax-declaration on-line, or managing e-health. In e-voting, identification plays a central role. On the one hand, each citizen must be strongly identified to prove his rights to vote and, on the other hand, the vote itself must be totally anonymous.

We can nevertheless see some persistent elements in e-government applications. Even when working in the virtual world, they always rely on a physical identification done in the physical world: for example, on the desk of an office, or by sending credentials (user-name/passwords or more sophisticated credentials) to a physical location using postal service. This physical link can also be used from the client side to the administration side by sending some information not only on the internet, when for example the real validation needs to be transmitted on paper (e.g., with a real hand-written signature)[11].

---

[8] <en.wikipedia.org/wiki/HTTPS>
[9] <www.ebay.com>
[10] <www.ricardo.ch>
[11] See for instance the tax declaration process used in the canton Bern (Switzerland) <www.taxme.ch>

This multi-channel approach can be used to add confidence in the electronic relation between the two partners.

2.3 Virtual Persons

What is a *virtual person*? What is it used for? What is its added value?

The term *virtual* in this deliverable does not mean digital. Something is considered to be *virtual* if it is the product of one's mind or imagination. The virtual image in a mirror illustrates well our definition. Virtual is not opposed to real (for example, the concept of virtual reality is compatible with our definition of virtual), but to physical. For any physical entity some sort of physical constituent is compulsory. Digital entities can be either physical or virtual: we observe physical digital entities (e.g., magnetic storage of bits on a hard drive) as well as virtual digital entities (e.g., one bit of information or a specific avatar in a computer game).

Virtual persons are sometimes used to describe avatars and new forms of identities in online games. They also appear in other contexts; some authors use them in the legal domain. Within FIDIS (cf. Section 2.4), the concept of virtual person has been extended in order to better describe and understand new forms of identities in the Information Society in relation to rights, duties, obligations and responsibilities.

Virtual persons, as other virtual entities, exist in the virtual world, the collection of all (abstract) entities which are or have been the product of the mind or imagination. A *virtual person* is a virtual entity that can have (not necessarily legal) rights, duties, obligations and/or responsibilities associated to it in a certain context.[12]

The virtual world –not to be confused with the digital world– allows a unified description of many identity-related concepts that are usually defined separately without taking into consideration their similarities: avatars, pseudonyms, categories, profiles, legal persons.

2.4 Link to Previous Deliverables

This deliverable revisits the concepts of trust and identification in the light of the model based on virtual persons. This model has been precisely defined in Subsection 5.1 of FIDIS deliverable D2.13 "Virtual persons and Identities". We will not repeat here the description of this model, its definitions and all context information with respect to these definitions. We strongly recommend the reader who does not know this model yet, to read first FIDIS deliverable D2.13. The interested reader might also want to study first the UML description of the model given in Section 7 of FIDIS deliverable D17.1: "Modelling New Forms of Identities: *Applicability of the Model Based on Virtual Persons*" or even read D17.1 completely in order to become more familiar with the use of the model and its potential. FIDIS deliverable D17.2: "New (Id)entities and the Law: Perspectives on Legal Personhood for Non-Humans" is the cornerstone of D17.3: "Bridging the Accountability Gap: Rights for New Entities in the Information Society?". Deliverable D17.3 observes that new entities in

[12] Avatars are a special kind of virtual persons. However, the concept of virtual person is much broader and should not be reduced to avatars only.

the information society, such as pseudonyms, avatars, software agents, and robots, create an 'accountability gap' because they operate at increasing distance from their principals. While D17.3 tackles legal issues related to this increasing distance and the corresponding accountability gap, the current deliverable discusses the impact of this increasing distance on trust and confidence, as well as on the identification process itself.

3 Case Studies

The purpose of this section is to accentuate the significance and importance of trust and identification in various web applications such as *DigiNotar*, *Amazon*, *eBay*, and *myProf (.de, .ch, or .at)*, or more generally in electronic mail. We will see that applications and services of that kind typically require various forms of trust, corresponding trust building mechanisms, and reliable user identification instruments. We will also see that trust and identification are often intrinsically interwoven, and that identification in a purely digital context is far more complex than in the physical world. Potential frauds and abuses in the absence of proper trust and identification mechanism will be exposed, and a description of possible counter-measures will be given. The diversity of the given examples together with the similarities of the underlying problems will then serve as a motivation for the more profound analysis of trust and identification in the remainder of this deliverable.

3.1 DigiNotar

In this first subsection, a description is given of the organization, products, and services of DigiNotar.[13] This is, as they call themselves, an Internet Trust Provider active in The Netherlands, Spain, and Scandinavia. They provide services to exchange valuable data or to carry out transactions through the Internet. DigiNotar as Internet Trust Provider intends to ensure all parties in electronic communications are identified, that all information reaches an addressee unaltered, and that the filing of digital documents is as reliable and legally secure as in the 'paper period'.

DigiNotar provides trust and security solutions such as identity management, electronic signatures, reliable document exchange and electronic archiving; all used for conducting business over the Internet. DigiNotar also supplies every type of identification resource, from simple username & password systems, via identification using a mobile telephone or bankcard, to the strong type in the shape of electronic signatures on smart cards.

DigiNotar offers services to governmental as well as commercial organizations.

The name DigiNotar refers to digital notary services, which the founders of the company had in mind. Notarial services are not only to be provided by DigiNotar, but also by notaries. Therefore DigiNotar has a strong link with the office of notary. This link which was relatively easy to make because one of the founders was a notary when DigiNotar started and is working as one still.

Below a short description is given of the products and services offered by this Internet Trust Provider.

[13] English website DigiNotar, Internet Trust Services: <www.diginotar.com>.

3.1.1 Products and Services - Authentication

DigiNotar has an authentication service called PASS, Personal Access Security Service. Via PASS users are given a means of authentication, varying from a simple username/password or a mobile phone to a DigiNotar OTP (One Time Password) token, a smartcard or a USB token. With PASS users can log onto websites that have accepted their authorisation using PASS.

One can link a website to the PASS authentication service via a standard available agent (routing software). Through the PASS admin room functionality one can determine which users one wants to invite and which users one wants to give access to specific (website) services. The admin room is in fact the main entrance to the house and the user of this service can indicate which rooms a client is allowed to enter.

The idea behind PASS is to avoid website owners having to implement their own authentication system. Also the registration and distribution of the usemames and passwords are organized for PASS customers. The logon takes place via a PASS authentication server, which is the location that checks the user name and password or other means of authorisation.

PASS is suited as a means of authorisation for electronic signatures, secure document exchange, and archiving.

Interestingly, via PASS, DigiNotar also offers the possibility to get a pseudo-identity, i.e. a pseudonym. This protects the privacy of the user, as his identity cannot be deduced. The user's identity will be checked by DigiNotar before the user gains access to a website or online service under a pseudonym. The thoroughness of this identity check depends on the type of service one is going to use. For optimal security one has to go to a DigiNotar office or a notary and identify oneself. In other cases the identification might take place online.

The exploiter of the site or service will not know the identity of this type of users, but they know that the user has authorised access.

3.1.2 Products and Services – Exchange of Documents

For the secure exchange of documents DigiNotar has developed a web application called DocProof. This service is launched to avoid (important) documents to be sent back and forth via (registered) post. This is not only costly; it also often takes days, or even weeks, before parties agree upon the text of a document.

DigiNotar has organized a digital environment called DocProof. To use DocProof, first of all customers have to register with DigiNotar. After that a customer is given a personal certificate to be able to access his own space on DocProof. Having received the certificate the customer can start registering clients and other partners he wants to communicate with. During this registration, he can choose the means of authentication for his clients and partners, after which DigiNotar will send them the chosen means of identification.

Next, the customer can start placing documents in his DocProof environment, and link them to a client. The client then will automatically receive an e-mail notification and can go to the DocProof environment (data room) and view, download, respond to, and (if necessary) electronically sign[14] the document. He can also add documents, a changed version, his response or attachments.

[14] See SCADPlus for the community framework for electronic signatures, <europa.eu/scadplus/leg/en/lvb/l24118.htm>.

The entire process - uploading, response, downloading, signing, comments – is being logged and will be archived for (at least) seven years. For both customer and client this logging will offer the opportunity to prove *which* documents are sent, received, viewed and signed *when*, and by *whom*.

Another advantage of this type of document exchange is that clients are able to sign documents electronically without having a public key infrastructure (PKI) certificate. Without investing in a PKI environment one is able to do secure document exchanges.

3.1.3 Products and Services – Electronic Mail Receipts

MailProof is a service, which tracks an e-mail message until it arrives at the recipient's server.[15] At that moment the sender of the message will receive an electronic confirmation of receipt. This way, the user of this service can be sure the message has arrived and has been received. The message and its attachments, the electronically signed log and the source are stored in DigiNotar's notarial, electronic archive. A hash value of all messages and attachments will be calculated and archived. This will also be displayed in the electronic source. If required, the evidence and the original message, including attachments, can be provided with a notarial statement.

3.1.4 Products and Services – Signing Service

An increasing number of official documents and forms (applications, orders, contracts, appeals, permits, etc.) are completed and processed via the Internet. The DigiNotar Signing Service makes it possible that those documents and forms can be signed electronically.

Via the signing service, signatures are put on documents in a secure, independent environment, which checks the validity of the means of authentication. It is also possible to archive the signed document and the transaction data. In this case, the customer receives a proof of receipt, which includes a time stamp.

The service can be fully integrated into a website. The signed documents can be provided in either PDF or XML format. Users have the option of receiving a hyperlink to the document stored in the DigiNotar archive.

3.1.5 Products and Services – PDF Signing Service

Many organizations use the Portable Document Format (PDF) for the exchange of electronic documents, and virtually any type of electronic document can be converted into a PDF document, which can then be read by the recipient using a PDF Reader. The PDF format has even become an ISO standard.[16]

DigiNotar has developed a special service for signing PDF documents: NotarSign.[17] Using this service, customers can automatically give PDF documents a legally acknowledged[18] signature, and e-mail these to their relations or the electronic archive. Providing PDF

---

[15] Technically, tracking an e-mail message until it arrives at the recipient's server seems impossible. The reliability of this service is therefore unclear .

[16] ISO 32000-1:2008, Document management -- Portable document format -- Part 1: PDF 1.7, ISO

[17] See SCADPlus for the community framework for electronic signatures, <europa.eu/scadplus/leg/en/lvb/l24118.htm>.

[18] Broderick, Martha A., Gibson, Virginia R., Tarasewich, Peter. 2001. Electronic Signatures: They're Legal, Now What?, Internet Research: Networking Applications and Policy, 2001, 11(5), 423-434.

documents with an electronic signature guarantees who the actual sender of the document is and that the document has not been changed.

3.1.6 Products and Services – Digital Certificates

Organizations increasingly communicate with their (business) relations over the Internet. Risks that relate to this type of communications can be reduced for users having a so-called digital certificate.[19] This certificate can provide secure communication, electronic signatures, encryption and authentication. However, it has not been easy to use digital certificates, and therefore the application is not widespread so far. More than average technical skills were required to use electronic signatures. Therefore DigiNotar has developed what they call a Trusted Identity Manager with which customers can issue their own digital certificates to relations and employees easily.

Figure 1: Trusted Identity Manager

This type of easy-to-use secure services might give a boost to the use of digital certificates and signatures, also because cybercrime more and more becomes an evermore serious threat to the digital world, which makes the demand for secure services greater.[20]

3.1.7 Products and Services – Extended Validation Secure Sockets Layer[21]

EV SSL, which stands for "Extended Validation SSL", is a new type of SSL certificate. This certificate improves online security by performing extra checks with regard to the identity of the relevant organization. With EV SSL a website exploiter assures users that they are using the genuine site, not a fake. DigiNotar as well as other trusted third parties[22] issues EV SSL certificates.

The EV SSL certificate prevents illegal copying of websites to obtain e.g. personal and credit card data, thus preventing so-called *phishing*.[23] Extended Validation certificates are based on

[19] Public key certificate, Wikipedia, <en.wikipedia.org/wiki/Public_key_certificate>.
[20] According to FBI 2007 losses as a result of Internet fraud reached an all-time high in 2007. See <recht.nl/32208>.
[21] Extended Validation Certificate, Wikipedia, <en.wikipedia.org/wiki/Extended_Validation_Certificate>.
[22] Other parties issuing EV SSLs are for example VeriSign, GlobalSign, DigiCert.
[23] Phishing, Wikipedia, <en.wikipedia.org/wiki/Phishing>.

an extensive set of organization and domain checks. Only when an organization has passed all checks will an Extended Validation SSL certificate will be awarded.

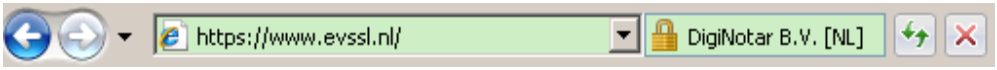

Figure 2: Extended Validation SSL

Extended Validation SSL certificates are indicated by the green address bar in the web browser. By showing the organization name and the issuing organization, a visitor can see immediately that he is dealing with the original website and not a copied phishing website.

3.1.8 Products and Services for the Virtual World?

The short description of the products and services of the trusted third party DigiNotar makes clear that this type of organizations offers interesting products and services for the e-government and e-commerce environment. In that e-spectrum they offer services as well for clients as website exploiters. They provide the opportunity for citizens to communicate with governments and webshops using a pseudo-identity. They also offer services preventing clients to become a victim of cybercrime activities like phishing. On the other hand they offer website exploiters services which create evidence and certainty in a virtual environment.

The services offered might also play a role in a world in which agents, avatars, and alike are active and in which less interaction takes place between real life persons and entrepreneurs/institutes. For agents and websites to be sure they deal with reliable entities, sockets layers, certificates and procedures may be even an absolute necessity. This is especially of importance for enabling interaction with unknown businesses and organizations worldwide. The use of sockets layers, certificates and procedures also helps to create a level playing field for smaller e-businesses and e-organizations. After all, it seems that consumers are highly conservative when shopping online, overwhelmingly shunning small online businesses in favour of bigger brands.[24]

3.2 Amazon

Amazon[25] is one of the best-known on-line trading companies where individuals can buy books and other goods, and also among those in business for a long time already (compared to the age of the internet, of course). In the following, we will focus on identification and trust between a user, i.e., a potential customer or buyer, of the Amazon website and Amazon or its partners etc. on the other side.

3.2.1 Conditions of Use

Consider now someone looking for a book on the subject of privacy at Amazon's site. In the process of the searching for the book, the potential customer is not typically interested

[24] Online Shoppers Getting Security Savvy, Paypoint.net, 8 July 2008
[25] In the sequel we consider Amazon.com, i.e., the American website. Similar considerations can be made when looking at Amazon's "international sites", e.g. in Germany at <amazon.de> .

necessarily in being identified by the site. However, trust does not need to be high in this phase.

This changes in several case when he

- is interested in the price of a specific book, because for example frequent shoppers might have a better price, etc.
- wants to actually buy the book
- wants to have personalised advertising, or personalised offers for related books
- sends (non-anonymized) reviews, comments, etc. for books
- etc.

Amazon specifies in its Conditions of Use that for such (and other) uses, one may not impersonate any person or entity:

"Visitors may post reviews, comments, photos, and other content; send e-cards and other communications; and submit suggestions, ideas, comments, questions, or other information, so long as the content is not illegal, obscene, threatening, defamatory, invasive of privacy, infringing of intellectual property rights, or otherwise injurious to third parties or objectionable and does not consist of or contain software viruses, political campaigning, commercial solicitation, chain letters, mass mailings, or any form of "spams." You may not use a false e-mail address, impersonate any person or entity, or otherwise mislead as to the origin of a card or other content. Amazon reserves the right (but not the obligation) to remove or edit such content, but does not regularly review posted content."[26]

Amazon wants therefore to have some basic level confidence in this data by not allowing impersonation or false e-mail addresses.[27] All potential information should not mislead but point somehow in the direction of the generator. Clearly, this gives no real identification possibilities to Amazon, but to some extent anonymity disappears. It seems however that there is little done to stop bots[28] registering on their own.

When the potential customer goes to Amazon's website, why should he trust it? There are two points which are important here:

- The customer may well had heard of Amazon (who has not...) and even usually knows some friend or somebody else who has successfully bought something from Amazon. Hence the brand is well known and trust towards this brand and the selling of books is to some extent well established
- On the other hand, why should he trust the website as really belonging to Amazon? An impostor might have brought up a fake website or some hijacking is in place. One approach is to download the website through a secured channel (from https://www.amazon.com) and check the corresponding certificate and hence trust some certificate authority certifying the information contained in the certificate.

[26] <www.amazon.com/gp/help/customer/display.html/?ie=UTF8&nodeId=508088> Amazon Conditions of Use (23.1.2009)
[27] What is a "false e-mail address" anyway? Can I use the one of my secretary or is this "false"?
[28] Bots are software applications that automatically do taks on the internet, a typical example is the web spiders which gather information from different web servers.

Probably as an attempt for making itself more trustworthy to potential customers, Amazon has – again in its Conditions of Use – specified a reference to its physical address:

"OUR ADDRESS Amazon.com, Inc., P.O. Box 81226 Seattle, WA 98108-1226[29]

However, looking closely at this address, it is "only" a P.O. Box address, hence how trustworthy is this really? Amazon is linking its virtual identity – its website – to some more or less physical existence, in order to increase trust of the client which is then able to somehow link the website with some physical instance. Whether an address on another continent really raises trustworthiness is another issue. On the other hand, there is clearly a link to the legal person "Amazon.com" here. The legal person itself is assumed to be listed in some register of companies where – in general – a link to one (or more) physical person is explicit. Therefore, an indirect link from the website to at least one physical person can be induced in this way.

3.2.2 Privacy Notice

At the bottom of the main web page, one can find Amazon's privacy notice and, as a preamble, the motivation for this notice says:

"Amazon.com knows that you care how information about you is used and shared, and we appreciate your trust that we will do so carefully and sensibly. This notice describes our privacy policy. By visiting Amazon.com, you are accepting the practices described in this Privacy Notice." [30]

The statement of the notice emphasizes the intention of Amazon to care about its clients' privacy. However, for the potential customer, there is no alternative than to "trust that we [Amazon] will do so carefully". Apparently, there is no independent supervisor taking care of the application of these privacy notices. Implicit trust of the customer is required here. Just leave it as a question whether this is a good starting point for privacy or not.

Amazon specifies the information it stores about its customers. There are four different means by which Amazon can get data:

"Information You Give Us: We receive and store any information you enter on our Website or give us in any other way. [...] You can choose not to provide certain information, but then you might not be able to take advantage of many of our features. We use the information that you provide for such purposes as responding to your requests, customizing future shopping for you, improving our stores, and communicating with you.

Automatic Information: We receive and store certain types of information whenever you interact with us. For example, like many Websites, we use "cookies," and we obtain certain types of information when your Web browser accesses Amazon.com or advertisements and other content served by or on behalf of Amazon.com on other Websites. [...]

E-mail Communications: To help us make e-mails more useful and interesting, we often receive a confirmation when you open e-mail from Amazon.com if your computer supports such capabilities. We also compare our customer list to lists received from other companies, in an effort to avoid sending unnecessary messages to our customers. If you do not want to

[29] Amazon conditions of use: <www.amazon.com/gp/help/customer/display.html/?ie=UTF8&nodeId=508088> (23.1.2009)
[30] Amazon Privacy Notice <www.amazon.com/gp/help/customer/display.html/?ie=UTF8&nodeId=468496> (23.1.2009)

receive e-mail or other mail from us, please adjust your Customer Communication Preferences.

Information from Other Sources: We might receive information about you from other sources and add it to our account information. [...]"[31]

The customer can understand to some extent from where the information Amazon gets about him can come from. Amazon is informing about the potential processing, sharing, etc. of the data it stores about its customers. Again, this is only a declaration of Amazon, implicit trust of the potential customer is assumed. On the other hand, this data will allow Amazon to identify customers more easily, and to do profiling (especially group profiling) on the customers in order to allow target-oriented advertising etc. This might even allow identifying through their behaviour customers to some extent, even when the customer does not actually want to be identified, while anonymously visiting Amazon's website.

"Does Amazon.com Share the Information It Receives?

Information about our customers is an important part of our business, and we are not in the business of selling it to others. We share customer information only as described below and with subsidiaries Amazon.com, Inc. controls that either are subject to this Privacy Notice or follow practices at least as protective as those described in this Privacy Notice." [32]

While this might be reassuring the customer about Amazon's good behaviour, and hence increase its trust in Amazon, one of the problems (even mentioned by Amazon itself) appears when parts of the business, say a subsidiary, is sold to some other company:

"Business Transfers: As we continue to develop our business, we might sell or buy stores, subsidiaries, or business units. In such transactions, customer information generally is one of the transferred business assets but remains subject to the promises made in any pre-existing Privacy Notice (unless, of course, the customer consents otherwise). Also, in the unlikely event that Amazon.com, Inc., or substantially all of its assets are acquired, customer information will of course be one of the transferred assets."[33]

The next paragraph opens quite a field of potential applications:

"Protection of Amazon.com and Others: We release account and other personal information when we believe release is appropriate to comply with the law; enforce or apply our Conditions of Use and other agreements; or protect the rights, property, or safety of Amazon.com, our users, or others. This includes exchanging information with other companies and organizations for fraud protection and credit risk reduction. Obviously, however, this does not include selling, renting, sharing, or otherwise disclosing personally identifiable information from customers for commercial purposes in violation of the commitments set forth in this Privacy Notice."[34]

Reassuring the customer is again here the main aspect. However, this paragraph clearly allows Amazon to do a lot with your data if they "believe" it to be appropriate in some sense. The customer again has to trust Amazon on this.

Finally, a general remark is reassuring the customer:

---

[31] idem
[32] idem
[33] idem
[34] idem

"Conditions of Use, Notices, and Revisions

If you choose to visit Amazon.com, your visit and any dispute over privacy is subject to this Notice and our Conditions of Use, including limitations on damages, resolution of disputes, and application of the law of the state of Washington. If you have any concern about privacy at Amazon.com, please contact us with a thorough description, and we will try to resolve it. Our business changes constantly, and our Privacy Notice and the Conditions of Use will change also. We may e-mail periodic reminders of our notices and conditions, unless you have instructed us not to, but you should check our Website frequently to see recent changes. Unless stated otherwise, our current Privacy Notice applies to all information that we have about you and your account. We stand behind the promises we make, however, and will never materially change our policies and practices to make them less protective of customer information collected in the past without the consent of affected customers." [35]

Again, this is reassuring but there is no evidence on how one can be sure this is the case. Indeed, a usual customer typically has no way to reliably detect problems in this context. The goal as well is just to increase customers' confidence in the good behaviour of Amazon.

For a frequent customer, it is easy to keep track of changes of this privacy notice:

"Amazon.com Privacy Notice. Last updated: October 1, 2008. To see what has changed, click here." [36]

So in principle, everyone can keep track of changes (under the assumption that the changes are reported correctly) and eventually consider to not buy books from Amazon anymore or, on the other hand, start buying if personal requirements are fulfilled by the privacy notice.[37]

On the positive side, there is much effort being done in trying to increase confidence of the future customer in the "good" behaviour of Amazon, especially with respect to privacy. On the negative side, there is little hard information about control of those efforts by – in some sense – independent (trusted) parties.

3.2.3 Registration Information

A new customer opens an account at Amazon. Some personal information is required to open an account successfully, however few things on this data are really required: only an e-mail address, a username as well as a password are required. The e-mail address is not even checked using for example an e-mail for validating the account creation. More personal data (e.g. name, address, billing information, etc.) is only required when the customer wants to really buy some goods. Therefore, Amazon gains little trust that the registering party really is the one she is pretending to be. The step from anonymously surfing on the Amazon's website (being followed by Amazon using a Cookie) to surfing as a registered user (without giving additional data) is not significant. Frequent uses of the same registered "identity" together with potential additional information – produced by the customer as discussed below – clearly changes this as Amazon can use this information for its business.

[35] idem
[36] idem
[37] One of the problems is that one often has not really good alternatives online, e.g. in the example of Amazon or eBay.

3.2.4 Profile

Amazon gives its customer the possibility to make a personal profile containing for example his wish list and reviews.

Figure 3: Create your profile

Creating a profile is easy for the customer and he can choose the data which is published about him. The minimal requirement appears to be the public name, chosen arbitrarily by the customer. Hence with respect to identification, the customer can restrict this according to his own preferences. On the other hand, little information also means probably few "friends" identifying you. The following picture shows what information a customer might typically provide in order to be "identifiable":

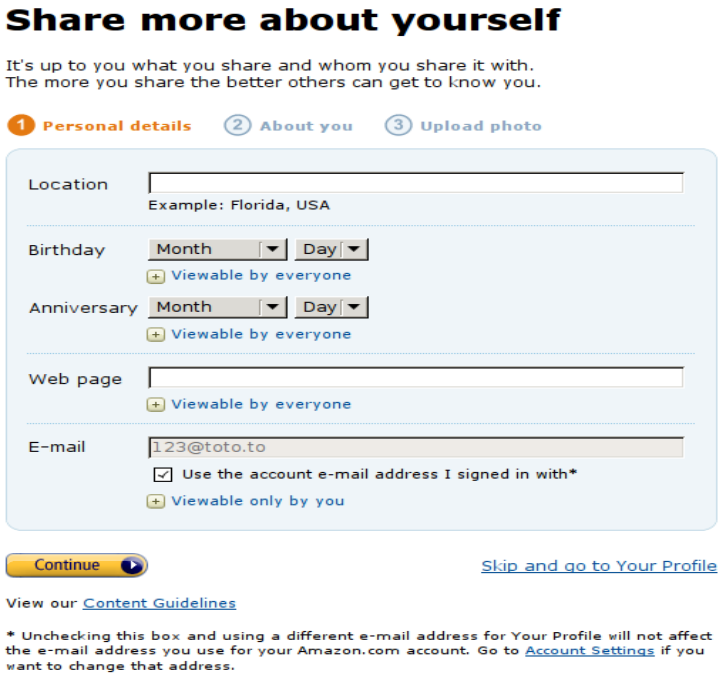

Figure 4: Share more about yourself

In order to not spread this information, one can choose to restrict the users who can access this type of data. The customer has to trust Amazon that it does behave as declared, i.e., does not let unauthorised persons access the information. On the other hand, Amazon clearly has access to this data and could potentially use it for identification purposes.

3.2.5 Feedback Information

For customers who additionally act as sellers[38], there is a possibility to leave feedback information for past transactions. This goes in a similar direction as eBay's reputation system which will be discussed in Subsection 3.3.3. To some extent, sellers have to trust that buyers make correct feedback statements, as feedback cannot be removed by the seller. Amazon is quite strict in not removing feedback information:

"How can I get feedback removed from my account?

Amazon will remove feedback only in the following cases:

- The feedback includes obscene language.
- The feedback includes personally identifiable information.
- The entire feedback comment is a product review.

That said, Amazon Marketplace buyers can remove feedback they have left. We encourage you to contact your buyer, and work with them to resolve any issues regarding the transaction. Your buyer could then remove the feedback if they feel it would be appropriate. See the question "How can I respond to my negative feedback?" for tips on how to deal with negative feedback." [39]

Hence, as a party in the process of selling, one has to trust the reliability of this feedback system, trust Amazon not influencing in any way the feedbacks (or simply cancelling some of them without one of the reasons specified above), and, as already mentioned, trust the one leaving the feedback. In a large system (as in Subsection 3.3.3 for eBay), the self-regulation of such a feedback system seems to work well.

3.3 EBay

This case study considers an online trading auction platform –eBay– with respect to trust, confidence and identification. Several phases and actions will be described from different point of views.

3.3.1 Registration phase

The very first time a user wants to interact on eBay, she has first to register and create his eBay account.[40]

---

[38] An additional problem with respect to trust here is that it is often not easy to determine whether Amazon itself is the seller or someone else, e.g. another Amazon user.

[39] <www.amazon.com/gp/help/customer/display.html?nodeId=1161284&#good> (23.1.2009)

[40] Note that the eBay company also gives the possibility to create a company account, i.e., an account for a legal person. In this case, a contact person (a human being) is also required.

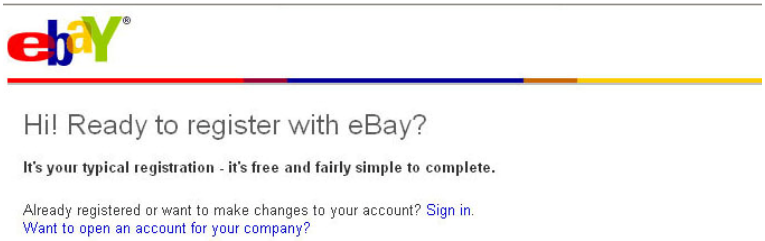

Figure 5: Registration phase

During the registration phase,[41] the user is required to provide personal identifying information, including real name and address. Even though the process in itself is made easy, several issues are relevant in term of trust, confidence and identification. The process is not only "free and fairly simple to complete" it is also *fairly complete* in term of identity-related information that is required.

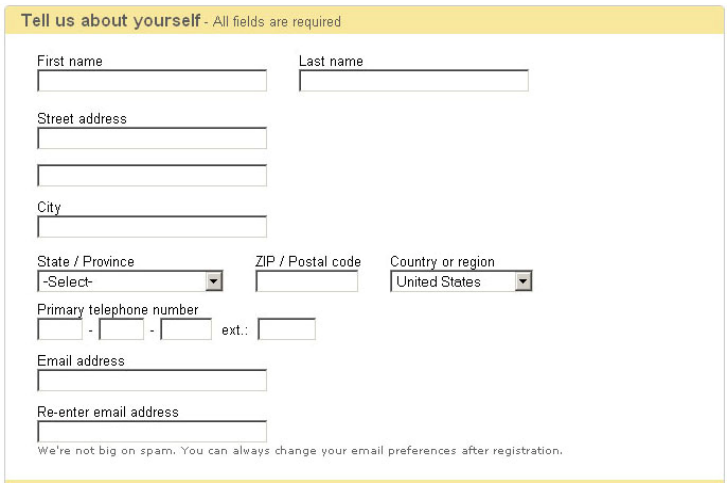

Figure 6: Identifying information

During registration, eBay must trust the user to give accurate information. To improve the confidence that eBay gains in the validity and accuracy of this personal information, the eBay user agreement[42] imposes in particular to enter no "false, inaccurate, misleading, […] personal information":

"While using eBay, you will not: […]

- post false, inaccurate, misleading, defamatory, or libellous content (including personal information);"

Some extra mechanisms might also be introduced to check this information or to require a certification authority to validate it.

---

[41] In this section, we consider <www.ebay.com>
[42] <pages.ebay.com/help/policies/user-agreement.html>

In particular, the eBay company wants to have confidence in the fact that the one who is registering is a human being that can be held liable for his actions and not a robot impersonating such a human being. In order to increase its confidence in the fact that the entity who is registering is a human being, eBay asks to give the answer to a challenge that is supposed to be hard to solve for non-human entities: the registering user has to type back a text that is drawn and hidden in a so-called CAPTCHA-picture. This can be seen as a human entity authentication.

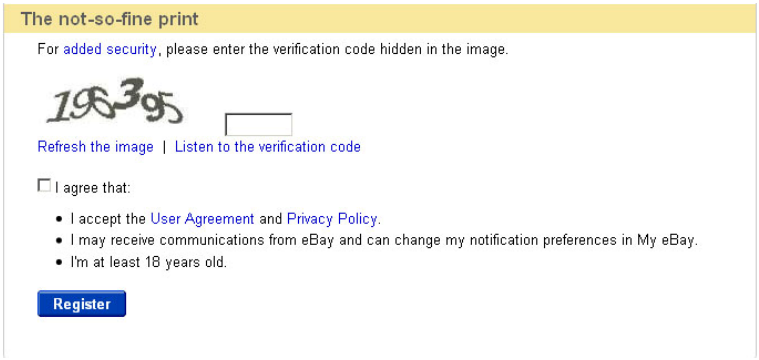

Figure 7: Human entity authentication

At the same time, the registering user must trust eBay to handle his personal data securely and carefully. In order to increase users' trust, eBay provides users with several evidences:

- The eBay Privacy Policy[43],
- A certification authority TRUSTe[44] that claims to have verified and validated eBay privacy statements,

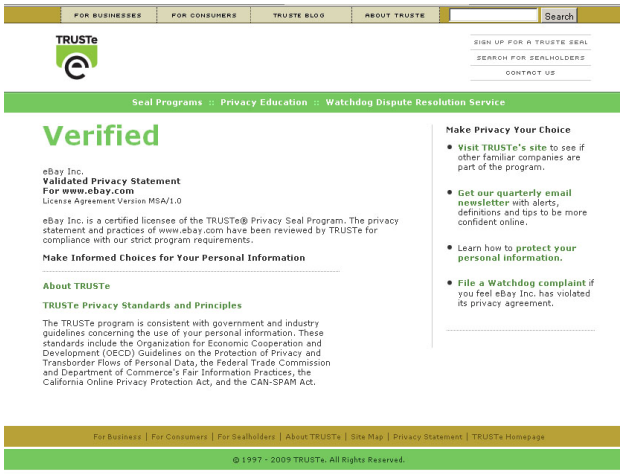

Figure 8: TRUSTe certification authority

[43] <pages.ebay.com/help/policies/privacy-policy.html>
[44] <www.truste.org/>
Trusting TRUSTe (which claims to have AOL, Apple, eBay, Microsoft amongst its clients) increases the trust one has in eBay.

- Information on where servers that deal with private information are located (they are located in the US) and commitments to physically and logically protect these servers:

  "Security:

  Your information is stored on our servers located in the United States. We treat data as an asset that must be protected and use lots of tools (encryption, passwords, physical security, etc.) to protect your personal information against unauthorized access and disclosure." [eBay Privacy Policy]

One of eBay's characteristics is to allow users to interact between each other using pseudonyms. During the registration phase, after having given his personal identifying information, the user is asked to create his eBay user ID –a self-chosen pseudonym– that will act later as an anonymizing tool towards the other users.[45]

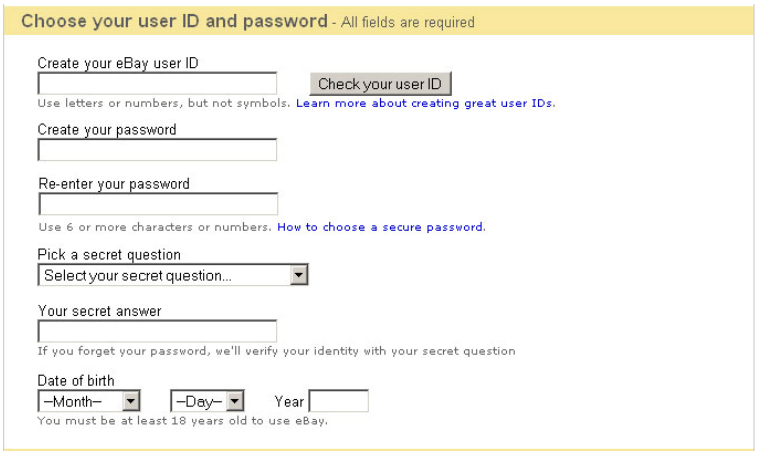

Figure 9: eBay user ID

3.3.2 Transactions

Transactions mainly occur between sellers (who offer items) and (buyers who try to acquire some of these items). If the transaction corresponds to an auction, eBay can automatically bid on the buyer's behalf up to his maximum bid. Furthermore, other actors may intervene: shipping companies, credit card companies, escrow services, etc.

Both buyers and sellers must in particular trust eBay

- to securely and consistently run the eBay platform and manage identity-related information.

(Potential) buyers must in particular trust eBay

[45] Cleary this is only the case if the users do not choose their real names as pseudonyms, nor sell anything to each other.

- to place the best bets on their behalf (even though eBay receives a final value fee which is proportional to the higher bet).

(Potential) buyers must in particular trust sellers

- to honestly describe their proposed items (e.g., model, condition, quality, etc.). Potential buyers must have confidence in the accuracy of the description.
- to ship bought items to the buyer in due time, and to wrap them correctly.
- not to misuse sensitive financial information (e.g., credit card number) if such information is divulged.
- not to misuse shipping information (name, address, etc.).

Sellers must in particular trust buyers

- to pay timely for bought items and to do it with a valid method.

Both buyers and sellers must in particular trust shipping companies

- to safely and timely deliver items from the seller to the buyer.

Different payment methods are available on eBay[46]. Every seller can choose the methods of payments that he offers. Most sellers offer PayPal which is the preferred payment method for most eBay users. But credit cards, for example, can be directly used too if both the seller and the buyer agree.

If the buyer communicates his credit card number to the seller, he must trust the seller not to misuse this information afterwards. On the other hand, in this case, the seller does not need to attribute trust directly to the buyer, only indirectly to the credit card company. Moreover, the seller must have confidence in the validity of the credit card information. Trusting the credit card company means to be confident that the credit card company will pay the due amount if the credit card information is correct.

The PayPal method presents several advantages in comparison to the credit card method. The credit card used by PayPal (if a credit card is used[47]) is not divulged to the seller. Therefore, the buyer only needs to trust PayPal not to misuse the credit card information. Moreover, PayPal offers "PayPal Buyer Protection"[48] that "helps you if you don't receive your item or the item is significantly different from its description in the seller's listing."

Escrow services can also be chosen to play the role of a trusted third party. However, it is hazardous for a buyer to trust any proposed escrow service. Indeed, they have been reports of fraudulent sellers using nonexistent escrow companies.

3.3.3 Reputation

Reputation is a key component when building trust. PayPal, for example, is more trusted than any escrow service in particular because it has a strong implicit positive reputation, just by being the preferred payment method for most eBay buyers and sellers.

---

[46] <pages.ebay.com/help/pay/methods.html>
[47] The buyer can choose not to give his credit card number to PayPal and make a fund transfer from his bank account.
[48] <pages.ebay.com/help/buy/paypal-buyer-protection.html>

The eBay platform provides a reputation system that allows the building of trust between eBay users who do not know each other, who have never interacted together and who are hidden behind pseudonyms.

To each eBay user ID is attached a so-called "feedback profile". The feedback profile of an eBay user ID measures the concordance between

- the actual behaviour of this eBay user ID during his previous transactions and
- the expected behaviour of this eBay user ID, according to other users who have already taken part into these transactions.

The eBay reputation system is fed by users themselves. It collects experiences of previous eBay transaction partners.

Feedback within the eBay reputation system is considered to be the most critical element for enhancing trust, according to a wide survey carried out on this matter by (Pavlou, 2002). Users expect eBay users' behaviour in former transactions to correspond with their behaviour in future transactions (Resnik, 2006).

What kind of confidence can we have in the eBay reputation system? Is it trustworthy? Can we trust users when they give feedback? In order for the eBay reputation system to work correctly, at least

- (most of the) users must be honest when they give their feedback,
- (most of the) users who give feedback must be competent to compare the actual behaviour of an eBay user ID and its expected behaviour.

The first condition is not always fulfilled. Some cases of abuse are regularly reported. For example, a professional eBay seller may try to undermine the reputation of a competitor. Or a group of fraudulent users (or maybe one user having succeeded in controlling many eBay user IDs) artificially improve their own feedback profiles by posting reciprocal or circular positive feedbacks.

Only users who have directly participated in a transaction can be competent to compare the expected and actual behaviour of an eBay user ID. In order to remove most of the non-competence, the eBay reputation system officially requires that only users who have been directly involved in a transaction can provide feedback about their transaction partner. Users have to trust eBay that this requirement is enforced.

3.4 The meinProf Web Sites

The web sites *meinProf* (.de, .ch, or .at)[49] propose to the students of all German-speaking Universities in Germany, Switzerland and Austria to evaluate their professors. The websites give the opportunity to any student to register. Once a student is registered, he has the possibility to evaluate his teachers.

The sites provide quite a broad list of teaching staff at universities. The student can evaluate an existing course or create a new one. The student can evaluate each course on various criteria: general opinion, fairness, support, material, comprehensiveness, fun, interest, ratio

[49] <www.meinprof.de>, <www.meinprof.at> and <www.meinprof.ch>

mark/effort. The most important criterion is recommendation: does the student recommend one of his fellows to follow this course?

For each course, a mean is computed for each of the evaluation criteria. Any future student can see, for each professor, which courses he gives and the evaluations of them. The site proposes also a « ranking » of professors and institutions.

3.4.1 Role of Identification in the Evaluation Process

While in a previous version of the sites any evaluation could have been done totally anonymously, this is not possible any more. In order to vote on a course, a student must now be registered. The registration process includes a validation of the e-mail. This new feature offers more possibilities to students; they can now modify their votes during a semester or after it. Professors can also register in the system. Once registered, the professor can access his courses. The registration step is composed of a manual validation of the e-mail. This validation only relies on the information given by the person registering. The only information available for the validation is the e-mail address of the new user, since the rest of the information (family name, first name, and institution) can easily be faked.

There is no strong identification of the professors. However, the manual registration is a step in this direction. For the students, the registration step does not include any manual identification; the site just verifies that a valid e-mail is given. The site sends an e-mail containing a link which has to be clicked by the user. There is no control on the identity of the student; it could be the professor himself for increasing his ranking or anybody having an interest in changing the results.

Once a professor is registered in the system, he can manage his courses. If a professor decides to participate in the process, the identification of students can become very strong.

The website gives the possibility to a professor to restrict the evaluation of his courses, so that only legitimate users can vote on a course. The professor can generate one time passwords for the site, and then distribute them among his students. A registered student will need such a password to give an evaluation for such a course.

3.4.2 Competitors

The *meinProf* web sites play two different roles. On one side, they use methods of traditional quality assessment process used in Universities for decades to evaluate courses: the teacher (resp. the University) asks the student for his opinion about a specific course, in order to improve its quality. On the other hand, the way the results are made publicly available intervenes refers on the ranking of the universities.

A competitor for the *meinProf* web sites for the evaluation of courses is, for instance, EvaSys[50]. This system is used to organize quality control in universities. Administrators can create new questionnaires, on paper or on the Internet. Each course can be evaluated. This system is internally managed by each university, so the workflow respects the policies of the institution. Teachers have the possibility to implement their own questionnaires, they can also be sure that the data concerning their course will remain protected. In evaluation systems managed by universities, students can give their opinion in an anonymous way, but the system

[50] See the Evasys for education web site: <www.electricpaper.biz/products/evasys-education.html>

is organized so that no student will give more than one opinion. There are two different ways to question students: using paper-form or using an Internet questionnaire. The paper-form is given to students and they have to fill it out inside the classroom. The forms are anonymous, but a teacher may see which student wrote what. In this case, the return rate is very high. Students can also have to fill an Internet form. In this case, the return rate may be smaller (depends on the organization). In case of EvaSys, both systems can be combined and handled with the same efficiency.

Such systems are very similar to *meinProf* from the point of view of the organization. They ask the students what they think about a given professor or course, but the results have not the same status. One important requirement for quality control is that data are not made publicly available outside the organization, the aim of *meinProf* is to publish them.

For the ranking of the institutions, the *meinProf* websites can also be compared with the various ranking of universities that are published every year. On the international level, the most famous is the ranking of the University of Shanghai[51,52]. This ranking has been created to rank the best universities all over the world on measurable criteria. The criteria used for this ranking are mainly bound to research. They are: number of alumni having acquired a Nobel prize or a Fields Medal in Mathematics (which is the best prestigious mathematic award); number of persons that received one of those prices while they were working at this university; number of researchers cited in the most prestigious journals; number of articles published in Nature and Science (the two most prestigious scientific journals); number of articles referenced in the Science Citation Index-expanded. A bonus is also given to small institutions. Such a ranking works for internationally oriented universities, but is useless to compare two small universities on the national level. It is unlikely that they have a Nobel award winner among their alumni and, since all indicators are directed towards research, they are not reflecting the teaching quality.

Some other rankings use the fame of an institution as main indicator of quality[53]. Such ranking will be very stable, since the fame of an institution does not change easily. For instance, even if the Sorbonne has not existed since 1968, it is still cited among the most prestigious universities of France by many foreigners.

So even if the ranking provided by *meinProf* is not scientific, it gives the point of view of the students (or of some of them) on their institution (through the evaluation of their professors).

3.4.3 Is Anonymity Augmenting the Efficiency of the Tool?

The website protects the privacy of the students, as a professor does not know the persons who evaluate him. Since the site is independent of the University, the student may be quite certain about this. Moreover, the website gives no possibility for a professor to opt out, so there is no way for a professor to restrict access to his results, neither can they be anonymized from the point of view of the teacher. The professors are therefore not motivated in the use of the service. They have no control on the way the information is published afterwards.

The main problem for the system is that some persons (that may not even be students) will be able to manipulate the results of one institution or one professor. It is easy to create a dozen of

[51] Academic ranking of world universities (Shanghai Jiao Tong University): <www.arwu.org>
[52] List of international university rankings relevant for Swiss universities: <www.universityrankings.ch>
[53] Swiss Up, THES: <www.swissupranking.ch>

new accounts and this is sufficient to ruin the score of any teacher. The professors can restrict the access to evaluate their courses, but this requires their active work.

The system could become successful if it finds a way to include the professors. With the help of the main actors, the websites would provide the right security, and could provide statistically relevant results. Once the results are significant from a scientific point of view, such a ranking will find its place in the various ways used by future students to select their university.

3.5 E-Mail and PGP

Electronic mail is one of the most successful applications of the Internet since the very beginning of its existence. It is the electronic equivalent of a traditional postal service[54] by which information in form of written documents and letters, typically enclosed in envelopes, are transferred from a source (the sender) to a destination (the recipient). The necessary organizational structure for providing such a postal service is called *postal system*. Modern mail is organized by national and privatized services, which are interlinked by international regulations and agreements. This makes sending letters between almost any two countries in the world relatively efficient, cheap, and reliable.

An important prerequisite for a postal system to function properly is an appropriate way of unambiguously addressing possible recipients. Any violation to the assumption of unique addresses may cause documents to be delivered to wrong recipients. As an example, consider two namesakes (same first and last name) who live in the same street, and suppose that a letter has been sent to one of them with a missing street number in the address.

3.5.1 Electronic Mail

The Internet has made the process of sending letter-like messages nearly instantaneous. As a consequence, correspondents are using electronic mail today not only where previously they would have used letters, but also to exchange messages where previously no correspondence would have taken place. Ambiguous addresses are less of a problem in an electronic mailing system, where addresses are usually unique by definition. This means that the delivery of a message with an incomplete or non-existent address is typically aborted by an error message.

One of the main differences between the traditional paper-based postal system and an electronic mailing system is the number of possible valid addresses of a single recipient. While traditionally the number of such addresses is very restricted, typically to one private and one office address, there is almost no upper bound in the electronic case. Due to the large amount of e-mail providers, many of them offering free accounts, obtaining a large amount of different e-mail addresses for various purposes is simple and cheap.

Another important difference is the fact that a traditional postal address usually provides some sort of a link to the physical location where the recipient normally lives or works (except in the case of P.O. Boxes). This is not the case in an electronic mailing system, where most addresses point to the providers respectively their mail servers, but not necessarily to the

[54] Historically, communicating by carrying written documents from a sender to a recipient is nearly as old as the invention of writing. However, the first documented use of an organized postal systems occur dates back to the ancient Egypt, where Pharaohs used couriers for the diffusion of their decrees (2400 BC).

users. Some people prefer to use their real names as a prefix of their e-mail addresses, but this is sometimes even mandatory. Depending on the purpose of a specific e-mail account, it is sometimes very useful to have an anonymous address, which cannot be tracked back to its user. In general, one can say that an e-mail address provides no (or at most a weak) link to its user, even if the prefix of the address contains a name.

The absence of a strong link between addresses and users creates an authenticity problem when sending or receiving electronic mail. For example, receiving e-mail from an address with a name in the prefix does in general not give enough evidence to assume that the message truly originates from a person with such a name. Moreover, since the standard e-mail protocol SMTP[55] does not provide any means to prevent faked sender addresses (the corresponding header entries "MAIL FROM:" and "From:" can be arbitrarily chosen), it is easy to send e-mails that seem to originate from somebody else. This opens the door for all sorts of misuse, e.g., it allows e-mail spammers to send mass mailings while hiding their identities.

Despite the numerous problems and vulnerabilities of simple SMTP-based electronic mail, it is one of the most widely used communication systems in the world today. Thus, the benefits of the simplicity and convenience of SMTP seem to outweigh the dangers and risks resulting from the above-mentioned authenticity problem. This implies that from the perspective of a single user, there must be a considerable amount of confidence in the correctness of the corresponding link each time an e-mail address is attributed to a sender or recipient. In some cases, this confidence may be grounded on direct evidence obtained from the holder of a given address (e.g., by exchanging business cards with the address printed on it), but if the designated holder of an address is unknown, such direct evidence may not exist or may be difficult to obtain. Another approach is to collect indirect evidence from a trustworthy third party who knows the designated holder of an address in person.

3.5.2 Pretty Good Privacy

Given the above considerations, it seems quite obvious that using ambiguous addresses in an electronic mailing system almost always (often implicitly) requires some sort of an evidence-based reasoning mechanism to establish the necessary confidence with respect to the links between addresses and their designated users. Unfortunately, this process is not supported by the SMTP-based e-mail system.

A simple solution to this problem is included in a component of the e-mail encryption software Pretty Good Privacy (PGP). The main purpose of PGP is to facilitate the encryption/decryption of electronic messages and the creation/verification of corresponding digital signatures (Zimmermann, 1994). Note that the layer on which PGP operates is on top of SMTP, i.e., only the content of the e-mail is encrypted and digitally signed, not the data transmitted in the header of the message. This means that the link between the author of the message and the sender address indicated in the SMTP header entries is still questionable. On the other hand, if a message sent over SMTP is digitally signed with PGP (or any other digital signature scheme), a strong link is established between the signed message and the person who is in control of the private key used to generate the signature. By verifying the signature using the corresponding public key, the link itself becomes verifiable. However, the problem

[55] *Simple Mail Transfer Protocol* (SMTP) is the Internet standard for e-mail transmission across IP networks. SMTP was first defined in RFC 821 and last updated by RFC 5321.

then is to ensure that the given public key belongs to the signer of the message. In other words, a digital signature scheme such as the one provided by PGP transforms the authenticity problem of electronic mail into an authenticity problem of public keys.

The authenticity problem of public keys is usually solved with digital certificates. There are two fundamentally different approaches. In a centralized approach, a central authority issues such certificates, by which the existence of a binding between the public key included in the certificate and a person with a certain name or e-mail address is ensured. The certificate itself is digitally signed by the issuing authority. Most e-commerce applications are based on a centralized approach, which presumes that all involved parties consider the central authority as fully trustworthy. The central authority plays thus the above-mentioned role of a trusted third party in the absence of direct evidence.

The PGP approach to public key authenticity problem is decentralized. This means that issuing certificates is no longer restricted to a central authority. In the context of PGP, the process of issuing a certificate is commonly called signing a public key. Someone with direct evidence about the validity of a link between a key pair and its holder is thus supposed to sign the corresponding public key. These signatures are then sent to a key server, which makes them available to other users who need to evaluate the validity of a given link. Note that the verification of a signed key requires authentic public keys of the issuers, which produces further instances of the same problem. The resulting certification graph represents the key mechanism of a PGP-like decentralized certification scheme.

3.5.3 The PGP Web of Trust

For a decentralized certification scheme to work, it presumes that at least some of the certificate issuers involved in a certification graph are partly or fully trusted. For this, PGP provides a tool called “Web of Trust”, which allows users to specify their own trust assumptions. These assumptions are then used to evaluate the validity of public keys based on a small set of pragmatic evaluation rules. For example, a public key is considered authentic, if it is either signed by a fully trusted issuer with a valid key or if it is signed by two partially trusted issuers with valid keys. The application of these rules and thus the evaluation of the public keys can be automated. Note that the simplicity of the rules makes the resulting algorithm simple and efficient, but also not very robust against counter-intuitive results. Such problems have been reported in the literature and corresponding improvements have been suggested (Haenni, 2005; Haenni et al. 2007).

To conclude, what is the main purpose of the PGP Web of Trust (and similar methods)? Primarily, it is a decentralized solution to the problem of linking an electronic message or document to a person. To do so, it first uses a digital signature to create a strong (and verifiable) link between the message and a public key. For the remaining problem of linking the public key to a person, it collects evidence in form of certificates issued by arbitrary third parties and combines them with individual trust assumptions of the one who wants to validate a link. The whole link validation is thus a reasoning process, which methods like the PGP Web of Trust are trying to automate.

3.6 Conclusion

The case studies presented above are a very small sample of applications in the field of trust and identification. However, they still let us infer some of the main problems to be considered in the sequel.

- There is the problem of defining the main ingredients and clarify the relation in-between, hence Section 4 will focus on trust and confidence as well as the processes of identification and authentication and the interaction between these concepts. Where in these processes do trust and confidence occur and to what extent are they indispensable for them to function adequately?
- In most situations, trust is – especially at first glance – the main issue, or more precisely, how and by what means do the parties involved get to be trusted by each other? Different means have been presented in the use cases, from an Internet trust provider to distributed trust in the sense of a web of trust. But in all cases, initial trust has to be established in some way.
- Who is the "real" trustee? Who is the "real" trustor? Sometimes this is not very clear or even not as important, but it is enough to consider for example the trust in a virtual entity. This brings us back to the problem of trust and virtual persons, which will be discussed in Section 7.
- Of importance is also the role of response mechanisms, consider for example the very simple means of feedback information as in the case of the Amazon or the eBay use cases.

4 Core Concepts and Definitions

The goal of this section is to pin down further the key terminology of this deliverable and to build up a common language for the remaining sections. Our discussion in respective subsections includes concepts such as trust, trustworthiness, confidence, identification, authentication, and virtual persons, which have all been objects of thorough investigation in previous FIDIS deliverables[56] or FIDIS publications[57]. This section gives a thorough introduction to these topics and recapitulates some of the most relevant findings. For further details, we refer to the respective deliverables and the FIDIS summit book (Rannenberg et al., 2009).

4.1 Trust and Confidence

In this subsection, we provide a compilation of definitions and concepts related to trust and confidence. Our main focus will be on the sociological understanding of trust as somebody's belief state about someone else's trustworthiness, but the discussion will be rooted in the computer science literature, particularly in papers on computational aspects of trust as found in so-called *trust metrics*, or more generally on *trust management* in distributed systems. For a more comprehensive overview of trust in computer science, we refer to several excellent surveys (Grandison, M. Sloman, 2000; Ruohomaa, Kutvonen, 2005; Artz, Gil, 2007). The goal of this subsection is to prepare the subsequent discussion in Sections 5 and 6, in which trust, confidence, and identification will be regarded from the perspective of virtual persons and vice versa.

4.1.1 The Difference between Trust and Confidence

Despite the wide variety of literature on *trust* with its diverging definitions and conceptualizations, there are a number of common themes and patterns. A widely adopted commitment is to view trust as a person's belief state about the future contingent actions and behaviour of others, and to understand the degree to which one party trusts another as the corresponding strength of belief (Castelfranchi, Falcone, 2001). Usually, the strength of belief depends on what the so-called *trustor* knows about the so-called *trustee*, which is why trust relationships between closely related persons (e.g., between family members, friends, or partners) tend to be stronger than trust relationships between less related or unrelated persons. In a strong trust relationship, a trustee is presumed to meet the trustor's expectations, which stem from interactions or explicit agreements or promises. Trust is therefore directed towards the trustee's actions or behaviour in the *future*, and it should thus be seen as a *prediction* of reliance in the absence of full knowledge or control (Sztompka, 1999). From this point of view, trust is a mechanism by means of which an individual compensates a shortage of knowledge to obtain a feeling of control, and therefore a way of dealing with social complexities and uncertainties that goes beyond rationalistic reasoning (Numan, 1998; Luhmann, 2000).

To enlarge the flexibility of this initial position of (predictive) *behavioural trust*, we also allow retrospective trust relations with respect to the authenticity or credibility of statements from a source of information. This particular form of *informational trust* is important in

[56] See in particular FIDIS deliverables D2.13, D17.1and D17.2.
[57] See for example the FIDIS booklet "Identity [R]evolution" (Jaquet-Chiffelle, 2009) and the FIDIS summit book (Rannenberg et al., 2009).

semantic web and PKI applications, where the truth of a statement or the claim of the origin of a piece of information is a crucial factor for drawing reasonable conclusions. Informational trust usually leads thus to the acceptance of the truth of a statement in the absence of conclusive evidence. As a mental state, both behavioural and informational trust is inherently subjective. Note that informational trust can always be interpreted as a retrospective form of behavioural trust.

In the English language, trust is almost synonymous with *confidence*. The origin of confidence is the Latin verb *confidere*, which is a composition of the prefix *con-* (expressing intensive force) and the Latin verb *fidere* (to trust). Confidence could therefore be translated simply as *strong trust*, which implies that confidence is less general and coarser than trust, but there are further, more subtle differences in its meaning and common use today. Note that a complicating issue in this respect is the fact that other European languages like Dutch, French or German do not provide separate words for trust and confidence.

One of the most visible differences between trust and confidence is the type of relationship they describe. While behavioural or informational trust in the above-mentioned understanding always refers to a relationship between *two* entities (the trustor and the trustee), only *one* entity is involved in a state of confidence (the holder of that state). In general, somebody's confidence is not restricted to another entity's expected behaviour, but could involve any hypothesis or claim about open questions or predictions of future events. From this perspective, confidence is more general than trust and can be described as a state of being certain that a hypothesis, claim, or prediction is correct. If the hypothesis, claim, or prediction involves a decision or action, then confidence refers to the belief that a chosen decision or course of action is the best or most effective.

Another conspicuous difference between trust and confidence is the granularity of their respective scales. While trust is usually perceived as a continuous quantity with infinitely many degrees, confidence is often regarded as a concept of "*strong trust*" (see Latin origin above) with two possible states (e.g. *yes*/*no*) only. We do not entirely adopt this dichotomous view of confidence in this deliverable, but we will talk about confidence *thresholds* in Subsection 4.2, which define the level of confidence for accepting the truth of a hypothesis

The next noticeable difference between trust and confidence results from their respective opposites. To capture the meaning and nuances of trust to the full extent, it is well recognized in the literature that distrust should be separated properly from no trust (Lewicki, 2006), but there is no such plurality of opposite positions with regard to confidence. This limitation is a direct consequence of the restricted granularity of confidence. We will further discuss the difference between distrust and no trust in Subsection 4.1.2.

Some authors do also exhibit further differences between trust and confidence. In (Cofta, 2007), for example, confidence is defined as the composition of control and trust. It is argued that a state of confidence is only possible if those aspects, which are not under the subject's own control, are covered by corresponding trust assumptions. This is another interesting and valuable perspective, but we will not further consider it in this deliverable.

4.1.2 Defining Trust

Instead of repeating some of the most general definitions from the trust literature, as it is done e.g., in (Artz, Gil, 2007), we suggest here an incremental approach, in which the meaning of trust is narrowed down from a small set of very general, higher-level concepts (e.g., trustor, trustee, trustworthiness) towards some more elementary, lower-level concepts (e.g.,

competence, honesty, belief). By doing so, we attempt gradually to approximate the principal characteristics of trust and to incorporate the most important aspect and trends mentioned in the literature, but we are aware that the result will be far from delivering a complete picture. For some of those concepts, we will give some rough ideas about a possible formal description, but we will try to keep the necessary mathematics as simple as possible. The whole hierarchy of concepts is depicted in Fig. 10. For a more comprehensive overview of trust definitions and trust-related concepts, we refer to the excellent classification in (McKnight, Chervany, 1996).

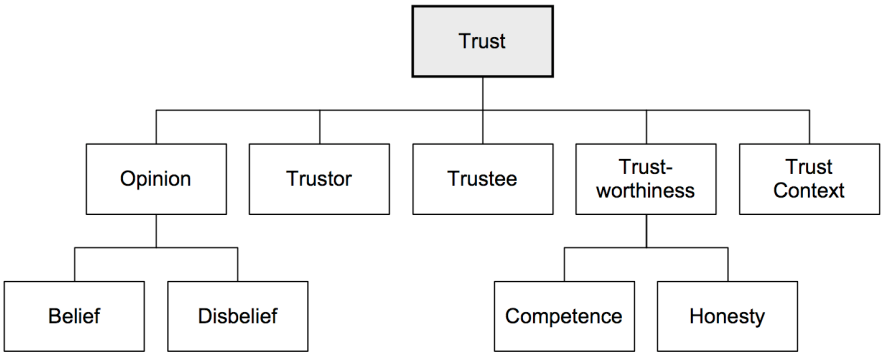

Figure 10: An incremental approach for trust-related concepts

Definition 1: *Trust (distrust) is the positive (negative) opinion of a trustor about a trustee's trustworthiness relative to some trust context.*

By saying that trust and distrust are *opinions*, we try to incorporate the above-mentioned idea of a mental state. This is consistent with the cognitive or epistemic view of trust, so-called *knowledge-based trust*, in which the strength of a trust relation depends on the trustor's available information or evidence about the trustee's trustworthiness.[58] It also includes the dynamic aspect of trust, because opinions may change non-monotonically over time when new information is accumulated. Finally, it also covers the aspect that most people experience trust as a matter of degree, because our conceptualization of an opinion as a composition of respective states of belief and disbelief is inherently a quantitative one (we will further pinpoint the relation between opinions and belief/disbelief states in Def. 8). Note that the three most basic positions of a qualitative approach (full trust, full distrust, zero trust/distrust) are included as extreme borderline cases. This allows us to distinguish properly between zero trust and distrust, an aspect that is well recognized in the literature (Marsh, 1994; Cho, 2006; Lewicki, 2006).

4.1.3 The Characteristics of Trust Relationships

Reduced to a relationship in the sense of an entity-relationship model (ERM), trust may thus be considered as a ternary relation, trust(*trustor*,*trustee*,*context*), which depends on three principal parameters (Jøsang et al., 2006; Hardin, 2006). Some authors suggest seeing trust as a binary, context-independent relation (Zhang et al., 2004), but this is only sufficient in applications with a general, predetermined trust context, or for extreme positions such as blind trust or paranoia. Other authors propose a quaternary relation with time or evidence as

[58] Some authors make a fundamental distinction between *cognitive* and *emotional* forms of trust (Lewis, Weigert, 1985).

an additional parameter (Gutscher, 2007), but those aspects are implicitly included in our ternary model as particular characteristics of the trustor's mental state. Other proposed parameters in the literature are the trustor's or trustee's future actions and goals, implied risks, or third-party authorities (Castelfranchi, Falcone, 2001). Some of those additional aspects are at least partly included in our general understanding of a trust context (see Def. 4).

An immediate question that arises from the ERM perspective of trust is whether certain properties hold for trust relationships. According to (Gutscher, 2007), the relation is *asymmetric* (*A* and *B* may not trust each other to the same extent), and *non-reflexive* (*A* does not always trust itself). In some specific situations, for example with respect to some indistinguishable members of a certain group, we may observe a one-to-many relation of equal trust between a trustor and several trustees. Many-to-one or even many-to-many trust relations are also thinkable, but only in some very particular situations (Grandison, Sloman, 2000). Those aspects are all well reflected in our cognitive view of knowledge-based trust.

A more controversial issue in the literature is the question of whether trust relations are *transitive* or not. In a transitive trust relation, it holds that if *A* trusts *B*, and *B* trusts *C*, then *A* trusts *C*. While transitivity of trust is the primary mechanism for deriving *indirect trust*, a key concept in most trust metrics or trust management systems (Mahoney et al., 2005), the majority of authors agree that general transitivity is only exhibited under very specific circumstances (Christianson, Harbison, 1997). For a given or fixed trust context, one such scenario results from the assumption that all agents share a common or centralized repository of trust-relevant information about each other. This may be achieved by mutually exchanging the information about existing direct trust relationships by means of recommendations or credentials (Jøsang et al., 2006). From a global perspective, interpreting a common or centralized trust repository corresponds to computing the transitive closure of the trust relations (Branchaud, Flinn, 2004). Our proposed perception of knowledge-based trust is generally not transitive, but it inherently supports such particular scenarios without difficulty.

4.1.4 Trustor and Trustee

To make the above-stated general definition of trust more accurate, let us first have a closer look at the possible subjects and objects of a trust relation. From the ERM perspective, it would be most natural to consider them generally as entities, which are not further specified except that they have a distinct existence. To emphasize that trust is a mental state about an entity's future behaviour or its credibility as a source of information, which requires entities with the capability to "think", "speak", and "act" (in a very broad sense), we prefer to call them agents, as it is common practice in economics and artificial intelligence. A cognitive agent is one that reasons, decides, speaks, and acts on the basis of the evidence from and knowledge about the agent's environment.

Definition 2: *A trustor is a cognitive agent with the ability to collect trust-related evidence and use it to form corresponding opinions about the trustworthiness of others.*[59]

Usually, we can think of a trustor as a living physical person, but our definition as a cognitive agent also includes software agents, autonomous robots, or intelligent machines as possible trusting parties. In a similar way, we may also consider groups or whole communities as subjects of a trust relation, as long as they act as if they were a closed unit with a common

[59] Note that by this definition we not impose any form of *rational* behaviour.

view of trust-related knowledge and respective opinions. More generally, we will allow the range of possible trustors to include various forms of virtual persons (see Section 7).

Definition 3: *A trustee is an agent whose actions and statements are (partly or fully) unpredictable, respectively unverifiable to others.*

As above, we may most typically think of a trustee as a living physical person, but our definition is again compatible with software agents or robots, groups or communities, and other virtual persons. As a borderline case, it is even compatible with all sorts of machines, systems, devices, or tools, as long as their future behaviour is not entirely predictable. In such cases, we may also replace in our consideration the machine itself by "the one(s) who built the machine". Nevertheless, we consider trust interpersonal with respect to the two agents involved, but not necessarily with respect to individual human beings.

As a final note, we want to stress that a trustee may not necessarily be different from the trustor. With this we explicitly allow reflexive trust relations, so-called *self-trust*. In many trust metrics, self-trust plays an important role as a base case on which recursive trust propagation methods are rooted.

4.1.5 Trust Context

The third parameter in the above-mentioned trust relation, trust(*trustor*,*trustee*,*context*), copes with the fact that a particular agent may not be equally dependable with respect to all different types of actions or statements. A trustee may thus be a perfect service provider or a reliable source of information in its own area of expertise (e.g., medical advice), but at the same time be completely unreliable in some other area (e.g., IT support). In other words, the level of trust attributed to an agent strongly depends on the context towards which the trust is directed. Some authors refer to it as context-specific trust (Branchaud, Flinn, 2004).

Definition 4: *A trust context is a particular class of actions or statements, which are not further distinguished when judging an agent's trustworthiness.*

A trust context limits thus the application of trust to a specific purpose or domain of action. In the literature, trust context is the most commonly used term, but essentially the same concept is sometimes called *trust scope* (Kohlas, 2007), *trust class* (Beth et al., 1994), *trust category* (Kinateder, Rothermel, 2003), *trust domain* (McLeod, 2006), or *subject-matter dependent trust* (Mahoney et al., 2005). A particular trust context is Maurer's concept of a *trust level* (Maurer, 1996), where a hierarchy of trust contexts reflects a particular form of inter-contextual dependencies.

The dimension of a particular trust context is application-dependent and may thus vary significantly. On one side of the spectrum, the context consists of a single, very specific action or statement (e.g., "to return the rental car in time"). Such a restricted context allows the trustor to draw very accurate conclusions, but it makes the process of collecting trust-relevant information more difficult. On the other side of the spectrum, we may consider a very broad and general class of actions or statements (e.g., "to follow orders"). The generality of such a context facilitates information gathering and enlarges the range of possible conclusions, but those will generally be less accurate. Choosing an appropriate specificity for a trust context is thus a trade-off between simplicity and accuracy.

In the computer science literature on trust metrics, many authors find it convenient to assume a single *generic trust context*, one that covers all possible types of actions and statements. Other authors prefer to work with a fixed *principal trust context* for the whole application. A

particular principal trust context, the so-called *issuer trust context* (Haenni et al., 2007), refers to an agent's ability to issue meaningful credentials about a third party's authenticity and trustworthiness. Issuer trust is thus the same type of trust that is required to accept public-key certificates in centralized and decentralized PKIs (see Subsection 3.5.3). In those types of applications, trust has a self-referral component, which imposes a cascade of related trust contexts on different layers (Maurer, 1996; Haenni et al., 2007). We refer to (Kohlas, 2007) for a more profound discussion of such interlinked trust contexts.

In various attempts to classify different forms of trust in the literature, the distinction between different trust contexts has often been a key parameter. One such proposal foresees the distinction between *service provision* trust, *resource access* trust, *delegation* trust, *certification* trust, and *infrastructure* trust (Grandison, Sloman, 2000). Further trust classes of that kind are *authentication* (or *identity*) trust and *system* trust (Jøsang et al., 2007). Another possible classification scheme distinguishes between *concrete* or *material* trust (e.g., "to pay the restaurant bill") and *abstract* trust (e.g., "to keep promises"), which is similar to the above-mentioned distinction between *behavioural* and *informational* trust.

4.1.6 Trustworthiness

The concepts of trust and trustworthiness are closely related, but they turn out to be quite distinct upon closer inspection. Most authors separate them sharply (Hardin, 2004; O'Hara, 2004; Ashraf et al., 2006). While trust is usually perceived as an attitude or mental state of the trustor, trustworthiness is a trustor-independent property of the trustee. A trustworthy person is thus someone in whom we can place our trust without any risk of being disappointed or betrayed. This position, which is based on the trustor's expectation of the trustee's trustworthiness, is the dominant view in the literature, especially in behavioural economics and psychology. Some authors refer to it as expectation-based trust or calculative trust (Rotter, 1980; Williamson, 1993). For example, psychological studies have shown that known trustworthiness implies trust but not vice versa (Chaudhuri, Gangadharan, 2007). Many of those studies have used variants of the trust game proposed in (Berg et al., 1995) to measure trust and trustworthiness.

Despite the obvious differences between trust and trustworthiness, one can always imagine a borderline case in which "*those whom we trust will be trustworthy, and those who are trustworthy will be trusted*" (McLeod, 2006). This would then be the ideal situation of so-called *objective trust* (Zhang et al., 2004), in which the trustor has full knowledge about whether or not, or to what degree, the trustee is actually trustworthy in some context. As it is usually not very realistic to assume full knowledge or even objectivity, we may see the difference between trust and trustworthiness to result from the trustor's incomplete epistemic state. The difference between trust and trustworthiness is thus another aspect of our subjective, knowledge-based perspective of trust, according to which various trustors may attribute to a particular trustee quite different levels of trust.

Perceiving trustworthiness as an agent's property gives rise to a number of implied questions. One of those questions concerns the existence of more fundamental components, on which trustworthiness is grounded. The following definition involves *competence* and *honesty* as necessary prerequisites for trustworthiness. While competence is a property that refers to the agent's capability to do and to say the truth about what it is trusted for, we use honesty for the agent's respective commitments and intentions. Therefore, competence mainly differs from honesty in the lack of a *motive*.

Definition 5: *Trustworthiness is an agent's compound property of being competent and honest with respect to the actions and statements in some trust context.*

Definition 6: *Competence is an agent's ability to act dependably and to make truthful and relevant statements in some trust context.*

Definition 7: *Honesty is an agent's will to act dependably and to make truthful and relevant statements in some trust context.*

In the discussion about the possible decomposition of trustworthiness, the absence or presence of motivational elements is the most common distinguishing feature in the literature (Castelfranchi, Falcon, 2001; McLeod, 2006). Competent but dishonest agents are sometimes called *malicious* (Kohlas, 2007). Note that some authors suggest a more detailed decomposition of trustworthiness with up to four different basic components (McKnight, Chervany, 1996; Oliver, Montgomery, 2001; Gefen, 2002).

To cover as many trust-related aspects as possible, both competence and honesty should be regarded as respective general terms for a large number of similar properties with subtle differences. While competence involves many motive-independent factors such as know-how, expertise, accuracy, efficiency, skill, proficiency, qualification, capability, dependability, power, strength, or experience, we use honesty as a general term for motive-dependent properties such as sincerity, benevolence, goodwill, loyalty, faithfulness, truthfulness, responsibility, adherence, incorruptibility, integrity, discreetness, or fairness. Note that the latter may depend on the intended *recipient* towards which the action or statement is addressed. This is a direct consequence of the fact that an agent's motives to act are highly recipient-dependent, which is most apparent in properties like loyalty or faithfulness. Some authors use recipient-dependence to distinguish trustworthiness from mere reliability. It follows then that "*people known or considered to be trustworthy have the power to betray us, whereas people known or considered to be merely reliable can only disappoint us*" (McLeod, 2006).

While most authors have recognized the importance of the motivational element that might underlie trustworthy behaviour, there are still various discrepancies with regard to its exact nature. The most dominant position discussed in the literature is the *goodwill view*, according to which a trustee acts out of goodwill toward the trustor (McLeod, 2006). Another type of motive is addressed in the *risk-assessment view*, in which the trustee's own risk determines its behaviour (Jones, 1999). This view is a generalized form of the *social contract view*, in which the force of social constraints compels the trustee, and the *encapsulated interest view*, where trustees are motivated by their own personal interests. Note that recipient-dependence is mainly present in the goodwill view.

Another important characteristic of trustworthiness follows from the observation that both competence and honesty are highly time-dependent. While somebody's level of competence may gradually improve with practice or new experiences, it may quickly decrease in the absence of such practice or as consequence of the agent's natural aging. Honesty may exhibit a similar non-monotonic behaviour, depending on whether the motives on which it is grounded change over time. An agent's trustworthiness is therefore a dynamic property, which is subject to constant changes. Note that the dynamics of trustworthiness as an objective property is quite different from the dynamics of trust as a subjective epistemic state.

To complete our analysis of trustworthiness, a few additional words need to be said about the observation that competence, honesty, and therefore trustworthiness are usually perceived as a

matter of degree. Note that the granularity of such degrees may depend on the context. In a very specific context, which consists of a single action or statement only, we may actually not need more than a pair of extreme values, e.g., 1 for "*fully trustworthy*" and 0 for "*not trustworthy*", but this may not suffice for a less specific context. A common approach in the trust metrics literature is to define the degree of trustworthiness as the proportion of action and statements, for which the trustee is actually trustworthy, relative to the total number of action and statements in the current trust context (Gutscher, 2007). This definition as a proportion allows degrees of trustworthiness to be interpreted as probabilities and to apply the probability calculus to compute all sorts of related quantities. For example, assuming probabilistic independence between competence and honesty allows degrees of trustworthiness to be computed as the product of respective degrees of competence and honesty (Kohlas, 2007).

4.1.7 Opinions

Most proponents of the knowledge-based perspective of trust, in which trust is regarded as a subjective attitude or mental state, agree that trust is more than a simple belief state with respect to the trustee's trustworthiness. The main concern comes from the observation that the absence of trust is clearly quite different from distrust. For example, one may not trust a stranger, but almost certainly distrust a notorious cheat or liar. In other words, trust has two different opposites, one in which the trustor's epistemic state is insufficient to draw any meaningful conclusion about the trustee's trustworthiness, and one in which the trustor's epistemic state leads to the conclusion that the trustee is actually untrustworthy. The absence of trust is sometimes called untrust (Marsh, Dibben, 2005). As most of the trust literature focuses on the positive aspect of trust, it is restricted to the simple dichotomy between trust and untrust. To include negative aspects of trust on an equal footing with positive aspects of trust, we suggest here a trichotomy between (positive) trust, untrust, and distrust (negative trust). This differentiation is necessary for agents to make use of distrust in their decision making in the same way they use trust or untrust, or more generally to establish a full symmetry between trust and distrust, with untrust as a neutral intermediate state.

An elegant way of capturing such a trichotomy is to detach the concepts of belief and disbelief. We depart thus from the dominant *probabilistic* (or *Bayesian*) view in the literature on representing uncertainty, in which degrees of belief and disbelief are represented by an *additive* pair of values $Bel(h) \in [0,1]$ and $Bel(\neg h) \in [0,1]$, respectively, for which $Bel(h)+Bel(\neg h) = 1$ holds for all hypotheses h and their negations $\neg h$, and thus implies that degrees of belief are determined by respective degrees of disbelief and vice versa. One simple way to detach them from each other is to relax the additivity requirement into the inequality $Bel(h)+Bel(\neg h) \leq 1$, or to remove the restriction altogether. The main advantage of such *sub-additive* or *non-additive* degrees of belief is their ability to properly represent states of partial or full ignorance (Haenni, 2009), and this is exactly what is needed to separate trust, untrust, and distrust. Recall from Def. 1 that we use the notion of an opinion as an additional concept to realise the separation between belief and disbelief. To make the following definition of an opinion more flexible and general, we prefer to talk about *states* of belief and disbelief rather than degrees of belief and disbelief, but without specifying what these states are and what granularity they have.

Definition 8: *An opinion about a hypothesis is the composition of an agent's respective states of belief and disbelief with respect to some hypothesis under consideration.*

Definition 9: *A belief state is an agent's cognitive attitude towards the truth of a hypothesis.*

Definition 10:*A disbelief state is an agent's cognitive attitude towards the falsity of a hypothesis.*

In our particular area of application of opinions as a means for defining trust, the hypotheses under consideration are statements like "*Agent X is trustworthy in context C at time T (towards recipient R)*", and the knowledge and evidence to consider includes both first-hand experience and second-hand credentials about the trustee's trustworthiness. The process of collecting and taking into consideration such trust-related evidence is an important feature of the proposed knowledge-based perspective of trust. It gives a concise explanation of the dynamics of trust, e.g., as a three-stage process of trust building, trust stability, and trust dissolution (Oliver, Montgomery, 2001).

Depending on the concrete way of representing states of belief and disbelief, we may obtain quite different forms of opinions. One of the simplest forms results from restricting the representation of belief and disbelief to their opposite extremities of full belief/disbelief and no belief/disbelief. If those extremities are represented by Boolean values 1 and 0, and without further restrictions, this delivers four different opinions (0,0), (1,0), (0,1), and (1,1), which may be interpreted as respective states of ignorance, belief, disbelief, and inconsistency. While it is common to exclude inconsistent opinions such as (1,1) by imposing the above-mentioned sub-additivity requirement, we can use the remaining consistent opinions (0,0), (1,0), and (0,1) to represent respective borderline cases of full untrust, full trust, and full distrust.

The most obvious generalization of this simple scheme is to relax the restriction to Boolean values. In the context of trust representations, there are various proposals for less restrictive discrete scales with values such as "*untrusted*", "*marginally trusted*", "*fully trusted*", and "*ultimately trusted*" (Abdul-Rahman, Hailes, 2000; Li, Singhal, 2000; Ruth et al., 2004), but most authors agree that they should be mapped into a continuous scale such as the unit interval [0,1] or the percentages scale [0,100]. The trust and distrust propagation method proposed in (Guha et al., 2004) uses such a general scheme, in which trust and distrust values are entirely detached from each other, and where trust and distrust calculations are performed independently on respective matrices. Each particular state of trust and distrust corresponds then to a point in the unit square [0,1]x[0,1].

If the above-mentioned sub-additivity property is imposed to exclude inconsistent belief states, half of the unit square is chopped off. The result is a so-called opinion triangle, which is bounded by the extreme opinions (0,0), (1,0), and (0,1). It is common to depict this space by an equilateral triangle and corresponding barycentric coordinates b, d, and $i=1-(b+d)$ for respective degrees of belief, disbelief, and ignorance (Jøsang, 2001; Haenni, 2009). This picture is shown on the left hand side of Fig. 11.

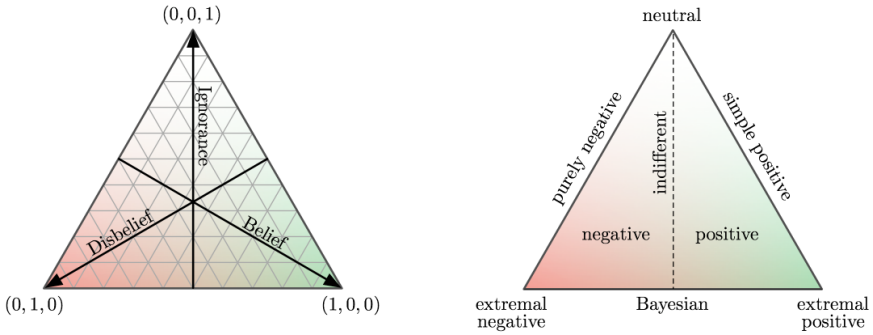

Figure 11: Left: The opinion triangle with its three coordinates for belief, disbelief, and ignorance; Right: different opinion classes

In a more mathematical setting, it is common to define opinions as additive triplets (*b*,*d*,*i*), for which $b,d,i \geq 0$ and $b+d+i = 1$ hold. According to the terminology proposed in (Hájek, Valdés, 1991), an opinion (*b*,*d*,*i*) is called *positive* if $b>d$, *negative* if $b<d$, *indifferent* if $b=d$, *Bayesian* if $i=0$, and *simple* if either $b=0$ or $i=0$. The borderline cases (1,0,0) and (0,1,0) are called *extremal*, whereas (0,0,1) is called *neutral*. All those particular types of opinions are depicted on the right hand side of Fig. 11. We propose to adopt the same terminology for corresponding states of trust and distrust.

Opinions as suggested above are the elementary structures of the *opinion algebra*, which has been applied as a calculus for indirect trust (Jøsang, 1999). Note that there are many strong links between opinions of that kind and various mathematical approaches to non-additive degrees of belief (Haenni, 2009). Some of them include the dominant Bayesian paradigm of representing uncertainty by probabilities as a special case. If applied to the problem of representing trust and distrust, they are thus compatible with the probabilistic interpretation of trust and trustworthiness.

As an alternative to the above opinion-based representation of trust and distrust, some authors suggest to apply other mathematical structures such as *fuzzy sets* (Griffiths, 2006), *Dempster-Shafer belief functions* (Yu, Singh, 2000), *probabilistic argumentation systems* (Haenni, 2005), or *imprecise* and *second-order probabilities* (Gutscher, 2007). We do not further address those directions here, but we want to point out that they are all motivated by the same fundamental problem of detaching distrust from untrust, or more generally uncertainty from ignorance. The same can be said about attempts to represent trust and distrust as a single continuous variable in a range like $[-1,+1]$ or $[-\infty,+\infty]$ and with 0 as a representation for untrust (Marsh, 1994), but those do not reach the full generality of our opinion-based definition of trust.

4.2 Identification

In the common language, *identification* means “the act of identifying” or “the state of being identified”. The etymological origin of *identify* comes from the medieval Latin “identificãre”: to make resemble. *To identify* means “to establish the identity of”, “to ascertain the origin,

nature, or definitive characteristics of", "to consider as identical or united; equate", or "to associate or affiliate (oneself) closely with a person or group".[60]

Identification of an entity can lead to an "equality": the unique characterization of this entity within a group (the act of establishing the identity of), or to an "inclusion": the membership of this entity to a certain group of entities sharing some properties (to associate or affiliate closely with a person or group).

In scientific literature, we do not see a universally accepted definition yet. The ISO draft[61] definition for identification is "process of an entity providing evidence to a level of confidence to an identity authority that this entity can be recognized within some context with unique identity references and/or any additional information that characterizes the entity". We read further, in the same document, "Identification is the process that will recognize the entity within this context by assigning identity references to an entity and observing additional entity attributes that in aggregate uniquely distinguish the entity from others within this context." The ISO approach only tackles the "equality" side of identification; the identification as a member of a group, the "inclusion" side of identification is ignored. Even though we essentially agree with the ISO draft definition (except for the missing "inclusion" side of identification), we think that time is an important parameter that should be explicitly recognized in the definition.

In (Kent et al., 2004), identification is "the process of using claimed or observed attributes of an individual to infer who the individual is". But who is an entity? What does it mean to know who an entity is? We perceive this definition to be either trivial or too vague. Very often, a person is considered to be identified when the person's name is known. But does it mean that we know who he *is*? As an example, in the movie "Meet Joe Black"[62], Bratt Pitt plays the role of Death. Bill (William Parish played by Anthony Hopkins) is chairman of the Parish Company. Drew is a board member:

Drew – There is a second question the board would like a response to.

Far simpler one.

Who is the man standing to your left?

Bill – I have already introduced Mr. Black to you all.

Drew – But who is he? What are his credentials? What is his relationship to you?

Moreover, contrarily to the ISO draft definition, Kent et al.'s definition only applies to individuals.

In (Adams, 2005), Carlisle Adams gives a very interesting overview of identification and identity related concepts in his article on pages 272 and 273, where he introduces authentication of a particular claim as a sub-process of the identification.

None of these definitions incorporate the time factor. All of these only cover the "equality" side of identification.

---

[60] American Heritage Dictionary, fourth edition

[61] ISO/IEC 4th Working Draft 24760 - Information technology - Security techniques - A framework for identity management.

[62] By Martin Brest, Universal Studios, 1998

4.2.1 Definition of Identification

Our definition is broader than the previous ones and emphasizes the time-dependency of identification:

Definition 11: *Identification is the process of establishing enough confidence in the fact that some identity-related information is valid and truly describes a specific entity in a given context or environment, at a certain time.*

If we want to restrict identification to its "equality" side, as mostly done in scientific literature, it is sufficient to replace "identity-related information"[63] by "identifying information"[64] in our definition. Our definition is therefore more general; it covers the "equality" side, as well as the "inclusion" side of identification.

Identity-related information might truly describe an entity in a given context, at a certain time, while being wrong for this same entity in another context, at the same time. For example, each country of a dual citizen might only recognize its own citizenship (and ignore the other one). A Swiss-American dual citizen is not allowed to enter the USA with his Swiss passport. In the USA context, his Swiss citizenship does not truly describe him. On the other hand, his Swiss citizenship truly describes him in Switzerland. When he travels in France, both citizenships truly describe him.

The context or environment in which the identification takes place can be purely virtual, i.e., the product of one's mind. This happens, for example, when we dream of somebody that we know. The identification happens in the context of our dream. We relate identity-related information that truly describes a human being with an image in our dream, an entity that only exists in our thoughts.

The validity of the identity-related information is often time-dependent. For example, a username is only valid after having been created in a system, and becomes invalid as soon as it has been revoked. In a PKI, there must be a revocation mechanism to invalidate a key if, for example, it has been compromised.

The fact that identity-related information truly describes a specific entity in a given context or environment, is also often time-dependent. Identity-related information might truly describe an entity in a given context, at a certain time, while being wrong for this same entity in the same context, at another time. For example, the last name of a person can change after his wedding; the age changes every year; the place of residence only stays constant as long as the person does not move. As another example, let us consider the context of a given family (the father, the mother and their two children), and "The tallest member in this family" as some identity-related information. The tallest member in this family was the father as long as both children were small; but when the eldest child was 14, he became the tallest member in the family until his younger brother grew even taller. Some other identity-related information (usually) stays the same during the whole life of an individual: place of birth, date of birth, etc.

An important category of identification schemes is related to authorization, access control, etc. In this case, the time of reference is the time when the identification takes place: *identification* is the process of establishing enough confidence in the fact that, when this

[63] *Identity-related information* is any information that characterizes (uniquely or not) an entity.
[64] *Identifying information* is any information which characterizes exactly one entity within a specific context or environment. (FIDIS deliverable "D2.13: Virtual persons and Identities").

process takes place, some identity-related information is valid and truly describes a specific entity in a given context or environment.

Identification might happen when the entity does not exist anymore. This is for example the case when DNA traces, recorded from a crime scene, lead to the identification of a dead criminal. In such a situation, the time of reference is in the past (e.g., the time when the crime was committed) and does not include the time when the identification takes place. The same might be true for the identification of victims in a cataclysm, when only parts of the bodies stay available, i.e., when the entity as a whole does not exist anymore. More generally, a trace is the result of a past activity. The entity that originated the trace can evolve (or even disappear): a shoe does not leave the same mark when it is new or when it has been used. The corresponding identity-related information is therefore not constant over time. The same is true for the traces left on a bullet by a gun.

We cannot exclude that the time of reference is in the future when compared to the time of identification, but it does not necessarily make the identification fail. As an example, the identity-related information "President of the USA on January 31st 2009" could have led to a successful identification of Barack Obama at Christmas 2008, even though it was not possible to completely exclude alternatives. The movie "Minority Report" is based on another example: premonition. In this movie, the character played by Tom Cruise can see crime scenes a few minutes or even a few hours before the crime actually takes place. This interval of time is used to identify the "future murder" and to arrest him before the actual crime takes place. Of course, arresting him makes his actual identification as a real murder fail (even in the future) because the future has been changed through the intervention of the crime prevention team.

4.2.2 Types of Identification

In order for an "identificator" – the one doing the identification – to start an identification process, at least one entity or some identity-related information must be given to him. This leads to four main types of identification:

Type of identification	Entity[65]	Information[66]
Verification of identity	x	x or x…x
Finding suitable identity-related information	x	
Finding a suitable entity or suitable entities		x
Profiling, categorizing	x…x	x…x

Figure 12: Types of identification

If the identificator is being convinced that some given identity-related information is valid and truly describes only one specific entity in the given context, then this entity is (fully)

[65] A "x" in this column means that exactly one entity is given. A "x…x" means that several different entities are given to the identificator.

[66] A "x" in this column means that all the information given to the identificator relates to one entity. A "x…x" means that the given information comes from several (at least 2) different entities.

identified, i.e. *individualized*, with this information in this context from the point of view of the identificator. If the identificator convinces himself that other entities can be truly described by this information in this context, then the identification is only partial from his point of view.

4.2.2.1 Verification of Identity

If both one entity and some identity-related information are given in a context, then the identificator is a verifier. He must try to establish enough confidence in the fact that the given information is valid and truly describes the given entity in this context, at a certain time. If the context is not given, the verifier must furthermore determine a suitable context. This is a one-to-one verification.

Typically, in such a situation, the entity actively participates in the process in explicitly claiming some alleged identity-related information. For example, a user gives his username to a server in order to access a service, or a traveller presents his passport at the border control.[67]

4.2.2.2 Finding Suitable Identity-Related Information

If only one entity in a context is given to the identificator, then he must try to find identity-related information for which he can establish enough confidence in the fact that this identity-related information is valid and truly describes the given entity in this context, at a certain time. This is a one-to-many identification. Usually, the participation of the entity is passive: the entity only gives implicit information.

A typical example is when the identificator sees a previously known person and tries to remember the person's name or role, and the context in which they met before.

If the situation requires it, the identificator might continue searching identity-related information that is valid and truly describes the given entity in this context, at a certain time, until the aggregation of all this information uniquely characterizes this entity in this context, at that time, i.e., until the entity is individualized.

4.2.2.3 Finding a Suitable Entity or Suitable Entities

If only some identity-related information in a context is given to the identificator, then he must try to discover an entity or entities for which he can establish enough confidence in the fact that the given identity-related information is valid and truly describes this entity or these entities in the given context, at a certain time. This is a many-to-one identification. If the context itself is unknown to the identificator, he must also determine the relevant possible contexts. The determination of the context or contexts –the "frame(s)" (Kind, 1987) – is part of the task of the forensic investigator.

A typical example comes from traces (identity-related information) retrieved from a crime scene: a fibre, a fingerprint, a hair, blood or some other biological traces, etc. The forensic investigator verifies the validity of the information and tries to find an entity (a person, maybe an animal or even an object) or entities that could be truly described by some or all of this identity-related information, at the time of the crime. Note that the exact time of the crime is often supposed or estimated; this is an extra source of uncertainty that should not be underestimated. In this context, an entity is truly described by some identity-related

[67] See also (Birch 2009) for an interesting discussion of a modern national identity scheme.

information if its participation in the crime, or its presence, could explain the existence or the origin of the trace.

4.2.2.4 Profiling and Categorizing

This type of identification differs from the previous ones, as entities or information has to be sorted into groups or categories. For example, many entities are given to the identificator, who needs to find a way to classify them, i.e., to find relevant identity-related information to define groups sharing some attributes, categories of entities, etc.

When a lot of information is given to the identificator (or profiler, in this case), he needs to sort out this information in order to infer knowledge either to define relevant categories or to find new expected identity-related information (e.g., expected behaviour, preferences, etc. of an individual), before linking this information to specific entities.

4.2.3 Sub-Processes

The identification process as described above can be subdivided into two sub-processes: a selection phase and an authentication phase.

4.2.3.1 Selection Phase

At this stage, the identificator has to determine one or more 4-tuples {entity, identity-related information, context, time of reference} that are potential candidates for a successful identification. The exact selection process depends on different factors: What is known or given to the identificator? Is the entity actively and explicitly involved in the process, or is it passive? Does the entity only give implicit information? Is it sufficient to reach a partial identification or not?

If the entity actively claims some identity-related information in a given context, the identificator (verifier) gets an alleged identity or an alleged partial identity of this entity in this context. Otherwise, the identity-related information is inferred, supposed or observed by the identificator himself. To do this, the identificator might rely on external sources of identity-related information.

When the identification consists in verifying the identity of the entity, only one 4-tuple is selected, and the time of reference is the time when the identification takes place.

For each 4-tuple {entity, identity-related information, context, time of reference} which has been selected, the identificator induces a corresponding claim:

> Claim: "This identity-related information is valid and truly describes this entity in this context, at that time."

The successful authentication of such a claim leads to a successful identification.

4.2.3.2 Authentication Phase

In (Kent et al., 2004), authentication is defined in a very general way, which goes beyond identification:

> "*Authentication* is the process of establishing confidence in the truth of some claim."

Note also that authentication is considered as a *process* too, not as a stable situation or the result of the process.

We prefer a slightly different formulation, which emphasizes the existence of different levels of confidence.

Definition 12: *Authentication is the process of establishing enough confidence in the truth of some claim.*

Our definition keeps the generality of Kent et al.'s definition. For instance, it is possible to authenticate a document as dating from the Middle Ages. In this case, the claim is "This document has been written in the 10th century", and authenticating it means establishing enough confidence in the truth of that claim, i.e., in the fact that the document was truly written in the 10^{th} century.

The level of confidence that needs to be established for the authentication to be successful depends on the application and on the consequences of a false authentication, or, in other words, on the risks that the identificator is ready to assume.

Therefore, during the authentication phase of identification, the identificator needs to estimate the risks related to an impersonation, i.e., to another entity being wrongly identified as an entity truly described by the identity-related information. Depending on the consequences, he adjusts the level of confidence that he requires for the authentication to be successful. For example, the level of confidence does not need to be the same to grant somebody access to an online newspaper or to a nuclear facility.

In order to establish enough confidence in the claims that he has induced, the identificator usually relies upon authentication technologies, i.e., upon some (hopefully) corroborating evidence produced either by the entity itself or by (trusted) third parties.

The evidence can be the password related to a username, a PKI certificate produced by a certification authority, a passport, a biometric feature, or, for example, a way of typing a passphrase on a keyboard.

Three basic classes of authentication technologies are commonly accepted:

- something you know (secret, password, etc.)
- something you have (certificate, passport, token, etc.)
- something you are (biometric feature: fingerprint, iris, DNA, etc.)

We (as many others) consider an additional one:

- something you do (pace, dynamic signature, etc.)

Those authentication technologies are often combined to increase the confidence of the identificator in the truth of his claim: token, like chip card, with a pin code (two-mode authentication), AXSionics Internet Passport™ (three-mode authentication), etc.

The relations between all four classes of authentication technologies are summarized in the following figure:

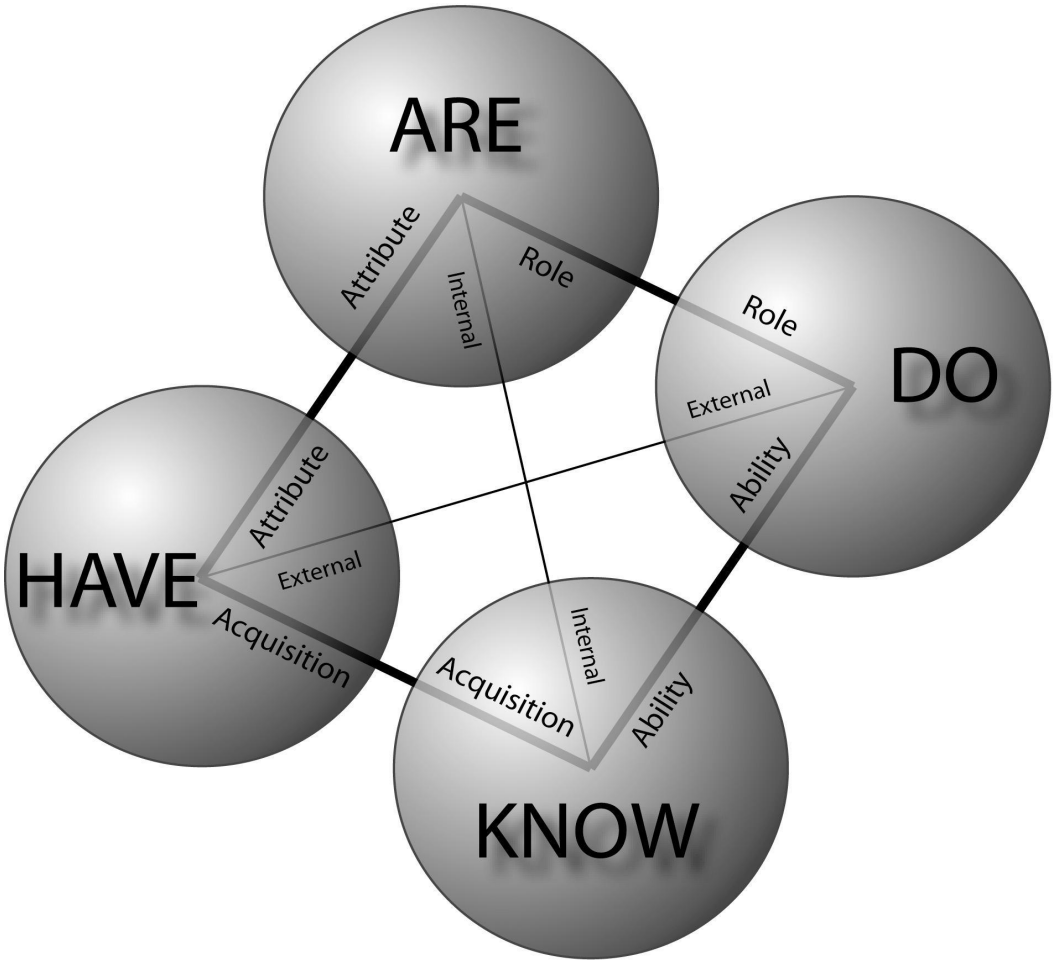

Figure 13: Relations between all four classes of authentication technologies

During the authentication phase, the identificator needs to answer the following fundamental question: Is this entity truly described by this information in this context, at that time?

The answer to this question ("yes" or "I do not know") is based on a threshold mechanism. The higher the level of confidence is, the closer to full confidence the threshold will be.

Fundamental question:

Is the entity truly described by this information, in this context, at that time?

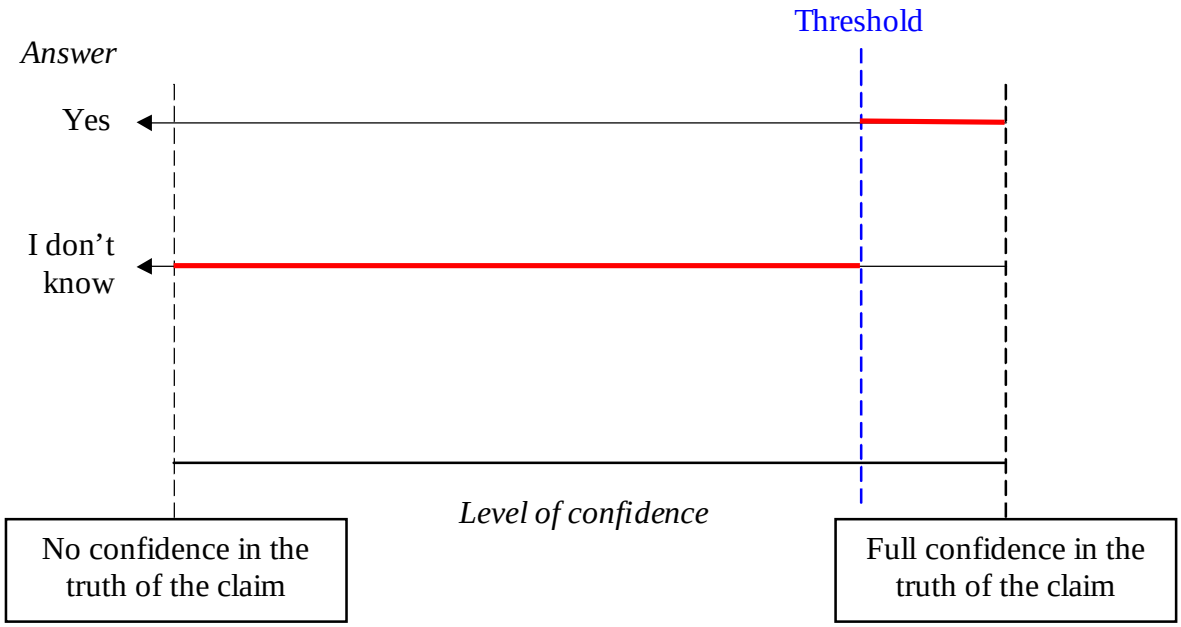

Figure 14: Decision with threshold (authentication)

When authentication fails, i.e., when the answer to the question is "I don't know", the identificator might try to be more precise and to answer another question: Is this entity *not* described by this information in this context, at that time? This is what we will call a refutation process:

Definition 13: *Refutation is the process of establishing enough confidence in the untruth of some claim.*

For example, passengers of an aircraft might be checked *not* to belong to a list of known terrorists. Exoneration is another example: a refutation process based on DNA can prove a suspect not to be at the source of the trace which was strongly supporting his participation to a crime.

Some authentication technologies, like face recognition, are actually more efficient when used as refutation technologies: for example, automatic face recognition can allow people that are proven *not* to belong to a certain list of individuals to pass the border with a simplified procedure, while others need to be checked further.

The higher the level of confidence is in the untruth of the claim, the closer to full confidence in the untruth of the claim the threshold will be.

Fundamental question (refutation process):

Is the entity truly described by this information, in this context, at that time?

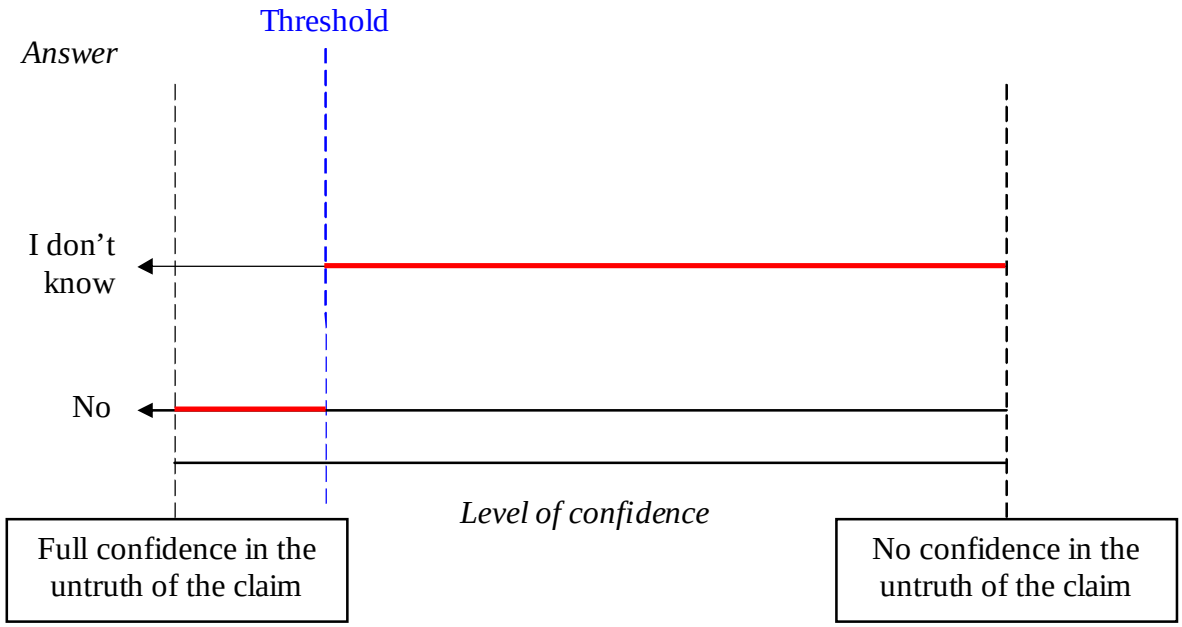

Figure 15: Decision with threshold (refutation)

Authentication and refutation processes might use different tests. However, with some important authentication technologies, the test is the same. This is in particular true with many authentication technologies based on biometrics. Indeed, if the correlation between two templates is high enough, they are supposed to come from the same entity, while if the correlation is low enough, they cannot come from the same entity. When the correlation is in between, no definitive conclusion can be drawn.

In this case, we have a decision with two thresholds: a positive threshold for a successful authentication and a negative one for a successful refutation.

Fundamental question (authentication and refutation combined):

Is the entity truly described by this information, in this context, at that time?

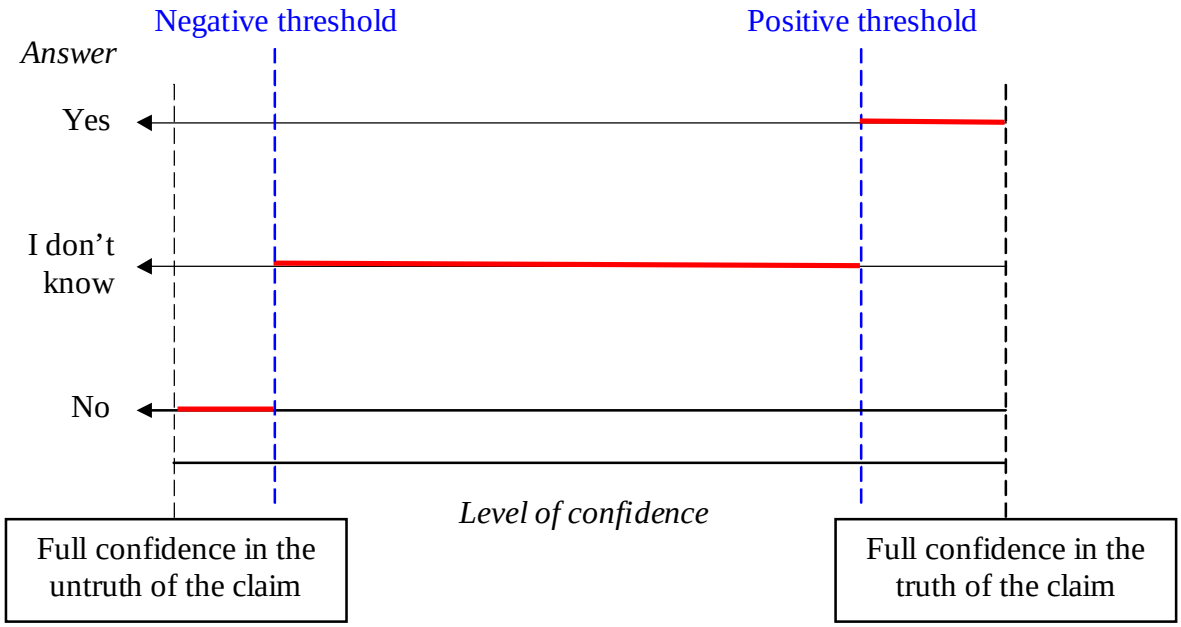

Figure 16: Decision with two thresholds

4.3 Virtual Persons and the Virtual World

In FIDIS deliverable "D2.13: Virtual Persons and Identities", we have defined a two-layer model. The basic conceptual elements in this model are entities. An entity is the fundamental "thing" that can be identified. Entities in the model are either physical or virtual. The nature of existence of a physical entity is material, i.e., some sort of physical constituent is compulsory.[68] The nature of existence of a virtual entity is abstract, which means essentially that the entity is or has been a product of the mind or imagination.[69]

Virtual persons sometimes describe avatars and new forms of identities in online games. They also appear in other contexts; some authors use them in the legal domain. Within FIDIS, the concept of virtual person has been extended in order to better describe and understand new forms of identities in the Information Society in relation to rights, duties, obligations and responsibilities.

Some virtual entities can have rights, duties, obligations and/or responsibilities associated to them: for example, *"the CEO of Apple Inc."*, *"the President of France"*, *"the owner of a house"*, avatars in online games, etc. Other virtual entities cannot have such rights, duties,

[68] The existence of a physical entity is time-dependent and the lifetime of a physical entity is usually bound in time. At any specific point in time, a physical entity either exists or not. For example, a physical person will no longer be considered as a physical entity after his death, when his body will have disappeared.

[69] Virtual entities are thus entirely detached from any physical reality. They typically belong to concepts, thoughts, perceptions, illusions, categories, or abstractions.

obligations and/or responsibilities associated to them: for example, *"the Sun"*, *"a white sheet of paper"*, etc.

This allows us to make a clear distinction between those two categories of virtual entities. A virtual entity that can have rights[70], duties, obligations and/or responsibilities associated to it in a certain context is called a "*virtual person*". This is one of the key concepts of our model that relates identities with rights, duties, obligations and/or responsibilities.

For any specific point in time, the collection of all existing physical entities is what we call the "*physical world*" at that specific time and the collection of all existing virtual entities is what we call the "*virtual world*" at that specific time.[71] The virtual world –not to be confused with the digital world– allows a unified description of many identity-related concepts that are usually defined separately without taking into consideration their similarities: avatars, pseudonyms, categories, profiles, legal persons, etc.

The legal system has a long experience of using abstract entities to define rules, categories, etc., in order to associate legal rights, obligations, and responsibilities to persons that can be considered instances of these abstract entities in specific situations. The model developed within FIDIS intentionally uses a similar construction.

4.4 Conclusion

In this section, we have defined and analyzed the core concepts used in this deliverable: trust, trustworthiness, confidence, identification, and authentication. We have tried to capture the essence of these concepts in the classic real world context. This thorough study led to an in-depth understanding of these concepts and of the relations between them that goes beyond the former state-of-the art. This study becomes eventually one of the main outputs of this deliverable.

The precise definition of *identification* brings the traditional IT approach closer to what is usually adopted in forensic sciences. In the IT community, we see too often confusion between identification and individualization. We hope that this precise, but general, definition of identification will bridge both communities on this core concept.

[70] Rights, duties, obligations and responsibilities are not restricted to the legal domain only: rights include access rights, for example.
[71] By assuming the existence of both a physical and a virtual (non-physical) world, our model implements a dualistic view of the reality.

5 Feedback or Response Mechanisms for Trust

5.1 General Response Mechanisms

5.1.1 Classic Response Mechanisms

Response mechanisms do exist in various environments: crisis management, the eco system, electricity, gaming, etc. In the context of this section we discuss response mechanisms, which enable trust or regulate behaviour. We discuss the mechanisms of social pressure, reputation, and liability as examples. First of all, we describe those mechanisms as they exist in the classic physical world, and in the next subsection they are referred to in the digital context.

Social pressure

Social pressure consists of comments, criticisms, attitudes and emotions of people directed against other people who have a deviating opinion on a certain issue. It is a method of changing and controlling the behaviour of other people, both on the "micro" level (i.e. among acquaintances, friends and family), and also on the "macro" level (as a technique of social engineering and control).

Reputation

In order to be trusted individuals as well as organizations strive to establish a favourable reputation. Although an implicit reputation doesn't necessarily imply direct interaction, it can nevertheless be used as data on which to base one's judgement of trustworthiness. This doesn't mean that information concerning a reputation will be interpreted the same way by different people. For instance, some people may have a rather negative perception of direct marketing companies. This entails that those people will approach such companies with a certain amount of distrust.

When it comes to international business, a similar line of reasoning can be held for associations one might have with particular regions. An individual or organization might have more trust in an organization based in Europe than in a politically unstable country.[72]

Liability

Liability is an important legal concept. It refers to the legal responsibility one has for one's acts or omissions. Failure of a person or entity to meet that responsibility leaves the person or entity open to a lawsuit for any resulting damages or a court order to perform. The awareness of the existence of a legal responsibility will influence the behaviour of an entity. However, it will not influence the behaviour of all sorts of entities in the same way. A person with less money to compensate damages might take other risks than a person with a healthy bank account.

[72] Florian Egger, Consumer Trust in E-Commerce: From Psychology to Interaction Design, in: Trust in Electronic Commerce, Kluwer Law International, Den Haag, 2002, p.32.

5.1.2 Response Mechanisms in the Digital World

The response mechanisms described in the previous sections have proven to be of value in the physical world. Will they be applicable in the digital world as well? Below, examples are given of the same response mechanisms applied in a digital environment with impact in the physical world.

Social pressure

Interesting examples of the use of social pressure in the digital world took place in Facebook, the social utility that connects people with friends and others who work, study and live around them.

One example is about HSBC, one of the bigger banking and financial services organisations in the world. This bank has a network of offices in Europe, the Asia-Pacific region, the Americas, the Middle East and Africa.

The HSBC story centres around the bank's decision to change the terms of the interest-free overdraft facility it offered its student customers. The change was met with strong resistance by a large number of students, who felt that the bank had reneged on the deal it had originally offered them. Shortly after HSBC´s decision in Facebook a co-ordinated group emerged which had one goal: to reverse the measure of the bank. The group in the end had around 6,000 members and threatened to boycott the bank unless it addressed their concern.

Soon after the first boycott threats of the Facebook group, and largely due to the growing pressure from the Facebook group, HSBC announced that it had decided to reverse its decision.

This HSBC story shows social pressure can be organized quickly via a social network utility like Facebook, and can even force a leading company to reverse decisions.[73]

Another example of public pressure via social networks resulted in the re-introduction on the market of the Wispa chocolate bar by Cadbury, a company behind some of the world's favourite chocolate confections.

The Wispa bar sold successfully for over 22 years until, in 2003, it was taken of the shelf. This withdrawal immediately caused uproar of Wispa-worshippers. At a certain moment in time Facebook counted 93 'Bring Back Wispa' groups, with almost 14,000 members, and further strong support from similar groups on Myspace and Bebo. Cadbury's couldn't resist this pressure and put the bar back on shop counters a short while later.[74]

Reputation

On auction sites users buy and sell things from others across the nation and around the world. Giving this opportunity, stress is put on some of the foundations of the traditional market place. In traditional markets, a buyer can usually inspect the shelves thoroughly before buying. Beyond this, a seller's reputation, possibly built over many years, including the cost of a physical presence and a place gained in a community, provides a signal about the seller's

[73] Students celebrate Facebook triumph over HSBC, Guardian, 30 August 2007
[74] Cadbury announces Wispa comeback, BBC News, 4 August 2008

ability and keeps the seller honest and diligent. Sales over the Internet lack these tools of reputation.

Internet marketplaces have been forced to find a substitute for the traditional reputation of a seller. Several systems have been introduced to enable the distribution of reputational information. These systems collect information on the past behaviour of a seller, and then make that information available to potential future transaction partners. Because people know that their behaviour today will affect their ability to transact in the future, opportunistic behaviour is deterred. Moreover, less reliable players are discouraged from joining the marketplace. Reputation systems seek to inform buyers about whether potential trading partners are trustworthy, and thereby avoid cheating.

Internet markets also have advantages in establishing reputations after all; information can be collected easily and continuously. Information can and will be disclosed to millions of potential customers, with the advantage that the information will not alter along the way as often is the case by spreading information by word of mouth.[75]

Liability

In section 5.2.2 below liability is discussed from the point of view of an entity, an avatar, roaming the Internet and causing possible damages. Who will be reliable for what under which circumstances? Liability, however, can also play a role in regulating behaviour in the digital world. As a matter of fact, liability played a negative role in the smooth functioning of the internal market of electronic commerce in Europe. Different liability rules in the Member States, and as a result different outcomes in case law, hindered the development of cross-border services and produced distortions of competition. To solve those problems with liability issues the European Commission decided to draw up a directive on electronic commerce, which, amongst others, had as a goal to streamline liability issues in the emerging European market of information society services.[76]

The directive, which had to be implemented in the Member States beginning of 2002, created exemptions from liability in specific circumstances. They cover cases where the activity of the information society service provider is limited to the technical process of operating and giving access to a communication network over which information made available by third parties is transmitted or temporarily stored, for the sole purpose of making the transmission more efficient. This activity is of a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored.

Intermediaries can benefit from the exemptions for mere conduit[77] and for caching[78] when they are in no way involved with the information transmitted. So they should not modify the information. Of course, this does not cover manipulations of a technical nature which take

[75] Paul Resnick, et al., The Value of Reputation on eBay: A Controlled Experiment, Experimental Economics, Volume 9, Issue 2, Jun 2006, Page 79-101.
[76] Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.
[77] See article 12 of the directive of e-commerce.
[78] See article 13 of the directive of e-commerce.

place in the course of the transmission as they do not alter the integrity of the information contained in the transmission.

An intermediary who deliberately collaborates with one of the recipients of his service in order to undertake illegal acts goes beyond the activities of mere conduit or caching. Therefore he cannot benefit from the liability exemptions established for these activities.

In order to benefit from a limitation of liability, the provider of an information society service, consisting of the storage of information, upon obtaining actual knowledge or awareness of illegal activities has to act expeditiously to remove or to disable access to the information concerned.

5.1.3 Conclusion

Traditional response mechanisms as social pressure, reputation, and liability seem to function in the digital world as well. The first and the latter work perfectly well to influence behaviour at a micro level, a product or service, but also at a macro level, the functioning of a market. For reputation, digital systems have been developed, which seem to be a perfect substitute for the traditional off-line reputation of sellers and buyers.

The functioning of these mechanisms in the digital world even has advantages compared to the physical world: a greater speed and scale can give the response mechanisms described an enormous impact.

5.2 Legal Response Mechanisms

Trust is central to all transactions. It is taken for granted that consumers meet their budget constraints and that producers always deliver the goods and services they said they would. The question is how can we be sure of the rectitude of these agents? We do not know if they are persons of honour or if there is a background agency which enforces contracts, credibly threatening to mete out punishment if obligations are not fulfilled. This punishment should be sufficiently stiff to deter consumers from ever failing to fulfil them. Legal response mechanisms in the real world context, as well as the digital world, resort to liability regulations as the response to that kind of behaviour. However, from a broader societal point of view this would come along with a loss of honour, social and economic ostracism. Nowadays we would call this naming and shaming. Our readiness to trust depends to a large extent on the nature of the setting in which we act. At the end of the line, society may nurture a trusting climate by setting in place, various forms of trust insurance to provide safety nets for those whose trust is betrayed. A simple example is the arrangement of liability for fraud. The next step is a legal response, which should be sought in the context of a term in a debtor's prison. Furthermore the enforcement agency itself must be credible. Otherwise the threat or the response mechanism is not functioning. The agency will only be trustworthy if it will do what it says and only what it says.

Another legal mechanism is that of contracts and agreements. In a certain way, these are in fact props for creating sufficient trust. In fact, no contract can detail every eventuality. Language cannot cope with unlimited refinement in all contingencies. Trust covers expectations about what others will do or have done in circumstances that are not explicitly covered in the agreement. In many cases there may not even be a contract. In the end, trust is much more frequently and additionally based on reputation. Reputation is ultimately acquired

through behaviour over time in well-understood circumstances. In order to acquire trust in the online as well as the offline world, certification is an important measure as well. Unfortunately the third party liability for the organisations providing these certificates has not been adequately settled in the law. Only the contracting party can enforce damage payment related to losses suffered relying on certificates. Nevertheless, article 6 of the Directive of the European Parliament and the Council of 13 December on a Community Framework for Electronic Signatures (99/93/EC) does set forth a third party liability clause for Certification Service Providers.[79]

5.2.1 Warranted Trust and TMO's in the Digital World

Putting it into a wider context of the fundamental goal of law, which finds its roots in the philosophies of Aristotle, it may be argued that law should seek to inculcate habits of good conduct and should support a social environment which will encourage citizens to pursue worthy goals and to lead valuable lives.[80] Thus law and ethics complement each other. Ethics sets the basic societal interests that law should guarantee. If we extend this principle to the codes of conduct in the digital world, it is easiest to take as a starting point the principle of 'what applies offline should apply online'.[81] The question is then, how can trust be guaranteed in the online setting? The question of trust lies at the heart of the transposition of the world as we know it, to the digital world. Guarantors of trust such as accountants, auditors, notaries, surveyors function in this realm. Trying to create a system of liability in the virtual world, which can be effectively enforced, is a daunting task. Organizations such as Trusted Third Parties (TTP's) and Trust Mark Organisation (TMO's) have tried to fill in the gap. A recent exponent of this theory has been Paolo Balboni.

In his thesis, Paolo Balboni, introduces the ethical theory of "warranted trust". This defines the trust relationship between a Trust Mark Organization (TMO) and an e-consumer. This aims to protect the trustor reliance on the trustee by implementing a regulative framework, which takes into consideration the interest of both parties and the influence of the specific context in which the trust relationship develops.[82] The concept of warranted trust is further developed as the trust relationship, which the TMO provides, as being information about the good intention of the trust-parties, their competence, their compliance with a regulative framework, and there being in place an adequate regulative framework for the TMO.[83] This leads Balboni to the conclusion that "Trust is warranted when a liability system which protects what e-consumers value (e.g., security, privacy and fair business practice) considering the TMO's context of action (i.e. the Internet) is in place."[84]

As argued by Rolf Haenni[85], trust is the opinion of a trustor about a trustee's trustworthiness relative to some trust context. Trustworthiness is an agent's compound property of being competent and honest, with respect to the actions and statements in some trust context. Pertaining to the trust context, it can be said that the digital world creates a setting, which is

---

[79] OJ L 13, 19 January 2000, p. 12.
[80] Craig, E., The Shorter Routledge Encyclopaedia of Philosophy. New York, Routledge, 2005.
[81] Koops, B.J., et al., ICT Law and Internationalisation. A Survey of Government Views, Law and Electronic Commerce. Vol. 10. The Hague, Kluwer Law International, 2000.
[82] P. Balboni. Trustmarks. Third-part liability of trustmark organisations in Europe. Thesis. Universiteit van Tilburg, 2008, p. 213.
[83] *Ibid, p.217.*
[84] *Ibid,* p. 222
[85] See section 4.1.2.

unfortunately hampered by an expectations gap. Dematerialization, internationalization and technological turbulence are specific features of the digital world. In this dynamic context, a virtual entity cannot be traced. E-consumer data can be transferred to some other location on the Internet with a simple mouse-click. Suppose the mouse-click is performed by an avatar, how can this avatar action be credited to an avatar endowed with the virtue of honesty? Competence seems a character trait which can be objectively measured. Honesty is much more subjective.[86] Law may function to expand the radius of trust and make it more likely that individuals will trust those outside their inner circle of family and close friends. In the e-commerce world, TMO's try to waive any possible liability towards e-consumers on the accuracy of the information they provide on e-merchant security, privacy and business practices through trustmarks. By the concept of warranted trust, it will be possible to take into consideration both TMO's and e-consumers' needs and expectations and to translate them into a legal provision, that is adequate for both sides.[87] As such, the aim of what applies offline should apply online can be said to be achieved. TMO's, by applying the concept of warranted trust, adopt a position comparable to those of accountants, surveyors, auditors and notaries. In this manner, the concept of liability is brought into play.

5.2.2 Liability in the Digital World

Increasingly independent agents, such as avatars, roam the Internet even though they do so often in compliance with the instructions given to them. Many potential instances of damage are conceivable. The agent may corrupt a database consulted by it. There may be a fault in its software, which is the reason why it sells rather than buys shares. It may become contaminated by a virus and, subsequently, it infects a website. It may be 'kidnapped' by a killer agent before reaching its owner. Factors such as the autonomy of the agent, the mobility of the agent and the dynamics of the environment, the agents operate in, are of particular relevance here.[88] The incorporation of a recall mechanism and the possibility for the user to destroy the agent by remote control could help to limit the risks. Its flexibility could be limited by increasing its degree of security. Certification may be assigned a role here as well. The agent and the objects sought by them are assigned a certain certificate containing data about authorization, security etc. The agent would thus be prevented from interacting with an object or other agent, that does not have a corresponding security level.

To solve the problem of the attribution of liability for damage in relation to the use of agents, Karnow[89] has proposed a system, by which agents only can be used when they are registered in a registry. In this registry, the certified certification authority files the assigned certificates of the agents. In the certification process, the risks, attached to the use of that particular agent, *e.g.* what is the risk of the agent operating outside its environment, what decisions can it make itself, are registered. The registry will issue a certificate on the basis of its evaluation. Karnow proposes that the agent will be able to interact only with systems that also bear a certificate,

[86] See subsection 4.1.6 and (McKnight, Chervany, 1996; McLeod, 2006)
[87] Balboni, p. 214.
[88] Heckman, C. and J.O. Wobbrock, Liability for autonomous agent design, in Autonomous agents and Multi-agent systems, Kluwer Academic Publishers, 1999, pp. 87-103.
[89] Karnow, C.E.A., Liability for distributed artificial intelligences, Berkeley Technology Law Journal, 1996, p. 98.

issued by the registry. If damage occurs, the registry will make payment without any investigation into fault or causal relation.[90]

An additional problem with respect to the question of liability, is, that liability in the law requires causal agency. Liability questions are often solved by referral to the principle of what is reasonably foreseeable. This depends on custom and common sense. In a fast changing environment such as the Internet, Karnow points out that in this context, reasonable foreseeability is a moving target.[91] But does not this situation therefore pose a huge problem for creating a trusted environment, because that cannot operate without foreseeability and predictability? And indeed, Karnow warns, the probability that Artificial Intelligences[92], that operate in the real world, with decision programs making decisions unforeseen by humans, will lead to insuperable difficulties, posed by the traditional tort system's reliance on the essential element of causation.

5.2.3 Legislative Efforts for the Digital World

In this subsection, the legal infrastructure for successful e-commerce will not be discussed. Only those aspects relevant from the point of view of creating a trusted environment for the e-commerce and the more digital environment of the gaming world, receive attention here. Nevertheless, elements, which are of significance in the e-commerce environment, are worth noting in this respect, because, in the Second Life setting, transactions are also concluded. The following topics play a role in relation to the conclusion of contracts with the help of agents:

- How can the identity of the consumer and the supplier be established?
- How can the authenticity of the exchanged information be determined?
- How can evidence of the transaction be provided?
- How can it be determined whether the parties are authorized to enter into a contract?
- How can the identity of parties who act through agents be established?[93]

New solutions will have to be provided in the case of autonomous agents given the difference between human actors and intelligent agents. What legislative efforts seem promising to provide a solution? An early effort will be described in the next paragraph.

In 2000, the Uniform Computer Information Transactions Act (UCITA)[94] was one of the first legislative efforts, which recognizes the existence of intelligent agents in carrying out transactions. UCITA uses the term 'electronic agent' to mean "a computer program, or electronic or other automated means, used by a person to initiate an action, or to respond to electronic messages or performances, on the person's behalf without review or action by an individual at the time of the action or response to the message or performance". In other words, the UCITA is based on the concept of an autonomous agent. Still, the UCITA proposes a simple resolution to the question of the attribution of responsibility for the actions of the agent. The person, who uses the agent, is bound by the consequences of the actions of

[90] See also Stuurman, K. and H. Wijnands, Sofware law, Intelligent agents: a curse or a blessing? A survey of the Legal aspects of the application of intelligent software agents, Computer Law and Security Report, vol. 17, no. 2, 2001, p. 96.
[91] FIDIS deliverable D17.2, p. 58.
[92] AI's are the predecessors of electronic agents and in some sense, avatars.
[93] Stuurman (2001), p. 98.
[94] <www.law.upenn.edu/bil/ulc/ucita/citam99.htm>.

the agent. From a point of view of the trust aspects of the contracting actions performed by the agent, it is interesting that the UCITA puts much emphasis on the communication skills of the agent. The UCITA provides that "The terms of a contract (…) do not include a term provided by the individual if the individual had reason to know that the electronic agent could not react to the term." Therefore the extent of the knowledge of the agent plays an important role. This, in turn, raises the question whether the electronic agent itself is able to indicate what it knows and does not know. It is suggested, that a system of certificates could also provide a solution for this. Moreover, the UCITA contains the term of "a reasonably configured electronic agent". When this agent is a party to a transaction, it should be able to react to conspicuous terms in a contract without review of the record by an individual. The UCITA can be said therefore to provide a lot of leeway already for an autonomous agent to act autonomously. A legal basis is thus laid for a trust context, in which an agent can be said to be competent to act and make statements which are trustworthy.

5.2.4 Criminal and Civil Law in the Digital World

Point of departure in the virtual world is, that problems are solved internally as much as possible. The provider almost always has the opportunity to intervene effectively, whenever undesirable situations arise. It is still extremely rare, that the digital world calls in the assistance of the real world. In April 2007 Linden Labs asked for the assistance of the FBI against persons who expressly overloaded the servers of Linden Labs.[95] This situation is no longer tenable because the digital and the real world are mingling more and more. For example, virtual objects are traded for real money via eBay. Regulation of the digital world takes place at three levels.

- The participant in a virtual world has a contract with the provider.
- The avatars in the virtual world, their mutual relationship and the relationship with the virtual objects. Social conventions still prevail here.
- Legal rules can be made to apply to the relationships between the avatars and with their objects. These are adopted or transposed from the real world.

A proper balance still needs to be struck between when a government feels the need to interfere and when the virtual environment can solve its own problems. This occurs in the case of theft of a virtual good.[96] A provider cannot act here because the object in question is still within its realm. The players however think different. The government will probably only really interfere when more people complain and the economic importance starts to assume larger proportions. Similarly the almost limitless privileges the provider assumes in his Terms of Service may soon be questioned successfully because it can be argued, that they are to be qualified as an unreasonably aggravating action.[97] Consumer protection may lead to the change of these Terms of Service. It can hardly be maintained, that a provider holds the copyright of all that has been developed in its virtual world. Linden Labs incidentally recognizes the Digital Millennium Copyright Act. This act has a Notice-and Take-Down procedure. This means, that Linden Labs is obliged to remove content that breaches copyright. This is a step in the right direction.

[95] 'Second Life calls the FBI', <www.sirlin.net/archive/second-life-calls-the-fbi/>
[96] <virtuallyblind.com/2007/11/14/habbo-theft/>
[97] Dutch Civil Law, art 6:236 BW (onredelijk bezwarend beding)

Most legal experts would agree that criminal law is not the avenue to follow to tackle successfully the problems around virtual identities.[98] The simple reason for this is that the perpetrator is only virtually known. Criminal pursuit is therefore difficult if not impossible. In the MUVE and MMORPG world, swindling is part of the game. Hacking has been recognized as a crime in the sense of trespassing into a computer system with the intention of changing or removing data. Another possible option in criminal law is libel or perhaps even discrimination. This can occur when data is connected in the course of profiling. Theft is also hard to prove in the case of stealing a virtual good. It is a valid question whether a virtual good is also a property. Furthermore, the investigative authorities often will not conduct an investigation because they give these investigations low priority, the information is very difficult to obtain and there are also judicial competence problems.[99]

In civil law, when there is a failure by one of the parties there can only be compensation. However, in the case of Hyves[100] and Second Life[101], the providers exclude their responsibility for any damage whatsoever in their Terms of Service. In general, it can be expected, that the parties involved act reasonably and fairly. At the very least, it can be expected, that the provider warns about the risks involved in using the site. If the problems in the virtual world are not in the contractual sphere, recourse can always be had to the principle of the unlawful action in civil law. An essential aspect in this context is that a breach of law can be demonstrated. This can be the case when copyright is harmed in the case of an avatar, which has been developed by someone. In the virtual world, however, rules of the game or codes of conduct play an important role. These are not normative but they do play an important role in judging whether, for example, there has been negligence. It appears as if the data protection regulations do not help much either. Very often, users of virtual worlds have accepted the contractual provisions, in which they surrender their personal rights.

5.3 Conclusion

Response mechanisms such as social pressure, reputation and liability play a role in the digital world as well as the real world in contributing towards trust. Creating trust can be an important element of ensuring truthful identification of an entity in the online world. The new environment of the digital world will not immediately have a full-fledged regulatory system in place to allow for the required trustworthy setting in which e.g. the performance of commercial transactions can take place. Online relations such as Facebook and Second Life but also the gaming world call for secure identification because the traditional face-to-face identification is impossible.

In this chapter it is shown that the role that traditional response mechanisms play in the real world, seems to work in the digital world as well. Examples such as the social pressure organised in the Facebook setting against a large bank, when it changed its terms, show this convincingly. Quite well-known cases in point are the systems to enable the distribution of reputational information of online shops such as eBay. An EU wide directive, providing the exemption of liability by limiting the activity of an information service provider to specific

[98] Fennell-van Esch, S., van der Hof, S. and Marbus, R., Virtuele identitieiten en juridische problemen. Kan de wet uitkomst bieden?, Ars Aequi, July/August 2008, p. 568.
[99] *Ibid*., p. 569.
[100] <<www.hyves.net/useragreement>>
[101] http://secondlife.com/corporate/tos.php

technical processes, is referred to, to show how at a macro-scale the functioning of the internal market can be influenced by creating a trustworthy environment.

In the traditional world guarantors of trust such as notaries, auditors and accountants provide a setting that encourages citizens and organisations to pursue worthy goals and lead valuable lives. In the digital world organisations such as Trust Mark Organisations (TMO) and TTPs have tried to fill in the gap. Reference is made to a study that has introduced the ethical concept of warranted trust. Based on this concept the TMO is put forward as the solution for providing trust in the digital world. However, the concept of trust is enriched to that of warranted trust which is defined as "Trust is warranted when a liability system which protects what e-consumers value (e.g., security, privacy and fair business practice) considering the TMO's context of action (i.e. the Internet) is in place."

The question of liability in the digital world is still not resolved. After all, an avatar may perform harmful actions which are not part of its instructions. By a fault in its software it may sell instead of buying shares. Certification could be a solution but the Karnow registry in which agents are registered in a registry which always makes payment when damages occur when a registered agent is involved, may solve the liability problem. Furthermore if artificial intelligences make decisions unforeseen by humans, this could lead to insuperable difficulties because the traditional tort system relies on causal agency. Legislative efforts such as the UCITA may provide a solution. Codes of conduct in the gaming worlds or other virtual worlds with avatars (Second Life) seem more promising than having recourse to criminal or civil law. The legislator remains rather hesitant in stepping into virtual worlds with avatars because the economic interests are currently not high enough to validate such uncertain actions. He will not feel that need as long as these virtual worlds can solve their own problems internally.

6 Relations between Trust and Identification

In the case studies of Section 3, we have already demonstrated that situations, in which both trust and identification are of crucial importance, are not untypical in today's internet and web applications. We have also seen that trust is often an important prerequisite for an identification mechanism to work properly, and conversely, that identification is often necessary to attribute the right weight to trust-related evidence. We will now analyze the role of identification for trust and the importance of trust for identification more systematically and more profoundly. By doing so, we do no longer restrict our focus to internet or web applications.

6.1 Identification Components in Trust

In our analysis of trust in Subsection 4.1, we have learnt that trust is a belief state about the future contingent actions and behaviour of others, and that this belief state depends on the available experience and knowledge about the others' competence and honesty. There are thus two general ways of establishing, updating (increasing or decreasing), or verifying trust towards another entity, either by augmenting the pool of direct experiences and interactions, or by expanding the available set of indirect evidence about the entity's trustworthiness with new recommendations or testimonials. We will see next, that proper identification mechanisms are necessary in both cases.

6.1.1 Is Identification Necessary for Trust?

Let us now discuss the importance of identification in each of the two basic trust building mechanisms. The first mechanism – the one based on direct experiences – requires some sort of personal interaction between the trustor and the trustee. The principle is very basic: the better two entities know each other, the more they are willing to trust (or distrust in case of negative experiences) each other. Family, close friends, or long lasting partners are usually among the most trusted (or untrusted). In a traditional real world context, establishing direct trust relationships of that kind naturally implies that the trustor and the trustee spend quite some time together (physically at the same location). The identification mechanism in such situations is usually based on the trustor's natural face and voice recognition abilities (except if the trustor is a visually or hearing impaired person). At first sight, the importance of a proper identification mechanism is not very apparent in this basic setting, but it is obviously crucial to avoid mixing up the experiences with different entities. Note that identification based on the face and voice recognition abilities of a human being is usually very reliable (it must have been beneficial at some point in the evolution).

The emergence of various communication channels (from simple postal services to telex, telephone, fax, e-mail, or online social networks) has introduced a new form of interaction between individuals, one that does not require a personal reunion at the same physical location. This is a complicating issue with respect to the identification problem of trust, especially for those communication channels, which disallow both face and voice recognition. While telephone conversations are still quite reliable in this respect, it is almost impossible to uniquely identify uniquely the sender of an electronic mail. We have already discussed this difficulty and possible solutions in Subsection 3.4, but what are the consequences of this to the process of establishing trust based on direct interactions?

In a worst-case scenario, we could think of a situation in which a betrayer misuses the identity of an unsuspecting victim, e.g. by creating a new e-mail address or a postal box under the victim's real name. The prevention of such an attack may be quite difficult in a real world context, which means that the betrayer will then be able to communicate under a false identity. By adding a second false identity of another unsuspecting victim, the betrayer may even become a so-called "man-in-middle", which is in full control of the communication between two arbitrary entities. The problem with respect to the question of trust is then at least two-fold:

- First, as long as the betrayer manages to remain incognito, he will essentially receive the same level of trust as the holder of the stolen identity, and the betrayer can thus exploit the trustor's benevolence that is usually restricted to the close circle of the most trusted ones. Attacks of that kind are typical for stealing secrets (such as passwords or access codes). The most prominent examples of that kind are "phishing" (password fishing) e-mails, but one could also think of a betrayed husband trying to get further evidence about the love affair of his wife, of a marketing analyst trying to get secret information about the competitor's future products, or of any other scenario of private, industrial, political, or military espionage.
- Second, if the betrayer's motivation is to discredit the holder of the abducted identity, then acting undetectably under the victim's identity may be a surprisingly powerful way of violating or destroying existing trust relationships. A reason for this lies in the asymmetric psychological nature of trust and the resulting asymmetry of effort needed for establishing or destroying trust (see Subsection 4.1). This type of attacking existing trust relationships is therefore a common way of discrediting and bullying rivals, e.g., in a private or job-related context. The result of it is a distorted pool of seemingly direct experiences, from which the trustor derives misguided trust decisions. Note that the same sort of attack may have the opposite effect of establishing new or improving existing trust relationships.

As both types of betrayal could easily be avoided with the aid of a more reliable identification mechanism, the answer to the question of whether identification is necessary for *direct* trust is thus clearly affirmative. This conclusion holds for all traditional forms of human interaction, but it has become much more evident with the emergence of new communication technologies, which enable entirely new forms of forging or stealing identities.

The second fundamental trust building mechanism – the one that aims at expanding the available set of indirect evidence about the entity's trustworthiness – is also vulnerable with respect to improper identification mechanisms. The main problem here is the fact that each additional piece of evidence, e.g. a new third-party recommendation, needs to be evaluated with respect to its own credibility. This implies that the reliability or reputation of the involved third-party needs to be checked first, and this usually means to verify its identity. Dealing with third-party evidence such as recommendations or discredits generates thus a cascade of further identification problems, and each of them should be solved before taking into account respective pieces of evidence. Note that attack scenarios similar to the ones mentioned above are again possible to tamper the whole process, e.g. by injecting false recommendations in the name of someone else. Typical examples of that kind are biased entries under wrong names in recommendation or rating systems such as the ones provided by eBay, Amazon, or the *meinProf* web sites (see section 3.4).

The answer to the question whether identification is necessary for *indirect* trust is thus again affirmative. We have even seen that the dependence of indirect trust from identification ranges over a single trust-identification layer, because the evaluation of third-party evidence requires further trust relationships and thus further successful identity verifications. Note that quite some research has been devoted to the formal analysis of so-called multi-layer trust networks (Haenni et al, 2007), in which trust and identification mechanisms are understood as two opposite characteristics of essentially the same problem.

6.1.2 Trust in the Case of Anonymity

Identification is necessary for establishing trust as we have seen in the previous subsection, it is nevertheless possible to trust a pseudonymous or even anonymous partner. Identification is often seen as a way to access the "official" identity of a person (his real name and surname). We will see that accessing this particular identity is not necessary to trust someone. In the non digital world, pseudonyms were already used and they did not prevent trust, they were sometime even a way to achieve good trust.

Pseudonyms were not created by the information society. In the past, some persons have mainly been known by their pseudonyms. And they were trusted only with this pseudonym. Let us take the example of World War II in France. Mr. Jacques Delmas joined the resistance and had to take a pseudonym (in fact he took different ones but was famous with one particularly). He was appointed General in the French Free Army under the pseudonym of Chaban. This name was famous while his official name was still unknown. In order to start a career as a politician, Jacques Delmas integrated his pseudonym in his family name, which became Chaban-Delmas. The people trusted the leader of the resistance called Chaban, they didn't know (and had therefore no trust in) the young politician Delmas. Jacques Chaban-Delmas was elected as a Member of Parliament and Mayor of Bordeaux, he was later Prime Minister.

The use of a pseudonym can also be a great component in the trust. When entering the French foreign legion (Légion Etrangère) a new soldier receives a new identity[102]. This identity is the only one he can use while he is serving. The new identity is part of the enrolment process and provides a way for the foreign legion to control its new soldiers. Thus it also enhances the trust the army has in those persons[103].

The information society offers plenty of new pseudonyms or pseudonym-like identities, starting with IP addresses, e-mail accounts, website accounts for e-commerce, social networking or gaming, phone number or credit card information. The trust a person has for the partner hiding behind a pseudonym is based on various criteria as explained in the previous subsection. Even if dealing with pseudonyms, trust remains a combination of the knowledge of previous experiences, the possibility to influence the future of the partner (means the existence of a feedback mechanism). The need for a secure identification stressed in the previous section can be viewed as the confidence in the fact that the person behind the pseudonym is really the one expected (means: the same person as in previous experiences). The first two parts are the same if we consider relationships to anonymous or fully known persons, but the latter introduce a new uncertainty level. The level of confidence in the link

[102] See the official site of the Légion Etrangère: <www.legion-recrute.com/> the legionaires must use a "declared identity" (a pseudonym).

[103] Such a process occurs also in church, popes change their names and use a "pseudonym" when they are elected.

between physical person and pseudonym is not the same for IP-address as for websites using biometric token. IP addresses can be easily spoofed whereas there exists some systems that can be considered as quite reliable for biometric recognition[104]. The trust in a pseudonymous user is therefore based on the same components as for a named user. It relies on the trust in the user mixed up with the trust in the identification system provided with the environment (website, protocol, etc).

Trust can also be totally anonymous,[105] even if it is almost never used; it is technically feasible to provide an infrastructure allowing a trustor to trust an anonymous trustee. Anonymous credentials[106] can be provided to users by a (trusted) third party. Then the trustee could provide this credential to the trustor that could accept it and trust the new partner. A trustor does not necessarily need to know the identity or any pseudonym of a trustee, he just wants to be convinced that the partner will do what is expected from him. Such a service can be offered by a trusted third party. It will play the role of the middleman between the two partners, and will be charged to protect the identity of the trustee as long as it is behaving correctly. For example, a very important issue in e-commerce is the transfer of money. The seller of the product wants to be paid for it. A consumer could provide an anonymous credential that would only work once to prove the trustor that he will be paid, without revealing one's identity (or even any pseudonym). The provider of the good or service can verify that the credential was really issued by the trusted third party and he is certain to be paid. The e-CarteBleue[107] is an application providing a similar service. A client using this service can obtain for each transaction he wants to make a new valid credit card number. This number is generated and new for each transaction. In this case, the holder name remains the one of the holder of the "real" credit card, but this service would technically work even if the name were generated at random.

Granting access to restricted areas is a major privacy issue. For example, operators of services which access is restricted to people older than 18 (or 21 depending on the country) need to trust their users to give the right information about their age. Anonymous credentials can be used for controlling a specific area whose access is restricted to people over 18. For instance, the solution used actually by night club owners in order to verify the age of the persons is to ask for an ID. While they are doing this, they could also notice the name, address or any other data contained on the ID or passport. This situation is problematic from the point of view of privacy protection. It would be possible to find a technical solution for the persons to provide an anonymous proof of their age without giving their ID. One can imagine a biometric token, issued by the government that would be able to produce an anonymous credential each time it is required, means entering night clubs, buying alcoholic beverages or visiting adult websites. Using such a device, the owner of the site or of the bar could trust the user without knowing him and without accessing even his pseudonym.

---

[104] See <www.axionics.ch> for example.
[105] In the next paragraph, we make the following distinction between pseudonymous and anonymous relations: pseudonymous relations allow the linking of interactions, while anonymous ones do not.
[106] See for example (Chaum, 1990) *Showing credentials without identification: Transferring signatures between unconditionally unlinkable pseudonyms.*
[107] The e-carte bleue website : <www.carte-bleue.com/page.asp?menu_id=22>

Anonymous trust is technically achievable (anonymous credentials, one time credit cards, biometric tokens already exists). It requires a technical architecture which is possible to deploy, but most of all it requires to trust this architecture and its administrators.

Totally anonymous one-time credit card numbers may be implemented if the trust in the bank issuing the numbers is big enough to compensate the lack of trust in an anonymous person[108]. The trust in biometric tokens providing a proof of age would also be placed in the government or organization issuing these tokens. The trustor wouldn't need any information about the user himself. Trust in an anonymous user can not rely any more on the past experiences done with him, it relies on the trust in the trusted third party issuing the credentials, as well as its possibility to force the trustee to behave as expected.[109]

6.1.3 Privacy Issues

In most cases, trust is achieved using identification of the user. If the only identity used were the official one, then each operator in the information society would access to a unique key for the data of a person; it would be possible to link all the data together. This will help in establishing the trustworthiness of a user, since all his previous actions would be registered. We have also seen that such identification is not required. Trust is possible even using pseudonyms. It requires a mechanism for checking the trustworthiness of a user. So the operator of the system must provide a way for the trustors to evaluate the trustworthiness of a user hidden behind a pseudonym. This can be done accessing the history or reputation of the trustee or one can offer a possibility to "change the future" of the trustee (sue the trustee if he did not act accordingly to the expectations for instance). In this case, the privacy issues arise on the side of the operator of the pseudonym service. It must provide a reliable way for a person to be identified, as well as a reputation mechanism and/or a response mechanism (feedback). At the same time, he must ensure that no external observer can link a pseudonym to the real person using it, except if this is required by a judge or the police for legal reasons. Indeed, privacy protection is an important component in the acceptance of trust mechanism in pseudonymous or anonymous systems.

Anonymous trust is more complicated to achieve. Usually, it requires a trusted third party that can issue anonymous credentials. It can be coupled with a reputation system or identity control system, but the trustor will only see the one time credentials. So the trustor has no way to attack the privacy of the trustee, it is anyway possible for the trusted third party to gather a lot of information on the trustee if the system of credentials requires the trustee to transfer information each time a credential is required. So even in such a privacy-friendly environment, privacy of users can be threatened. Details of the implementation are very important. Such a system can become very dangerous by concentrating in one single hand some information that was shared among many operators in other systems. The aggregation of different databases may be much more difficult for the different trustors, but the central role played by the system delivering the anonymous credentials is a risk factor. It requires therefore a large amount of trust in this system from all the actors, trustors and trustees.

---

[108] This can be compared with bank notes. The introduction of such notes requires a great trust of the people in the bank emitting the notes.

[109] For instance in the "Watergate" scandal, trust in "deep throat" was not based on previous experience the readers had with him, but rather on the trust the two journalists placed into him. The relation uses a trusted third party (the journalists in this case).

Legal systems place a great deal of emphasis on the recognition of the person. Identity management however could potentially undermine personhood, if a person is not in control of his identity information. Human rights are attached to the person and not to the profiles that may be built from identity information about him. A person becomes increasingly recognized by the profiles that begin to accrue for him. It should be emphasized that the past actions, which constitute a person's profile, are not the source from which his human rights derive. His state of being a person gives rise to those rights. Identity information is detached from the person's control in the Internet environment. Therefore the ability to control the use of one's identity information is crucial for reminding others that there is a person behind data and enabling that person to have full status when dealing with others.[110] A person's freedom might increase the more he or she interacts with others through digital identities that remain under his or her control. It can even be argued, that society as a whole suffers because people begin to fear, that data are inaccurate and that decisions are arbitrary. User control over identity information therefore generates accountability and trust.

The aforementioned disjunction of the personal identity and the digital identity, which occurs in the real world, is of course exactly what makes the virtual world the virtual world. However, the challenge of the technological changes with which society is confronted, gives us at the same time the opportunity to organize this control over personal data. IDM promises to be a unifying thread for these emerging technologies. In the ubiquitous information environment, a healthy information society is possible, if the user enjoys control of identity information, relating to him.[111] Privacy guidelines, such as those of international organizations (*e.g.* the OECD) have drawn up some time ago are of help here. User-centric IDM in itself establishes trust and increases confidence in e-commerce and e-government. International organizations, such as ISO and ITU, should take note of the data protection guidelines in drawing up an identity infrastructure that fosters accountability and trust through user control.[112] When information is shared between jurisdictions, such as is at hand in the digital and Internet world, governments should take care, that accountability is incorporated in the system. An ombudsman role can play an important role in this respect. Democratically accountable governing bodies could assume the role of the guardian of citizen's interest in an open economy and society. When identity disputes are social disputes, the processes that need to be designed should be social and legal processes, with the usual provisions for fairness and due process. In summary, a free and open social order requires trust, and trust hinges on accountability; accountability, in turn, hinges on user control; user control, in turn, hinges on data protection; and data protection hinges on respect for Properties of Identity in law and technology. The Properties of Identity, if followed, will allow for robust personhood and digital identity going forward.[113]

---

[110] OECD. At a crossroads: "Personhood"and digital identity in the information society. STI Working Paper 2007/7, p. 10. <www.oecd.org/sti/ict/reports>

[111] Recently a small internet company, Yuntaa, offers online backup services using cloud computing. It advertises special features such as security, respect for privacy. The data is stored using encryption. In addition all data on the servers is insured and an amount of € 250.000 can be compensated. "Sociaal netwerken met respect voor uw privacy", De Standaard August 20, 2008.

[112] OECD, p.37.

[113] *Ibid*., p. 50.

6.2 Trust Component in Identification

After having considered in the previous subsection identification components in trust, we look the other way around and determine the trust components in identification. To start with, we propose some basic questions related to trust in identification:

- Can I trust your ID card?
- Can I trust your certificate?
- Can I trust the issuer of your ID card / certificate?
- Are you really the person you pretend to be?

Those four questions are typical issues one has to consider when faced with identification, all being quite focused on identification of human beings; but one might easily generalize these statements to the identification of non-human persons (legal persons, virtual persons). The keyword trust is of major interest in all of the questions, explicit in the first three, somewhat hidden in the last one. Note however that in all those questions trust is focused on the identificator, the one *doing* the identification, whereas trust of all other entities – e.g. the one being identified – is not mentioned at all. But we will see in the sequel that even for these other entities trust is of major interest as well.

6.2.1 Entities and Identification

For clarifying the issues, reconsider now the basic definition of identification given before:

> ***Identification*** *is the process of establishing enough confidence in the fact that some identity-related information is valid and truly describes a specific entity in a given context or environment, at a certain time.*

Let us first specify the two basic entities participating in this process when identification is a verification of identity:

1. The *verifier* is the one doing the process of identification, and also typically the one mainly interested in the result of the process itself,
2. the *entity whose identity is to be identified*.

When the verification of identity is based on certified information, there is a third entity involved in the process:

3. the *certificate authority*[114] *(CA)* – an entity which certifies data.

Whether the verifier trusts this CA or not has a direct impact on the outcome of the identification process.

In case a CA is involved, the identification must have been preceded by a registration process. There are special cases where one is implicitly forced to trust a CA, consider for example the case of the CA "state" delivering passports: typically, the border police (of the same country; things might look different for other countries) will have to trust this CA. Or the case of an employee in a company which has to trust implicitly the CA of the company (after having signed his employment contract).

[114] Note that we use the term *certificate authority* in a general meaning here, not only focusing on digital certificates in the sense of e.g. X.509.

The following figure illustrates the potential trust relations between all parties involved. It is a bi-directional complete graph.

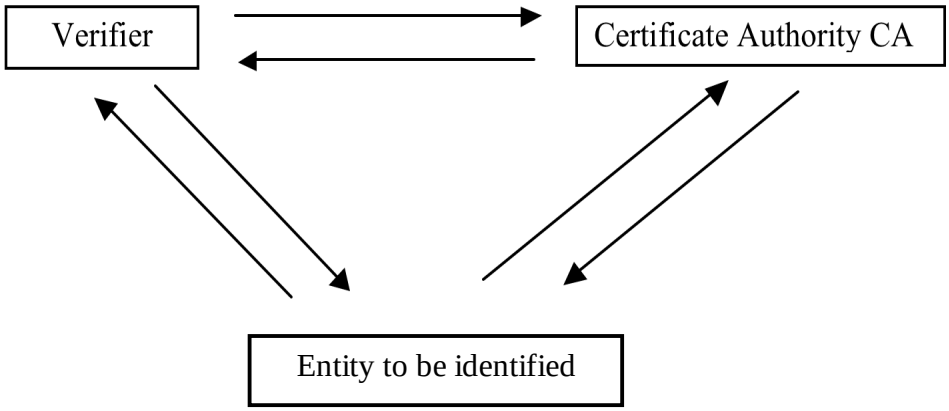

Figure 17: Potential Trust Relations

Trust typically is bound to some expected comportment of an entity. In the above figure, there are three possible trustors which each might potentially have trust in two other entities – trustees – (omitting here the case of self-trust). Let us consider the six possible "trust arrows" of the figure separately:

- **Trust of the verifier in the CA:** This is probably the most essential part of trust needed for a successful identification process, successfulness here clearly interpreted with respect to the goals and expectations of the verifier. The verifier must trust the CA to work as expected: a CA must deliver its certificates only by rigorously respecting the legislation and/or its own rules. Rules and regulations which might incorporate for example detailed information about which data is needed and how to check its authenticity, the accuracy of the processes themselves, the reliability of the systems, etc. These rules must be made available to the verifier and to some extent the verifier should be able to check that to check that these rules are applied.
- **Trust of the verifier in the entity to be identified:** Here, trust is not relevant. If however it is already present, it may have some (typically positive) influence in establishing confidence as described in the definition.
- **Trust of the entity to be identified in the verifier:** Typically, identification means giving (personal) data to the verifier. Hence privacy issues are of major concern when the data contains sensitive information with respect to the entity to be identified. In this case, the entity to be identified must trust the verifier not to misuse this information, use it only for the identification process for which it has been provided, and delete it afterwards, or make it inaccessible for other entities (or for other processes, or at a later time, etc.) using appropriate cryptographic tools for example. Additionally, if the entity to be identified is itself interested in the positive outcome e.g. when presenting the passport at the border, the entity needs trust that the verifier does its job correctly and delivers the according rights.

- **Trust of the entity to be identified in the CA:** Several aspects of trust are of interest in this relation. First, the corresponding privacy issues as just mentioned in the case above are central here as well. Second, the entity to be identified must trust the CA that it delivers certificates, which are of a certain "quality" and only contain the information needed (i.e., no "hidden" side-information). Typically, the entity to be identified should also be able to verify its certificate (just like the verifier can do) and have access to all information contained therein. Other trust aspects might be possible or advisable in the sense of a CA in a public-key infrastructure, e.g., trust in the CA functioning as expected – for example – keeping the revocation lists up to date.
- **Trust of the CA in the verifier:** Even though it is usually irrelevant, there are nevertheless some cases, where trust of the CA in the verifier is required. Consider, for example, the biometric passport to be implemented in Switzerland. The key for accessing the biometric data stored in the electronic chip within the passport is only given to "trusted countries" which then may access and compare the electronically stored biometric data in the passport to the person present at the border.
- **Trust of the CA in the entity to be identified:** Trust is needed here as the CA must be sure that the data provided to issue the certificate is valid, correct and belongs to the very entity providing it. This trust can again be based on other certificates, consider for example the case where I prove my identity to a CA using my passport, and then the CA delivers a X.509 certificate for me.

Given the diagram of the figure above, one might try to derive trust paths from it, for example: the verifier has trust in the CA that has trust in the entity to be identified (creating confidence in the entity's identity); hence the verifier has trust in the entity to be identified and has confidence in the entity's identity. However, we then face the general problem whether or not transitivity in combination with trust is appropriate or not (see also Section 4.1.3. on transitivity of trust relations).

If we reconsider the definition of identification above, the verifier has to determine if he can have enough *confidence* (not trust) in the fact that some identity-related information is valid and truly describes a specific entity in a given context or environment, at a certain time. This is illustrated be the following diagram:

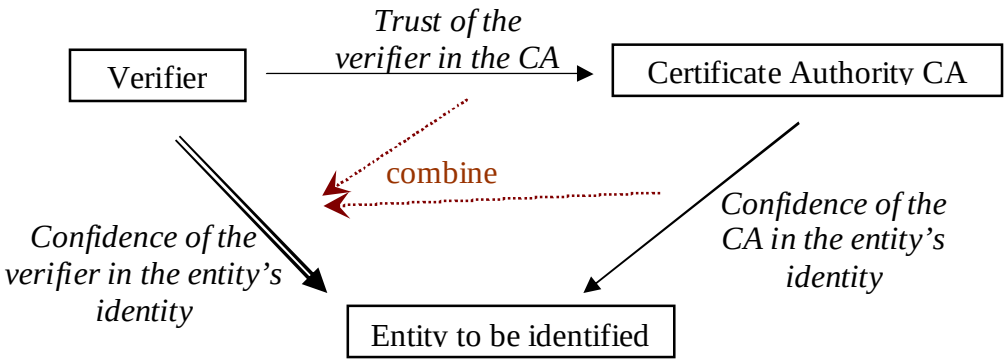

Figure 18: Inferred Confidence

We reconsider some special trust situations from above, in order to clarify our interpretations.

The CA – and after the identification process to some extent also the verifier in case of the implied confidence mentioned above – has confidence in the identity information provided by the entity to be identified. As said, this can be based on other certificates, etc. However, this confidence has several aspects:

- Is the identity information correct?
- Is the identity information valid?

Especially for the determination of correctness and validity, the CA has to make additional processing and possibly query other data providers. The case of validity is also dependent on time and hence validity of certificates should typically be bound to some time slot (valid from – to).[115]

As mentioned, the CA can only influence to some extent which entities can use the certificates it provides. The verifier chooses the appropriate context to do the identification, and the entity to be verified has to trust that the verifier makes an appropriate choice, as typically no enforcement is possible here.

6.3 Conclusion

In the first part of this section, we have seen that trust relies for a significant part in the capacity of the trustor to recognize the trustee. This requires the use of an identification process. But the resulting identification does not necessarily rely on the official identity of the trustee, since the trustor may trust a pseudonym and just needs to be sure that the person hidden behind this pseudonym really is the one expected. The trustor may even trust an anonymous trustee if a third party provides anonymous credentials allowing the trustor to have trust in the fact that the trustee will act accordingly to what is expected.

In the second part of the section, we have diametrically changed the viewpoint and looked at trust components in identification. In case of a certification authority (CA) being a relevant actor, we have discussed the six (oriented) trust relations between the verifier, the CA, and the entity to be identified, as well as their respective implications. In the example of an identification process, trust of the verifier in the CA together with confidence of the CA in the entity's identity may result in confidence of the verifier in the entity's identity. Hence, we experience some limited form of transitivity.

In general, both concepts of identification and trust – together with confidence – are closely interwoven; each one is of main interest for the other one. Trust is typically needed in the process of identification, and identification is – in many situations needed for trust. However as we have discussed, trust and anonymity are not incompatible.

[115] Note that we do not consider completeness here, which brings us to the distinction between full identification (or individualization) vs. partial identification.

7 Models of Trust and Identification in the Virtual World

With the growing virtualization of many real world processes and activities, together with the resulting shift of personal interactions in the physical world to virtual interactions in the digital world, we are faced today with a number of entirely new problems and challenges. Many of these encompass sociological, psychological, philosophical, or even legal aspects, and are thus not purely technological. Two particular classes of problems result from the question of how to establish trust and how to verify identities in a digital context, in which virtual identities are the front-ends of almost all interactions. The majority of existing trust and identity management systems are based on some simplifying assumptions, often implicitly, but those are mostly designated to avoid rather than to overcome such problems.

In this section, we are trying to expose some of the key problems related to trust and identification in applications of the virtual world. The analysis will be based on the definitions of trust and identification given in Section 4, and it will incorporate the conclusions and findings of the preceding discussions of Sections 5 and 6. The common viewpoint throughout this section is the virtual person model from the FIDIS deliverables D2.13 and D17.1 (see summary in Subsection 4.3).

7.1 How Can We Build Response Mechanisms in the Virtual World?

7.1.1 Social Pressure, Reputation, Liability in the Virtual World

As we have seen in section 5.1.2 traditional response mechanisms can be applied, can have their influence in the digital world as well. Behaviour can be regulated by social pressure or liability, and digital reputations suit perfectly well to create trust in the digital world.

Will this be different in a virtual world in which the distance increases between abstract entities and the persons who employ them?

Social pressure

Social pressure as described in the cases of the HSBC bank and the Wispa chocolate bar can be created by virtual persons as well, considering natural persons taking part in society are virtual persons because they have rights, duties, etc.[116] This means virtual persons can create an uproar online, which is initiated in the physical world.

This will be different in a world of autonomous acting agents. This type of virtual person will not be triggered by an event in the physical world. Apart from this, such a virtual person will not be capable to communicate with other agents on an unexpected topic as the change of the terms of an interest-free overdraft facility. Being unable to interpret unexpected events (online and off line) and to communicate about such events and take actions, makes it impossible for electronic agents to organize social pressure.

[116] See FIDIS deliverable "D2.13: Virtual Persons and Identities"

Reputation

A virtual person can build up a reputation. An entity buying and selling goods on the Internet can be given rates. Once these ratings are good, another virtual person will not doubt to do a transaction with the particular entity.

Once such persons start to act more autonomously, ratings can still be used to inform other entities on the reputation of other entities. For example, once a transaction has been completed successfully a positive rating can be given to the trading partner. In fact, such a more autonomous reputation mechanism could even be more reliable, if no human interaction is involved. After all, in that case ratings are not influenced by friends or enemies of flesh and blood, but based on transactions, good or bad.

Liability

In the virtual world behaviour can be influenced by liability issues as well. Abstract entities can decide only to do business with other entities when those entities are connected to, for example, branch organizations having reasonable liability conditions. In the world of autonomous agents this would mean that there needs to be a check whether or not a potential trade partner is connected to a branch organization having such conditions. Of course, it is not possible yet for agents to interpret such conditions as being reasonable, but the persons employing agents can, and they can instruct the agents to act according to their wishes if contacts with entities arise, being connected to specific organizations.

The same will be true for a virtual entity if a potential virtual client knocks at the door. If such a client is connected to a trusted third party which is considered to be reliable amongst others because of their reasonable liability conditions, that aspect will not prevent a transaction taking place.

This does not mean virtual entities need to be given legal personhood to enable contracting or paying damages. As the FIDIS deliverable on legal personhood for non-humans (D.17.2) concluded, current legal constructions still seem to suffice to solve potential conflicts that arise through the increasing distance between abstract entities and the persons who employ them.

7.1.2 Legal Infrastructure in the Virtual World

In this subsection we will build on the work that has been done in the workpackage "Abstract persons" (WP17). WP17 in turn builds on the definition of the model of virtual persons as defined in FIDIS deliverable 2.13 *Virtual Persons and Identities*[117]. Moreover we refer to Subsection 5.2 Legal consequences in civil and criminal law of this deliverable. In that subsection some fundamental questions of law in relation to the virtual world are already made. In this subsection we will confine our analysis to the possibilities of providing a legal infrastructure for virtual identities. The definition of the model of the virtual person of FIDIS deliverable D2.13 fits nicely as a basis for the application of law to the virtual world. The model of FIDIS deliverable D2.13 is defined with two layers: the physical world and the virtual world[118]. It allows a precise and unifying representation of new forms of identities in

[117] Jaquet-Chiffelle 2008.

[118] This is an abstract environment, which is a product of the mind rather than of matter.

the information society. Thus a unified description of many identity-related concepts that are usually defined separately without taking into consideration their similarities: avatars, pseudonyms, agents, profiles, legal persons, etc. is made possible.

As noted in FIDIS deliverable D17.2, laws have a long experience of using abstract entities to define rules, categories, etc., in order to associate legal rights, obligations, and responsibilities to persons that can in concrete situations be considered instances of these abstract entities. The application of the law in a specific situation makes an entity with legal personhood the bearer of the legal rights, legal obligations, legal responsibilities associated to one of these abstract entities that the law uses. The model of FIDIS deliverable D2.13 may profit from the long experience of handling abstract entities in law to refine some of its concepts specifically for the legal framework. Reciprocally, the legal framework might use this generic model to represent its abstract entities as well as new abstract entities together. The analysis of FIDIS deliverable D17.2 provides a tentative conclusion saying that current laws need to be adapted to encompass new paradigms, such as the rise of autonomically[119] acting entities, to better understand if and when new laws or even new legal persons have to be created as a response to new technological developments. After all, law has managed to tackle and incorporate these new concepts quite successfully for a long time. The abstract layer in the model is particularly well-suited to describe (new) entities operating at an increasing distance from the physical or legal persons behind them. It recognises the existence of these (new) intermediate entities and explicitly incorporates them in the model. Especially the concept of agents has been well recognised as persons in law. Still the conundrum of the legal personhood of a computer agent deserves much consideration. A computer agent is not considered a natural person in the present legal framework.

In itself, the fact that an entity is not a natural person does not preclude the attribution of legal rights. The question is however, *under which conditions* it makes sense to attribute legal subjectivity and *to what extent* legal subjectivity should be granted. As noted in Subsection 5.2, this depends on the relevant legal domain, because strict liability for harm caused may be adequate in private law but out of bounds in criminal law. While investigating under which conditions and to what extent legal subjectivity could and perhaps should be granted to new creatures, we must keep in mind that in the present legal framework the producer and other entities involved in the production and use of a smart technology may be imputed strict liability for harm caused. In the context of agents, the question should be posed if creating new legal abstract persons solves problems that cannot be solved by imputing liability to the relevant actors? What is the added value of creating new legal abstract persons? On the one hand, it might simplify or even avoid disputes between the parties involved when liability is hard, if ever possible, to attribute to exactly one of them. On the other hand, creating new legal abstract persons may generate new problems that are avoided when the person at a distance is held liable.

If a legal infrastructure in the world of virtual entities is to make sense and be viable, an answer will have to be formulated whether virtual persons can be attributed legal personhood. The research for an approach to this problem has long been Solum's work.[120] He defines the conditions for legal personhood in terms of the capacity to perform complex actions and/or the capacity to act intentionally and with (self-)consciousness. This seems to comply with the traditional idea that personhood implies the capacity to act in a deliberate way. Legal

---

[119] On the concept of autonomic entity, see FIDIS deliverable D17.2.

[120] Solum, L., „Legal personhood for artificial intelligences", North Carolina Law Review (70) 1992, 1231-1287.

personhood is often attributed to entities that do *not* qualify for such personhood (like ships, corporations etc.). Legal theory refers to this as a legal fiction: the law attributes personhood though in "normal" life we would not think of the relevant entity as a person. Solum argues for a pragmatic approach to legal personhood. He argues that legal personhood is empirically dependent on the measure of independence of the artificial intelligence. Such independence depends on the capability to perform complex actions (reducing the need for human intervention).

In the most concrete example viz that of agency instances are known where the law has been changed using *specific*, well-known constructions. For example, as pointed out in the Subsection 5.2 with the example of the UCITA framework rules can be drafted for electronic agents, stipulating under which conditions contracts are valid and who is liable for which actions of agents. Such sector-specific rules provide legal certainty, and they can also – if the need to do so is felt – deviate from 'off-line' legal constructs, for example, limiting liability in order to stimulate the market for promising new technologies, or on the contrary, introducing strict liability for electronic agents if their unpredictable actions are felt to be too risky for business or consumers. A simple solution seems to be to develop registers for agents, which could allow contractants to find the identity of an agent's principal, or, alternatively, to lay a claim on insurance for damages in case a registered agent goes haywire. Trustmark organisations which apply the concept of warranted trust may in the future also allow for a legal infrastructure that is viable in the virtual world.[121]

Perhaps a solution for the present evolving state of virtual entities is granting them limited legal personhood. In the aforementioned register of electronic agents, they may be listed as agents with a limited type of personhood. The agent itself is responsible for its contracts and potential mishaps (outside of the moral or criminal sphere). The agent could have money itself, for example by earning a small provision for each transaction he makes for his principal, and use this money – perhaps via an insurance – to pay civil damages or administrative fines.[122] At some point in the future, it might be possible that electronic agents could be attributed moral personhood, if they gain the ability to make decisions that are functionally equivalent to moral decisions, in other words, when they acquire self-consciousness. For the time being, current legal constructions suffice to solve potential conflicts that arise through the increasing distance between emerging abstract entities and the persons who employ them, as concluded in deliverable 17.2. By the time electronic agents become more autonomic and acquire a 'personality' of their own, it is considered to be useful to treat them as new entities with their own identities in themselves, with certain legal rights, duties, obligations, and/or responsibilities. If, and when, electronic agents with some sort of personality will ever exist, remains to be seen.[123]

As far as trust is concerned, now we will have to make do with the practical solutions that for example guarantors of trust such as notaries and auditors but also trustmark organisations already put forward. Law by its nature will also in the virtual world expand the radius of trust and make e-commerce increasingly possible. It might even never be necessary to create a legal abstract person. Recent history shows that new technologies have led to a certain

---

[121] See section 5.2 especially the references to Balboni's thesis.

[122] A solution should be found to the question what happens when the agent 'dies'. It would seem logical that the money left in the account of the agent is returned to the principal.

[123] See in this respect for example Setargew Kenaw, *Hubert L. Dreyfus's Critique of Classical AI and its Rationalist Assumptions*, Minds and Machines, Springer Netherlands, Volume 18, Number 2 / June, 2008.

amount of new rules, directives, etc., but at the same time maybe even more often lead to the conclusion that proven legal concepts can be applied adequately to those technologies. Existing rules seem to be able to create trust in new technologies as well.

Another phenomenon, which can enhance trust in a world in which there is greater distance between entities and the persons who employ them, is online dispute resolution. After all, if online dispute resolution schemas do exist, the abstract entity as well as his employer knows that solutions are sought for if problems do arise. This is especially of value not only when distance exists between an abstract entity and his principal, but also when there is a greater distance between the buying entity and the selling one.

The European Consumer Centres Network (ECC-Net)[124] and the European Judicial Network in civil and commercial matters[125] are networks, which deal with all the ingredients necessary for what might once be a European online dispute resolution schema. The first one has been created as a tool to have better informed and educated consumers and also to help those consumers in getting the appropriate redress in case of a violation of their rights in cross-border transactions. The second network consists of representatives of the Member States' judicial and administrative authorities and meets several times each year to exchange information and experience and boost cooperation between the Member States as regards civil and commercial law. The main objective of the latter is to make life easier for people facing litigation of whatever kind where there is a transnational element - i.e. where it involves more than one Member State.

An ideal European online dispute resolution schema should give entities the opportunity to solve their problems online and fully automatically. In such system jurisdictional problems should not play a role. It should be a problem-solving system in which the rules of the schema need to be obeyed and not the rules of the countries of the parties involved. Being connected to such a schema gives organizations a status in which entities will trust, and as result it will give a boost to cross-border e-commerce.

An example of an online dispute resolution system is the American Cybersettle.[126] This company offers services that are entirely online and focus primarily on negotiating monetary settlements. Their website serves as a neutral arena to exchange settlement offers.

The World Intellectual Property Organization has an online dispute resolution schema for domain names. Via their Arbitration and Mediation Centre it is possible to file complaints online. Those complaints are submitted under the Uniform Domain Name Dispute Resolution Policy (UDRP) of the ICANN.

EBay also offers online dispute resolution services to help their clients to resolve disputes. SquareTrade is eBay's preferred dispute resolution provider. This company offers two services: a free web-based forum which allows users to attempt to resolve their differences on their own or if necessary, the use of a professional mediator.[127] These various efforts show an increasing possibility to solve the problems of virtual world with legal infrastructural regulatory measures from the classic real world. As such they help bridge the accountability gap noted earlier which rises through the increasing distance between the principal and the agent in the virtual world setting.

---

[124] Website ECC-Net <ec.europa.eu/consumers/redress_cons/>.
[125] Website European Judicial Network <ec.europa.eu/civiljustice>.
[126] Website Cybersettle <www.cybersettle.com>.
[127] Website SquareTrade <www.squaretrade.com>.

7.2 Trust in the Virtual World

The goal of this subsection is to shed some light on the problems that arise, if trust concepts are transferred from a traditional physical world context to the context of the virtual world and the model of virtual persons. We will first try to expose some key problems, which are intrinsically tied to the basic conceptualization of the virtual person model with its interlinked structure between physical and virtual entities. We will then discuss some of the existing trust management techniques and see whether they provide satisfactory solutions to the underlying trust problems.

7.2.1 Trusting Virtual Persons

One particular type of trust problem results from the disguised link between a virtual person and the physical person(s) by which it is controlled. Similar problems arise in situations in which the same physical person is hidden behind multiple virtual identities. Problems of that kind are quite fundamental for applications situated in the digital world. As an example, consider the problem of paying for an auctioned item on eBay (see Subsection 3.3), where the unknown seller is only visible as a subscribed eBay user with a corresponding account, but not as a tangible physical person within reach. We may use other users' ratings to judge the seller's trustworthiness, but considering their own trustworthiness with respect to issuing such ratings makes the problem even more complicated (see Subsection 6.1.1).

To analyze the application of trust to virtual persons more systematically, let us look at the basic model proposed in the FIDIS deliverables D2.13 and D17.1 (see summary in Subsection 4.3), in which the connection between physical and virtual persons is described as a *n*-to-*m* relationship: one physical person may be linked to several virtual persons, and several physical persons may be linked to the same virtual person. Now if we permit virtual persons or general virtual entities as objects of a trust relationship, as suggested in Definition 3 (see Subsection 4.1.4), we obtain four entirely new, but quite fundamental questions:

- Q1: Can a trust relationship between a trustor and a physical person be transferred to a virtual person?
- Q2: Can a trust relationship between a trustor and a virtual person be transferred to a physical person?
- Q3: Can a trust relationship between a trustor and a physical person be transferred to another physical person?
- Q4: Can a trust relationship between a trustor and a virtual person be transferred to another virtual person?

It is obvious that transferring trust from physical to virtual persons (Q1), or vice versa Q2), requires respective links in the model. The problem is that the existence of these links is generally unknown to the trustor, and transferring trust may thus be impossible even in cases in which such links exist. In other words, the transfer of trust between physical and virtual persons may only occur after a process of link verification has given the trustor sufficient evidence to become confident in its existence.[128] This means that, for example, we should

[128] This "link verification" is nothing but an identification process (see Section 7.3).

only rely on the e-mail content apparently received from a trustworthy friend if we have enough evidence to believe that our friend is in control of that particular e-mail account and appears thus to be the author of the message. Note that a complicating issue in this particular example is the possibility of forging the address field of SMTP-based e-mails (see Subsection 3.5).

To model formally the link verification (identification) process, we may look at it as another particular instance of the evidence-based process of forming respective opinions, as exposed in the last part of Subsection 4.1. In its most general form, we may thus come out with an additive triplet (*b*,*d*,*i*) of respective degrees of belief, disbelief, and ignorance, which may then be combined with the opinion included in the original trust relation to obtain an adapted opinion for the transferred trust relation. A complicating issue of this general method is the possibility of the available evidence to include statements from other physical or virtual entities, which may themselves depend on further trust relationships. Assuming that these complicating issues do not entirely prevent the belief formation process, it is then a question of defining meaningful thresholds for accepting or rejecting corresponding trust assumptions.

The most challenging questions in the above list are the problems of transferring a trust relationship from a physical person to another physical person (Q3), or similarly, from a virtual person to another virtual person (Q4). For this to take place, it is compulsory that at least two links exists. These links are needed to connect the two entities under consideration via an intermediate virtual entity (in the case of Q3) or an intermediate physical entity (in the case of Q4). This intermediate entity can then be regarded "trust transfer hub", over which the original trust relationship is transferred to the target entity. This mechanism is illustrated in Fig. 19 for each of the two cases.

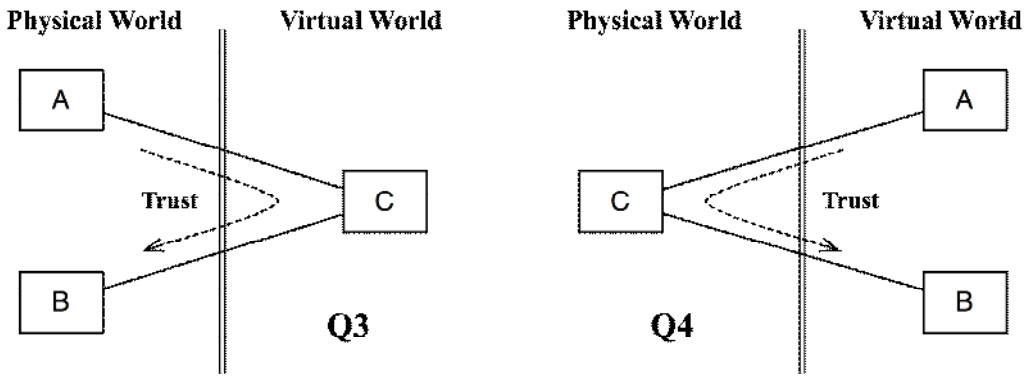

Figure 19: Transferring trust from A to B over C in the situations Q3 and Q4.

Note that the intermediate entity may possibly be unknown to the trustor. In the case of Q4, such situations are not untypical though. They arise for example when two mutually unknown eBay users switch from the messaging service provided by eBay to traditional E-mail communication, e.g. to arrange the payment details of an auctioned item.

If we assume that the transfer of trust between physical and virtual persons can be handled properly, then we may additionally pose questions related to the transitivity of trust. Then we may also try to apply or to extend the methods from existing trust management systems to virtual persons. This will be the topic of the next subsection. Note that the fact that a physical

person may be linked to various virtual persons has then an important impact on whether certain independence assumptions are still justified or not. Such independence assumptions are fundamental in many trust metrics and management systems.

7.2.2 Trust Management Systems

Designing general computational models of trust has always been a principal goal since the scientific investigation of trust has started in various areas. The mathematical foundation of such a computational model, together with associated algorithms, is what is usually called *trust metric*. In sociology and psychology, trust metrics are used as a measure of how somebody is trusted by others. Usually, they use quantitative statements about *direct* trust relationships between two individuals as input data to compute quantitative estimates of *indirect* or *derived* trust relations. This general idea has been adopted in many e-commerce, network security, peer-to-peer, web-services, or online community applications to draw conclusions about the trustworthiness of their users. One of the key problems there is the fact that the objects of the involved trust relations are not the users of the applications themselves, but corresponding entities of the virtual world. The real users are thus hidden behind the chosen usernames, pseudonyms, virtual addresses specified by the application. This means that all the fundamental questions of the previous subsection are directly applicable and should thus be taken into account. As we will see next, many trust metrics and trust management systems are based on simplifying assumptions to avoid these problems from the beginning.

Due to the wide range of different application areas, each with its own characteristics and singularities, there is no general agreement on what is the "best" trust metric. To provide a rough guideline when judging the appropriateness of a concrete trust metric, some authors tried to identify different soundness properties that one would expect from a reasonable trust metric (Degerlund, 2007). Trust metrics are often embedded in so-called *trust management systems* (TMS), which support trust decision processes in distributed systems. Research on trust management systems is rooted in authentication based on public-key certificates, see Subsection 3.5 or (Zimmermann, 1994; Blaze et al., 1996), but it is nowadays established in many other application areas of distributed systems. We refer to (Ruohomaa, Kutvonen, 2005) for an excellent survey.

7.2.2.1 Classification of Trust Management Systems

Despite the diversity and wide variety of proposed trust metrics and trust management systems, a few principal dimensions with distinctive features have been identified as major axes for a possible classification (Ziegler, Lausen, 2005; Wang, Vassileva, 2007). One of them concerns the origin and availability of the input data, and another one the place and the so-called trust view of the evaluator. These axes are not entirely orthogonal, as the following discussion will show.

A centralized trust management system is based on one or several central authorities, which are responsible for making judgments and decisions about the trustworthiness of the users (virtual persons). To place such trust decisions at everybody's disposal, they are usually stored in central repositories. For such a system to work, it is necessary that all users accept the central authorities as trustworthy with respect to the task of making such trust decisions and for properly maintaining the accuracy and data consistency of the repository. In this way, the direct trust relation between a particular user and a central authority is transformed into an indirect trust relation between two users. The classic X.509 PKI is one of the most prominent

examples of such a centralized system. What X.509 and similar approaches have in common is that users are actually virtual persons with digital identities in form of WWW or e-mail addresses. While a X.509 PKI creates strong links between these digital identities (or their related virtual persons) and corresponding cryptographic key pairs (similar to PGP, see Subsection 3.5), they do not provide similarly strong links between the virtual persons and the physical persons or organizations behind them. Trust mechanisms based on PKIs are therefore naturally and exclusively located in the virtual world.

In opposition to centralized approaches, most research on trust metrics focuses explicitly on a *decentralized* or *distributed* trust decision process, in which individual users are empowered to make their own decisions and to communicate them to others in form of credentials. Note that the mechanisms in distributed systems are usually more complex to implement than those in centralized systems (Wang, Vassileva, 2007). This is a consequence of the fact that a particular subject in a trust relation usually requires its own individual repository of trust-related evidence (see next subsection). Distributing this evidence and managing the local repositories is obviously more complex than in the case of a single central repository. It also means that more entities are involves in the whole process, and that these entities are again located in the virtual world. The resulting possibility of mix-ups or ambiguities creates an additional potential of failures.

The second major dimension to classify trust management system concerns the place and the corresponding *trust view*, from which indirect trust relations are evaluated and established (Ruth et al., 2004). In a *global* system, the information relevant for making trust decisions is public and visible to all members of the system. For a given trust metric, the evaluation of such a global trust view is the same for all users and can thus be delegated to a single user or central authority. Note that centralized trust management systems are usually global and vice versa. In a *local* (or *personalized*) trust management system, each user collects its own repository of trust-related information. Depending on the users' connections and interactions in the network, this can lead to quite different trust views and corresponding conclusions. Local trust management systems are designed to implement the subjective aspect of trust.

In the context of large online communities, in which users are allowed to issue ratings or recommendations about each other, trust management systems are sometimes called *reputation systems*. They differ from *recommendation systems* in their purpose of establishing interpersonal trust among users rather than trust towards external resources such as books, music, services, or web pages. The object type of the intended trust relations is thus another major dimension for the classification of trust management systems. It allows us to distinguish between *interpersonal* and *resource-oriented* trust management systems.

Another important distinguishing feature of trust management systems is the type of the underlying trust metrics. Some of the existing trust metrics, so-called *scalar metrics*, are designed to quantitatively evaluate trust relations of particular pairs of users, whereas *group metrics* are designed to compute each user's trustworthiness individually. Note that scalar metrics are inherently local (Ziegler, Lausen, 2005). Another distinctive characteristic of the underlying trust metric is the actual choice of the mathematical representation of trust (see last part of Subsection 4.1) and the adopted mechanism to derive indirect trust relations from direct trust relations. A short outline of such *trust propagation algorithms* is given in the following subsection.

7.2.2.2 Web of Trust

In a distributed trust management system, we can assume that a particular user (a virtual person) issues credentials about some of the other users (virtual persons too), but not about all of them. Such a credential, which describes a direct trust relationship between its issuer and recipient, can then be regarded as a link between two nodes in a trust network called web of trust. To represent different degrees of trust and to distinguish them properly from respective degrees of distrust, we assume here that an opinion is assigned to each link in the network (see Subsection 4.1.7). A missing link between two nodes means that no trust opinion has yet been formed. An exemplary web of trust with virtual persons (users) *A*, *B*, *C*, *D*, and *E* with corresponding links and opinions is shown in Fig. 20. Note that most of the existing webs of trust are far less general than the proposed opinion-based scheme.

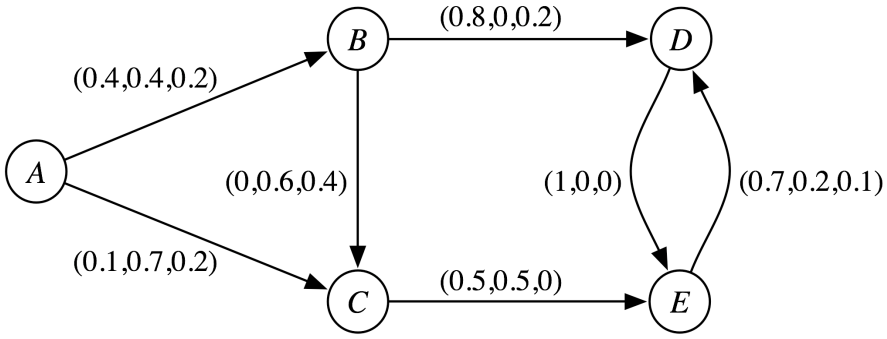

Figure 20: A simple web of trust with users *A*, *B*, *C*, *D*, *E* and corresponding opinions.

For a given web of trust like the one in Figure 13, the principal question now is how to use the given links and opinions to evaluate indirect trust relations for a pair of virtual persons with no direct link, e.g., for *A* and *E*. To make such calculations, a couple of assumptions with regard to the transitivity of trust and distrust need to be imposed (Ziegler, Lausen, 2005; Jøsang et al., 2006). As mentioned before, trust relations are only transitive under very particular semantic constraints. It is thus important for a meaningful trust propagation method to prove that such constraints are given in the intended application and to properly expose the necessary conditions and parameters for those constraints.

One particular set of such conditions, which is known as *conditional transitivity* (Abdul-Rahman, Hailes, 1997), constitutes the foundation for many existing trust metrics. Its key requirement for a credential to be used to derive transitive trust is that the issuer of the credential is considered trustworthy as a recommender. For this, we may either assume a generic trust context, which includes the activity of issuing credentials, or we may distinguish at least two trust contexts, one for the application-specific activity and another one for the activity of issuing credentials (Haenni et al., 2007).

Another important, but less cited condition for transitive trust results from the problem of unambiguously attributing a credential to its issuer. This is an identification problem of informational trust with respect to the origin of a piece of information. In a closed system with registered users, e.g., in eBay's so-called "feedback forum", this problem is usually solved by implementing various security measures and policies, which prevent users from issuing credentials about others from the outside of their own user account. Such systems however can often not guarantee that each participant is only registered once, i.e., as a unique

user. This can easily be exploited to circumvent negative reputation. This is a typical weakness of many centralized or global trust management systems. It results from the more general problem of attributing trust in the presence of persons with multiple virtual identities. It seems that this problem is inherent to all sorts of trust management systems (Kohlas, 2007).

7.2.2.3 Trust Propagation

A decentralized solution for the problem of validating the origin of a credential is to request a digital signature. The corresponding trust decision is then a question of verifying the digital signature using the issuer's public-key certificate. Note that certificates themselves are a particular form of credentials, i.e., accepting the authenticity of a piece of information after successfully verifying a digital signature requires an additional trust decision towards the issuer's ability and will of issuing certificates only after carefully checking the recipient's identity. The problems of authentication and trust are therefore intrinsically intertwined. This observation has been the motivation of many PKI-related trust management systems, which are thus followers of PGP's original proposal of a web of trust (Zimmerman, 1994; Maurer, 1996; Levien, Aiken, 1998; Jonczy, Haenni, 2005; Kohlas, 2007). For a general analysis of such two- or three-layer models we refer to (Haenni et al., 2007).

If a system and the intended application are such that all the necessary conditions for transitive trust are satisfied, then the problem of the "right" trust propagation method arises. The proposal of a network with opinions attributed to the links is one of the most general schemes, and it is best analyzed from an algebraic perspective (Jøsang, 1999; Theodorakopoulos, 2004). Most other approaches are restricted to single-valued trust representations and are therefore not designed to cope with some of the most important aspects of trust such as the aforementioned distinction between distrust and untrust (see Subsection 4.1). Their advantage however is the reduced mathematical complexity. For example, if each value assigned to a link (or node) of a web of trust is interpreted as a probability of operation of that link (or node), then it is possible to transform the problem of computing transitive trust into a *network reliability* problem, for which a wide variety of general solutions exist (Mahoney, 2005; Haenni, Jonczy 2007). Many other single-valued trust propagation schemes use local trust aggregation functions like weighted sums or averages to update the degree of trustworthiness of a user according to its incoming links in the web of trust. Some of them iterate through the loops of arbitrarily long trust chains. The simplest systems are usually those with discrete trust values. They often do no more than generating simple statistics or applying some pragmatic trust propagation rules. The most popular systems of that kind are eBay's "feedback forum" and the web of trust in PGP (see Subsections 3.3 and 3.5).

Another popular family of approaches to solve the trust propagation problem is based on logical inference rules. Those rules are designed to formally describe the process of building up transitive trust relations. Usually, they are based on some first-order predicate logic (Beth, 1994, Maurer, 1996; Jonczy, Haenni, 2005; Gutscher, 2007; Kohlas, 2007). Their strength is the preciseness and transparency obtained from using logical predicates to encode fundamental concepts like trust and trustworthiness. However, most of those systems are computationally not very efficient and thus not scalable to large real world applications.

7.2.3 Conclusion

Trust mechanisms based on PKIs, an important class of today's trust mechanism, appear to be naturally and exclusively located in the virtual world. Trust models usually make simplifying

assumptions about the relations (or independence) between the virtual persons that are involved. In particular, usually, they do not take into consideration information that could be retrieved, for example, from the knowledge that two different virtual persons are likely to represent the same physical person. Those simplified assumptions are required to make the trust models applicable in real-life scenarios. As discussed in this deliverable, the model based on virtual persons would allow describing the reality much more faithfully. The trust models could be refined, taking into consideration the potential relations between physical entities and virtual ones. Nevertheless, from a practical point of view, such a refinement might lead to inapplicable models, especially for large-scale systems, due to an overwhelming complexity.

7.3 Identification in the Virtual World

In Subsection 4.2, we have produced an in-depth understanding of the identification process in the classic real world context. We repeat here the general definition of identification:

Definition: *Identification is the process of establishing enough confidence in the fact that some identity-related information is valid and truly describes a specific entity in a given context or environment, at a certain time.*

Any identity-related information in a given context or environment is the tautological identity of its corresponding virtual entity.

An entity is represented by this virtual entity at a certain time, if the corresponding identity-related information in the given context or environment is valid and truly describes this specific entity at that time. In other words, identification of an entity means the verification of the existence of a oriented link, at a certain time, from a virtual entity –tautologically defined by the identity-related information in the given context or environment– to this entity. The entity to identify might be a physical entity or a virtual one.

The oriented link between two entities indicates that the first entity, a virtual one, represents the second one. This is in line with the UML-description in FIDIS deliverable D17.1.

Definition There is a (direct) *link* between two entities, if one entity represents the other one.

Model

Link

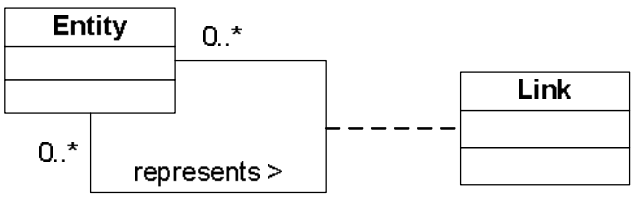

Figure 21: Oriented link

Intuitively speaking, the orientation goes from the information to the entity to identify, i.e., from the virtual entity defined by this information to the entity to identify. The verification of the existence of the oriented link means the verification that the information related to the virtual entity is also valid and truly describes the entity to identify. After a successful identification, the identified entity inherits the information related to the virtual entity from the point of view of the identificator or verifier. This creates a dynamic in the links from or to the identified entity as the existence of these links has to be re-evaluated in the light of the new information inherited by the identified entity.

As our definition of identification encompasses both the equality side and the inclusion side of identification, its transposition in the model based on virtual persons encompasses both sides too.

In the following subsections, we will consider three typical identification/authentication scenarios and represent them in the model based on virtual persons.

7.3.1 Username-Password

We consider here the classic scheme where a user wants to access a service and is authenticated by a username/password protocol. If the authentication is successful, the user is granted access to the service. What are the typical entities involved in this protocol in the model based on virtual persons from the point of view of the verifier:

- The physical person registered with the username
- The physical person that tries to access a service
- A corresponding virtual person: the one trying to access the service
- Another virtual person: the one who knows the password related to the given username

The right to access the service is granted to the virtual person *"The one who knows the password related to the username"* and is inherited by any entity that is identified as knowing the password related to the username.

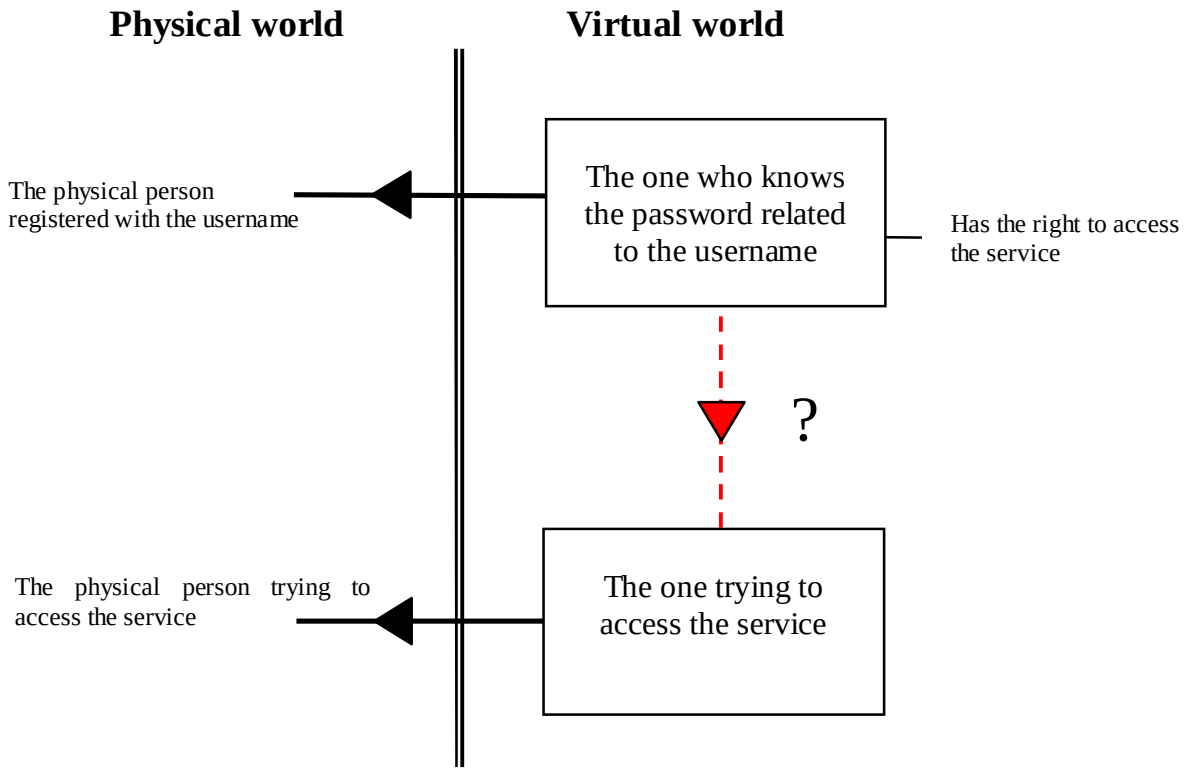

Figure 22: Username – password protocol

The physical person registered with the username is supposed to know the related password. The corresponding oriented link is not verified (forgetting the password means, in the model, that the link disappears). The oriented link between the virtual person *"The one trying to access the service"* and the physical person trying to access the service is tautological and does not need to be verified. Therefore, the verifier actually only checks the existence of an oriented link from *"The one who knows the password related to the username"* and *"The one trying to access the service"* in the virtual world. If this identification succeeds in the virtual world, then the right to access the service are inherited first by *"The one trying to access the service"* and then, as a consequence, by the physical person trying to access the service.

The whole identification process actually occurs in the virtual world: *"The one trying to access the service"* is identified as *"The one who knows the password related to the username"*. In particular, at the end, the verifier cannot conclude that the physical person trying to access the system is the physical person registered with the username.

Actually, in general, the verifier is not even sure that the oriented link from *"The one trying to access the service"* points to a physical person. It could be a non-human physical entity (a computer, a robot, a dog, etc.) or even another virtual person (for example, a software-agent).

Moreover, the one who knows the password related to the username is not necessarily a single physical person: it could be a group of persons (a virtual person), a computer or a smartcard, etc.

Therefore, the following simplified diagram focuses on what really happens when running a username – password protocol:

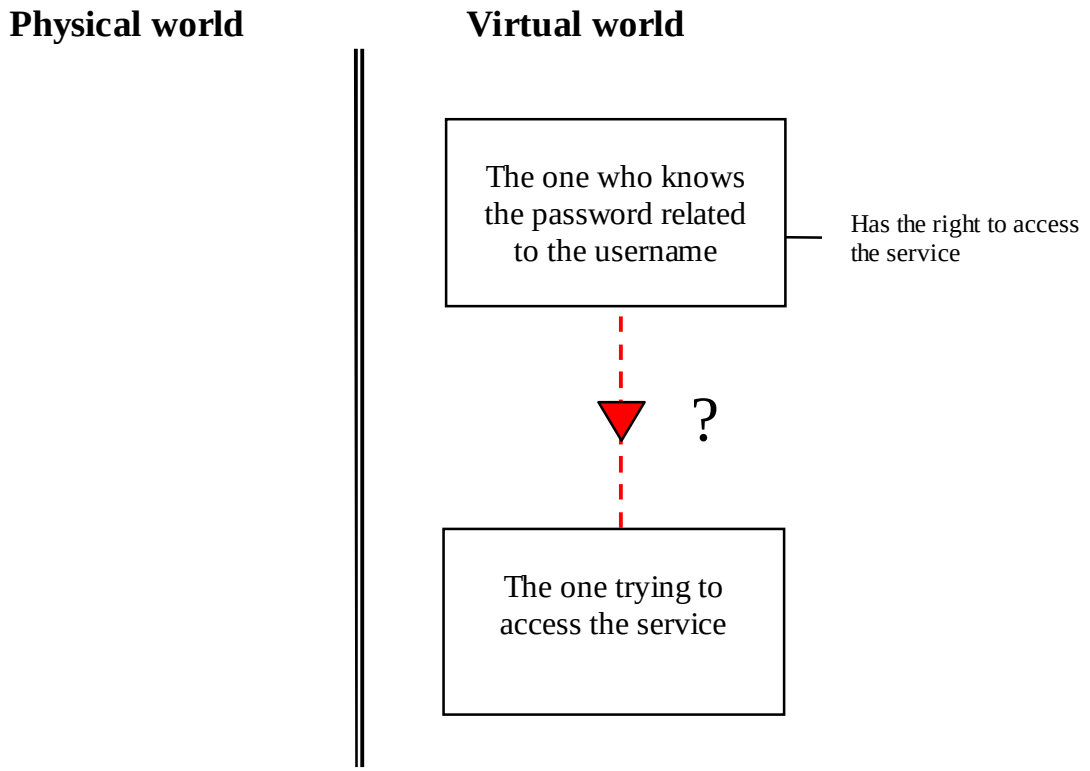

Figure 23: Username – password protocol (core process)

We observe in particular that what really happens exclusively occurs in the virtual world.

7.3.2 Biometric authentication

In this subsection, we describe the typical situation of a physical person using some biometric feature in order to be authenticated.

During the registration, we have two entities: a physical one *"The physical person being registered"* and a virtual one *"The one who is the source of the registered template"*.

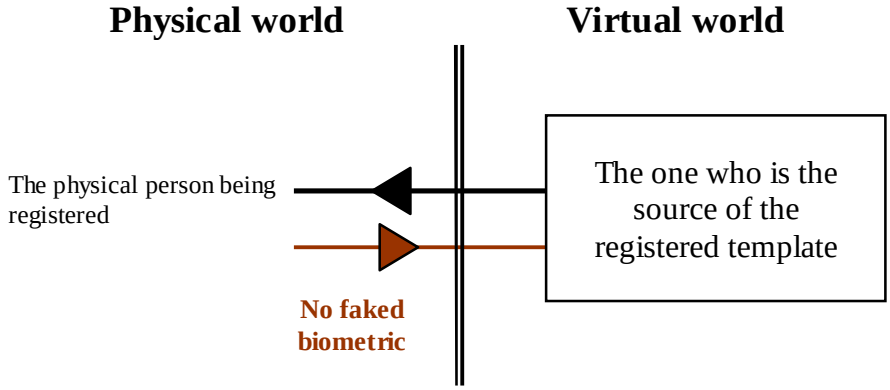

Figure 24: Biometric authentication (registration)

The oriented link from the virtual person to the physical one is trivial. However, as we will see later, we also need an oriented link from the physical one to the virtual one (in brown in the above figure). This link is only valid if the physical person uses his true biometric feature (not a faked one) during the registration process. When necessary, the registration can be done in presence of a trusted observer in order to avoid this kind of fraud, so that the verifier later can have enough confidence in the existence of this link.

Not all biometric features are suited for use as an authentication technique. Indeed, we try to find biometric features so that, ideally, the virtual person *"The one who is the source of the registered template"* and the virtual person *"The one matching with the registered template"* are the same, i.e., they both can be identified to the other one.

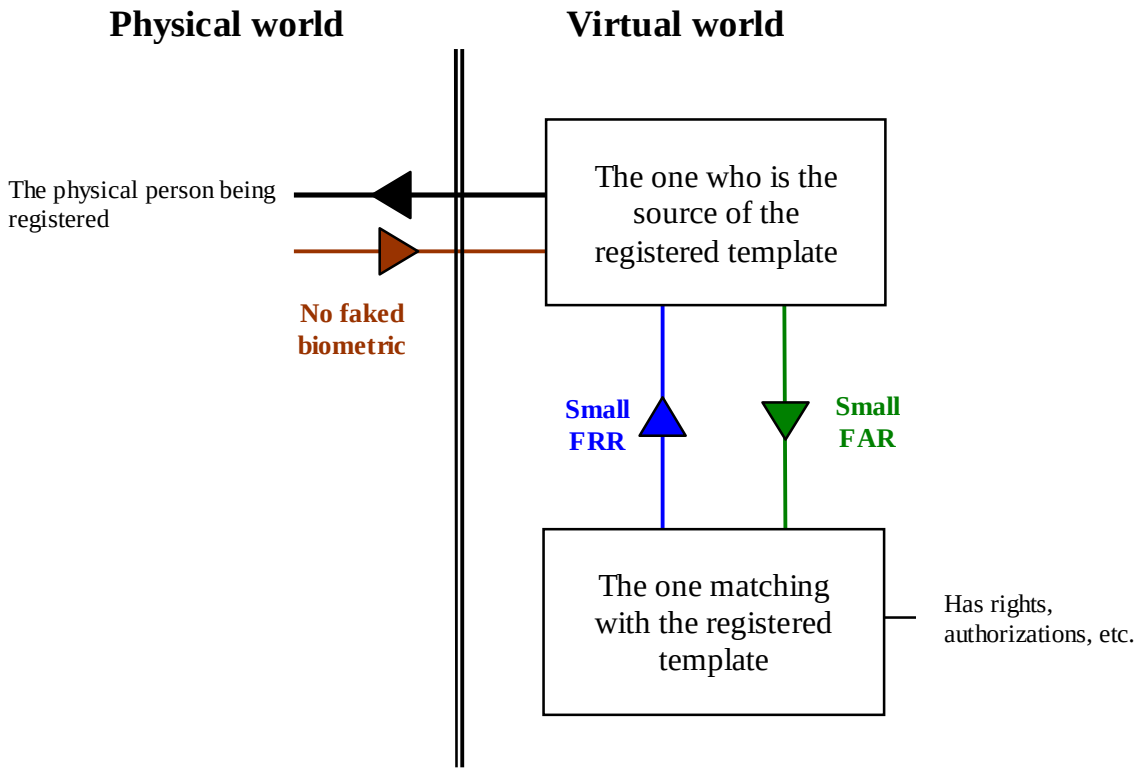

Figure 25: Biometric authentication (registration)

The verification of the existence of an oriented link from *"The one matching with the registered template"* to *"The one who is the source of the registered template"* requires a small false rejection rate (FRR), while the verification of the existence of an oriented link from *"The one who is the source of the registered template"* to *"The one matching with the registered template"* requires a small false acceptance rate (FAR). The small FRR is necessary for the biometric to be user-friendly and convenient. The small FAR is necessary for the verifier to accept the existence of the corresponding oriented link (green) that belongs to the chain of confidence as we will see later.

Rights, authorizations, etc. are associated with *"The one matching with the registered template"*.

During the actual authentication process, we have two other entities, a physical one *"The physical person trying to be authenticated"* and a virtual one *"The one who is the source of the ongoing template"*.

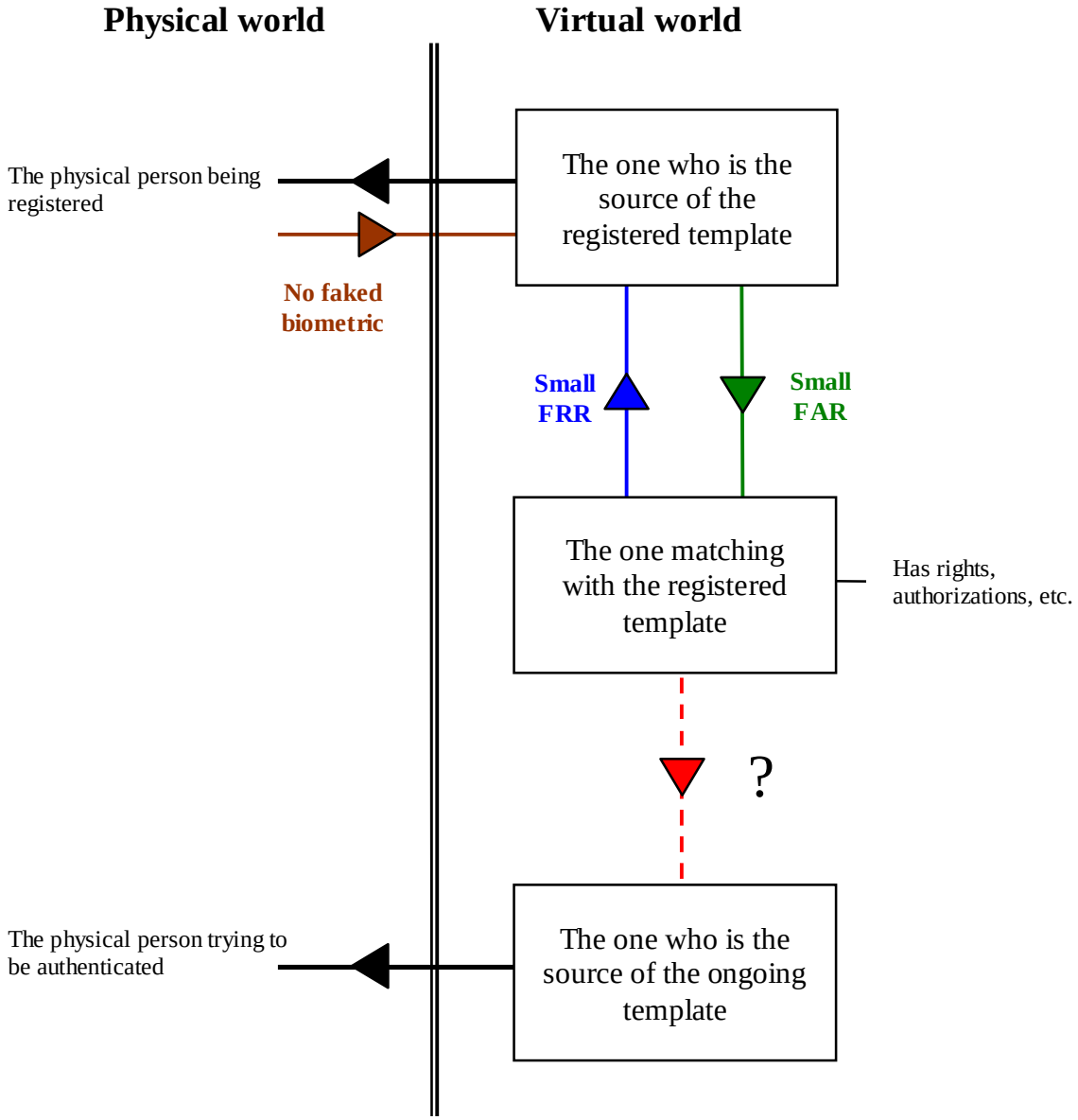

Figure 26: Biometric authentication (actual authentication)

The oriented link from *"The one who is the source of the ongoing template"* and *"The physical person trying to be authenticated"* is trivial.

The matching process actually verifies the existence of the oriented link from *"The one matching with the registered template"* to *"The one who is the source of the ongoing template"*. In case of a match (successful verification of the existence of the red oriented link), i.e. if the virtual person *"The one who is the source of the ongoing template"* is identified as *"The one matching with the registered template"*, the authentication process is considered to be successful. This closes an oriented path, a chain of confidence, from *"The*

physical person being registered" to *"The physical person trying to be authenticated"*, under the assumption that the registered biometric was not faked and that the FAR is sufficiently small.

It is interesting to point out that the model based on virtual persons emphasizes the indirection between a physical person and his biometric templates. The model clearly shows assumptions that are too often implicit (or even forgotten) and represents in a visual way the chain of confidence that is at stake. This illustrates also the different points of attacks with respect to biometric authentication.

7.3.3 Passport

In this subsection, we describe the typical situation of a physical person presenting his passport at the border. The border officer must verify the identity of the person holding the passport. The passport is used as a (hopefully) corroborating evidence of the true identity of its holder.

If somebody has received a valid passport, we have two entities: a physical one *"The physical person owning the valid passport"* and a virtual one *"The one who is described in the valid passport"*. Strong procedures are developed to build confidence in the fact that only true citizens are issued a valid passport and that the information in the passport about its owner is correct. As already mentioned in Subsection 6.2.1, the border officer must fully trust his own state CA who issues the passports.

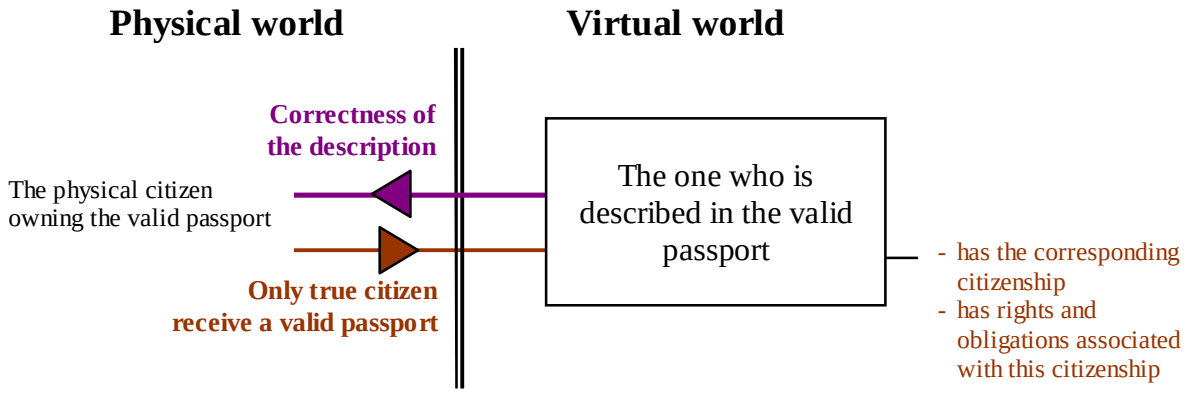

Figure 27: Passport issuance

If the passport is issued by another country, the border officer might question the existence of an oriented link between *"The physical citizen owning the valid passport"* and *"The one who is described in the valid passport"*. If this identification fails from the point of view of the border officer, he might deny the owner of the passport the rights associated with this citizenship. Indeed, the virtual person *"The one who is described in the valid passport"* would not inherit in this case the citizenship property from the physical person owning the valid passport.

In the following, we will suppose that the border officer has enough confidence in the issuance process.

When somebody presents his passport at the border, we have two other entities: a physical one *"The physical person who is present at the border"* and a virtual one *"The one who is described in the passport"*.

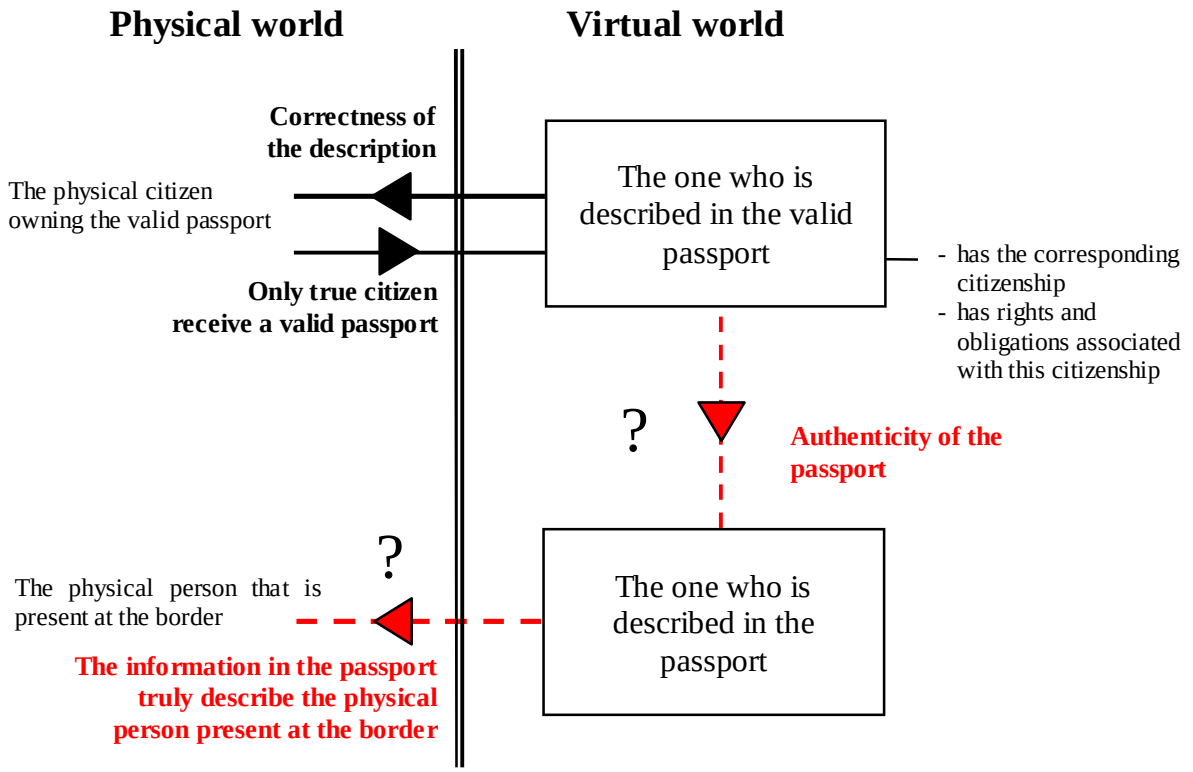

Figure 28: Border control

In order to grant the physical person that is present at the border rights and obligations associated with his citizenship, the border officer must perform two successful identifications.

The first identification consists in verifying the existence of an oriented link from *"The one who is described in the valid passport"* to *"The one who is described in the passport"*. This consists in authenticating the validity of the passport (date of expiry, not a stolen passport, the passport is not a forged one, etc.).

The second identification consists in verifying the existence of an oriented link from *"The one who is described in the passport"* to *"The physical person who is present at the border"*. The border officer must have enough confidence in the fact that the information (picture, age, gender, fingerprints, etc.) contained in the passport truly describe the physical person who is present at the border.

Actually, the border officer usually tries a third identification to see if the physical person present at the border belongs to the list of wanted criminals. More precisely, he tries either to identify or to exclude *"The one described in the passport"* and *"The one who is wanted"*. In case of a successful identification, the physical person at the border might be refused the entrance in the country or even arrested.

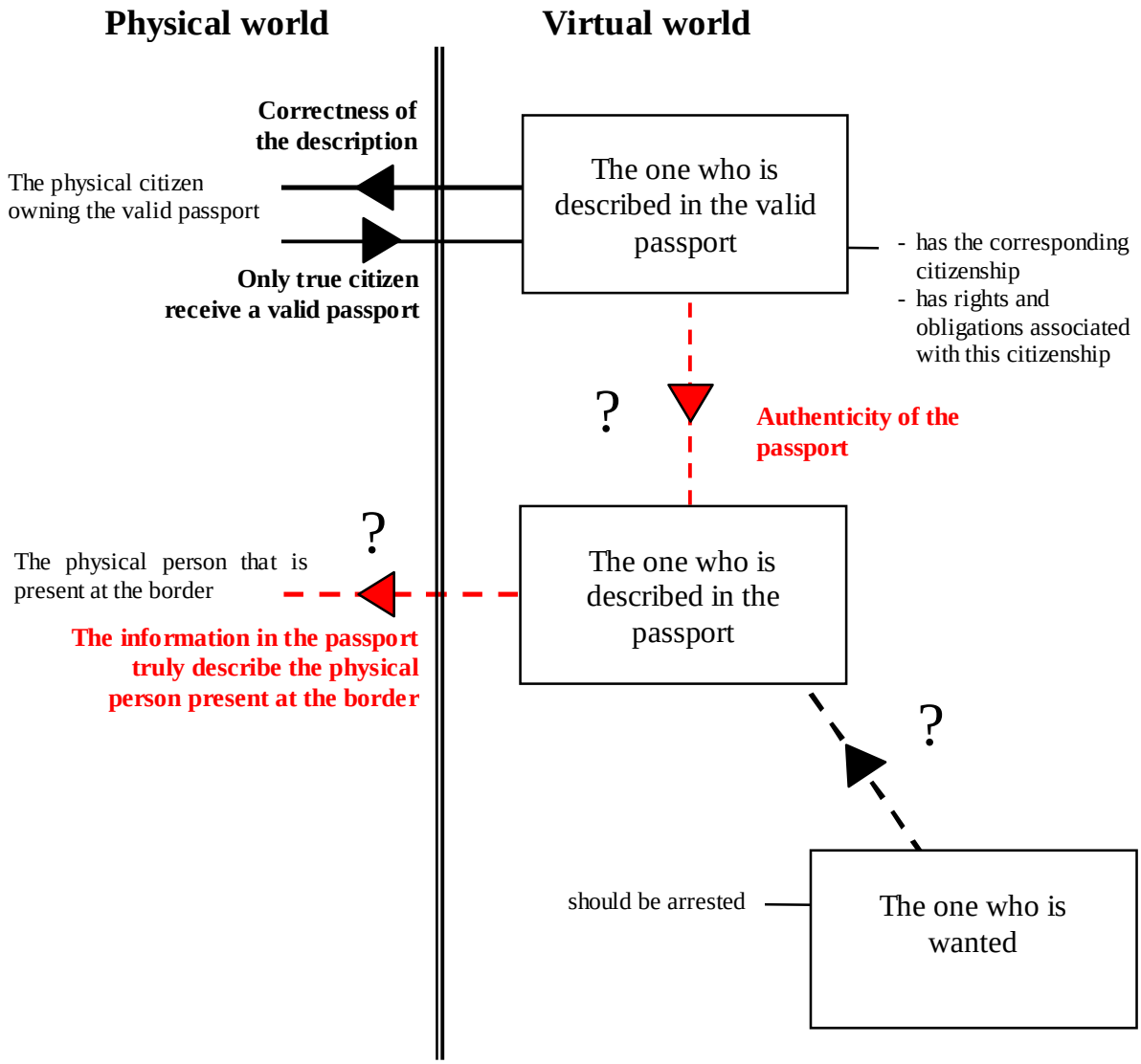

Figure 29: Border control: extra check

7.3.4 Conclusion

Identification in the light of virtual persons unifies both sides of identification, namely the equality side and the inclusion side. It leads therefore to a conceptual simplification. Contrarily to what we observed for trust, the refinement brought by the model based on virtual persons does not make it difficult to apply to identification. It efficiently allows a more precise decomposition of practical identification schemes (username/password, biometrics, passports, etc.). As a direct consequence, possible points of attack appear more clearly. This might also help to detect wrong simplified assumptions leading to inappropriate conclusions.

Most of the identification processes actually occur in the virtual world. The identification of a physical person and a virtual person verifies the existence of a corresponding link between the physical world and the virtual one. Moreover, these links between the physical world and the virtual one play a central role when transferring trust from one virtual person to another one.

8 Conclusion

The model based on virtual persons allows a very precise and faithful description of many interactions and processes that happen in particular in the information society. Because of this precision, we quickly realized that before trying to illuminate and analyze the concepts of trust, confidence and identification from the perspective of the "virtual person" model in the virtual world, we needed first to capture the essence of these concepts in the classic real world context (Section 4). This thorough study led to an in-depth understanding of these concepts and of the relations between them that goes beyond the former state-of-the art. This preliminary study becomes eventually one of the main outputs of this deliverable.

The precise definition of *identification* (Subsection 4.2) brings the traditional IT approach closer to what is usually adopted in forensic sciences. In the IT community, we see too often confusion between identification and individualization. We hope that this precise, but general, definition of identification will bridge both communities on this core concept.

Trust mechanisms based on PKIs, an important class of today's trust mechanism, appear to be naturally and exclusively located in the virtual world (Section 7). Trust models usually make simplifying assumptions about the relations (or independence) between the virtual persons that are involved. In particular, usually, they do not take into consideration information that could be retrieved, for example, from the knowledge that two different virtual persons are likely to represent the same physical person. Those simplified assumptions are required to make the trust models applicable in real-life scenarios. As discussed in this deliverable, the model based on virtual persons would allow describing the reality much more faithfully. The trust models could be refined, taking into consideration the potential relations between physical entities and virtual ones. Nevertheless, from a practical point of view, such a refinement might lead to inapplicable models, especially for large-scale systems, due to an overwhelming complexity.

Identification in the light of virtual persons (Subsection 7.3) unifies both sides of identification, namely the equality side and the inclusion side. It leads therefore to a conceptual simplification. Contrarily to what we observed for trust, the refinement brought by the model based on virtual persons does not make it difficult to apply to identification. It efficiently allows a more precise decomposition of practical identification schemes (username/password, biometrics, passports, etc.). As a direct consequence, possible points of attack appear more clearly. This might also help to detect wrong simplified assumptions leading to inappropriate conclusions.

Most of the identification processes actually occur in the virtual world. The identification of a physical person and a virtual person verifies the existence of a corresponding link between the physical world and the virtual one. Moreover, these links between the physical world and the virtual one play a central role when transferring trust from one virtual person to another one.

9 References

Abdul-Rahman, A., Hailes, S. 2000. Supporting trust in virtual communities, *HICSS-33, 33rd Hawaii International Conference on System Sciences*, pp 1769–1777, Maui, USA.

Adams, C., 2005. Identification. In van Tilburg, Henk C. A. (editor-in-chief), *Encyclopedia of Cryptography and Security*, pages 272–273. Springer.

Artz, D., Gil, Z. 2007. A survey of trust in computer science and the semantic web, *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):58–71.

Ashraf, N., Bohnet, I., Piankov, N. 2006. Decomposing trust and trustworthiness, *Experimental Economics*, 9(3):193–208.

Berg, J., Dickhaut, J., McCabe, K. 1995. Trust, reciprocity and social history, *Games and Economic Behaviour*, 10:122–142.

Beth, T., Borcherding, M., Klein, B. 1994. Valuation of trust in open networks, *ESORICS'94, 3rd European Symposium on Research in Computer Security*, LNCS 875, pp 3–18.

Birch, D.G.W. 2009. Psychic ID: Ablueprint for a modern national identity scheme, IDIS Identity in the Information Society, DOI 10.1007/s12394-009-0014-6, Springer.

Blaze, M., Feigenbaum, J., Lacy, J. 1996, Decentralized trust management, *SP'96, IEEE Symposium on Security and Privacy*, pp 164–173, Oakland, USA.

Branchaud, M., Flinn, S. 2004. x-Trust: A scalable trust management infrastructure, *PST'04, 2nd Annual Conference on Privacy, Security and Trust*, pp 207–218, Fredericton, Canada.

Castelfranchi, C., Falcone, R. 2001. Social trust: a cognitive approach, Castelfranchi, C., Tan, Y. H. (eds), *Trust and Deception in Virtual Societies*, pp 55–90, Kluwer Academic Publishers.

Chaudhuri, A., Gangadharan, L. 2007. An experimental analysis of trust and trustworthiness, *Southern Economic Journal*, 73(4):959–985.

Chaum, D., 1990. Showing credentials without identification: Transferring signatures between unconditionally unlinkable pseudonyms, *Auscrypt '90, LNCS 453*, pp 246-264, Springer, Berlin.

Cho, J. 2006. The mechanism of trust and distrust formation and their relational outcomes, *Journal of Retailing*, 82(1):25–35.

Christianson, B., Harbison, W.S. 1997. Why isn't trust transitive?, in Christianson, B., Crispo, B., Lomas, T.M.A., Roe, M. (eds), *IWSP'97, 5th International Workshop on Security Protocols*, LNCS 1189, pp 171–176, Paris, France.

Cofta, P. 2007. Confidence, trust and identity, *BT Technology Journal*, 25(2):173–178.

Degerlund, F. 2007. Trust mass, volume and density – a novel approach to reasoning about trust, *Electronic Notes in Theoretical Computer Science*, 179:87–96.

Flowerday, S., von Solms, R. 2006. Trust: An element of information security, Fischer-Hübner, S., Rannenberg, K., Yngström, L., Lindskog, S. (eds), *SEC'06, 21st International Information Security Conference*, pp 87–98, Karlstad, Sweden.

Gefen, D. 2002. Reflections on the dimensions of trust and trustworthiness among online consumers, *ACM SIGMIS Database*, 33(3):38–53.

Gerck, E. 2002. Trust as qualified reliance on information, *The COOK Report on Internet*, X(10):19–24.

Grandison, T., Sloman, M. 2000. A survey of trust in internet applications, *IEEE Communications Surveys and Tutorials*, 3(4).

Griffiths, N. 2006, A fuzzy approach to reasoning with trust, distrust and insufficient trust, Klusch, M., Rovatsos, M., Payne, T.R. (eds), *CIA'06, 10th International Workshop on Cooperative Information Agents*, volume 4149 of LNCS 4149, pp 360–374, Edinburgh, U.K.

Guha, R., Kumar, R., Raghavan, P., Tomkins, A. 2004, Propagation of trust and distrust, in Feldman, S.I., Uretsky, M., Najork, M., Wills, C.S. (eds), *WWW'04, 13th International Conference on World Wide Web*, pp 403–412, New York, USA.

Gutscher, A. 2007. A trust model for an open, decentralized reputation system, in Etalle, S., Marsh, S. (eds), *IFIPTM'07, 1st Joint iTrust and PST Conferences on Privacy Trust Management and Security*, pp 285–300, Moncton, Canada.

Haenni, R. 2005. Using probabilistic argumentation for key validation in public-key cryptography. International Journal of Approximate Reasoning, 38(3):355–376.

Haenni, R. 2009. Non-additive degrees of belief, Huber F, Schmidt-Petri C (eds), *Degrees of Belief*, pp 121–160, Springer.

Haenni, R., Jonczy, J., Kohlas, R. 2007. Two-layer models for managing authenticity and trust, in Song, R., Korba, L., Yee, G. (eds), *Trust in E-Services: Technologies, Practices and Challenges*, section VI, pp 140–167, Idea Group Publishing.

Haenni, R., Jonczy, J. 2007. A new approach to PGP's web of trust. *EEMA'07, European e-Identity Conference*, Paris, France.

Hájek, P., Valdés, J.J. 1991, Generalized algebraic approach to uncertainty processing in rule-based expert systems (dempsteroids), *Computers and Artificial Intelligence*, 10:29–42.

Hardin, R. 2006, Trust (Key Concepts), Polity.

Jaquet-Chiffelle, D. O., Benoist, E., Haenni, R., Wenger, F., Zwingelberg, H. 2009. Virtual persons and identities. In Rannenberg, K., Royer, D., Deuker, A. (eds), *The Future of Identity in the Information Society (FIDIS) – Challenges and Opportunities*, pages 75–122. Springer.

Jonczy, J., Haenni, R. 2005. Credential networks: a general model for distributed trust and authenticity management, in Ghorbani, A., Marsh, S. (eds), *PST'05, 3rd Annual Conference on Privacy, Security and Trust*, pp 101–112, St. Andrews, Canada.

Jøsang, A. 1999. An algebra for assessing trust in certification chains, *NDSS'99: 6th Annual Symposium on Network and Distributed System Security*, San Diego, USA.

Jøsang, A. 2001. A logic for uncertain probabilities, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311.

Jøsang, A., Gray, E., Kinateder, M. 2006. Simplification and analysis of transitive trust networks, *Web Intelligence and Agent Systems*, 4(2):139–161.

Kinateder, M., Rothermel, K. 2003, Architecture and algorithms for a distributed reputation system, *iTrust'03: 1st International Conference on Trust Management*, LNCS 2692, pp 1–16, Heraklion, Greece.

Kind, S. 1987. The scientific investigation of crime. Forensic Science Services.

Kohlas, R. 2007. Decentralized Trust Evaluation and Public-Key Authentication, PhD thesis, University of Bern, Switzerland.

Lee, J., Moray, N. 1992. Trust, control strategies and allocation of function in human-machine systems, *Ergonomics*, 35:1243–1270.

Levien, R., Aiken, A. 1998, Attack-resistant trust metrics for public key certification, *Security'98, 7th USENIX Security Symposium*, pp 229–242, San Antonio, USA.

Lewicki, R.J. 2006. Trust and distrust, in Kupfer Schneider, A., Honeyman, C. (eds), *The Negotiator's Fieldbook: The Desk Reference for the Experienced Negotiator*, ch. 22, pp 191–202, American Bar Association.

Lewis, J.D., Weigert_ A. 1985. Trust as a Social Reality, *Social Forces*, 63(4): 967–985.

Li, H., Singhal, M. 2000. Trust management in distributed systems, Computer, 40(2):45–53.

Luhmann, N. 2000. Vertrauen – ein Mechanismus der Reduktion sozialer Komplexität, UTP, Stuttgart, Germany, 4th ed.

Mahoney, G., Myrvold, W., Shoja, G.C. 2005. *Generic reliability trust model*, in Ghorbani, A., Marsh, S. (eds), *PST'05: 3rd Annual Conference on Privacy, Security and Trust*, pp 113–120, St. Andrews, Canada.

Marsh, S.P. 1994. Formalising Trust as a Computational Concept, PhD thesis, University of Stirling, Scotland, U.K.

Marsh, S., Dibben, M.R. 2005. Trust, untrust, distrust and mistrust – an exploration of the dark(er) side, in Herrmann, P., Issarny, V., Shiu, S. (eds), *iTrust'05: 3rd International Conference on Trust Management*, LNCS 3477, pp 17–33, Rocquencourt, France.

Maurer, U. 1996. Modelling a public-key infrastructure, in Bertino, E., Kurth, H., Martella, G., Montolivo, E. (eds), *ESORICS, European Symposium on Research in Computer Security*, LNCS 1146, pp 324–350.

McKnight, H.D., Chervany, N.L. 1996. The meanings of trust, Technical report, Carlson School of Management, University of Minnesota, Minneapolis, USA.

McLeod, C. 2006. Trust, in Zalta, E.N. (ed), *The Stanford Encyclopedia of Philosophy*, Center for the Study of Language and Information, Stanford University, USA.

Muir, B.M. 1987. Trust between humans and machines, and the design of decision aids, International Journal of Man-Machine Studies, 27(5-6):527–539.

Numan, J. H. 1998. Knowledge-Based Systems as Companions: Trust, Human Computer Interaction and Complex Systems, PhD thesis, University of Groningen, The Netherlands.

Oliver, A., Montgomery, K. 2001. A system cybernetic approach to the dynamics of individual- and organizational-level trust, *Human relations*, 54(8):1045–1063.

O'Hara, K., 2004. *Trust – From Socrates to Spin*. Icon Book Ltd.

Pavlou, P. A. 2002. Institution-based trust in interorganizational exchange relationships: the role of online B2B marketplaces on trust formation. *Journal of strategic information systems*. Vol. 11, no 3–4, 2002, pp. 215–243.

Rannenberg, K., Royer, D., Deuker, A (eds). 2009. The Future of Identity in the Information Society (FIDIS) – Challenges and Opportunities. Springer.

Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K. 2006. The Value of Reputation on eBay: A Controlled Experiment. Experimental Economics. Volume 9, Issue 2, Jun 2006, Pages 79-101.

Rotter, J.B. 1980. Interpersonal trust, trustworthiness, and gullibility, *American Psychologist*, 35:1–7.

Ruohomaa, S., Kutvonen, L. 2005. Trust management survey, Herrmann P, Issarny V, Shiu S (eds), *iTrust'05: 3rd International Conference on Trust Management*, LNCS 3477, pp 77–92, Rocquencourt, France.

Ruth, P., Xu, D., Bhargava, B., Regnier, F. 2004, E-notebook middleware for accountability and reputation based trust in distributed data sharing communities, in Jensen, C.D., Poslad, S., Dimitrakos, T. (eds), *iTrust'04, 2rd International Conference on Trust Management*, LNCS 2995, Oxford, U.K.

Schlechtriem, P. 2000. Restitution und Bereicherungsausgleich in Europa: eine rechtsvergleichende Darstellung, Tübingen, Mohr Siebeck.

Steinbrecher, S., Gross, S., Meichau, M. 2009. Jason: A scalable reputation system for the semantic web, in Emerging challenges for security, privacy and trust, Gritzalis, D., Lopez, J. (Eds), 24th IFIP TC 11 International Conference, Security Conference, SEC2009, Proceedings, Springer.

Sztompka, P. 1999. Trust: a Sociological Theory, Cambridge University Press.

Theodorakopoulos, G. 2004. *Distributed trust evaluation in ad-hoc networks*, Master's thesis, University of Maryland, College Park, USA.

Wang, Y., Vassileva, J. 2007. Toward trust and reputation based web-service selection: A survey, *International Transactions on Systems Science and Applications*, 3(2):118–132.

Williamson, O.E. 1993. Calculativeness, trust, and economic organization, *Journal of Law & Economics*, 36(1):453–486.

Yu, B., Singh, M.P. 2002. Distributed reputation management for electronic commerce, *Computational Intelligence*, 18(4):535–549.

Zhang, Q., Yu, T., Irwin, K. 2004. A classification scheme for trust functions in reputation-based trust management, *ISWC'04, 3rd International Semantic Web Conference*, Workshop on *Trust, Security, and Reputation on the Semantic Web*, Hiroshima, Japan.

Ziegler, C.N., Lausen, G. 2005. Propagation models for trust and distrust in social networks, *Information Systems Frontiers*, 7(4–5):337–358.

Zimmermann, P. R. 1994. The Official PGP User's Guide. MIT Press.