Chapter 1

TRUST-BASED DECISION MAKING FOR ELECTRONIC TRANSACTIONS[1]

Audun Jøsang
Norwegian University of Science and Technology, N-7491 Trondheim, Norway
ajos@item.ntnu.no

Abstract Financial transactions that are made in an environment of imperfect knowledge will always contain a degree of risk. When dealing with humans or human organisations the relative knowledge about the co-operative behaviour of others can be perceived as trust, and trust therefore is a crucial factor in the decision making process. However, assessing trust becomes a problem in electronic transactions due to the impersonal aspects of computer networks. This paper proposes a scheme for propagating trust through computer networks based on public key certificates and trust relationships, and demonstrates how the resulting measures of trust can be used for making decisions about electronic transactions.

1 INTRODUCTION

The concept of trust has received considerable attention in the information security literature, but it is rarely clear what people exactly mean by it. In a way, trust and security are just two sides of the same thing, because if a system is (in)secure, it is (dis)trusted, and if it is (dis)trusted, then it must be (in)secure. This observation can be explained by considering security as a property of a system in a given environment, and trust as a subjective belief resulting from assessing a system and its environment. Philosophically seen it is impossible to determine security perfectly because we are condemned to live in a subjective world and do not have direct access to reality. As such trust is a subjective measure of a system's assumed real security.

Not only systems can be (dis)trusted, but also humans and human organisations. However, it would be meaningless to say that a human is (in)secure, so

[1] Appears in the Proceedings of the Fourth Nordic Workshop on Secure Computer Systems (NORDSEC'99), Stockholm University, Sweden, 1999

this must be a different type of trust. In [Jøs96] the trust in systems and trust in humans for the purpose of information security was defined as follows:

- Trust in a system is the belief that it will resist malicious attacks.
- Trust in a human is the belief that he or she will co-operate and not defect.

In both cases trust is a belief, and a suitable metric for belief also becomes a suitable metric for trust [Jøs99]. The trust model which will be described in Sec.2 uses the term '*opinion*' to express belief, and a metric for opinions that can express measures of belief ranging from absolute belief to absolute disbelief as well as uncertainty. Sec.3 describes a set of operators for combining opinions that can be used to logically analyse trust relationships and to infer new trust values.

We will use a traditional method for decision making based on expected utility of the outcomes of transactions between (human) agents or organisations. However, we will not attach probabilities to the possible outcomes of a transaction, but instead assume that the outcome of the transaction depends entirely on whether the other agent defects or co-operates, making it possible to use trust measures to determine rational choice, i.e. whether or not to enter into a transaction with a particular agent. For electronic transactions across computer networks, the problem is essentially to determine reliable trust measures about an agent you want to transact with, but whom you for example have never interacted with before and will never interact with again after the transaction.

In the definition of trust above, the term 'system' is not limited to cover physical computer equipment, but can also mean abstract entities such as cryptographic algorithms or keys. Cryptographic mechanisms always rely on initial trust assumptions, and have as purpose to create trust in certain system entities such as the confidentiality of a message or the authenticity of a certificate. As such, cryptographic mechanisms can be said to propagate trust from where it exists to where it is needed.

An algebra for assessing trust in certification chains is described in [Jøs99], and we will extend this method in order to compute trust values about remote agents. Other methods for valuating authenticity have been analysed in [RS97, Jøs98]. By combining this method for propagating trust through computer networks with a traditional utility based decision making model, we are able to determine rational choice regarding electronic transactions with remote agents.

2 THE TRUST MODEL

The metric that will be used to represent trust is based on a belief model similar to the belief model of the Dempster-Shafer theory. The Dempster-Shafer theory of evidence was first set forth by Dempster in the 1960s and subsequently extended by Shafer who in 1976 published *A Mathematical Theory of Evidence*[Sha76]. A more recent description cab be found in e.g. [SK94].

A proposition such as "*the agent will co-operate*" is assumed to be either true or false, and not something in between, and is therefore a binary proposition. However, due to our imperfect knowledge it is impossible to know with certainty whether it is true or false, so that we can only have an *opinion* about it, which translates into degrees of belief or disbelief as well as uncertainty in case both belief and disbelief are lacking. We express this mathematically as:

$$b + d + u = 1, \quad b, d, u \in [0, 1] \tag{1.1}$$

where b, d and u designate belief, disbelief and uncertainty respectively.

Uncertainty is caused by the lack of evidence to support either belief or disbelief. In order to illustrate the interpretation of the uncertainty component we will use the following example, which is cited from [Ell61].

"Let us suppose that you confront two urns containing red and black balls, from one of which a ball will be drawn at random. To 'bet on Red_{I}' will mean that you choose to draw from Urn I; and that you will receive a prize a (say \$100) if you draw a red ball and a smaller amount b (say \$0) if you draw a black. You have the following information: Urn I contains 100 red and black balls, but in ratio entirely unknown to you; there may be from 0 to 100 red balls. In Urn II, you confirm that there are exactly 50 red and 50 black balls."

For Urn II, most people would agree that the probability of drawing a red ball is 0.5, because the numbers of red and black balls are equal. For Urn I however, it is not obvious. If one was forced to make a bet on Red_{I}, most people would agree that the probabilities of drawing a red ball and a black ball are equal, so that the probability of drawing a red ball also in this case must be 0.5.

This example illustrates extreme cases of probability, one which is totally certain, and the other which is totally uncertain, but interestingly they are both 0.5 The uncertain probability of 0.5 is intuitively determined by the binary set of 2 possible colours $\Theta = \{\text{red, black}\}$ and that $|\{\text{red}\}| = 1/2 \times |\Theta|$. Had there for example been an uknown number of red, black, yellow and blue balls in the urn so that $\Theta = \{\text{red, black, yellow, blue}\}$, then $|\{\text{red}\}| = 1/4 \times |\Theta|$ and the uncertain probability of drawing a red ball would have been 0.25.

The term *relative atomicity*, denoted by a, will be used to describe the atomicity of a state relative to the atomicity of the full state space. With the same set of 4 colours, the relative atomicity of $\{\text{red} \cup \text{black}\}$ is $a_{\text{red}\cup\text{black}} = 1/2$ and the uncertain probability of drawing a red or a black ball would have been 0.5.

With the concepts defined so far we can define the metric that will be used to express trust.

Definition 1 Opinion
Let $\omega = (b, d, u, a)$ be a quadruple where the components correspond to belief, disbelief, uncertainty and relative atomicity respectively in the same order, and where (b, d, u) satisfy Eq. (1.1). Then ω is called an opinion.

□

The three coordinates (b, d, u) are dependent through Eq.(1.1) so that one is reduntant. As such they represent nothing more than the traditional (*Belief*, *Plausibility*) pair of Shaferian belief theory [Sha76]. However, it is useful to keep all three coordinates in order to obtain simple expressions when introducing operators on opinions in Sec.3. Allthough an opinion has 4 coordinates it is in fact a 3 dimensional metrics for representing uncertain probabilities. Def. 1 differs from that in previous papers ([Jøs99] and earlier) by the inclusion of the relative atomicity. This is necessary for the computation of a consistent probability expectation value of opinions.

Definition 2 (Probability expectation) Let $\omega = (b, d, u, a)$ be an opinion. Then the probability expeectation of ω, denoted by $\mathrm{E}(\omega)$, is defined by:

$$\mathrm{E}(\omega) = b + au \tag{1.2}$$

□

This definition corresponds to the pignistic probability described in e.g. [SK94], and is based on the principle of insufficient reason: uncertainty about n atomic states is split equally among these n states.

The probability expectation of a given state is thus determined by the belief, uncertainty and relative atomicity components. The probability expectation function removes information so that there can be infinitely many opinions that correspond to the same probability expectation value. Eq.(1.1) defines a triangle that can be used to graphically illustrate opinions as shown in Fig.1.1.

As an example the position of the opinion $\omega_x = (0.40,\ 0.10,\ 0.50,\ 0.60)$ is indicated as a point in the triangle. Also shown are the probability expectation value and the relative atomicity. The relative atomicity should be interpreted such that for example the atomicity of x is 6 and the atomicity of the full state space is 10 producing $a_x = 0.60$.

The horizontal bottom line between the belief and disbelief corners in Fig.1.1 is called the *projection plane*. The relative atomicity can be graphically represented as a point on the projection plane. The line joining the top corner of the triangle and the relative atomicity point becomes the *director*. In Fig.1.1 $a_x = 0.60$ is represented as a point, and the dashed line touching it represents the director.

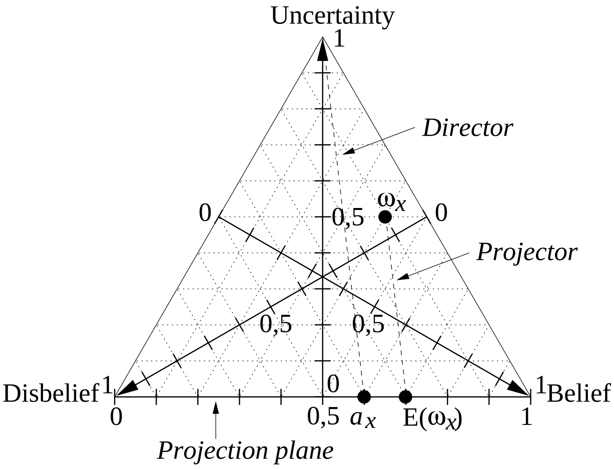

Figure 1.1 Opinion triangle with ω_x as example

The *projector* is parallel to the director and passes through the opinion point. Its intersection with the projection plane defines the probability expectation value which otherwise can be computed by the formula of Def.2. The position of the probability expectation $\mathrm{E}(\omega_x) = 0.70$ is shown.

Opinions situated in the left or right corner, i.e. with either $b = 1$ or $d = 1$ are called *absolute opinions*. They represent situations where it is absolutely certain that a state is true or that a state is false, and correspond to TRUE or FALSE proposition in binary logic.

Opinions situated on the projection plane are called *dogmatic opinions*. They represent situations without uncertainty and correspond to traditional probabilities. The distance between an opinion point and the projection plane can be interpreted as the degree of uncertainty.

In real situations, a probability estimate can never be certain, and a single valued probability estimate is always inadequate for expressing an observer's subjective belief regarding a real situation. By using opinions the degree of (un)certainty can easily be expressed such that the opinions about Red$_{\mathrm{I}}$ and Red$_{\mathrm{II}}$ from the example of the urns above become $\omega_{\mathrm{I}} = (0.0,\ 0.0,\ 1.0,\ 0.5)$ and $\omega_{\mathrm{II}} = (0.5,\ 0.5,\ 0.0,\ 0.5)$ respectively.

Opinions can be strictly ordered by first ordering opinions according to probability expectation, and subsequently ordering those with the same probability expectation according to certainty. By taking the examples with the urns, we have for example that $\mathrm{E}(\omega_{\mathrm{I}}) = \mathrm{E}(\omega_{\mathrm{II}})$ but $\omega_{\mathrm{I}} < \omega_{\mathrm{II}}$.

It is of course crucial that opinions can be consistently determined in order for the model to be useful in practical applications. It can be shown

[Jøs98] that opinions can be deterministically determined if the evidence at hand can be analysed statistically. If on the other hand the evidence can only be intuitively and qualitatively assessed, it is sufficient to determine the probability of truth of a proposition like for example: x : *The agent will co-operate in the next transaction* and to compute the corresponding opinion by assuming the proposition to which the opinion applies to be binary, and to maximise the uncertainty by using Eq.(1.2) and Eq.(1.1). For example $p = 0.9 \Rightarrow \omega = (0.8,\ 0.0,\ 0.2,\ 0.5)$. Because most people are familiar with probability assessment, this constitutes a practical way of determining intuitive opinions.

3 SUBJECTIVE LOGIC

Standard binary logic operates on binary propositions that can take the values TRUE or FALSE. *Subjective logic* which we will present in this section operates on opinions about binary propositions, i.e. opinions about propositions that are assumed to be either true or false. Binary logic and subjective logic are thus very similar, and are in fact compatible.

Opinions are considered individual, and will therefore have an ownership assigned whenever relevant. In our notation, superscripts indicate ownership, and subscripts indicate the proposition to which the opinion applies. For example ω_x^A is an opinion held by agent A about the truth of proposition x. Below we will define the set of operators for opinions which constitutes subjective logic.

3.1 CONJUNCTION AND DISJUNCTION

A conjunction of two opinions about propositions from different state spaces consists of determining from the two opinions a new opinion reflecting the truth of both propositions simultaneously.

Definition 3 (Conjunction)
Let $\omega_x = (b_x, d_x, u_x, a_x)$ and $\omega_y = (b_y, d_y, u_y, a_y)$ be an agent's opinions about two distinct propositions x and y. Let $\omega_{x\wedge y} = (b_{x\wedge y}, d_{x\wedge y}, u_{x\wedge y}, a_{x\wedge y})$ be the opinion such that

$$\begin{array}{ll}
\mathbf{1}. & b_{x\wedge y} = b_x b_y \\
\mathbf{2}. & d_{x\wedge y} = d_x + d_y - d_x d_y \\
\mathbf{3}. & u_{x\wedge y} = b_x u_y + u_x b_y + u_x u_y \\
\mathbf{4}. & a_{x\wedge y} = \frac{b_x u_y a_y + u_x a_x b_y + u_x a_x u_y a_y}{b_x u_y + u_x b_y + u_x u_y}
\end{array}$$

where $u_{x\wedge y} \neq 0$. Then $\omega_{x\wedge y}$ is called the conjunction of ω_x and ω_y, representing the agent's opinion about both x and y being true. By using the symbol "$\wedge$" to designate this operation, we define $\omega_{x\wedge y} \equiv \omega_x \wedge \omega_y$.

□

A disjunction of two opinions about propositions from different state spaces consists of determining from the two opinions a new opinion reflecting the truth of one or the other or both propositions.

Definition 4 (Disjunction)
Let $\omega_x = (b_x, d_x, u_x, a_x)$ and $\omega_y = (b_y, d_y, u_y, a_y)$ be an agent's opinions about two distinct propositions x and y. Let $\omega_{x\vee y} = (b_{x\vee y}, d_{x\vee y}, u_{x\vee y}, a_{x\vee y})$ be the opinion such that

1. $b_{x\vee y} = b_x + b_y - b_x b_y$
2. $d_{x\vee y} = d_x d_y$
3. $u_{x\vee y} = d_x u_y + u_x d_y + u_x u_y$
4. $a_{x\vee y} = \frac{u_x a_x + u_y a_y - b_x u_y a_y - u_x a_x b_y - u_x a_x u_y a_y}{u_x + u_y - b_x u_y - u_x b_y - u_x u_y}$

where $u_{x\vee y} \neq 0$. Then $\omega_{x\vee y}$ is called the disjunction of ω_x and ω_y, representing the agents opinion about x or y or both being true. By using the symbol "$\vee$" to designate this operation, we define $\omega_{x\vee y} \equiv \omega_x \vee \omega_y$.

□

As would be expected, conjunction and disjunction of opinions are both commutative and associative. Idempotence is not defined because it would assume that the arguments are identical and therefore belong to the same state space. It must always be assumed that the arguments are independent.

In case x and y belong to binary state spaces the conjunction and disjunction operators are equivalent to the 'AND' and 'OR' operators of Baldwin's support logic [Bal86] except for the relative atomicity parameter. When applied to absolute opinions, i.e with either $b = 1$ or $d = 1$, the conjunction and disjunction operators are equivalent to 'AND' and 'OR' of binary logic, that is; they produce the truth tables of logical 'AND' and 'OR' respectively. When applied to dogmatic opinions, i.e opinions with zero uncertainty, they produce the same results as multiplication and co-multiplication of probabilities respectively. It can be observed that when the uncertainty is zero the relative atomicity is not defined in Def. 3 and Def. 4. In practice this poses no problem because in that case the relative atomicity is irrelevant and can therefore be skipped in the calculations. See also comment about dogmatic opinions in Sec.3.4 below.

Conjunction and disjunction of opinions are not distributive on each other. If for example ω_x, ω_y and ω_z are opinions we have:

$$\omega_x \wedge (\omega_y \vee \omega_z) \quad \neq \quad (\omega_x \wedge \omega_y) \vee (\omega_x \wedge \omega_z) \tag{1.3}$$

This result which may seem surprising is due to the fact that that ω_x appears twice in the expression on the right side so that it in fact represents the disjunction of partially dependent arguments. Only the expression on the left side is thus correct.

Conjunction decreases the relative atomicity whereas disjunction increases it. What really happens is that the product of the two state spaces produces a new state space with atomicity equal to the product of the respective atomicities. However, as opinions only apply to binary state spaces, a new binary state space is formed of the conjunction (or the disjunction) and its complement. The expressions for relative atomicity in Def. 3 and Def. 4 are in fact obtained by forming the product of the two state spaces and applying Def. 2 and Eq. (1.1).

The following theorem shows that subjective logic is compatible with probability calculus regarding product and co-product of probabilities.

Theorem 1 (Product and co-product)
Let $\omega_x = (b_x, d_x, u_x, a_x)$ and $\omega_y = (b_y, d_y, u_y, a_y)$ be an agent's opinions about two distinct propositions x and y, and let $\omega_{x\wedge y} = (b_{x\wedge y}, d_{x\wedge y}, u_{x\wedge y}, a_{x\wedge y})$ and $\omega_{x\vee y} = (b_{x\vee y}, d_{x\vee y}, u_{x\vee y}, a_{x\vee y})$ be their conjunction and disjunction respectively. The probability expectation function E satisfies:

$$\begin{array}{ll} \mathbf{1}. & \mathrm{E}(\omega_{x\wedge y}) = \mathrm{E}(\omega_x)\mathrm{E}(\omega_y) \\ \mathbf{2}. & \mathrm{E}(\omega_{x\vee y}) = \mathrm{E}(\omega_x) + \mathrm{E}(\omega_y) - \mathrm{E}(\omega_x)\mathrm{E}(\omega_y) \end{array}$$

Proof 1 *Each property can be proved separately.*

1. *Equation 1 corresponds to the product of probabilities. By using Def.2 and Def. 3 we get:*

$$\begin{aligned} & E(\omega_{x\wedge y}) \\ & = b_{x\wedge y} + u_{x\wedge y}\ a_{x\wedge y} \\ & = b_x b_y + b_x u_y a_y + u_x a_x b_y + u_x a_x u_y a_y \\ & = (b_x + u_x a_x)(b_y + u_y a_y) \\ & = E(\omega_x)E(\omega_y) \end{aligned} \tag{1.4}$$

2. *Equation 2 corresponds to the co-product of probabilities. By using Def.2, Def. 4 and Eq.(1.1) we get:*

$$\begin{aligned} & E(\omega_{x\vee y}) \\ & = b_{x\vee y} + u_{x\vee y}\ a_{x\vee y} \\ & = b_{x\vee y} + (d_x u_y + u_x d_y + u_x u_y)\ a_{x\vee y} \\ & = b_{x\vee y} + (u_x + u_y - b_x u_y - u_x b_y - u_x u_y)\ a_{x\vee y} \\ & = b_x + b_y - b_x b_y + u_x a_x + u_y a_y - b_x u_y a_y - u_x a_x b_y - u_x a_x u_y a_y \\ & = b_x + u_x a_x + b_y + u_y a_y - (b_x + u_x a_x)(b_y + u_y a_y) \\ & = E(\omega_x) + E(\omega_y) - E(\omega_x)E(\omega_y) \end{aligned} \tag{1.5}$$

□

3.2 NEGATION

The negation of an opinion about proposition x represents the opinion about x being false.

Definition 5 (Negation)
Let $\omega_x = (b_x, d_x, u_x, a_x)$ be an opinion about the proposition x. Then $\omega_{\neg x} = (b_{\neg x}, d_{\neg x}, u_{\neg x}, a_{\neg x})$ is the negation of ω_x where:

$$\begin{array}{ll} \mathbf{1}. & b_{\neg x} = d_x \\ \mathbf{2}. & d_{\neg x} = b_x \\ \mathbf{3}. & u_{\neg x} = u_x \\ \mathbf{4}. & a_{\neg x} = 1 - a_x \end{array}$$

□

Negation can be applied to expressions containing conjunction and disjunction, and it can be shown that De Morgans's laws are valid.

3.3 RECOMMENDATION

Assume two agents A and B where A has an opinion about B in the form of the proposition: '*B will co-operate*'. In addition B has an opinion about a proposition x. A recommendation then consists of combining A's opinion about B with B's opinion about x in order for A to get an opinion about x. There is no such thing as 'physical recommendation', and recommendation therefore lends itself to different interpretations. The main difficulty lies with describing the effect of A disbelieving that B will give a good advice. This we will interpret so that if A thinks that B ignores the truth value of x then A ignores the truth value of x no matter what B's actual recommendation to A is.

Definition 6 (Recommendation)
Let A, B and be two agents where $\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$ is A's opinion about B's recommendations, and let x be a proposition where $\omega_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$ is B's opinion about x expressed in a recommendation to A. Let $\omega_x^{AB} = (b_x^{AB}, d_x^{AB}, u_x^{AB}, a_x^{AB})$ be the opinion such that

$$\begin{array}{ll} \mathbf{1}. & b_x^{AB} = b_B^A b_x^B, \\ \mathbf{2}. & d_x^{AB} = b_B^A d_x^B \\ \mathbf{3}. & u_x^{AB} = d_B^A + u_B^A + b_B^A u_x^B \\ \mathbf{4}. & a_x^{AB} = a_x^B \end{array}$$

then ω_x^{AB} is called the recommendation between ω_B^A and ω_x^B expressing A's opinion about x as a result of the recommendation from B. By using the symbol $\otimes$ to designate this operation, we define $\omega_x^{AB} \equiv \omega_B^A \otimes \omega_x^B$.

□

It is easy to prove that $\otimes$ is associative but not commutative. This means that the combination of opinions can start in either end of the chain, but that the order in which opinions are combined is significant. In a chain with more than one recommending entity, opinion independence must be assumed, which for example translates into not allowing the same entity to appear more than once in a chain. This operator is the same as the *discounting operator* defined in [Sha76].

3.4 CONSENSUS

The consensus opinion of two opinions is an opinion that reflects both opinions in a fair and equal way. For example if two agents have observed a machine over two different time intervals they might have different opinions about its reliability depending on the behaviour of the machine in the respective periods. According to the Bayesian approach the consensus must then be the opinion that a single agent would have after having observed the machine during both periods. It can be shown ([Jøs98]) that the following definition of consensus between opinions corresponds to this approach and is based on Bayesian calculus.

Definition 7 (Consensus)
Let $\omega_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)$ and $\omega_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$ be opinions respectively held by agents A and B about the same proposition x. Let $\omega_x^{A,B} = (b_x^{A,B}, d_x^{A,B}, u_x^{A,B}, a_x^{A,B})$ be the opinion such that

$$\begin{array}{ll}
\mathbf{1}. & b_x^{A,B} = (b_x^A u_x^B + b_x^B u_x^A)/\kappa \\
\mathbf{2}. & d_x^{A,B} = (d_x^A u_x^B + d_x^B u_x^A)/\kappa \\
\mathbf{3}. & u_x^{A,B} = (u_x^A u_x^B)/\kappa \\
\mathbf{4}. & a_x^{A,B} = \frac{a_x^B u_x^A + a_x^A u_x^B - (a_x^A + a_x^B) u_x^A u_x^B}{u_x^A + u_x^B - 2u_x^A u_x^B}
\end{array}$$

where $\kappa = u_x^A + u_x^B - u_x^A u_x^B$ and where $u_x^A = u_x^B \neq 0$ and $u_x^A = u_x^B \neq 1$. Then $\omega_x^{A,B}$ is called the consensus between ω_x^A and ω_x^B, representing an imaginary agent $[A, B]$'s opinion about x, as if she represented both A and B. By using the symbol $\oplus$ to designate this operation, we define $\omega_x^{A,B} \equiv \omega_x^A \oplus \omega_x^B$.

□

It is easy to prove that $\oplus$ is both commutative and associative which means that the order in which opinions are combined has no importance. Opinion independence is must be assumed, which obviously translates into not allowing an agent's opinion to be counted more than once

The effect of the consensus operator is to reduce the uncertainty. For example the case where several witnesses give consistent testimony should amplify the judge's opinion, and that is exactly what the operator does. Consensus between an infinite number of independent non-dogmatic opinions would necessarily produce a dogmatic consensus opinion, i.e. an opinion with zero uncertainty.

Two dogmatic opinions can not be combined according to Def.7. This can be explained by interpreting uncertainty as *room for influence*, meaning that it is only possible to reach consensus with somebody who maintains some uncertainty. A situation with conflicting dogmatic opinions is philosophically counterintuitive, primarily because opinions about real situations can never be certain, and secondly, because if they were they would necessarily be equal. The consensus of two absolutely uncertain opinions results in a new absolutely uncertain opinion, although the relative atomicity is not well defined. The limit of the relative atomicity when both $u_x^A, u_x^B \to 1$ is $(a_x^A + a_x^B)/2$, i.e. the average of the two relative atomicities, which intuitively makes sense.

The consensus operator will normally be used in combination with the recommendation operator, so that if dogmatic opinions are recommended, the recipient should not have absolute trust in the recommenders and thereby introduce ignorance before combining the recommendations by the consensus operator.

The consensus operator has the same purpose as Dempster's rule [Sha76], but is quite different from it. Dempster's rule can be criticised for producing counterintuitive results (see e.g. [Coh86]), but the same criticism does not apply to our operator.

3.5 THE PROBLEM OF DEPENDENCY

It is possible that several recommendation chains produce opinions about the same proposition. Under the condition of opinion independence, these opinions can be combined with the consensus rule to produce a single opinion about the target agent. An example of mixed consensus and recommendation is illustrated in Fig.1.2.

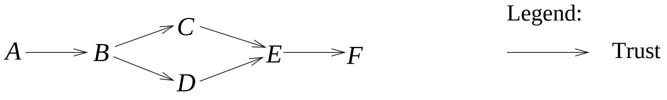

Figure 1.2 Mixing consensus and recommendation

The recommendation rule is not distributive relative to the consensus rule. Let ω^A_B, ω^B_C, ω^B_D, ω^C_E, ω^D_E and ω^E_F represent the opinion relationships in Fig.1.2. We then have

$$\begin{gathered} \omega^A_B \otimes ((\omega^B_C \otimes \omega^C_E) \oplus (\omega^B_D \otimes \omega^D_E)) \otimes \omega^E_F \\ \neq \\ (\omega^A_B \otimes \omega^B_C \otimes \omega^C_E \otimes \omega^E_F) \oplus (\omega^A_B \otimes \omega^B_D \otimes \omega^D_E \otimes \omega^E_F) \end{gathered} \tag{1.6}$$

which according to the short notation in Defs.6 and 7 can be written as

$$\omega_F^{AB(C,D)E} \neq \omega_F^{ABCE,ABDE} \quad . \tag{1.7}$$

The not-equal sign may seem surprising, but the right sides of (1.6) and (1.7) violate the requirement of independent opinions because both ω^A_B and ω^E_F appear twice. Only the left sides of (1.6) and (1.7) represent the graph of Fig.1.2 correctly. Note that the recommendation goes in the opposite direction of the trust arrows.

There will always be cases which can not be analysed completely. Fig.1.3 illustrates a situation where agent A needs to determine her opinion about F, of which she only has second-hand evidence trough a network of agents.

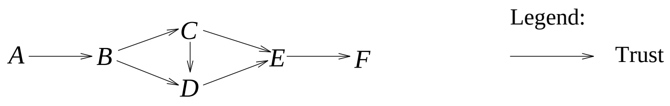

Figure 1.3 Intractable trust network

Whether the recommendations from D to C is ignored and thereby leaving out some of the evidence, or included and thereby violating the independence requirement, the result will never be as correct as one could wish.

4 TRUST-BASED DECISION MAKING

We assume transactions to have only two possible outcomes depending on whether the remote agent co-operates or defects. This is typically the case where the payment of goods and services is disconnected from the delivery. In case of defection, the seller might deliver without being paid. If payment comes first the buyer might pay without receiving any goods or services in return.

We also assume that each agent can attach utilities to the possible outcomes of a transaction, such that for example for agent A and transaction q, $\mathrm{U}^A_+(q)$ money units are gained if the other agent co-operates and $\mathrm{U}^A_-(q)$ money units are lost (i.e. negative value) if the other agent defects. We will not consider utilities of defection as seen from the defective agent's side, but the model can easily be extended for that purpose if desirable.

Let for example agent A's trust in agent B be expressed by ω_B^A. Now A considers a transaction q with B, for which A's gain and loss utilities are $\mathrm{U}_+^A(q)$ and $\mathrm{U}_-^A(q)$. Agent A can now compute the total expected utility $\mathrm{U}_B^A(q)$ of the transaction q with B as a function of the probability expectation values E of co-operation/defection and the corresponding transaction utilities U:

$$\begin{aligned} \mathrm{U}_B^A(q) &= \mathrm{E}(\omega_B^A)\ \mathrm{U}_+^A(q) + \mathrm{E}(\neg\omega_B^A)\ \mathrm{U}_-^A(q) \\ &= \mathrm{E}(\omega_B^A)\ (\mathrm{U}_+^A(q) - \mathrm{U}_-^A(q)) + \mathrm{U}_-^A(q) \end{aligned} \tag{1.8}$$

The rational choice of A will be to execute transaction q with B only if $\mathrm{U}_B^A(q) > 0$. The only difference between this method and a standard probability based method is that this method requires trust to be converted to a scalar expectation value (= probability) before computing the total expected utility.

As an example let A's trust in B be $\omega_B^A = (0.75,\ 0.05,\ 0.20,\ 0.50)$ ($\rightarrow$ $\mathrm{E}(\omega_B^A) = 0.85$) and the utilities for transaction q be $\mathrm{U}_+^A(q) = \$25$ and $\mathrm{U}_-^A(q) = -\$75$. By using Eq.(1.8) we get $\mathrm{U}_B^A(q) = \$10.00$ giving a small expected gain, indicating that the rational choice for A would be to execute the transaction although the risk would be relatively high.

The next section describes how trust values can be obtained through computer networks allowing the same method to be used for computing expected gain from electronic transactions with remote and a priori unknown agents.

5 PROPAGATION OF TRUST IN DATA NETWORKS

The mechanisms for propagating trust will be based on authentication by public key certificates combined with the trust relationships between users. We first describe a simple algebra [Jøs99] around these elements and subsequently describe how it can be used to compute measures of trust to be used for decision making in electronic transactions.

5.1 THE AUTHENTICATION ALGEBRA

Public keys can be exchanged manually or electronically. For manual distribution, agent $A1$ can for example meet agent $A2$ physically and give him a diskette containing her public key k_{A1}, and $A2$ can give his public key k_{A2} to her in return. The keys can then be considered authenticated through the persons' mutual physical recognition.

For electronic key distribution, keys need to be recommended and certified by someone whom the recipient trusts for recommending and certifying keys, and who's authenticated public key the recipient knows. For example if $A1$ knows $A2$'s public key k_{A2} and $A2$ knows $A3$'s public key k_{A3}, then $A2$ can send $A3$'s public key to $A1$, certified by his private key k_{A2}^{-1}. Upon reception, $A1$ will verify $A2$'s certificate, and if correct, will know that the received public key of $A3$ is authentic and can be used for secure communication with $A3$.

However, in order for a certificate to be meaningful, that is in order to really trust the authenticity of the public key it certifies, the recipient of the certificate must have an opinion $\omega^{A1}_{\text{KA}(k_{A2})}$ about the *Key Authenticity* (KA) of the key used to certify, that is, her opinion about the binding between the certifier and his public key. In addition, the recipient must have an opinion $\omega^{A1}_{\text{AC}(A2)}$ about *Agent Co-operation* (AC), that is, how much she trusts the certifier to actually recommend and certify other keys. Finally, the certifier must actually recommend to the recipient his own opinion $\omega^{A2}_{\text{KA}(k_{A3})}$ about the authenticity of the certified key by embedding it in the certificate to $A1$.

There are of course other considerations, such as e.g. that the cryptographic algorithm can not be broken, but it is assumed that these conditions are met.

We introduce the '*conjunctive recommendation term*' $(\omega^{A1}_{\text{AC}(A2)} \wedge \omega^{A1}_{\text{KA}(k_{A2}})$ which we will give the following short notation:

$$\omega^{A1}_{A2} \equiv (\omega^{A1}_{\text{AC}(A2)} \wedge \omega^{A1}_{\text{KA}(k_{A2})}) \tag{1.9}$$

In an environment of electronic message exchange, an agent's recommendation can only be trusted to the degree that both the AC and the KA can be trusted. The conjunctive recommendation term thus represents what in a normal interpersonal environment would be trust in a person or an organisation. The formal expression for trust-based authenticity of certified keys can then be defined.

Definition 8 Simple Key Authentication
$A1$, $A2$ and $A3$ are three agents, k_{A1}, k_{A2} and k_{A3} their respective public keys. Let $\omega^{A1}_{\text{KA}(k_{A2})}$ and $\omega^{A1}_{\text{AC}(A2)}$ be $A1$'s opinion about the authenticity of k_{A2}, and about $A2$'s co-operation trustworthiness respectively. Let $\omega^{A2}_{\text{KA}(k_{A3})}$ be $A2$'s opinion about the authenticity of k_{A3}. Then $A1$'s opinion about the authenticity of k_{A3} is defined by:

$$\omega^{A1\,A2}_{\text{KA}(k_{A3})} \quad = \omega^{A1}_{A2} \otimes \omega^{A2}_{\text{KA}(k_{A3})}$$

□

In case the certification path goes through intermediate certifiers opinions about trust in the recommending agent, denoted by ω_{AC}, must also be recommended and embedded in the certificate together with the certified key.

Definition 9 Chained Key Authentication
Let the agents $A1$, ..., An–1, An have chained trust and certification relationships in the same manner as for simple authentication. $A1$'s opinion about

the authenticity of k_{An} can then be expressed as:

$$\omega^{A1...An-1}_{\mathrm{KA}(k_{An})} = \omega^{A1}_{A2} \otimes \cdots \otimes \omega^{An-2}_{An-1} \otimes \omega^{An-1}_{\mathrm{KA}(k_{An})}$$

□

In order to assess trust in remote agents regarding correct execution of transactions a simple addition is needed, as described next.

5.2 PROPAGATING TRANSACTION TRUST

Although an authenticated key can be used to establish secure communication, the correct execution of a transaction with the owner of that key will also depend on his or her behaviour during the execution of the transaction. It is thus necessary to combine both AC-trust with KA-trust, or in plain words, the trust in the agent must be combined with the trust in the authenticity of her key. For this purpose the conjunctive recommendation term defined in (1.9) above can again be used because it contains both AC-trust and KA-trust.

Definition 10 Transaction Trust
Let the agents $A1,\ ...,\ An-1,\ An$ have chained trust and certification relationships in the same manner as in Def.9. $A1$'s trust in An regarding correct execution of transactions can then be expressed as:

$$\omega^{A1...An-1}_{An} = \omega^{A1}_{A2} \otimes \cdots \otimes \omega^{An-2}_{An-1} \otimes \omega^{An-1}_{An}$$

□

The framework defined above can now be used to asses trust in transactions with remote users. The uncertainty component is essential for the recommendation and consensus operators to be meaningful. Without the uncertainty component recommendation would be reduced to simple multiplication of probabilities which in our view is counterintuitive, and we claim that a meaningful consensus operator would be impossible to define. For a discussion of these issues see [Jøs98].

Before we apply this framework to computations of expected utility of particular transactions we will in the next section explain why recommendation of trust should be based on first-hand evidence only.

5.3 FIRST-HAND AND SECOND-HAND EVIDENCE

Whenever an agent sends certificates to other agents, opinions about key authenticity and co-operation trustworthiness must always be included. However, opinions based on recommendations from other agents, i.e. second-hand evidence, should in principle never be passed to other agents. This is because the recipient may receive recommendations from the same agents, causing opinion dependency when using the consensus operator. Only opinions based on first-hand evidence and experience should thus be recommended to other agents.

The problem can occur for example in the situation illustrated in Fig.1.4 where agents B and C have a second-hand opinion about agent E and his public key based on a recommendation from D.

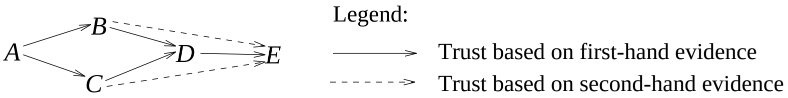
A
B
C
D
E
Legend:
Trust based on first-hand evidence
Trust based on second-hand evidence

Figure 1.4 Trust relationships based on first-hand and second-hand evidence

If B and C recommend their opinions about E to A as if they were based on first-hand evidence, i.e. without telling that they were based on recommendations from D, A would compute the following trust in E:

Incorrect:

$$\omega_E^{A(BD,CD)} = (\omega_B^A \otimes \omega_D^B \otimes \omega_E^D) \oplus (\omega_C^A \otimes \omega_D^C \otimes \omega_E^D) \tag{1.10}$$

The fact that the term ω_E^D appears twice in the expression and thereby violates the independence requirement would in fact be hidden for A, causing her to compute an incorrect key authenticity.

Instead, B and C should only recommend D to A, and D should recommend E to A. Alternatively B and C can pass the recommendations they received from D unmodified to A, because it does not matter who sent it as long as it is certified by D. With this information, A is able to compute the correct authenticity:

Correct:

$$\omega_E^{A(B,C)D} = ((\omega_B^A \otimes \omega_D^B) \oplus (\omega_C^A \otimes \omega_D^C)) \otimes \omega_E^D \tag{1.11}$$

To recapitulate, the rule for passing recommendations between agents is that recommendations must always be based on first-hand evidence.

6 EXAMPLE: DECISION MAKING WITH RECOMMENDED TRUST

Reliable authentication of public keys must always be based on an unbroken chain of certificates and recommendations. However, a path may be difficult to find even if theoretically it exists. Introducing hierarchies of certification authorities (CA) can be used to overcome these problems without excluding private trust and certification relationships, and each user should be allowed to choose which CA or trusted friend he or she wants to use.

Fig.1.5 shows a network of users (G, H, I, J, K, L, M, N) and certification authorities (A, B, C, D, E, F). In this example we require that every CA must at least be related to one CA on a superior plane, except for those already on the top plane, and that CAs on the top plane must all be related.

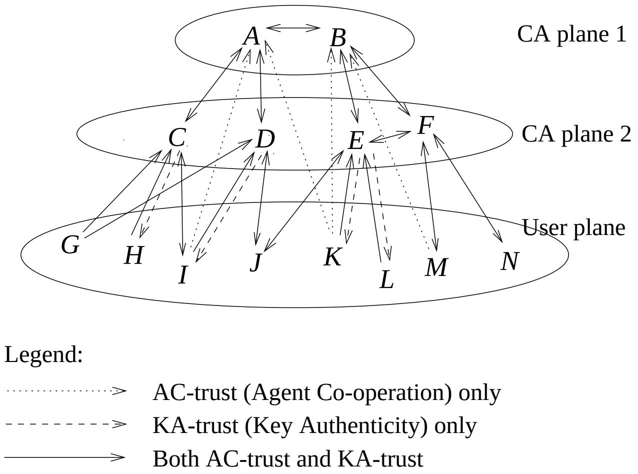
A
B
CA plane 1
C
D
E
F
CA plane 2
User plane
G
H
I
J
K
L
M
N
Legend:
AC-trust (Agent Co-operation) only
KA-trust (Key Authenticity) only
Both AC-trust and KA-trust

Figure 1.5 Trust based on first-hand evidence

The plain arrows indicate trust both for the purpose of agent co-operation and key authenticity (AC-trust and KA-trust). The dashed arrows indicate trust in key authenticity (KA-trust). The dotted arrows indicate trust in agent co-operation (AC-trust). Note that certification is directed in the opposite direction of the trust arrows.

Plain one-way arrows from user to CA thus indicate that a user trusts a CA to certify and recommend, but not the opposite. Dashed arrows from CA to user indicate that a CA has an opinion about the authenticity of a user's public key. Plain arrows from CA to user indicates that a CA has an opinion about the authenticity of a user's public key and about the honesty of the user. CAs that are connected with plain two-way arrows trust each other mutually. The dotted arrows from user to CA indicate that a user can have an opinion about a CA without the CA knowing anything about the user. Agents or CAs that are not connected with either plain, dashed or dotted arrows indicates that they are totally ignorant about each other, i.e. that they have the opinion $(0.0,\ 0.0,\ 1.0,\ 0.5)$ about each other regarding AC and KA trust. It should be noted that the arrows in Fig.1.5 perfectly well can represent distrust, so that users and CAs can use the same model to blacklist other users and CAs.

Assume that user J in Fig.1.5 is considering a transaction q with user M. For this purpose there is an obvious but rather long path via $DABF$ as well as a relatively short path via EF. Although other paths exist, only these two are considered here. Using the short notation, J's trust in M can be expressed as:

ω_{M}^{JDABF}	Trust via $DABF$ only
ω_{M}^{JEF}	Trust via EF only
$\omega_{M}^{J(DAB,E)F}$	Trust via both paths combined

It is assumed that J has previously stored the public keys of D and E in her private database and has opinions about them and their public keys according to Tab.1.1.

Trust in agent co-operation	Trust in key authenticity
$\omega^{J}_{\text{AC}(D)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{J}_{\text{KA}(k_D)}$ = (0.99, 0.00, 0.01, 0.50)
$\omega^{J}_{\text{AC}(E)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{J}_{\text{KA}(k_E)}$ = (0.99, 0.00, 0.01, 0.50)

Table 1.1 Table of AC-trust and KA-trust previously stored by J

Let J receive the public keys of agents A, B, F, and M electronically. Embedded in the certificates are also the certifying agents' opinions about the key authenticity and co-operation trustworthiness according to Tab.1.2.

Trust in agent co-operation	Trust in key authenticity
$\omega^{D}_{\text{AC}(A)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{D}_{\text{KA}(k_A)}$ = (0.99, 0.00, 0.01, 0.50)
$\omega^{A}_{\text{AC}(B)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{A}_{\text{KA}(k_B)}$ = (0.99, 0.00, 0.01, 0.50)
$\omega^{B}_{\text{AC}(F)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{B}_{\text{KA}(k_F)}$ = (0.99, 0.00, 0.01, 0.50)
$\omega^{F}_{\text{AC}(M)}$ = (0.99, 0.00, 0.01, 0.50)	$\omega^{F}_{\text{KA}(k_M)}$ = (0.99, 0.00, 0.01, 0.50)
$\omega^{E}_{\text{AC}(F)}$ = (0.00, 0.90, 0.10, 0.50)	$\omega^{E}_{\text{KA}(k_F)}$ = (0.99, 0.00, 0.01, 0.50)

Table 1.2 Table of AC-trust and KA-trust received by J

It can be seen that all opinions are quite positive and certain except E's opinion about F expressed by $\omega^{E}_{\text{AC}(F)}$ which is clearly negative, for example because E has received evidence that F has been involved in fraud. The trust values according to the three different path alternatives can now be computed.

$$\begin{aligned}
\omega_M^{JDABF} &= (0.904,\ 0.000,\ 0.096,\ 0.500)\\
\omega_M^{JEF} &= (0.941,\ 0.000,\ 0.059,\ 0.500)\\
\omega_M^{J(DAB,E)F} &= (0.573,\ 0.000,\ 0.427,\ 0.500)
\end{aligned}$$

This gives expectation values $\mathrm{E}(\omega_M^{JDABF}) = 0.952$ for the path via $DABF$ only, $\mathrm{E}(\omega_M^{JEF}) = 0.5$ for the path via EF only, and $\mathrm{E}(\omega_M^{J(DAB,E)F}) = 0.786$ for both paths combined.

J's opinion about M as a transaction partner is clearly positive for the path $JDABFM$ due to the recommended opinions all being positive. For the path $JEFM$ the resulting opinion is totally uncertain because E has nothing positive to recommend about F so recommendations from F are worthless and thus result in uncertainty. The combined opinion resulting from the consensus between the two paths is clearly influenced by E's negative opinion about F's recommendations.

These results can now be used in the utility model described in Sec.4. Let J's utilities regarding success and failure of q be $\mathrm{U}_+^J(q) = \$100$ and $\mathrm{U}_-^J(q) = -\$200$ respectively. In case J only is aware of the path via $DABF$ she would compute the rather high expected utility $\mathrm{U}_M^{JDABF}(q) = \$85.62$. If however J includes the recommendation from E, the expected utility would be reduced to $\mathrm{U}_M^{J(DAB,E)F}(q) = \$35.81$, indicating a relatively small utility and a high risk so that J might decide to cancel the transaction. By only taking the advice via EF and not the recommendation received through $DABF$ the expected utility would drop to $\mathrm{U}_M^{JEF}(q) = -\$50.38$.

7 CONCLUSION

We have shown how trust in remote agents can be determined by embedding trust recommendations inside public key certificates. By assuming that the utility of a transaction with an agent depends on whether he co-operates or not, and that the likelihood of this happening can be assessed as trust, we have described how decisions about electronic transactions with remote and unknown agents can be based on trust.

The scheme's main problem seems to be that of determine trust consistently, and we have simply mentioned that this can be done either statistically or intuitively by assessing probability expectation measures. It should be noted that the same problem exists for determining measures of probability in for example safety and risk analysis where human factors must be included and that guidelines and methods for probability assessment exist for such purposes. It should thus be feasible to establish similar practical guidelines for assessing trust.

References

[Bal86] J.F. Baldwin. Support logic programming. In A.I Jones et al., editors, *Fuzzy Sets: Theory and Applications*. Reidel, 1986.

[Coh86] M.S. Cohen. An expert system framework for non-monotonic reasoning about probabilistic assumptions. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 1*. North-Holland, 1986.

[Ell61] Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Ecomonics*, 75:643–669, 1961.

[Jøs96] A. Jøsang. The right type of trust for distributed systems. In C. Meadows, editor, *Proc. of the 1996 New Security Paradigms Workshop*. ACM, 1996.

[Jøs98] A. Jøsang. A Subjective Metric of Authentication. In J. Quisquater et al., editors, *Proceedings of ESORICS'98*, Louvain-la-Neuve, Belgium, 1998. Springer.

[Jøs99] A. Jøsang. An Algebra for Assessing Trust in Certification Chains. In J. Kochmar, editor, *Proceedings of the Network and Distributed Systems Security Symposium (NDSS'99)*. The Internet Society, 1999.

[RS97] Michael K. Reiter and Stuart G. Stubblebine. Toward acceptable metrics of authentication. In *Proceedings of the 1997 IEEE Symposium on Research in Security and Privacy*, Oakland, CA, 1997.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[SK94] Ph. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.