# An Algebra for Assessing Trust in Certification Chains

Audun Jøsang

Telenor R&D, N-7005 Trondheim, Norway, email:audun.josang@fou.telenor.no

## Abstract

*Open networks allow users to communicate without any prior arrangements such as contractual agreement or organisation membership. However, the very nature of open networks makes authenticity difficult to verify. We show that authentication can not be based on public key certificates alone, but also needs to include the binding between the key used for certification and it's owner, as well as the trust relationships between users. We develop a simple algebra around these elements and describe how it can be used to compute measures of authenticity.*

## 1   Introduction

For the distribution of public keys in open networks it is not conceivable to have a single global authority that is trusted for key generation and distribution because there will always be different administrative domains which typically will have conflicting economical and political interests. In this situation, each agent has to decide for herself which other agents she wants to trust for key distribution, and based on this determine the legitimacy of received certificates and the authenticity of keys. In this paper we propose a simple algebra for trust than can be used to determine authenticity of received keys. The algebra builds on the authenticity metric described in [5]. Previously proposed metrics and algebras of authentication have been discussed in [8, 5].

Technically seen, humans do not sign cryptographic certificates, keys do. However, it is usually assumed that human agents are using cryptographic keys as a tool to make certificates so that practically speaking humans do sign certificates. For this assumption to be correct it is essential to explicitly express trust in the binding between the key used for certification and it's owner, because failing to do so would deprive any authentication scheme of it's relationship to humans, and would turn the scheme into authentication for and

by keys. The key-to-owner binding can not be objectively assessed, and necessarily becomes a subjective measure, meaning that two individuals can have different opinions about any particular binding.

To have established the binding between a key and it's owner is not enough for accepting certificates produced by it if for example the key owner deliberately certifies flawed keys. Another essential element of the algebra is therefore to consider the trustworthiness of the certifying agents themselves for the purpose of recommending keys by certification. As for the binding, the recommendation trustworthiness also becomes a subjective measure, meaning that an agent who is trusted by me does not have to be trusted by you.

In [3] we argued that trust simply is a human belief, involving a subject (the trusting party) and the object (the trusted party). Trust in the key-to-owner binding can for example be expressed as believing that: "the key is authentic", whereas trust in the certifier is to believe that "he will only certify keys that he considers authentic".

In can here be added that the security of a system never can be objectively and universally assessed. It is always done by some individuals who may be qualified for that purpose, and the rest of us simply have to believe them. In that sense, trust in a system is a subjective measure of that system's security, and trust in a key is a subjective measure of its authenticity. We claim that there can be no other measure for security and authenticity than subjective trust.

## 2   The Trust Model

The trust model is based on a general model for expressing beliefs, or more precisely for expressing relatively uncertain beliefs about the truth of statements. The statements themselves must be crisp, i.e. they must be assumed to be either true or false, and not something in between. This way of modelling uncertainty is almost the exact opposite to fuzzy set theory where a fuzzy statement such as for example "*tall person*" defines the fuzzy set of tall persons, and a crisp

measure such as for example the height of a person measured in foot or cm combined with a membership function determines a person's degree of membership in the fuzzy set. Although *"trust"* is a fuzzy statement we do not see how fuzzy set theory could be used to model trust because there can be no crisp and reliable measure associated with trust.

In our model we focus on crisp statements that describe particular types of trust. A statement such as: *"the key is authentic"* can be assumed to be either true or false, and not something in between, and is therefore a crisp binary statement. The same can be said about the statement *"the agent will cooperate during our next interaction"*, and we will interpret belief in such statements as trust. However, we will not attempt to use crisp measures to assert the validity of these statements. Because of our imperfect knowledge about reality it is in fact impossible to know with certainty whether such statements are 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 which fills the void in the absence of both belief and disbelief. We express this mathematically as:

$$b + d + u = 1, \quad \{b, d, u\} \in [0, 1]^3 \qquad (1)$$

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

**Definition 1 Opinion**
*Let $\omega = \{b, d, u\}$ be a triplet satisfying (1) where the first, second and third component correspond to belief, disbelief and uncertainty respectively. Then $\omega$ is called an opinion.* □

Eq.(1) defines the triangle of Fig.1, and an opinion can be uniquely described as a point $\{b, d, u\}$ in the triangle. As an example, the opinion $\omega = \{0.8, 0.1, 0.1\}$ is represented as a point in the triangle.

The horizontal bottom line between belief and disbelief in Fig.1 represents situations without uncertainty and is equivalent to a traditional probability model. 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 [2].

"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
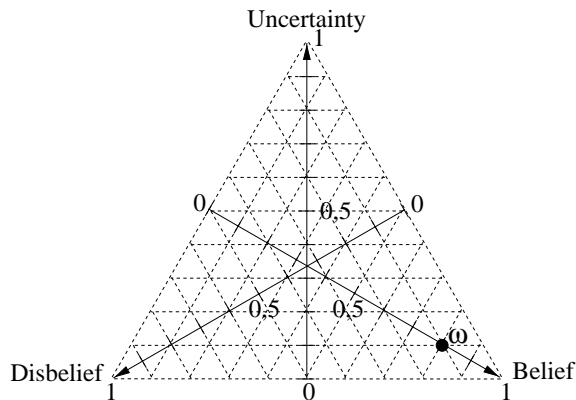


Figure 1: Opinion Triangle

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 chances of winning or loosing a bet on Red$_{II}$ are equal. For Urn I however, it is not obvious. If however one was forced to make a bet on Red$_I$, most people would agree that the chances also 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. In real situations, a probability estimate can never be absolutely 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$_I$ and Red$_{II}$ become $\omega_I = \{0, 0, 1\}$ and $\omega_{II} = \{0.5, 0.5, 0.0\}$ respectively.

Opinions as defined in Def.1 are in fact 2-dimensional measures, consisting of a probability dimension and an uncertainty dimension. By hiding the uncertainty dimension, opinions can be projected onto a 1-dimensional probability space to produce a probability expectance value given by

$$E(\{b, d, u\}) = \frac{b + u}{b + d + 2u}. \qquad (2)$$

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

# 3 Subjective Logic

The algebra for determining trust in certification chains will be based on a framework for artificial reasoning called *Subjective Logic* which has already been described in [4, 7, 6, 5]. Subjective Logic defines various logical operators for combining opinions. Since an opinion can be interpreted as an uncertain probability measure Subjective Logic can be called a calculus for uncertain probabilities.

Subjective Logic contains the equivalent of the traditional logical operators such as *conjunction* (AND), *disjunction* (OR) and *negation* (NOT), as well as some non-traditional operators such as *recommendation* and *consensus*. In the certification algebra described in the next section only the operators conjunction, recommendation and consensus are needed. For simplicity only these three operators will be defined here.

The symbol $\omega$ will be used to denote trust. According to the subject-object duality of trust, we will in addition use superscripts to indicate the subject and subscripts to indicate the believed statement, so that

$$\omega_p^A = \{b_p^A, d_p^A, u_p^A\}$$

represents agent $A$'s belief about $p$, where for example $p$: "*the key is authentic*", meaning that "*A believes that the key is authentic*" to the degree expressed by the belief, disbelief and uncertainty components $b_p^A$, $d_p^A$, and $u_p^A$ respectively. Such opinions are the input and output parameter for the operators defined below.

## Definition 2 Conjunction

*Let $\omega_p^A = \{b_p^A, d_p^A, u_p^A\}$ and $\omega_q^A = \{b_q^A, d_q^A, u_q^A\}$ be agent A's opinions about two distinct binary statements p and q. Then the conjunction of $\omega_p^A$ and $\omega_q^A$, representing A's opinion about both p and q being true is defined by*

$$\omega_{p \wedge q}^A \quad = \omega_p^A \wedge \omega_q^A$$
$$= \{b_{p \wedge q}^A, \ d_{p \wedge q}^A, \ u_{p \wedge q}^A\}$$

*where*

$$\begin{cases} b_{p \wedge q}^A = b_p^A b_q^A, \\ d_{p \wedge q}^A = d_p^A + d_q^A - d_p^A d_q^A, \\ u_{p \wedge q}^A = b_p^A u_q^A + u_p^A b_q^A + u_p^A u_q^A. \end{cases}$$

$\square$

Conjunction of opinions is commutative and associative and requires independent arguments so that the conjunction of an opinion with itself is meaningless. When applied to opinions with zero uncertainty, it is the same as serial multiplication of probabilities. When applied to opinions with absolute belief or disbelief (i.e. $b = 1$ or $d = 1$), it produces the truth table of logical binary AND.

## Definition 3 Recommendation

*Let $A$ and $B$ be two agents where $\omega_B^A = \{b_B^A, d_B^A, u_B^A\}$ is A's opinion about B's recommendations, and let $p$ be a binary statement where $\omega_p^B = \{b_p^B, d_p^B, u_p^B\}$ is B's opinion about p expressed in a recommendation to A. Then A's opinion about p as a result of the recommendation from B is defined by:*

$$\omega_p^{AB} \quad = \omega_B^A \otimes \omega_p^B$$
$$= \{b_p^{AB}, \ d_p^{AB}, \ u_p^{AB}\}$$

*where*

$$\begin{cases} b_p^{AB} = b_B^A b_p^B, \\ d_p^{AB} = b_B^A d_p^B, \\ u_p^{AB} = d_B^A + u_B^A + b_B^A u_p^B. \end{cases}$$

$\square$

$B$'s recommendation must be interpreted as what $B$ actually recommends to $A$, and *not* necessarily as $B$'s real opinion. It is obvious that these can be totally different if $B$ for example defects. The recommendation operator can only be justified when it can be assumed that recommendation is transitive, or more precisely that the agents in a recommendation chain do not change their behaviour (i.e. what they recommend) as a function of which entities they interact with. However, as pointed out in [3] and [1] this can not always be assumed, because defection can be motivated for example by antagonism between certain agents.

## Definition 4 Consensus

*Let $\omega_p^A = \{b_p^A, d_p^A, u_p^A\}$ and $\omega_p^B = \{b_p^B, d_p^B, u_p^B\}$ be opinions respectively held by agents A and B about the same binary statement p. Then the consensus opinion held by an imaginary agent $[A, B]$ representing both A and B is defined by:*

$$\omega_p^{A,B} \quad = \omega_p^A \oplus \omega_p^B$$
$$= \{b_p^{A,B}, \ d_p^{A,B}, \ u_p^{A,B}\}$$

*where*

$$\begin{cases} b_p^{A,B} = (b_p^A u_p^B + b_p^B u_p^A)/(u_p^A + u_p^B - u_p^A u_p^B), \\ d_p^{A,B} = (d_p^A u_p^B + d_p^B u_p^A)/(u_p^A + u_p^B - u_p^A u_p^B), \\ u_p^{A,B} = (u_p^A u_p^B)/(u_p^A + u_p^B - u_p^A u_p^B). \end{cases}$$

$\square$

Consensus is commutative and associative, and requires independent opinion arguments so that consensus of an opinion with itself is meaningless. The effect of the consensus operator is to reduce the uncertainty. Opinions containing zero uncertainty can not be combined, but in practice consensus will normally be mixed with the recommendation operator, so that an

agent receiving absolutely certain but conflicting recommendations will introduce uncertainty by taking her opinions about the recommenders into account before making the consensus. However, two agents that hold conflicting opinions will only be able reach a common consensus if their opinions contain uncertainty.

## 3.1 The Problem of Dependence

It is possible that several recommendation chains produce opinions about the same statement. Under the condition of opinion independence, these opinions can be combined with the consensus rule to produce a single opinion about the target statement. An example of mixed consensus and recommendation is illustrated in Fig.2.
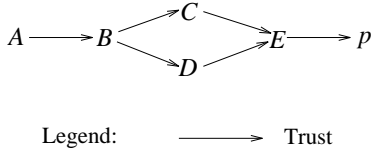


Figure 2: Mixing consensus and recommendation

The recommendation rule is not distributive relative to the consensus rule. Let $\omega_B^A$, $\omega_C^B$, $\omega_D^B$, $\omega_E^C$, $\omega_E^D$ and $\omega_p^E$ represent the opinion relationships in Fig.2. We then have

$$
\begin{array}{c}
\omega_B^A \otimes ((\omega_C^B \otimes \omega_E^C) \oplus (\omega_D^B \otimes \omega_E^D)) \otimes \omega_p^E \\
\neq \\
(\omega_B^A \otimes \omega_C^B \otimes \omega_E^C \otimes \omega_p^E) \oplus (\omega_B^A \otimes \omega_D^B \otimes \omega_E^D \otimes \omega_p^E)
\end{array}
\tag{3}
$$

which according to the short notation in Defs.3 and 4 can be written as

$$
\omega_p^{A(BC,BD)E} \neq \omega_p^{ABCE,ABDE} \quad . \tag{4}
$$

The not-equal sign may seem surprising, but the right sides of (3) and (4) violate the requirement of independent opinions because both $\omega_B^A$ and $\omega_p^E$ appear twice. Only the left sides of (3) and (4) represent the graph of Fig.2 correctly.

Explained differently, there are (at least) two ways of analysing this graph; According to the first method the two trust paths from $A$ to $p$ (the recommendation of course goes in the opposite direction) are analysed separately and finally combined by consensus. This method corresponds to the right sides of (3) and (4). According to the second method the sub-graph containing the nodes $B$, $C$, $D$ and $E$ is reduced to a single node by mixing recommendation and consensus. This sub-graph is analysed separately and the result is used as a sub-expression in the final analysis. This method

corresponds to the left sides of (3) and (4) and is the only correct way to analyse the graph because it avoids opinion dependence.

There will always be cases which can not be analysed directly. Fig.3 illustrates a situation where agent $A$ needs to determine her opinion about statement $p$, of which she only has second-hand evidence trough a network of agents.
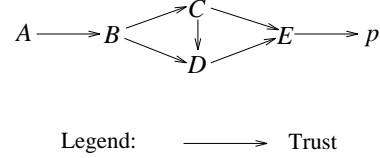


Figure 3: Network of trust that can not be completely analysed

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 Authentication and Certification in Open Networks

## 4.1 The Certification 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, and can be used for establishing secure communication and for certification of other keys.

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 possesses. For example if $A1$ possesses $A2$'s public key $k_{A2}$ and $A2$ possesses $A3$'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 then establish secure communication with $A3$.

However, certificates are not enough. In order to get a binding between keys and key owners, the recipient of the certificate must have an opinion $\omega_{KA(k_{A2})}^{A1}$ 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 $\omega_{\text{RT}(A2)}^{A1}$ about the certifier's recommendation trustworthiness (RT), that is how much she trusts him to actually recommend and certify other keys. Finally, the certifier must actually recommend to the recipient his own opinion $\omega_{\text{KA}(k_{A3})}^{A2}$ about the authenticity of the certified key. This opinion must be embedded in the certificate sent 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* $(\omega_{\text{RT}(A2)}^{A1} \wedge \omega_{\text{KA}(k_{A2})}^{A1})$ which we will give the following short notation:

$$\omega_{A2}^{A1} = (\omega_{\text{RT}(A2)}^{A1} \wedge \omega_{\text{KA}(k_{A2})}^{A1}) \tag{5}$$

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

### Definition 5 Simple Authentication
*$A1$, $A2$ and $A3$ are three agents, $k_{A1}$, $k_{A2}$ and $k_{A3}$ their respective public keys. Let $\omega_{KA(k_{A2})}^{A1}$ and $\omega_{RT(A2)}^{A1}$ be $A1$'s opinions about the authenticity of $k_{A2}$, and about $A2$'s recommendation trustworthiness respectively. Let $\omega_{KA(k_{A3})}^{A2}$ be $A2$'s opinion about the authenticity of $k_{A3}$. Then $A1$'s opinion about the authenticity of $k_{A3}$ is defined by:*

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

$$= (\omega_{RT(A2)}^{A1} \wedge \omega_{\text{KA}(k_{A2})}^{A1}) \otimes \omega_{\text{KA}(k_{A3})}^{A2}$$

□

In case the certification path goes through intermediate certifiers opinions about recommendation trustworthiness $\omega_{\text{RT}}$ must also be recommended along the path and embedded in the certificate together with the certified key. The recommendation trustworthiness RT not only applies to immediate certification of keys, but also to the recommendation of other agents for further recommendations. In [7] these two types of trustworthiness were treated separately and called CT (certification trustworthiness) and RT respectively. However, since they necessarily are dependent, separate treatment would lead to computational inconsistencies, and we therefore use only RT to denote both types of trustworthiness.

### Definition 6 Chained Authentication
*Let the agents $A1$, ..., $An-1$, $An$, have chained trust and certification relationships. $A1$'s opinion about the authenticity of $k_{An}$ can then be expressed by simply inserting the intermediate terms into the expression:*

$$\omega_{\text{KA}(k_{An})}^{A1...An-1} = \omega_{A2}^{A1} \otimes \cdots \otimes \omega_{An-1}^{An-2} \otimes \omega_{\text{KA}(k_{An})}^{An-1}$$

$$= (\omega_{\text{RT}(A2)}^{A1} \wedge \omega_{\text{KA}(k_{A2})}^{A1}) \otimes \cdots \otimes (\omega_{\text{RT}(An-1)}^{An-2} \wedge \omega_{\text{KA}(k_{An-1})}^{An-2}) \otimes \omega_{\text{KA}(k_{An})}^{An-1}$$

□

The framework defined above can now be used to compute the relative authenticity of keys received through an open computer network. If desirable, the algebra can be reduced to a one-dimensional probabilistic calculus by using opinions without uncertainty, i.e. $u = 0$ (but in this case the consensus operator must be modified). The algebra can also be reduced to binary logic by only allowing binary belief components, i.e. $b = 0$ or $b = 1$ (also requiring a modified consensus operator). The full two-dimensional algebra will be used in the examples below.

### 4.1.1 Example: Receiving Certificates.

Fig.4 illustrates a possible structure of certified public keys as stored in agent $A$'s private database. The structure above the dotted line represents the situation before any keys are received electronically, whereas the structure underneath is added after receiving keys electronically. The dotted line also indicates the separation between the keys for which the trust is based on first-hand and second-hand evidence, as seen by $A$.



Legend:


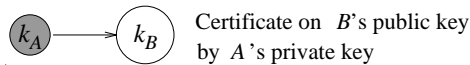Certificate on $B$'s public key by $A$'s private key
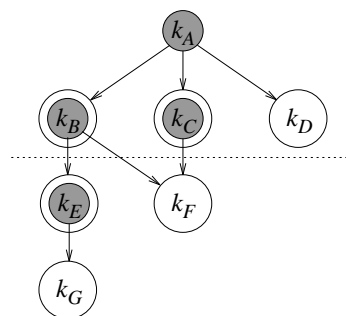
Figure 4: Structure of keys and certificates in agent $A$'s database

This structure makes no assumption about any binding between key owners and certificates. In addition agent $A$ must therefore keep a list of her opinions $\omega_{\text{KA}}^A$ about key authenticity, that is, her opinions about binding between keys and key owners. Tab.1 below gives an example of possible opinion values. Although it is not shown, a one-to-many binding between an agent and her different keys can perfectly well be accommodated within this structure.

| Key $k_X$ | Key owner $X$ | Key Authenticity $\omega_{\text{KA}(k_X)}^A$ |
|---|---|---|
| $k_A$ | $A$ | $\{1.00, 0.00, 0.00\}$ |
| $k_B$ | $B$ | $\{0.98, 0.00, 0.02\}$ |
| $k_C$ | $C$ | $\{0.97, 0.00, 0.03\}$ |
| $k_D$ | $D$ | $\{0.98, 0.00, 0.02\}$ |

Table 1: $A$'s first-hand opinions about the binding between keys and their owners

Agent $A$ must also keep a list of her opinions $\omega_{\text{RT}}^A$ about recommendation trustworthiness, that is how much she trusts the key owners to actually recommend other keys and other agents. Tab.2 below gives an example of possible opinion values.

| Key owner $X$ | Recommendation Trustworthiness $\omega_{\text{RT}(X)}^A$ |
|---|---|
| $A$ | $\{1.00, 0.00, 0.00\}$ |
| $B$ | $\{0.96, 0.02, 0.02\}$ |
| $C$ | $\{0.97, 0.01, 0.02\}$ |
| $D$ | $\{0.90, 0.00, 0.10\}$ |

Table 2: $A$'s first-hand opinions about agent trustworthiness

It is assumed that $A$ knows $B$, $C$ and $D$ personally and therefore has first-hand evidence about their recommendation trustworthiness. It is also assumed that $A$'s opinions about key authenticity is based on having physically exchanged public keys with them.

Let $A$ receive the public keys of agents $E$, $F$ and $G$ electronically. Embedded in the certificates are also the certifying agents' opinions about the key authenticity and recommendation trustworthiness according to Tabs.3 and 4.

The authenticity of for example $k_E$ as seen by $A$ can now be computed by using Def.5:

| Key Authenticity |
|---|
| $\omega_{\text{KA}(k_E)}^B = \{0.98, 0.00, 0.02\}$ |
| $\omega_{\text{KA}(k_F)}^B = \{0.95, 0.01, 0.04\}$ |
| $\omega_{\text{KA}(k_F)}^C = \{0.98, 0.00, 0.02\}$ |
| $\omega_{\text{KA}(k_G)}^E = \{0.90, 0.05, 0.05\}$ |

Table 3: Recommended key authenticity received by $A$

| Recommendation Trustworthiness |
|---|
| $\omega_{\text{RT}(E)}^B = \{0.99, 0.00, 0.01\}$ |
| $\omega_{\text{RT}(F)}^B = \{0.98, 0.01, 0.01\}$ |
| $\omega_{\text{RT}(F)}^C = \{0.90, 0.00, 0.10\}$ |
| $\omega_{\text{RT}(G)}^E = \{0.99, 0.00, 0.01\}$ |

Table 4: Recommended agent trustworthiness received by $A$

$$
\begin{aligned}
\omega_{\text{KA}(k_E)}^{AB} \quad &= (\omega_{\text{RT}(B)}^A \wedge \omega_{\text{KA}(k_B)}^A) \otimes \omega_{\text{KA}(k_E)}^B \\
&= \{0.922, 0.000, 0.078\}
\end{aligned}
\tag{6}
$$

When there are several certification paths to the same key, the authenticity can be computed as the consensus between the authenticities obtained for each path. The authenticity of $k_F$ as seen by $A$ can then be computed as:

$$
\begin{aligned}
\omega_{\text{KA}(k_F)}^{AB,AC} \quad &= ((\omega_{\text{RT}(B)}^A \wedge \omega_{\text{KA}(k_B)}^A) \otimes \omega_{\text{KA}(k_F)}^B) \oplus \\
&\quad ((\omega_{\text{RT}(C)}^A \wedge \omega_{\text{KA}(k_C)}^A) \otimes \omega_{\text{KA}(k_F)}^C) \\
&= \{0.951, 0.004, 0.045\}
\end{aligned}
\tag{7}
$$

When certificates pass through a chain of nodes, recommendation of each node must be included in the expression. The authenticity of $k_G$ as seen by $A$ can be computed as:

$$
\begin{aligned}
\omega_{\text{KA}(k_G)}^{ABE} \quad &= (\omega_{\text{RT}(B)}^A \wedge \omega_{\text{KA}(k_B)}^A) \otimes \\
&\quad (\omega_{\text{RT}(E)}^B \wedge \omega_{\text{KA}(k_E)}^B) \otimes \omega_{\text{KA}(k_G)}^E \\
&= \{0.821, 0.046, 0.133\}
\end{aligned}
\tag{8}
$$

The added structure of new certificates is illustrated in lower part of Fig.4. Although $A$ now has opinions about the authenticity of the public keys of $E$, $F$ and $G$, these opinions should never be passed to other agents. This will be explained in Sec.4.2 below.

## 4.2 First-Hand and Second-Hand Evidence

Whenever an agent sends certificates to other agents, opinions about key authenticity and recommendation 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 dependence 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.5 where agents $B$ and $C$ have a second-hand opinion about agent $E$ and his public key based on a recommendation from $D$.
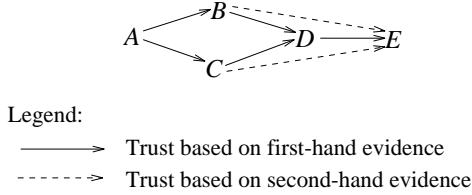


Legend:

$\longrightarrow$ Trust based on first-hand evidence

$----\rightarrow$ Trust based on second-hand evidence

Figure 5: 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 key authenticity for $k_E$:

**Incorrect:**

$$\omega_{\text{KA}(k_E)}^{ABD,ACD} =$$

$$
\begin{aligned}
&((\omega_{\text{RT}(B)}^A \wedge \omega_{\text{KA}(k_B)}^A) \otimes \\
&((\omega_{\text{RT}(D)}^B \wedge \omega_{\text{KA}(k_D)}^B) \otimes \omega_{\text{KA}(k_E)}^D)) \oplus \\
&((\omega_{\text{RT}(C)}^A \wedge \omega_{\text{KA}(k_C)}^A) \otimes \\
&((\omega_{\text{RT}(D)}^C \wedge \omega_{\text{KA}(k_D)}^C) \otimes \omega_{\text{KA}(k_E)}^D)) \; .
\end{aligned}
\tag{9}
$$

The fact that the term $\omega_{\text{KA}(k_E)}^D$ 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:**

$$\omega_{\text{KA}(k_E)}^{(AB,AC)D} =$$

$$
\begin{aligned}
&(((\omega_{\text{RT}(B)}^A \wedge \omega_{\text{KA}(k_B)}^A) \otimes (\omega_{\text{RT}(D)}^B \wedge \omega_{\text{KA}(k_D)}^B)) \oplus \\
&((\omega_{\text{RT}(C)}^A \wedge \omega_{\text{KA}(k_C)}^A) \otimes (\omega_{\text{RT}(D)}^C \wedge \omega_{\text{KA}(k_D)}^C))) \otimes \\
&\omega_{\text{KA}(k_E)}^D \; .
\end{aligned}
\tag{10}
$$

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

## 4.3 Trust-based Navigation on Open Networks

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 being in conflict with the philosophy of open networks, and each user should be allowed to choose which CA he or she wants to use.

According to the scenario described in Sec.4.1 first-hand evidence is obtained by having had direct experience with an agent and physically exchanging keys. This means that the relationship between CAs and users needs to be rather intimate, for example similar to a bank's relationship with it's customers.

By requiring recommendations to be based on first-hand evidence only, the problem of certificate revocation is drastically reduced because the recommender will always have full overview of every recipient of a particular certificate, and is thereby able to inform them efficiently in case of revocation. In addition, users never need to worry about trust intransitivity, or in other words that the CA they trust trusts another CA which they would not trust, because a user is always informed about the identity of every intermediate node in a chain and may in fact override the received recommendation trustworthiness value if he happens to have an opinion about that particular CA.

### 4.3.1 Example: Establishing Certification Paths.

Fig.6 shows a network of users $(G, H, I, J, K, L, M, N, O, P)$ 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.
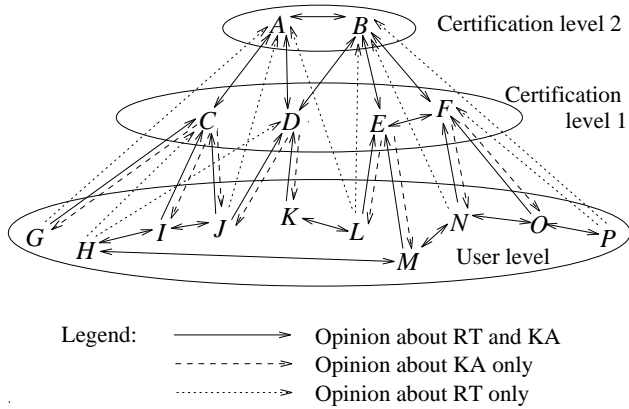
Figure 6: Trust based on first-hand evidence

The plain arrows indicate trust for the purpose of recommendation and for key authenticity. The plain one-way arrows between users and CAs indicate that a user trusts a CA to certify and recommend, but not the opposite. The dashed arrows between CAs and users indicate that a CA has an opinion about the authenticity of a user's public key. CAs that are connected with plain two-way arrows trust each other mutually, and so do the users. This means that CAs can certify public keys of users and other CAs, but can only recommend CAs for further recommendation, whereas users can certify public keys and recommend both CAs and other users to each other. The dotted arrows indicate trust for the purpose of recommendation, meaning that a user can have an opinion about a CA without the CA knowing anything about the user. Two agents 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 $\{0, 0, 1\}$ about each other regarding RT and KA. It should be noted that the arrows in Fig.6 perfectly well can represent distrust, so that users and CAs can use the same model to blacklist other users and CAs.

We will use the short notation to give a few examples of how key authenticity can be expressed.

- $\omega_{\mathrm{KA}(k_J)}^{G(C,CAD)}$ is $G$'s trust in $k_J$ based on recommendations via $C$ and via $CAD$. However, $G$ has a first-hand opinion $\omega_{\mathrm{RT}(A)}^{G}$ about $A$'s recommendation trustworthiness, and may use it to replace the one received from $C$, or may ignore the path via $A$ altogether if he distrusts $A$.

- $\omega_{\mathrm{KA}(k_P)}^{H(MN,MEF)O}$ is $H$'s trust in $k_P$ based on recommendations via $MNO$ and via $MEFO$. If $H$ knew that $P$ also could be reached via $I$ and via $J$, he

could have obtained recommendations for $C$ and $D$ from them, and further for $A$ and $B$.

- $\omega_{\mathrm{KA}(k_J)}^{MEBD}$ is $M$'s trust in $k_J$ based on recommendations via $EBD$. If $M$ could find out that there is a potential path to $J$ via $H$, he could have obtained $\omega_{\mathrm{KA}(k_J)}^{MEBD,MHI}$.

- $\omega_{\mathrm{KA}(k_M)}^{(KDB,KL)E}$ is $K$'s trust in $k_M$.

- $\omega_{\mathrm{KA}(k_K)}^{MEBD}$ is $M$'s trust in $k_K$. A recommendation from $L$ is not possible to obtain, because $E$ does not trust $L$ for that purpose.

## 4.4 Comparison with PGP

This final section will be used to compare the model described here with PGP [9] which is a well known method for handling authentication in the Internet. The trust model of PGP is perfectly compatible with ours, whereas the algebra differs, and we will show that an inherent weakness in the way PGP computes trust can make users get a false impression of key authenticity.

### 4.4.1 Compatible Trust Models.

The PGP electronic public key ring is used to store the public keys of other users, as well as certificates attached to each particular public key. Trust values are assigned to three aspects of each key. These aspects are the *Owner Trust*, i.e. how the owner of the key is trusted to certify and recommend other keys, the *Signature Trust*, or the trust on the owner of each key that was used to certify the key, and finally the *Key Legitimacy* or the actual key authenticity.

Owner Trust and Signature Trust are measured as *undefined, unknown user, usually not trusted, usually trusted, always trusted* and *ultimate* (the owner is me), and the value is always equal for a particular certificate and the owner of the key that signed it. These discrete measures can easily be represented as points in the opinion triangle as suggested in Fig.7.a.

Key Legitimacy is measured as *unknown, not trusted, marginally trusted* and *completely trusted*. These discrete measures can also be represented as points in the opinion triangle as suggested in Fig.7.b.

The Key Legitimacy is calculated on the basis of the signature trust fields as follows: If at least one Signature Trust has value *ultimate*, the Key Legitimacy is set to *complete*. Otherwise, PGP computes a weighted sum of the Signature Trust values. A weight of $1/x$ is given to signatures that are always trusted and $1/y$ to signatures that are usually trusted, where $x$ and $y$

a) "Owner Trust" and "Signature Trust"
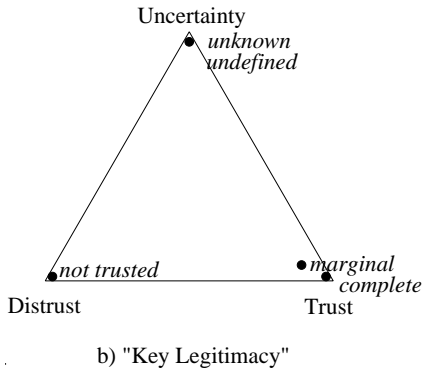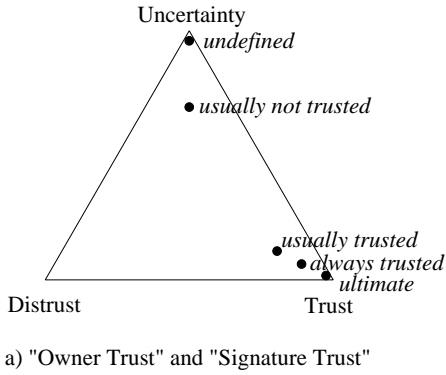


b) "Key Legitimacy"

Figure 7: Discrete trust values of PGP expressed as opinions

are user configurable parameters. When the total of weights reaches 1, the Key Legitimacy is set to *complete*, otherwise it is set to *marginal*.

The difference between our model and PGP, is that in our model the computed Key Authenticity is kept as such, instead of using thresholds and adjusting the trust to a discrete value. In our model, a threshold value can be determined for the use of a key in a particular situation, instead of accepting a key as *completely trusted* or *marginally trusted* once for all. After all, different situations involve different risk, and thereby require different trust.

### 4.4.2 Hidden Dependencies in PGP Trust Values.

As was mentioned in Sec.3.1, dependence between arguments in an expression for trust leads to incorrect computational results. The way in which webs of trust expand with PGP causes recommendations based on second-hand evidence to be transferred between users, and we will show how this can lead to dependence.

Let user $A$ receive the public key of user $G$ certified by the users $B$, $C$, $D$ and $E$ whom she trusts with value *usually trusted*. Suppose that $A$ has specified that 4 *usually trusted* or 2 *always trusted* certificates are required to accept the received public key as *completely trusted*, in which case $A$ will have *complete trust* in $G$'s key. Let us now suppose that the certifiers $B$, $C$, $D$ and $E$ all had received $G$'s key certified by the same user $F$. In that case, their recommendations about $G$ are highly dependent, and the certificates sent to $A$ can not be considered as coming from different sources.

Fig.8 illustrates this situation, where what $A$ sees and what really takes place leads to $A$ getting a false impression of trust in $G$'s key.
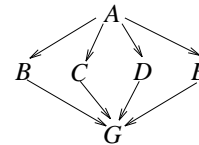


a) The situation that $A$ sees
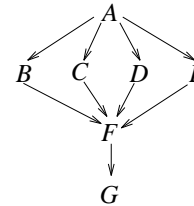


b) The real situation which is hidden for $A$

Figure 8: Apparent and real trust relationships

What $A$ in fact computes is ultimately based on only one recommendation, namely the one from $F$, so that $A$'s requirement of at least 4 *usually trusted* or 2 *always trusted* certificates has been violated.

The problem is caused by the way users use certificates to compute their own trust in received keys which they in turn certify and pass to others as if the authenticity of those keys was based on first-hand evidence, whereas in reality, it is based on second-hand evidence. The only way this can be solved is to only certify keys that are trusted with first-hand evidence, or else always pass the original certificates unmodified to other users, so they themselves can determine their trustworthiness.

In comparison, a correct analysis of the graph of Fig.8.b as seen by $A$ using the certification algebra described in Sec.4.1 results in the opinion

$$\omega_{KA(k_G)}^{(AB,AC,AD,AE)F}.$$

This would require that $A$ has received certificates di-

rectly from $B$, $C$, $D$ and $E$ containing the public key of $F$ with corresponding recommended key authenticity and agent trustworthiness, as well as a certificate from $F$ containing the public key of $G$ with a recommended key authenticity.

On the other hand, if $B$, $C$, $D$ and $E$ recommend to $A$ their second-hand opinions about $G$ according to Fig.8.a, then $A$ would compute the opinion

$$\omega_{\text{KA}(k_G)}^{ABF,ACF,ADF,AEF}$$

which as explained in Sec.4.2 is incorrect because $F$'s opinion about $\text{KA}(k_G)$ appears four times in the expression and thereby violates the requirement of independent opinions.

## 5 Conclusion

In traditional authentication schemes, the key-to-owner binding as well as the recommendation trustworthiness are trust aspects that are usually part of the initial assumptions. However, in the real world these aspects can never be absolutely trusted, and assuming absolute trust can then be dangerous. We have introduced an authentication algebra that takes relative trust in the key-to-owner binding and trust in the ability to recommend into consideration. In order to avoid undesirable dependencies, the algebra requires recommendations to be based on first-hand evidence only. This does however not put any restriction on possible certification paths, but simply enforces a particular way of establishing such paths. The algebra provides a practical solution to the problem of authentication in open networks, and is ready to be implemented in systems.

## References

[1] Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized trust management. In *Proceedings of the 1996 IEEE Conference on Security and Privacy*, Oakland, CA, 1996.

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

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

[4] A. Jøsang. Artificial reasoning with subjective logic. In Abhaya Nayak, editor, *Proceedings of the Second Australian Workshop on Commonsense Reasoning*, 1997.

[5] A. Jøsang. A subjective metric of authentication. In J.-J. Quisquater, Y. Deswarte, C. Meadows, and D. Gollmann, editors, *ESORICS'98. Louvain-la-Neuve, Belgium*. Springer-Verlag, 1998.

[6] A. Jøsang and S.J. Knapskog. A metric for trusted systems. In *Proceedings of the 21st National Security Conference*. NSA, 1998.

[7] Audun Jøsang. *Modelling Trust in Information Security*. PhD thesis, Norwegian University of Science and Technology, 1997.

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

[9] P.R. Zimmermann. *The Official PGP User's Guide*. MIT Press, 1995.