A Subjective Metric of Authentication

Audun Jøsang

Telenor R&D, 7005 Trondheim, Norway
audun.josang@fou.telenor.no

Abstract. Determining the authenticity of public keys in large-scale open networks can not be based on certificates alone, but must also include the binding between the key used for certification and it's owner, as well as the trust relationships between individual agents. This paper describes a method for computing authenticity measures based on certificates, on key binding, and on trust relationships. Two essential elements of the method are the *opinion model* which is a radically new way of representing trust, and *subjective logic* which consists of a set of logical operators for combining opinions. We show that our method for computing authenticity measures can be applied to both anarchic and hierarchic authentication networks.

1 Introduction

Public key cryptography seems to be the technical solution for securing global open telecommunication networks. The problem however is to find a reliable way of determining the authenticity of public keys in a large-scale open network. For this purpose, pure cryptography is not enough. It is for example not conceivable to have a single global authority that is trusted for key generation and distribution. 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, and based on this determine the legitimacy of certificates and the authenticity of keys.

Several metrics of authentication have been proposed in the literature, such as the BBK model [BBK94], the PGP model [Zim95], the Maurer model [Mau96], and the Reiter-Stubblebine model [RS97]. These models will be briefly discussed in the next section, but we can already mention that they all represent trust either as a discrete or as a continuous parameter in the range $[0, 1]$.

All the mentioned models have drawbacks which we will try to overcome. In particular, we find neither the discrete, nor the continuous approaches satisfactory. The discrete are insufficient because they only provide a small set of possible trust values. The continuous and probability oriented models fall short because the operators for combining trust values often seem counterintuitive. This indicates that by modelling trust as probability, important aspects of trust as a human cognitive phenomenon is missing. In our view, the missing component is ignorance and uncertainty which can not be reflected by probabilities.

Recently, Reiter and Stubblebine[RS97] have proposed design principles for defining a metric for trust in authentication schemes. In our opinion, our model satisfies all principles. In particular we have solved the problem of the binding between key and key owner, which is only implicitly assumed in many previous approaches, but explicitly expressed in our approach.

The computational model expresses trust as a two-dimensional metric, and uses special logical operators which are both intuitive and which have a sound mathematical basis. The two dimensional metric is called an *opinion* and can also be interpreted as a measure of uncertain probabilities. As such the opinion model and the set of logical operators together represent a logic of uncertain probabilities. The logic which is called *subjective logic*[1] seems particularly suitable for modelling authentication networks. In particular we are able to model recommendation of certificates along paths of agents, and the combination of possibly conflicting recommendations in order to reach a consensus.

2 Previously Proposed Metrics

2.1 The BBK Model

The BBK model[BBK94] represents trust relationships between agents as edges, so that $A \longrightarrow B$ means that agent A trusts agent B for example to determine the authenticity of public keys, and $A - - \rightarrow B$ that agent A trusts agent B to recommend other agents for the purpose of determining the authenticity of public keys, where trust is expressed as a value in the range $[0, 1]$. The BBK model has several weaknesses that have been pointed out in [Jøs97c,RS97]. Its recommendation and consensus operators are found to be easily manipulable and thereby inadequate for determining authenticity.

2.2 The Maurer Model

As in the BBK model, the Maurer model[Mau96] is based on direct and recommended trust represented as edges in a graph, but the interpretation of the edges is different. A direct edge $A \longrightarrow B$ means that the user evaluating the metric "holds a certificate for B's public key(allegedly) issued and signed by entity A". Similarly, a recommendation edge $A - - \rightarrow B$ denotes that the user is in possession of a recommendation (for recommending or authenticating other entities) for B allegedly signed by entity A. Associated with each direct edge and recommendation edge is a value $[0, 1]$, called a *confidence parameter*, that is assigned by the entity that created (the construct represented by) the edge.

The Maurer model depends on an assumed binding between certificate and certifier. However, as pointed out in [RS97] *agents do not sign certificates, keys do.* To repeat from the previous paragraph, the edge $A \longrightarrow B$ exists in the Maurer model if the user evaluating the metric "holds a certificate for B's public key (allegedly) issued and signed by entity A" ([Mau96] Definition 3.1). Maurer

[1] A more detailed description of subject logic can be found in [Jøs97a,Jøs97d].

uses the word "allegedly" because "without verification, there exists no evidence that the certificate was indeed issued by the claimed entity". Put another way, when the entity that allegedly signed the certificate is claimed with the certificate, this claim is at best a hint, and at worst an opportunity to be misled. It is therefore ambiguous how certificates should be represented in the Maurer model.

A similar concern arises in the BBK model described in the previous section. Evaluating a metric requires the user to collect values from other entities for the various direct and recommendation edges. However, before the user can safely assign a value to the edge $A \longrightarrow B$ or $A - - \rightarrow B$, the user must authenticate this value as having come from A. Assuming that this authentication is performed cryptographically, (e.g. via certificates), again the user is asked to determine a key that can be used to authenticate A. As a result, the answer to the original authentication problem is a new and identical authentication problem, which therefore never can be solved. This fundamental flaw in both the Maurer and the BBK model makes them totally unsuitable for their intended usage.

2.3 The PGP Model

PGP (Pretty Good Privacy [Zim95]) contains one of the most popular civilian public key management systems in the world today. It is based on a representation of public keys and their certificates as a graph where the nodes are keys, and an edge $K_1 \longrightarrow K_2$ represent a certificate binding the signing private part of key K_1 to the signed public part of key K_2. The binding is verifiable by the public part of K_1. Associated with each public key in the user's database is a trust value which can be *unknown, untrusted, marginally trusted* or *fully trusted.*

PGP computes the legitimacy of each key as follows: PGP first declares legitimate the node K_0 representing the users own public key and any node K_x such that $K_0 \longrightarrow K_x$ is an edge in the graph. Now, if for some node K_y, there is an edge to K_y from a legitimate fully trusted node, then K_y is considered legitimate also. Alternatively, if for some node K_z, there are a minimum number of edges to K_z from marginally trusted nodes, then K_z is considered legitimate. The number of edges required from fully trusted or marginally trusted nodes can be adjusted, but 1 and 2 are the defaults. In practice, determination of node legitimacy is interwoven with assigning trust values to nodes. That is, a trust value for certifying other keys is assigned to a key only after it has been determined to be legitimate, and thus its owner assumed to be known.

The PGP trust model has several merits and does not have the defect of the BBK and Maurer models that the binding between a key owner and the certificate is assumed as part of the certificate. PGP first determines the binding between the owner and the key in the form of legitimacy, and then lets the user specify the trust in the certificates produced by the key. This makes PGP practical in real application, as its widespread usage also testifies. However, the limitations of the trust model in PGP is that the rules fore determining legitimacy and the trust values themselves are discrete. In Sec.5, we will build on the PGP approach by separating certification and binding, and use a two-dimensional continuous metric which will be described in Sec.3.

2.4 The Reiter-Stubblebine Model

As in PGP, the Reiter-Stubblebine model[RS97] is based on representing keys and their certificates as nodes and edges in a directed graph, so that the edge $K_1 \longrightarrow K_2$ represents that the user is in possession of a certificate that signs K_2, and who's signature can be verified by using the public part of key K_1. Each edge also has a numeric value associated that represents the amount of money for which the owner of K_1 insures the attributes of K_2, i.e. the value for which the owner of K_1 will be liable to the user if the attributes bound to K_2 in the certificate are incorrect, or if the private part of K_2 is used to mislead the user, intentionally or otherwise. In particular, if the private key corresponding to K_2 is compromised and used maliciously, then the owner of K_1 is liable for the stated amount. This form of insurance is called *surety bonding.*

The insurance label of the edge $K_1 \longrightarrow K_2$ must be obtained from the owner of K_1 in some reliable way, and so this value is stored in the certificate that $K_1 \longrightarrow K_2$ represents. This does not force the user to make assumptions about the true owner of K_1 as in the BBK and Maurer models. In fact it is possible that K_1 has been compromised and used to forge the certificate $K_1 \longrightarrow K_2$, including its attributes and insurance value. In this case, whoever certified K_1 would be liable, and this can regress along a path arbitrarily far.

By assuming the existence of an underlying framework for contract enforcement and insurance collection, they introduce an element which could have been used to enforce the security policy of the key authentication scheme in the first place, and thereby making the proposed model obsolete.

What the authors try to achieve with the model is to shift all the burden of damages incurred by a misused key to whoever has certified the key. As a result, certifiers would have to be extremely careful, undoubtedly resulting in a very conservative behaviour. As such, this is not a model for determining trust, but a method for making certifiers behave carefully and correctly, and thereby for creating more trust in the certified keys.

As already mentioned, by requiring contract and insurance payment enforcement, the method seems to be impractical at best, and to make itself obsolete at worst.

3 Representing Trust Mathematically

The evidence space and the opinion space are two equivalent models of representing uncertain probabilities, and this duality will be very useful for determining uncertain probabilities and for giving them intuitive interpretations.

3.1 The Evidence Space

The mathematical analysis leading to the expression for posteriori probability estimates of binary events can be found in many text books on probability theory, e.g. [CB90] p.298, and we will only present the results here.

It can be shown that posteriori probabilities of binary events can be represented by the beta distribution function. The beta-family of distributions is a continuous family of functions indexed by the two parameters α and β. The beta(α, β) distribution can be expressed using the gamma function Γ as:

$$f(\theta \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\ ,\ \ 0 \leq \theta \leq 1,\ \alpha > 0,\ \beta > 0 \qquad (1)$$

with the restriction that $\theta \neq 0$ if $\alpha < 1$, and $\theta \neq 1$ if $\beta < 1$. The mean value of the beta distribution is given by $\mathrm{E}(\theta) = \alpha/(\alpha+\beta)$.

We will in the following only consider the subclass of beta distributions which we will call *probability certainty density functions* or pcdf for short. We will denote by Φ the set of pcdfs.

In our notation, pcdfs will be characterised by the parameters $\{r, s\}$ instead of $\{\alpha, \beta\}$ through the following correspondence:

$$\begin{aligned} &\alpha = r+1,\ r \geq 0 \quad \text{and} \\ &\beta = s+1,\ s \geq 0. \end{aligned} \qquad (2)$$

Let φ be a pcdf over the probability variable θ. In our notation φ will then be characterised by r and s according to:

$$\varphi(\theta \mid r, s) = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)}\theta^{r}(1-\theta)^{s}\ ,\ \ 0 \leq \theta \leq 1,\ r \geq 0,\ s \geq 0 \qquad (3)$$

The mean value of a pcdf is given by $\mathrm{E}(\theta) = (r+1)/(r+s+2)$.

As an example, assume that an entity has produced $r = 8$ positive and $s = 1$ negative events. The pcdf expressed as $\varphi(\theta \mid 8, 1)$ is plotted in Fig.1.

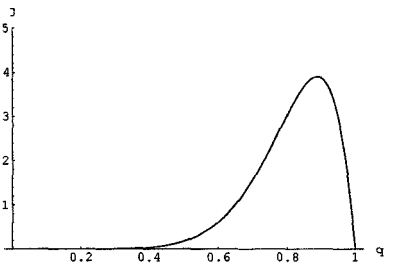

Fig. 1. Posteriori pcdf after 8 positive and 1 negative results

The curve plotted in Fig.1 must not be confused with an ordinary probability density function. A pcdf represents the certainty density regarding the expected probability of a binary event, and not the distribution of probabilities. This is explained in more detail in [Jøs97d].

3.2 The Opinion Space

For the purpose of believing a binary proposition such as for example: *the key is authentic*, we assume that the proposition will either be true or false, and not something in between. However, because of 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 which fills the void in the absence of both belief and disbelief. We express this mathematically as:

$$b + d + u = 1, \quad \{b, d, u\} \in [0, 1]^3 \tag{4}$$

where b, d and u designate belief, disbelief and uncertainty respectively. Eq.(4) defines the triangle of Fig.2, and an opinion can be uniquely described as a point $\{b, d, u\}$ in the triangle.

Definition 1. *Let $\omega = \{b, d, u\}$ be a triplet satisfying (4) where the first, second and third component correspond to belief, disbelief and uncertainty respectively. Then ω is called an opinion. We will denote by Ω the set of opinions.* □

As an example, the opinion $\omega = \{0.8, 0.1, 0.1\}$ which corresponds the the pcdf of Fig.1 is represented as a point in the figure.

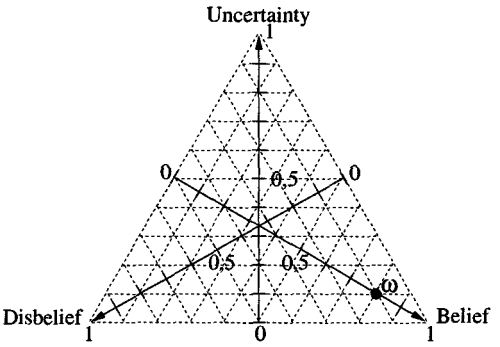

Fig. 2. Opinion Triangle

The horizontal bottom line between belief and disbelief in Fig.2 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 [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 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 $\omega_{\text{I}} = \{0,\ 0,\ 1\}$ and $\omega_{\text{II}} = \{0.5,\ 0.5,\ 0.0\}$ respectively.

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 apply. For example ω_p^A is an opinion held by agent A about the truth of proposition p.

3.3 Equivalence between the Evidence and Opinion Spaces

We have defined Φ to be the class of pcdfs, and Ω to be the class of opinions. Let $\omega_p = \{b_p, d_p, u_p\}$ be an agent's opinion about a binary event p, and let $\varphi_p(r_p, s_p)$ be the same agent's probability estimate regarding p expressed as a pcdf. We now define ω_p as a function of $\varphi_p(r_p, s_p)$ according to:

$$\begin{cases} b_p = \frac{r_p}{r_p+s_p+1} \\ d_p = \frac{s_p}{r_p+s_p+1} \\ u_p = \frac{1}{r_p+s_p+1} \end{cases} \tag{5}$$

We see for example that the uniform $\varphi(0,0)$ corresponds to $\omega = \{0,0,1\}$ which expresses total uncertainty, that $\varphi(\infty,0)$ or the absolute probability corresponds to $\omega = \{1,0,0\}$ which expresses absolute belief, and that $\varphi(0,\infty)$ or the zero probability corresponds to $\omega = \{0,1,0\}$ which expresses absolute disbelief. By defining ω as a function of φ according to (5), the interpretation of ω corresponds exactly to the interpretation of φ.

Strictly speaking, opinions without uncertainty, such as for example $\omega = \{0.5, 0.5, 0\}$, do not have a clear equivalent representation as pcdf because the $\{r, s\}$ parameters would explode. In order to avoid this problem, we can define Ω' to be the class of opinions with non-zero uncertainty, that is with $u \neq 0$.

Eq.(5) defines a bijective mapping between the evidence space and the opinion space so that any pcdf has an equivalent mathematical and interpretative representation as an opinion and vice versa, making it possible to produce opinions based on statistical evidence.

3.4 Representing Trust

From an information security point of view, it can be observed that humans are trusted because they are believed to be honest whereas systems are trusted because they are believed to be secure [Jøs96], and this will form the basis for our definition of trust.

Imagine an observer A who is considering her trust in a particular system. She can form the proposition p: "*The system will resist malicious attacks.*" Now, her trust in the system will be her belief in p, expressed as ω_p^A.

Let the same observer A consider her trust in a particular human agent. She must assume that the agent will either cooperate or defect. She can form the proposition q: "*The agent will cooperate.*" Her trust in the agent can simply be expressed as ω_q^A, which is the belief that he will cooperate.

In a similar way, trust in the authenticity of a cryptographic key can be expressed by defining r: "*The key is authentic.*" and express the opinion ω_r^A.

These simple examples demonstrate that trust easily can be expressed as an opinion. The whole framework for artificial reasoning based on subjective logic can therefore be used for reasoning about trust.

4 The Operators of Subjective Logic

Standard propositional logic operates on binary variables which can only take the values *TRUE* and *FALSE*. Subjective logic operates on our subjective perception about binary propositions, represented as opinions. Presently, subjective logic contains about 10 operators[Jøs97b], but due to limited space, we only describe the subset consisting of *conjunction*, *consensus* and *recommendation* here.

4.1 Conjunction

A conjunction of two opinions about two distinct propositions consists of determining from the two opinions a new opinion reflecting the conjunctive truth of both propositions. This corresponds to the logical binary "AND" operation in standard logic.

Definition 2. *Let* $\omega_p = \{b_p, d_p, u_p\}$ *and* $\omega_q = \{b_q, d_q, u_q\}$ *be an agent's opinions about two distinct propositions* p *and* q. *Let* $\omega_{p\wedge q} = \{b_{p\wedge q}, d_{p\wedge q}, u_{p\wedge q}\}$ *be the opinion such that*

1. $b_{p\wedge q} = b_p b_q$
2. $d_{p\wedge q} = d_p + d_q - d_p d_q$
3. $u_{p\wedge q} = b_p u_q + u_p b_q + u_p u_q$

Then $\omega_{p\wedge q}$ *is called the conjunction of* ω_p *and* ω_q, *representing the agents opinion about both* p *and* q *being true. By using the symbol "*$\wedge$*" to designate this operation, we get* $\omega_{p\wedge q} = \omega_p \wedge \omega_q$. □

As would be expected, conjunction of opinions is both commutative and associative. It must be assumed that the opinion arguments in a conjunction are independent. This means for example that the conjunction of an opinion with itself will be meaningless, because the conjunction rule will see them as if they were opinions about distinct propositions.

When applied to opinions with absolute belief or disbelief, it produces the same results as the conjunction rule in standard logic, that is; it produces the truth table of logical "AND". In addition, when applied to opinions with zero uncertainty, it produces the same results as serial multiplication of probabilities.

4.2 Consensus of Independent Opinions

This operator is most naturally expressed in the evidence space, so we will define it there first and subsequently map it over to the opinion space.

Assume two agents A and B having observed an entity produce binary events over two different periods respectively, with A having observed r^A positive and s^A negative events, and B having observed r^B positive and s^B negative events. According to (3), their respective pcdfs are then $\varphi(r^A, s^A)$ and $\varphi(r^B, s^B)$. Imagine now that they combine their observations to form a better estimate of the events' probability. This is equivalent to an imaginary agent $[A, B]$ having made all the observations and who therefore is able to form the pcdf defined by $\varphi(r^A + r^B, s^A + s^B)$.

Definition 3. *Let $\varphi(r_p^A, s_p^A)$ and $\varphi(r_p^B, s_p^B)$ be two pcdfs respectively held by the agents A and B regarding the truth of a proposition p. The pcdf $\varphi(r_p^{A,B}, s_p^{A,B})$ defined by*

1. $r_p^{A,B} = r_p^A + r_p^B$
2. $s_p^{A,B} = s_p^A + s_p^B$

is then called the consensus rule for combining A's and B's estimates, as if it was an estimate held by an imaginary agent $[A, B]$. By using the symbol $\oplus$ to designate this operation, we get $\varphi(r_p^{A,B}, s_p^{A,B}) = \varphi(r_p^A, s_p^A) \oplus \varphi(r_p^B, s_p^B)$. □

The consensus rule for combining independent opinions is easily obtained by using Def.3 above and the evidence-opinion mapping of Eq.(5).

Theorem 1. *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 proposition p. Let $\omega_p^{A,B} = \{b_p^{A,B}, d_p^{A,B}, u_p^{A,B}\}$ be the opinion such that*

1. $b_p^{A,B} = (b_p^A u_p^B + b_p^B u_p^A)/\kappa$
2. $d_p^{A,B} = (d_p^A u_p^B + d_p^B u_p^A)/\kappa$
3. $u_p^{A,B} = (u_p^A u_p^B)/\kappa$

where $\kappa = u_p^A + u_p^B - u_p^A u_p^B$ such that $\kappa \neq 0$. Then $\omega_p^{A,B}$ is called the Bayesian consensus between ω_p^A and ω_p^B, representing an imaginary agent $[A, B]$'s opinion about p, as if she represented both A and B. By using the symbol $\oplus$ to designate this operation, we get $\omega_p^{A,B} = \omega_p^A \oplus \omega_p^B$. □

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 entity's opinion to be counted more than once

The effect of independent consensus 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 not totally uncertain (i.e. $u < 1$) opinions would necessarily produce a consensus opinion with $u = 0$.

Two opinions which both contain zero uncertainty can not be combined according to Def.1. This can be explained by interpreting uncertainty as *room for influence*, meaning that it is only possible to influence an opinion which has not yet been committed to belief or disbelief. An opinion containing zero uncertainty can only influence opinions which do contain uncertainty, not the opposite, and the result will always be the total elimination of uncertainty. In reality, opinions which do not include uncertainty are usually counterintuitive, except in specially designed situations such as games with carefully controlled probabilities.

4.3 Recommendation

Assume two agents A and B where A has an opinion about B, and B has an opinion about a proposition p. A recommendation of these two opinions consists of combining A's opinion about B with B's opinion about p in order for A to get an opinion about p.

There is no such thing as physical recommendation, and recommendation of opinions therefore lends itself to different interpretations. The main difficulty lies with describing the effect of A disbelieving that B will give a good advice. For the definition of the recommendation operator, A's disbelief in the recommending agent B means that A thinks that B is uncertain about the truth value of p. As a result A is also uncertain about the truth value of p.

Definition 4. *Let A, B and 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 proposition 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. Let $\omega_p^{AB} = \{b_p^{AB}, d_p^{AB}, u_p^{AB}\}$ be the opinion such that*

$$\begin{array}{l} \textbf{1.}\ b_p^{AB} = b_B^A b_p^B, \\ \textbf{2.}\ d_p^{AB} = b_B^A d_p^B \\ \textbf{3.}\ u_p^{AB} = d_B^A + u_B^A + b_B^A u_p^B \end{array}$$

then ω_p^{AB} is called the recommendation rule for combining ω_B^A and ω_p^B expressing A's opinion about p as a result of the recommendation from B. By using the symbol $\otimes$ to designate this operation, we get $\omega_p^{AB} = \omega_B^A \otimes \omega_p^B$. □

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, and 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.

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.

It is important to notice that the recommendation rule can only be justified when it can be assumed that recommendation is transitive. More precisely it must be assumed that the agents in the 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 [Jøs96] and [BFL96] this can not always be assumed, because defection can be motivated for example by antagonism between certain agents. The recommendation rule must therefore be used with care, and can only be applied in environments where behaviour invariance can be assumed.

5 Evaluating Authenticity

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. These keys can then be trusted and used for confidential message exchange, or 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 communicate confidentially with $A3$. This simple certification chain is illustrated in Fig.3.

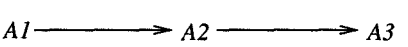

Fig. 3. Simple certification path

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 $\omega^{A1}_{\mathrm{KA}(k_{A2})}$ about the authenticity 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 $\omega^{A1}_{\mathrm{RT}(A2)}$ about the certifiers 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 $\omega^{A2}_{\mathrm{KA}(k_{A3})}$ about the authenticity of the certified key. This opinion must be embedded in the certificate sent to $A1$. The definition of authenticity of a certified key can then be defined.

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

$$\omega^{A1}_{\mathrm{KA}(k_{A3})} = (\omega^{A1}_{\mathrm{RT}(A2)} \wedge \omega^{A1}_{\mathrm{KA}(k_{A2})}) \otimes \omega^{A2}_{\mathrm{KA}(k_{A3})}$$

□

In case there is a path through intermediate certifiers, as illustrated in Fig.4, opinions about recommendation trustworthiness ω_{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 [Jøs97d] 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.

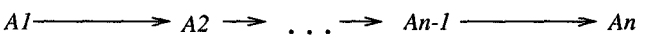

Fig. 4. Chained certification path

Definition 6. *Let the agents $A1$, ... An have chained trust and certification relationships according to Fig.4. $A1$'s opinion about the authenticity of k_{An} can then be expressed as:*

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

□

The framework defined above can now be used to compute authenticity in a given network. If desirable, subjective logic can be reduced to a one-dimensional probabilistic logic by using opinions without uncertainty, i.e. $u = 0$. The logic can also be made binary by only allowing binary belief components, i.e. $b = 0$ or $b = 1$. The full two-dimensional logic will be used in the examples below.

We will start by describing how authenticity can be computed in an anarchic structure, and subsequently show how the same method can be used in a hierarchic structure.

5.1 Anarchic Authentication Network

In anarchic authentication networks each agent decides individually which other agents she will trust to produce certificates. Fig.5 illustrates a possible structure of public keys and their certificates as stored in agent A's private database.

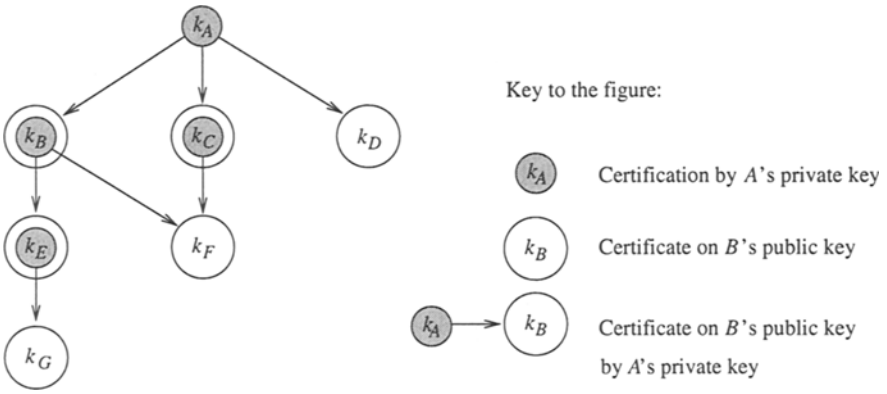

Fig. 5. Structure of keys and certificates in agent A's database

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 ω^A_{KA} about key authenticity, that is, her opinions about binding between keys and key owners. Agent A must also keep a list of her opinions ω^A_{RT} about recommendation trustworthiness, that is how much she trusts the key owners to actually recommend other keys and other agents.

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 Authenticity $\omega^A_{\mathrm{KA}(k_X)}$	Key owner X	Recommendation Trustworthiness $\omega^A_{\mathrm{RT}(X)}$
k_A	{1.00, 0.00, 0.00}	A	{1.00, 0.00, 0.00}
k_B	{0.98, 0.00, 0.02}	B	{0.96, 0.02, 0.02}
k_C	{0.97, 0.00, 0.03}	C	{0.97, 0.01, 0.02}
k_D	{0.98, 0.00, 0.02}	D	{0.90, 0.00, 0.10}

Table 1. Table of A's opinions about public keys and their owners

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 Tab.2.

Key Authenticity	Recommendation Trustworthiness
$\omega^B_{\mathrm{KA}(k_E)} = \{0.98, 0.00, 0.02\}$	$\omega^B_{\mathrm{RT}(E)} = \{0.99, 0.00, 0.01\}$
$\omega^B_{\mathrm{KA}(k_F)} = \{0.95, 0.01, 0.04\}$	$\omega^B_{\mathrm{RT}(F)} = \{0.98, 0.01, 0.01\}$
$\omega^C_{\mathrm{KA}(k_F)} = \{0.98, 0.00, 0.02\}$	$\omega^C_{\mathrm{RT}(F)} = \{0.90, 0.00, 0.10\}$
$\omega^E_{\mathrm{KA}(k_G)} = \{0.90, 0.05, 0.05\}$	$\omega^E_{\mathrm{RT}(G)} = \{0.99, 0.00, 0.01\}$

Table 2. Table of opinions received by A

A key which is received electronically can be considered authentic if it has been certified by someone who is considered trustworthy, and who's public key is considered authentic and if the certifier recommends the key to be authentic. 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. The authenticity of for example k_E as seen by A can then be expressed as:

$$\begin{aligned}\omega^A_{\mathrm{KA}(k_E)} &= (\omega^A_{\mathrm{RT}(B)} \wedge \omega^A_{\mathrm{KA}(k_B)}) \otimes \omega^B_{\mathrm{KA}(k_E)} \\ &= \{0.922,\ 0.000,\ 0.078\}\end{aligned} \tag{6}$$

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:

$$\begin{aligned}\omega^A_{\mathrm{KA}(k_F)} &= ((\omega^A_{\mathrm{RT}(B)} \wedge \omega^A_{\mathrm{KA}(k_B)}) \otimes \omega^B_{\mathrm{KA}(k_F)}) \oplus ((\omega^A_{\mathrm{RT}(C)} \wedge \omega^A_{\mathrm{KA}(k_C)}) \otimes \omega^C_{\mathrm{KA}(k_F)}) \\ &= \{0.951,\ 0.004,\ 0.045\}\end{aligned} \tag{7}$$

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:

$$\begin{aligned}\omega^A_{\mathrm{KA}(k_G)} &= (\omega^A_{\mathrm{RT}(B)} \wedge \omega^A_{\mathrm{KA}(k_B)}) \otimes (\omega^B_{\mathrm{RT}(E)} \wedge \omega^B_{\mathrm{KA}(k_E)}) \otimes \omega^E_{\mathrm{KA}(k_G)} \\ &= \{0.821,\ 0.046,\ 0.133\}\end{aligned} \tag{8}$$

Opinions about key authenticity and recommendation trustworthiness must always be included when sending certificates to other agents. However, opinions based on recommendations from other agents, i.e. based on second-hand evidence, should never be passed to other agents. This is because the recipient may receive recommendations from the same agents, causing opinion dependence and computational inconsistencies. Only opinions based on first-hand evidence and experience should thus be recommended to other agents.

5.2 Hierarchic Authentication Network

The X.509 authentication framework [ITU89] defines a typical hierarchic network structure in which Certification Authorities (*CA*s) distribute certified cryptographic keys to the users. Fig.6 illustrates two partly overlapping hierarchies where user $U3$ belongs to both $CA2$ and $CA3$.

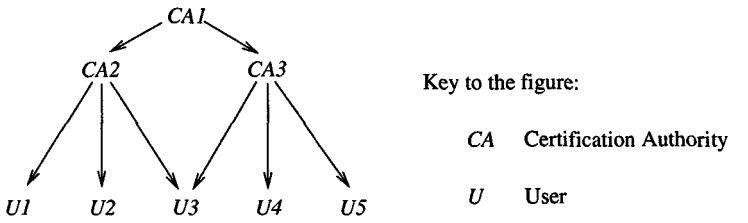

Fig. 6. Hierarchic Authentication Network

It is assumed that every user U has a certain opinion $\omega^{U}_{\mathrm{RT}(CA)}$ about its superior CA regarding recommendation trustworthiness. The same applies to trust between *CA*s both in upwards and downwards direction.

Every user U possesses the public key k_{CA} of his superior CA, to which is attached an opinion $\omega^{U}_{\mathrm{KA}(k_{CA})}$ about the key authenticity. The same applies to the *CA*s in upwards and downwards direction, as well as to the *CA*s regarding their users.

The computations for determining key authenticities are analogous to the computations in case of an anarchic network. User $U1$'s opinion about the key authenticity of user $U2$ can for example be expressed as:

$$\omega^{U1}_{\mathrm{KA}(k_{U2})} = (\omega^{U1}_{\mathrm{RT}(CA2)} \wedge \omega^{U1}_{\mathrm{KA}(k_{CA2})}) \otimes \omega^{CA2}_{\mathrm{KA}(k_{U2})} \tag{9}$$

The key authenticity of $U3$ as seen by $U1$ can be computed using both the short path via $CA2$ as well as the long path via $CA1$ and $CA3$ to obtain:

$$\begin{aligned} \omega^{U1}_{\mathrm{KA}(k_{U3})} = \quad & (\omega^{U1}_{\mathrm{RT}(CA2)} \wedge \omega^{U1}_{\mathrm{KA}(k_{CA2})}) \\ & \otimes (\omega^{CA2}_{\mathrm{KA}(k_{U3})} \oplus ((\omega^{CA2}_{\mathrm{RT}(CA1)} \wedge \omega^{CA2}_{\mathrm{KA}(k_{CA1})}) \\ & \otimes (\omega^{CA1}_{\mathrm{RT}(CA3)} \wedge \omega^{CA1}_{\mathrm{KA}(k_{CA3})}) \otimes \omega^{CA3}_{\mathrm{KA}(k_{U3})})) \end{aligned} \tag{10}$$

When the computation is based on intermediate CA's, the recommendations can be embedded inside a multi layer certificate, corresponding to the CA's through which it has passed, or each CA can send a certificate regarding the neighbouring CA directly to the recipient user.

When the computation is based on multiple paths, the consensus must be computed by the agent where the paths meet which is $CA2$ in our example. Otherwise, opinions about the node where the paths meet will appear more than once in the expression, and thereby introduce dependence.

6 Conclusion

We have proposed a radically new way of reasoning about trust and authentication using subjective logic and have described how this can be used to define a metric for authenticity as well as a framework for computing authenticity values. An essential aspect of our approach is to include the binding between keys and their owners, without which a certificate would have no real meaning to the users. We believe that this approach can be directly applied in real implementations of authentication networks.

References

[BBK94] T. Beth, M. Borcherding, and B. Klein. Valuation of trust in open networks. In *ESORICS 94. Brighton, UK*, November 1994.

[BFL96] Matt Blaze, Joan Feigenbaum, and Jack Lacy. Decentralized trust management. In *Proceedings of the 1996 IEEE Conference on Security and Privacy*, Oakland, CA, 1996.

[CB90] George Casella and Roger L. Berger. *Statistical Inference.* Duxbury Press, 1990.

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

[ITU89] ITU. *Recommendation X.509, The Directory Authentication Framework.* International Telecommunications Union, 1989.

[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.

[Jøs97a] A. Jøsang. Artificial reasoning with subjective logic. In Abhaya Nayak, editor, *Proceedings of the Second Australian Workshop on Commonsense Reasoning*, 1997.

[Jøs97b] A. Jøsang. A model for trust in security systems. In Arto Karila and Timo Aalto, editors, *Proceedings of the Second Second Nordic Workshop on Secure Computer Systems.* Helsinki University of Technology, November 1997.

[Jøs97c] A. Jøsang. Prospectives for modelling trust in information security. In Vijay Varadharajan, editor, *Proceedings of the 1997 Australasian Conference on Information Security and Privacy.* Springer-Verlag, 1997.

[Jøs97d] Audun Jøsang. *Modelling Trust in Information Security.* PhD thesis, Norwegian University of Science and Technology, 1997.

[Mau96] Ueli Maurer. Modelling a public-key infrastructure. In *Computer Security - ESORICS'96.* Springer-Verlag, 1996.

[RS97] 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.

[Zim95] P.R. Zimmermann. *The Official PGP User's Guide.* MIT Press, 1995.