# A Metric for Trusted Systems*

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

S.J. Knapskog
The Norwegian University of Science and Technology
N-7034 Trondheim, Norway, email: svein.knapskog@item.ntnu.no

**Abstract.** This paper proposes a model for quantifying and reasoning about trust in IT equipment. Trust is considered to be a subjective belief, and the model consists of a belief model and set of operators for combining beliefs. Security evaluation is being discussed as a method for determining trust. Trust may also be based on other types of evidence such as for example ISO 9000 certification, and the model can be used to quantify and compare the contribution to the total trust each type of evidence provides.

## 1 Introduction

Security evaluation is an example of a well established method for determining trust in implemented system components. The method is based on a set of evaluation criteria like e.g. TCSEC [USD85], ITSEC [EC92], CC [ISO98] or similar, and accredited evaluation laboratories which perform the evaluation under supervision of a national authority. A successful evaluation leads to the determination of an assurance level which shall reflect to which degree the TOE or system component can be trusted.

As mentioned in [JVLKV97], evaluation assurance does not represent the users own trust in the actual system component, but rather a recommendation from a supposedly trusted authority. In addition, the evaluation assurance is only one of several factors supporting the user's own trust in the product.

The traditional IT security evaluation has been regarded as an activity exclusive for the higher end of the assurance scale, and always performed by some external body, guaranteeing the potential user an impartial assessment of the correctness and effectiveness of the security measures taken during the development of the system, and the strength and suitability of the security services implemented in the system itself. The traditional security evaluation effort has, quite justifiably, been regarded as costly and extremely time consuming - the evaluation effort in the majority of cases demanding as many resources as the development itself. Government driven evaluation schemes traditionally encompass four roles, being *accreditor*, *certifier*, *evaluator* and *sponsor* (usually the developer).

The accreditor is a government body that accredits the certifier, the evaluator, and also, in some cases, evaluated IT systems. The discrimination of an IT system and an IT product is that for the system, all operational parameters are known, and taken into consideration during an evaluation, while for the IT product, the operational parameters encountered during its future use can only be assumed. The accreditation of the certifier is done based on the documented competence level, skill and resources the certifier possesses for the purpose of overseeing the IT security evaluations performed by the evaluators, and the issuing of appropriate security certificates based on the evaluation results.

The certifier is also a government body issuing security certificates based on the evaluation reports from the evaluators. The certificate will contain a description of the security functionality of the IT product of system in question, and give an assurance rating for the same. Both the functionality and assurance will be characterised by terms used in the security evaluation criteria on which this evaluation scheme rests.

In the traditional scheme, the evaluator is yet another government body or agency, working independently of, but in close cooperation with both the developers and the certifier. The evaluator is accredited by the accreditor, and the quality of the evaluator's effort will be supervised by the certifier. The evaluator works with the IT products and systems put up for evaluation by the sponsors, taking into consideration both the Target of Evaluation (TOE) itself, its documentation, and for the IT product case, assumptions of it's future operational environment. The evaluation methods used will vary with the assurance level aimed at, from

* Published as short paper in [JK98]

simple checks to sophisticated independent tests of the implementations. The checks are performed against the standard set of evaluation criteria used for the scheme, but the detailed evaluation methods are not in any way standardised. The overall quality of the evaluation methods, and the competence with which they are applied, will, however, undoubtedly influence the end results, and should ideally be standardised, or as a minimum, be brought to the certifiers knowledge through the evaluation report.

The evaluation sponsors are the parties interested in having the TOE evaluated. A manufacturer's motivation would be to obtain a security certificate, helpful for marketing purposes, and in some circumstances demanded by his customers, typically for a delivery under a government type of contract. Another sponsor role will be played by the end users of IT products or systems for security critical applications - they would become evaluation sponsors because they can not rely on their own expertise when it comes to assessing the IT security functionality and judge if the assurance components implemented in the product are adequate for its intended use. In the existing transition period, where many IT products and systems exist which claim to have adequate security, without having a certificate to support the validity of the claims, the end user will have to sponsor the majority of evaluations, especially for IT systems.

In a commercially based IT security evaluation scheme, the evaluators will be entrepreneurs performing IT security evaluation on a commercial contract basis, and there will be competition between different evaluators. To retain a certain trust for the end user in such a scheme, beyond the possibilities created by the commercial contracts themselves, the accreditation and certification authority roles need to be maintained, and possibly strengthened. The common basis for the evaluations performed will be the evaluation criteria on which the scheme is based, but the evaluation methods will probably become the most important competitive factor decisive of success or failure of an evaluator in the commercial IT security evaluation market, and as such, not publicly known. This places a significant responsibility on the accreditation and certification bodies, both with respect to their professional integrity as such, and their ability to convince the end user society that their integrity and competence in this area are unquestionable.

In this paper, we propose a formal model for reasoning about trust and security evaluation. Our approach is based on *subjective logic*[Jøs97] which consists of the *opinion model* for representing beliefs and a set of operators for combining opinions. We show that trust can be represented as an opinion so that situations involving trust and trust relationships can be modelled with subjective logic.

## 2 The Opinion Model

The *evidence space* and *opinion space* are two equivalent models for representing human beliefs, and we will in Sec.5 show how they can be used to represent trust mathematically.

### 2.1 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 $\alpha$ and $\beta$. The beta$(\alpha, \beta)$ distribution can be expressed using the gamma function $\Gamma$ as:

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

with the restriction that $\theta \neq 0$ if $\alpha < 1$, and $\theta \neq 1$ if $\beta < 1$.

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

$$\begin{array}{l} \alpha = r + 1,\ r \geq 0 \quad \text{and} \\ \beta = s + 1,\ s \geq 0. \end{array} \tag{2}$$

Let $\varphi$ be a pcdf over the probability variable $\theta$. In our notation $\varphi$ can then be characterised by $r$ and $s$ according to:

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

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.

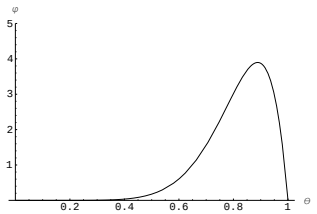


**Fig. 1.** Posteriori pcdf after 8 positive and 1 negative results

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 [Jøs97].

### 2.2 The Opinion Space

For the purpose of believing a binary proposition such as for example: "*The system will resist malicious attacks*", 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. This can be mathematically expressed 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. 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. We will denote by $\Omega$ the set of opinions defined in this way.

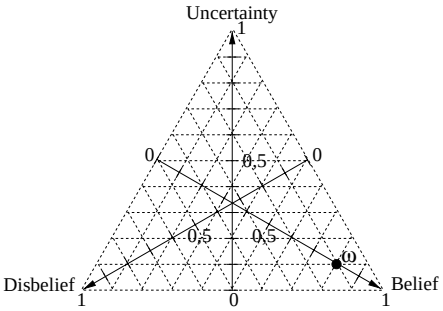


**Fig. 2.** Opinion Triangle

The horizontal bottom line between belief and disbelief in Fig.2 represents situations without uncertainty and is equivalent to a traditional binary probability model. The degree of uncertainty can be interpreted

as 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 $\text{Red}_\text{I}$" 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 $\text{Red}_\text{II}$ are equal. For Urn I however, it is not obvious. If however one was forced to make a bet on $\text{Red}_\text{I}$, 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 $\text{Red}_\text{I}$ and $\text{Red}_\text{II}$ become $\omega_\text{I} = \{0,\ 0,\ 1\}$ and $\omega_\text{II} = \{0.5,\ 0.5,\ 0.0\}$ respectively.

### 2.3 Equivalence between the Opinion Space and the Evidence Space

We have defined $\Phi$ to be the class of pcdfs, and $\Omega$ 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 $\varphi(r_p, s_p)$ be the same agent's pcdf regarding $p$ expressed as a pcdf. We now define $\omega_p$ as a function of $\varphi(r_p, s_p)$ 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 $\omega$ as a function of $\varphi$ according to (5), the interpretation of $\omega$ corresponds exactly to the interpretation of $\varphi$.

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 $\Omega'$ 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 Subjective Logic

Standard propositional logic operates on binary variables that can take the discrete values of TRUE or FALSE. Subjective logic which we will present in this section operates on opinions as described in Sec.2.2, or equivalently on evidence based pcdfs as defined in Sec.2.1.

Opinions, as well as pcdfs, 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 $\omega_p^A$ is an opinion held by agent $A$ about the truth of proposition $p$.

Subjective logic contains about 10 operators[Jøs97], but only the subset consisting of *conjunction*, *independent consensus*, *dependent consensus*, *partly dependent consensus* and *recommendation* will be described here.

### 3.1 Conjunction

A conjunction of two opinions about 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 1.** *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*

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

*Then $\omega_{p\wedge q}$ is called the conjunction of $\omega_p$ and $\omega_q$, 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.

### 3.2 Consensus between Independent Opinions

Assume two agents $A$ and $B$ having observed an entity produce a binary event 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 Eq.(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 event's probability. This is equivalent to an imaginary agent $[A, B]$ having made all the observations and who therefore can form the pcdf defined by $\varphi(r^A + r^B, s^A + s^B)$.

**Definition 2.** *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*

$$\begin{array}{l} \mathbf{1}.\ r_p^{A,B} = r_p^A + r_p^B \\ \mathbf{2}.\ s_p^{A,B} = s_p^A + s_p^B \end{array}$$

*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 equivalent operator for opinions is easily obtained by using Def.2 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*

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

*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 $\omega_p^A$ and $\omega_p^B$, 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

Two opinions which both contain zero uncertainty can not be combined according to Th.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. A situation with conflicting certain opinions is philosophically meaningless, primarily because opinions about real situations can never be absolutely certain, and secondly, because if they were they would necessarily be equal.

### 3.3 Consensus between Dependent Opinions

Assume two agents $A$ and $B$ having simultaneously observed an entity produce a binary event a certain number of times. Because their observations are identical, their respective opinions will necessarily be dependent, and a consensus according to Def.2 would be meaningless.

We will define a consensus rule for dependent pcdfs based on the average of recorded positive and negative observations. If their respective dependent pcdfs are $\varphi(r^A, s^A)$ and $\varphi(r^B, s^B)$, we will let the consensus estimate be defined by the pcdf $\varphi((r^A + r^B)/2,\ (s^A + s^B)/2)$. When the consensus between $n$ dependent opinion is to be computed, the general expression can be defined as follows:

**Definition 3.** *Let $\varphi(r_p^{A_i}, s_p^{A_i})$, where $i \in \{1, .., n\}$, be $n$ dependent pcdfs respectively held by the agents $A_1, .., A_n$ regarding the truth of proposition $p$. The pcdf $\varphi(r_p^{\overline{A_1,..,A_n}}, s_p^{\overline{A_1,..,A_n}})$ defined by*

$$\mathbf{1}.\ r_p^{\overline{A_1,..,A_n}} = \frac{\sum_{i=1}^{n} r_p^{A_i}}{n}$$

$$\mathbf{2}.\ s_p^{\overline{A_1,..,A_n}} = \frac{\sum_{i=1}^{n} s_p^{A_i}}{n}$$

*is then called the consensus rule for combining dependent pcdfs. By using the symbol $\overline{\oplus}$ to designate this operation, we get $\varphi(r_p^{\overline{A_1,..,A_n}}, s_p^{\overline{A_1,..,A_n}}) = \varphi(r_p^{A_1}, s_p^{A_1}) \overline{\oplus} \ldots \overline{\oplus} \varphi(r_p^{A_n}, s_p^{A_n})$.* $\square$

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

**Theorem 2.** *Let $\omega_p^{A_i} = \{b_p^{A_i}, d_p^{A_i}, u_p^{A_i}\}$ where $i \in \{1, .., n\}$, be $n$ dependent opinions respectively held by agents $A_1, .., A_n$ about the same proposition $p$. Let $\omega_p^{\overline{A_1,..,A_n}} = \{b_p^{\overline{A_1,..,A_n}}, d_p^{\overline{A_1,..,A_n}}, u_p^{\overline{A_1,..,A_n}}\}$ be the opinion such that*

$$\mathbf{1}.\ b_p^{\overline{A_1,..,A_n}} = \frac{\sum_{i=1}^{n} (b_p^{A_i}/u_p^{A_i})}{\sum_{i=1}^{n} (b_p^{A_i}/u_p^{A_i}) + \sum_{i=1}^{n} (d_p^{A_i}/u_p^{A_i}) + n}$$

$$\mathbf{2}.\ d_p^{\overline{A_1,..,A_n}} = \frac{\sum_{i=1}^{n} (d_p^{A_i}/u_p^{A_i})}{\sum_{i=1}^{n} (b_p^{A_i}/u_p^{A_i}) + \sum_{i=1}^{n} (d_p^{A_i}/u_p^{A_i}) + n}$$

$$\mathbf{3}.\ u_p^{\overline{A_1,..,A_n}} = \frac{n}{\sum_{i=1}^{n} (b_p^{A_i}/u_p^{A_i}) + \sum_{i=1}^{n} (d_p^{A_i}/u_p^{A_i}) + n}$$

*Then $\omega_p^{\overline{A_1,..,A_n}}$ is called the dependent consensus between all the $\omega_p^{A_i}$. By using the symbol $\overline{\oplus}$ to designate this operation, we get $\omega_p^{\overline{A_1,..,A_n}} = \omega_p^{A_1} \overline{\oplus} \ldots \overline{\oplus} \omega_p^{A_n}$.*

It is easy to prove that $\overline{\oplus}$ is both commutative and associative which means that the order in which opinions are combined has no importance. Opinions without uncertainty can not be combined according to Th.2 for the same reason as for consensus between independent opinions.

### 3.4 Consensus between Partially Dependent Opinions

If two agents $A$ and $B$ have observed the same events during two partially overlapping periods, their respective pcdfs will be partially dependent. A situation where the pcdfs are based on partly dependent observations is illustrated in Fig.3. In the figure, $\varphi_p^{A\mathrm{i}(B)}$ and $\varphi_p^{B\mathrm{i}(A)}$ represent the independent parts of $A$ and $B$'s pcdfs, whereas $\varphi_p^{A\mathrm{d}(B)}$ and $\varphi_p^{B\mathrm{d}(A)}$ represent their dependent parts. On the condition that the fraction of simultaneously observed events is known, it is possible to isolate the dependent and the independent parts of their observations and thereby also of their pcdfs. The dependent estimates can then be combined according to the dependent consensus rule, and this outcome can be further combined with the remaining two independent estimates according to the independent consensus rule.

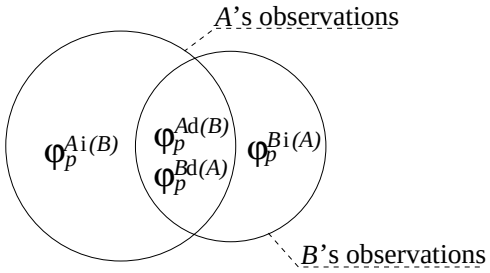


**Fig. 3.** Pcdfs based on partly overlapping observations

Let $\varphi_p^A$'s fraction of dependence with $\varphi_p^B$ and vice versa be represented by the dependence factors $\gamma_p^{AB}$ and $\gamma_p^{BA}$. The dependent and independent pcdfs can then be defined as a function of the dependence factors.

$$
\begin{array}{ll}
\textbf{1}.\ \varphi_p^{A\mathrm{i}(B)}: \begin{cases} r_p^{A\mathrm{i}(B)} = r_p^A(1-\gamma_p^{AB}) \\ s_p^{A\mathrm{i}(B)} = s_p^A(1-\gamma_p^{AB}) \end{cases} & \textbf{3}.\ \varphi_p^{B\mathrm{i}(A)}: \begin{cases} r_p^{B\mathrm{i}(A)} = r_p^B(1-\gamma_p^{BA}) \\ s_p^{B\mathrm{i}(A)} = s_p^B(1-\gamma_p^{BA}) \end{cases} \\
\textbf{2}.\ \varphi_p^{A\mathrm{d}(B)}: \begin{cases} r_p^{A\mathrm{d}(B)} = r_p^A\gamma_p^{AB} \\ s_p^{A\mathrm{d}(B)} = s_p^A\gamma_p^{AB} \end{cases} & \textbf{4}.\ \varphi_p^{B\mathrm{d}(A)}: \begin{cases} r_p^{B\mathrm{d}(A)} = r_p^B\gamma_p^{BA} \\ s_p^{B\mathrm{d}(A)} = s_p^B\gamma_p^{BA} \end{cases}
\end{array}
\tag{6}
$$

We will use the symbol $\widetilde{\oplus}$ to designate consensus between partially dependent pcdfs. As before $\overline{\oplus}$ is the operator for entirely dependent pcdfs. The consensus of $A$ and $B$'s pcdfs can then be written as:

$$
\begin{aligned}
\varphi_p^A \widetilde{\oplus} \varphi_p^B &= \varphi_{\underline{p}}^{\widetilde{A,B}} \\
&= \varphi_p^{\overline{A\mathrm{d}(B),B\mathrm{d}(A)},A\mathrm{i}(B),B\mathrm{i}(A)} \\
&= \left(\varphi_p^{A\mathrm{d}(B)} \overline{\oplus} \varphi_p^{B\mathrm{d}(A)}\right) \oplus \varphi_p^{A\mathrm{i}(B)} \oplus \varphi_p^{B\mathrm{i}(A)}
\end{aligned}
\tag{7}
$$

The equivalent representation of dependent and independent opinions can be obtained by using Eq.(6) and the evidence-opinion mapping Eq.(5). The reciprocal dependence factors are as before denoted by $\gamma^{AB}$ and $\gamma^{BA}$.

$$
\begin{array}{lll}
\mathbf{1}.\ \omega_p^{A\mathrm{i}(B)}: & \begin{cases} b_p^{A\mathrm{i}(B)} = b_p^A \mu_p^{AB} \\ d_p^{A\mathrm{i}(B)} = d_p^A \mu_p^{AB} \\ u_p^{A\mathrm{i}(B)} = u_p^A \mu_p^{AB} / (1-\gamma_p^{AB}) \end{cases} & \text{where } \mu_p^{AB} = \frac{1-\gamma_p^{AB}}{(1-\gamma_p^{AB})(b_p^A + d_p^A) + u_p^A} \\
\mathbf{2}.\ \omega_p^{A\mathrm{d}(B)}: & \begin{cases} b_p^{A\mathrm{d}(B)} = b_p^A \nu_p^{AB} \\ d_p^{A\mathrm{d}(B)} = d_p^A \nu_p^{AB} \\ u_p^{A\mathrm{d}(B)} = u_p^A \nu_p^{AB} / \gamma_p^{AB} \end{cases} & \text{where } \nu_p^{AB} = \frac{\gamma_p^{AB}}{\gamma_p^{AB}(b_p^A + d_p^A) + u_p^A} \\
\mathbf{3}.\ \omega_p^{B\mathrm{i}(A)}: & \begin{cases} b_p^{B\mathrm{i}(A)} = b_p^B \mu_p^{BA} \\ d_p^{B\mathrm{i}(A)} = d_p^B \mu_p^{BA} \\ u_p^{B\mathrm{i}(A)} = u_p^B \mu_p^{BA} / (1-\gamma_p^{BA}) \end{cases} & \text{where } \mu_p^{BA} = \frac{1-\gamma_p^{BA}}{(1-\gamma_p^{BA})(b_p^B + d_p^B) + u_p^B} \\
\mathbf{4}.\ \omega_p^{B\mathrm{d}(A)}: & \begin{cases} b_p^{B\mathrm{d}(A)} = b_p^B \nu_p^{BA} \\ d_p^{B\mathrm{d}(A)} = d_p^B \nu_p^{BA} \\ u_p^{B\mathrm{d}(A)} = u_p^B \nu_p^{BA} / \gamma_p^{BA} \end{cases} & \text{where } \nu_p^{BA} = \frac{\gamma_p^{BA}}{\gamma_p^{BA}(b_p^B + d_p^B) + u_p^B}
\end{array}
\tag{8}
$$

We will use the symbol $\widetilde{\oplus}$ to designate consensus between partially dependent opinions. As before $\overline{\oplus}$ is the operator for entirely dependent opinions. The consensus of $A$ and $B$'s opinions can then be written as:

$$
\begin{aligned}
\omega_p^A \widetilde{\oplus} \omega_p^B &= \omega_p^{\widetilde{A,B}} \\
&= \omega_p^{\overline{A\mathrm{d}(B), B\mathrm{d}(A)}, A\mathrm{i}(B), B\mathrm{i}(A)} \\
&= (\omega_p^{A\mathrm{d}(B)} \overline{\oplus} \omega_p^{B\mathrm{d}(A)}) \oplus \omega_p^{A\mathrm{i}(B)} \oplus \omega_p^{B\mathrm{i}(A)}
\end{aligned}
\tag{9}
$$

It is easy to prove that for any opinion $\omega_p^A$ with a dependence factor $\gamma_p^{AB}$ to any other opinion $\omega_p^B$ the following equality holds:

$$
\omega_p^A = \omega_p^{A\mathrm{i}(B)} \oplus \omega_p^{A\mathrm{d}(B)} \tag{10}
$$

### 3.5 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*

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

*then $\omega_p^{AB}$ is called the recommendation rule for combining $\omega_B^A$ and $\omega_p^B$ 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 $\omega_p^{AB} = \omega_B^A \otimes \omega_p^B$.* $\square$

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.

## 4 How to Determine Opinions

The major difficulty with applying subjective logic is to find a way to consistently determine opinions to be used as input parameters. People may find the opinion model unfamiliar, and different individuals may produce conflicting opinions when faced with the same evidence.

For basing opinions on statistical evidence, Sec.2.1 described how it is possible to define formal rules for determining opinions. If on the other hand the evidence can only be analysed intuitively, guidelines for determining opinions may be useful. In this section we will attempt to formulate a questionnaire for guiding people in expressing their beliefs as opinions. The idea behind the questionnaire is to let the observer consider each component of her opinion separately.

An opinion is always about a binary proposition, so the first task when trying to determine an opinion intuitively is to express the proposition clearly. The subject should be informed that it is assumed that nobody can be absolutely sure about anything, so that opinions with $u = 0$ should never be specified. The questionnaire below will help observers isolate the components of their opinions in the form of belief, disbelief and uncertainty.

**Questionnaire for Determining Intuitive Opinions.**

1. *Is the proposition clearly expressed?*
   *Yes*: $\rightarrow$ (2)
   *No*: *Do it, and start again.* $\rightarrow$ (1)
2. *Is there any evidence, or do you have an intuitive feeling in favour of or against the proposition?*
   *Yes*: $\rightarrow$ (3)
   *No*: *You are totally uncertain.* $b := 0$, $d := 0$, $u := 1$. $\rightarrow$ (7)
3. *How conclusive is this evidence or how strong is this feeling?*
   *Give a value* $0 \leq x < 1$.
   $\rightarrow$ (4)
4. *How strong is the evidence or the intuitive feeling against the proposition?*
   *Give a value* $0 \leq y < 1$.
   $\rightarrow$ (5)
5. *How strong is the evidence or the intuitive feeling in favour of the proposition? Give a value* $0 \leq z \leq 1$.
   $\rightarrow$ (6)
6. Normalisation of results:

$$b := \frac{z}{z+y+(1-x)}$$
$$d := \frac{y}{z+y+(1-x)}$$
$$i := \frac{1-x}{z+y+(1-x)}$$

   $\rightarrow$ (7)
7. $\omega = \{b,\ d,\ u\}$

## 5 Modelling Trust

Trust can be defined as a subjective belief. In particular, trust in a system is the belief that the system will resist malicious attacks, and trust in a human agent is the belief that he will cooperate [Jøs96].

An observer $A$ who is assessing the security of particular system can form the proposition $p$: "*The system will resist malicious attacks.*" Now, her trust in the system will be her opinion about $p$, expressed as $\omega_p^A$.

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 $\omega_q^A$, 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 $\omega_r^A$.

These simple examples demonstrate that trust easily can be expressed as an opinion. The whole framework for subjective logic described above can therefore also be used for reasoning about trust. In the following, we will demonstrate how subjective logic can be applied to modelling trust in systems.

### 5.1 Modelling the Evaluation Scheme

In this section we will attempt to analyse trust resutling from a traditional security evaluation scheme as described in Sec.1. For this analysis, we will distinguish between the establishment of the trust relationships, and their subsequent role during security evaluations.

**The Set-Up Phase.** An authority assigns the role of accreditor to a suitable organisation which is assumed to be trustworthy by all future participants in the scheme.

Organisations which want to be certifiers can apply to the accreditor and provide evidence to prove that they fulfil the necessary requirements. If the accreditor is satisfied it will grant a licence to the new certifier. This fact becomes evidence for everyone interested in order to trust the new certifier, as illustrated in Figure 4.a. A similar process takes place for establishing trust in evaluators, as illustrated in Figure 4.b.

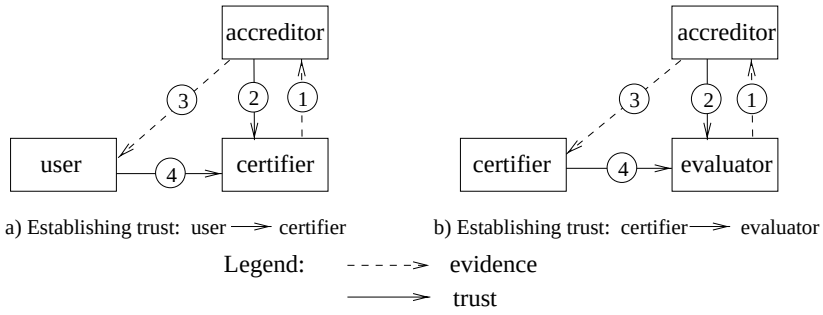


**Fig. 4.** Set-up of security evaluation scheme

The accreditor is only checking that the applicant is fulfilling a set of requirements, and strictly speaking does not need to trust the certifier or the evaluator. However, we find that individual human members of the accrediting organisation should actually trust an applicant before granting a licence. The licensing can therefore be seen as a recommendation to the public to trust the services provided by the certifier and the evaluator.

**The Evaluation Phase.** The developer provides evidence to the evaluator who checks that the set of criteria are fulfilled. Note that the evaluator does not need to trust the actual system. It is for example possible that individual employees of the evaluation laboratory find the specified criteria insufficient to cover the security well, but nevertheless can testify that the actual criteria are met. Based on the evaluation report,

the certifier decides whether or not a certificate of evaluation shall be issued. Note that here too, the certifier does not technically need to trust the system, but simply certifies that the evaluation has been performed correctly and that the issued certificate is consistent with the findings of the evaluation. The certificate is not a recommendation to trust the system, but only certifies that the system has been checked against a set of criteria. The user must therefore consider to which degree the certificate will make her trust the system. This chained process is illustrated in Figure 5.

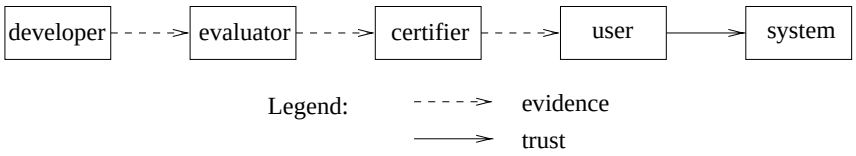


**Fig. 5.** Evaluation and certification

A certificate indicating a security evaluation assurance level can not be directly translated into trust. The establishment of the user's trust in the system includes considering the appropriateness of the criteria themselves, and the quality of the evaluation scheme. Very few users will have the necessary expertise to consider these issues, and therefore have to simply accept them as recommendations. This illustrates that we live in a second-hand world, and very rarely are able to get first-hand evidence about any product or relationship with more than basic complexity.

**Formal Model.** We will apply subjective logic to model the evaluation scheme described above. Subjective logic works on opinions about propositions about the real world, so the first task when applying the logic is to define a suitable set of propositions which reflect the user's point of view.

$a$: "*The system will be sufficiently resistant against malicious attacks.*"
$b$: "*A correctly evaluated and certified system will be sufficiently resistant against malicious attacks.*"
$c$: "*The system is correctly evaluated and certified.*"
$d$: "*The certifier will only certify systems which have been correctly evaluated.*"
$e$: "*The system is certified.*"
$f$: "*The accreditor will only license certifiers which satisfy d.*"
$g$: "*The certifier has a license.*"

Let $A$ be the user. Her trust in the system can be expressed as a function of her opinions about the propositions above.

$$\omega_a^A = \omega_b^A \wedge \omega_c^A$$

$$\omega_c^A = \omega_d^A \wedge \omega_e^A$$
$$\omega_d^A = \omega_f^A \wedge \omega_g^A$$

$$\omega_a^A = \omega_b^A \wedge \omega_e^A \wedge \omega_f^A \wedge \omega_g^A$$

The propositions $e$ and $g$ will normally be non-controversial, and the corresponding opinions can thereby be omitted in the calculation, so that we get:

$$\omega_a^A = \omega_b^A \wedge \omega_f^A \tag{11}$$

which in plain language translates into: "*User A trusts the system to the degree that she believes that a correctly evaluated system will resist malicious attacks, and that she believes that the accreditor will only licence certifiers which truly only certify correctly evaluated systems.*"

The analysis has shown that the user's trust in the system depends on her trust in the appropriateness of the criteria themselves, and her trust in the evaluation scheme in general, here represented by the accreditor which is its highest authority.

**Numerical Example.** Assume the opinions about the propositions $b$ and $f$ above have been determined on an intuitive basis. Let for example

$$\omega_b^A = \{0.80,\ 0.10,\ 0.10\}$$
$$\omega_f^A = \{0.99,\ 0.00,\ 0.01\}$$

The trust in an evaluated and certified system then becomes:

$$\begin{aligned}\omega_{b\wedge f}^A &= \omega_b^A \wedge \omega_f^A \\ &= \{0.792,\ 0.100,\ 0.108\}\end{aligned}$$

### 5.2 Modelling the Context

As pointed out in [JVLKV97] the user's total trust in a system will always be based on evidence from different sources of which security evaluation is only one. Some of the possible sources are are illustrated in Fig.6 and briefly described below.

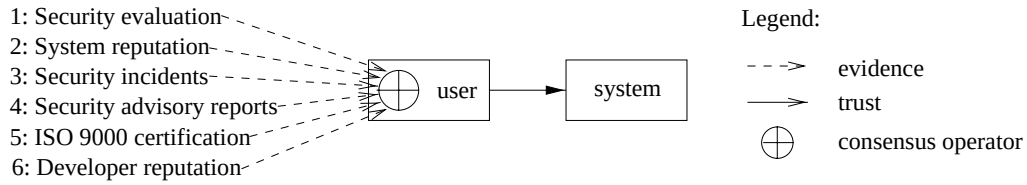


**Fig. 6.** Sources of trust in systems

1. **Security evaluation**. Trust resulting from security evaluation was analysed on the previous section, but each individual user will have to determine this trust, or the underlying opinions, on an intuitive basis. This trust will to a large extent be incorporated in, and thereby dependent on, the trust resulting from the system reputation which will be described next.
2. **System reputation**. This trust can be based on advice from IT consultants or positive critics e.g. from technical journals, and must be determined on an intuitive basis by the user. If the system has been evaluated, the system reputation will be partly dependent on the evaluation itself.
3. **Security incidents**. Security incidents will be considered to only create distrust. It can be determined on a statistical basis, and will be independent of all the other types of evidence.
4. **Security advisory reports**. As for incidents, advisory reports can only create distrust as long as the reported problem has not been fixed. After the problem has been fixed, this opinion should be neutral (uncertain) as long as the number of reported problems is not very high indicating that the system quality in general is low. Each user must determine the trust resulting from this type of evidence on an intuitive basis, and this trust component can be considered independent from the others.
5. **ISO 9000 certification**. Users will trust to a varying degree systems which have been developed and manufactured by ISO 9000 certified companies. Each user must determine this trust component on an intuitive basis. This component can be considered incorporated in, and therefore dependent on, the trust resulting from the developer reputation.
6. **Developer reputation**. This trust is based on what the user has been told or has read in the press about other products produced by the same developer, and about the developer itself. Each user must determine this trust component on an intuitive basis. However, this component is likely to be partly dependent on an eventual ISO 9000 certification.

Faced with the above described types of evidence the user has to determine what she believes is the level of security of the system. The consensus operator described in Sec.3 models this mental process. In the following analysis, the user will be considered as consisting of different personalities, each faced with one

type of evidence, and based on this, having a different opinion about the system's security. Finally all the opinions are combined using the consensus rule to obtain what is expected to be the user's real opinion.

Our goal is to analyse and determine the user's opinion about the proposition $a$: "*The system will be sufficiently resistant against malicious attacks*". Let the user be called $A$ and the sub-personalities $A_1$ ... $A_6$ corresponding to each type of evidence described above. We then have

$$\begin{aligned}\omega_a^A &= (\omega_a^{A_1} \widetilde{\oplus} \omega_a^{A_2}) \oplus \omega_a^{A_3} \oplus \omega_a^{A_4} \oplus (\omega_a^{A_5} \widetilde{\oplus} \omega_a^{A_6}) \\ &= (\omega_a^{A_1 \mathrm{d}(A_2)} \overline{\oplus} \omega_a^{A_2 \mathrm{d}(A_1)}) \oplus \omega_a^{A_2 \mathrm{i}(A_1)} \oplus \omega_a^{A_3} \oplus \omega_a^{A_4} \oplus (\omega_a^{A_5 \mathrm{d}(A_6)} \overline{\oplus} \omega_a^{A_6 \mathrm{d}(A_5)}) \oplus \omega_a^{A_6 \mathrm{i}(A_5)}\end{aligned}$$

We have assumed that $\omega_a^{A_1}$ is totally dependent on $\omega_a^{A_2}$, and that $\omega_a^{A_5}$ is totally dependent on $\omega_a^{A_6}$. However, the reciprocal factors of dependence $\gamma^{A_2 A_1}$ and $\gamma^{A_6 A_5}$ will not be absolute, as the system reputation normally comes from more than the security evaluation, and the developer reputation results from many sources, including e.g. ISO 9000 certification.

**Numerical Example.** Let the trust resulting from security evaluation be called $\omega_a^{A_1}$. In Sec.5.1 it was computed as:

$$\omega_a^{A_1} = \{0.792,\ 0.100,\ 0.108\}.$$

In addition, let for example:

$$\begin{aligned}\omega_a^{A_2} &= \{0.820,\ 0.100,\ 0.080\} \\ \omega_a^{A_3} &= \{0.000,\ 0.000,\ 1.000\} \\ \omega_a^{A_4} &= \{0.000,\ 0.000,\ 1.000\} \\ \omega_a^{A_5} &= \{0.600,\ 0.100,\ 0.300\} \\ \omega_a^{A_6} &= \{0.850,\ 0.050,\ 0.100\},\end{aligned}$$

and let the dependence factors between security evaluation and system reputation be

$$\begin{aligned}\gamma^{A_1 A_2} &= 1.0, \text{ full dependence} \\ \gamma^{A_2 A_1} &= 0.5, \text{ partial dependence} \\ \gamma^{A_5 A_6} &= 1.0, \text{ full dependence} \\ \gamma^{A_6 A_5} &= 0.5, \text{ partial dependence.}\end{aligned}$$

$\omega_a^{A_3}$ and $\omega_a^{A_4}$ can here be omitted in the calculations because the opinions are totally uncertain. Consensus between the remaining opinions gives:

$$\begin{aligned}\omega_a^{A_1} \widetilde{\oplus} \omega_a^{A_2} &= \{0.825,\ 0.102,\ 0.073\} \\ \omega_a^{A_5} \widetilde{\oplus} \omega_a^{A_6} &= \{0.827,\ 0.061,\ 0.112\}\end{aligned}$$

$$\begin{aligned}\omega_a^A &= (\omega_a^{A_1} \widetilde{\oplus} \omega_a^{A_2}) \oplus (\omega_a^{A_5} \widetilde{\oplus} \omega_a^{A_6}) \\ &= \{0.864,\ 0.090,\ 0.046\}.\end{aligned}$$

In order to see the effect on the trust of for example a security incident, let $\omega_a^{A_3} = \{0.0,\ 0.9,\ 0.1\}$ as a result of a newly discovered security breach. We then get:

$$\begin{aligned}\omega_a^A &= (\omega_a^{A_1} \widetilde{\oplus} \omega_a^{A_2}) \oplus \omega_a^{A_3} \oplus (\omega_a^{A_5} \widetilde{\oplus} \omega_a^{A_6}) \\ &= \{0.825,\ 0.102,\ 0.073\} \oplus \{0.000,\ 0.900,\ 0.100\} \oplus \{0.827,\ 0.061,\ 0.112\} \\ &= \{0.611,\ 0.357,\ 0.032\}\end{aligned}$$

which shows that the overall trust is clearly influenced by the security incident.

In order to see the effect of security evaluation on the total trust, we will compute the overall trust without $\omega_a^{A_1}$ (and no security incident). In this case, that is with $\omega_a^{A_1} = \{0,\ 0,\ 1\}$, $\omega_a^{A_2}$ would be reduced according to its dependence with $\omega_a^{A_1}$. To find the new $\omega_a^{A_2}$, we have to solve for $\omega_a^{A_2 \mathrm{i}(A_1)}$ the equation:

$$\omega_a^{A_2} = \omega_a^{A_2 \mathrm{i}(A_1)} \oplus \omega_a^{A_2 \mathrm{d}(A_1)}$$

and set the $\omega_{a,\mathrm{new}}^{A_2} = \omega_a^{A_2 \mathrm{i}(A_1)}$. This is easily done using Eq.(8), giving $\omega_{a,\mathrm{new}}^{A_2} = \{0.759,\ 0.093,\ 0.148\}$. The new overall trust becomes:

$$\begin{aligned}\omega_a^A &= \omega_{a,\mathrm{new}}^{A_2} \oplus (\omega_a^{A_5} \widetilde{\oplus} \omega_a^{A_6}) \\ &= \{0.759,\ 0.093,\ 0.148\} \oplus \{0.827,\ 0.061,\ 0.112\} \\ &= \{0.852,\ 0.080,\ 0.068\}\end{aligned}$$

which is only slightly worse than the trust including the contribution from security evaluation. This indicates that trust resulting from security evaluation is not significantly greater than trust resulting from other sources. But of course, this all depends on the input values, and we invite the readers to try out trust values which in their opinion are meaningful.

This example shows that subjective logic can be used to provide a metric for trusted systems. As would be expected, security incidents can have a great impact in the overall trust, and trust resulting from security evaluation alone is not significantly more important than trust from other sources.

## 6 Conclusion

The presented model provides a metric and a method for reasoning about trust in the security of IT systems. We believe that the presented model can be integrated in standardisation efforts for security evaluation criteria for IT systems. This will make it possible to see the total effect of the assurance from the different types of evidence.

The biggest problem for the model to be useful is to be able to consistently determine opinions to be used as input. We have shown how opinions can be formally determined if the evidence can be analysed statistically. For situations where the evidence can only be assessed intuitively, we have proposed a simple questionnaire to be used as a guideline.

## References


[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.

[EC92] EC. *Information Technology Security Evaluation Criteria (ITSEC)*. The European Commission, 1992.

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

[ISO98] ISO. *Evaluation Criteria for IT Security (Common Criteria), documents N-2052, N-2053, N-2054*. ISO/IEC JTC1/SC 27, May 1998.

[JK98] A. Jøsang and S.J. Knapskog. A metric for trusted systems (short paper). In Reinhard Posch, editor, *Proceedings of the 15th IFIP/SEC International Information Security Conference*. IFIP, 1998.

[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.

[Jøs97] Audun Jøsang. *Modelling Trust in Information Security*. PhD thesis, Norwegian University of Science and Technology, 1997.

[JVLKV97] A. Jøsang, F. Van Laenen, S.J. Knapskog, and J. Vandewalle. How to trust systems. In L. Yngström, editor, *Proceedings of the 1997 IFIP/SEC International Information Security Conference*. IFIP, 1997.

[USD85] USDoD. *Trusted Computer System Evaluation Criteria (TCSEC)*. US Department of Defence, 1985.