A Formalization of Computational Trust

Çiçek Güven, Mike Holenderski, Tanır Özçelebi, Johan Lukkien
Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

***Abstract*—Computational trust aims to quantify trust and is studied by many disciplines including computer science, social sciences and business science. We propose a formal computational trust model, including its parameters and operations on these parameters, as well as a step by step guide to compute trust in a real application. We make a distinction between trust statements that aim to capture the truth of a dynamic phenomenon and trust statements that aim to capture the truth of a static phenomenon. We elaborate on how this difference should be reflected in trust computation. To this end we apply a dynamic base rate (prior trust) as an alternative to the widely used fixed base rate.**

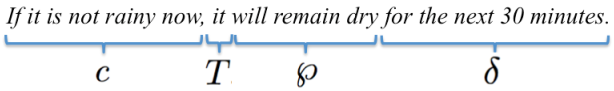
I. Introduction

In life we always have to make decisions to make progress. For example, when leaving home, we need to decide whether we should take an umbrella or not. This decision is not just about whether it is raining right now, it is also about whether it is going to be rainy in the rest of the day. Consider Alice, who has to walk 30 minutes from home to work every morning, and who makes such judgment call by peeking outside through a window to see if it is currently raining or not. In doing so, Alice intrinsically trusts that "The weather, if it is not rainy right now, will remain dry for the next half hour". When Alice experiences no rain during a walk towards work, her trust in the above statement will be reinforced. If Alice finds out, after a few days, that she is getting wet even though it was dry when leaving home, she will lose trust. Hence, Alice builds and loses trust in this statement based on direct observations on the truth of the statement.

Many times an assessment of the truth of a statement is not readily available and we make use of what we consider to be relevant knowledge, that is knowledge coming from observations and our understanding of the system, to make an assessment. For this case, the following example illustrates a different way in which trust can be built and lost. Assume that the chief of surgery Bob in a hospital is interested in the truth of the statement "Doctor Yang is a brilliant heart surgeon." Bob will then probably attend operations performed by Dr. Yang and, over time, will build trust or distrust in the statement. In doing so, in every operation Bob may look at factors like the difficulty of the operation, the behavior of Dr. Yang in the operation room and the outcome (failure, success, partial success). Thus, Bob does not directly observe the excellence, but instead observes clues about the phenomenon and has to make a judgment based on that. Note that Bob may value a partial success in a difficult operation more than a complete success in a trivial operation. Bob may disregard an operation with a negative outcome if he thinks that some other factor than the doctor's ability played a dominant role. On the other hand, if the chief of surgery was another person and not Bob, she could have approached this completely differently.

978-1-5386-3197-3/17/$31.00

As also illustrated by this example, trust is not (necessarily) a property of the phenomenon under consideration like probability, but rather a subjective interpretation of it. It is subjective, i.e. different trustors can build and lose trust in different ways, even though trust is most of the time tied to observations and even experiments. In [4] we defined trust as "*the degree of justifiable belief a trustor has that, in a given context, a trustee will live up to a given set of statements about its behavior.*" Computational trust provides a framework for an assessment by the trustor of the truth of a *trust statement* P that looks like "The trustee T satisfies a predicate $\wp$ within context c in the time period δ". Hence, a trust statement is given by the tuple $P = (T, \wp, c, \delta)$. The notation can be mapped to an example trust statement as follows.



In practical settings, observing the truth of P directly may not be possible and only clues about the truth of P may be available through observations, as illustrated by the hospital example.

In this paper we present a formalization of computational trust, that is applicable in such practical settings. The proposed computational trust model describes the details of the individual steps that need to be taken in computing trust for practical problems and mimics the properties of real-life trust.

The paper is organized as follows. Section II lays out the requirements for the proposed trust model, its model parameters and operations on these parameters. Section III gives a short background of Bayesian inference. Section IV provides a review of the literature. Section V discusses the concepts of the proposed computational trust model; trust statement, evidence, evidence update and uncertainty in detail. Section VI gives examples of the proposed trust model. Section VII concludes the paper and gives future research directions.

II. Requirements for Computational Trust Model

The proposed computational trust model mimics the following properties of real-life trust.

- Trust is maintained by a trustor on the truth of a statement describing a phenomenon whose subject is a trustee, tied

to a specific context and time frame. The trust statement may describe a fixed phenomenon (e.g. fairness of a dice) or a dynamic phenomenon (e.g. the excellence of Dr. Yang) which may change over time.
- The trustor relies on direct or indirect observations on the truth of such statement to build *evidence* and reevaluate trust. Observed successes and failures are not necessarily treated equally in updating the evidence. For some trust statements failures count heavier than successes, while for others it is the other way around.
- A lack of evidence (e.g. initially) implies *uncertainty*.
- The trustor has an initial (default) level of trust that loses importance with increasing evidence. Eventually, the trust converges to a value independent of the initial state.
- The impact of a given observation on trust may change over time: it may gradually become less relevant as time goes by. Alternatively, it may also stick forever.
- Trust is subjective. Different trustors may have different observations on the same phenomenon or they may even interpret the same observations differently; e.g. what seems like a success to one trustor may seem like a failure to another trustor. As a result they may develop different levels of trust. The interpretation of uncertainty by the trustor is also subjective.

Based on these properties, the proposed computational trust model shall have the following parameters and the operations on these parameters.

- The model quantifies trust of a trustor A on the truth of a trust statement $P = (T, \wp, c, \delta)$. In practice, the context c and the time interval δ may be implicit.
- The inputs to this model are domain knowledge and measurements. The domain knowledge specifies the types of measurements that are relevant for determining the truth of P and helps to evaluate these measurements. A trustor is free to choose any one of these measurement types. Measurements $(m^A_{P,i})_{i=1}^N$ (i: measurement index) are not necessarily in the form of direct experiments on the truth of P. They can also be on parameters that give clues about the truth of P. Across different trustors the actual values of measurements and even the measurement type chosen (one of the alternatives given by domain knowledge) may vary.
- Measurements $(m^A_{P,i})_{i=1}^N$ are mapped to observations $(o^A_{P,i})_{i=1}^N$ about P. Such observation is subjective to trustor A and aims to determine to what extent the measurement hints at the truth of P. An observation is a real number in the range $[0, 1]$, with 1 representing a strong hint.
$$o^A_{P,i} = V^A(m^A_{P,i})$$
- A history of observations $(o^A_{P,i})_{i=1}^N$ is mapped to what we call positive evidence $p^A_{P,i} \in \mathbb{R}_0^+ = [0, \infty)$, and negative evidence $n^A_{P,i} \in \mathbb{R}_0^+$. Upon the i^{th} observation the pair $(p^A_{P,i}, n^A_{P,i})$ is a function of $p^A_{P,i-1}$, $n^A_{P,i-1}$ and $o^A_{P,i}$. The amounts of positive and negative evidence taken from an observation are subjective, i.e. these may differ across trust models of individual trustors.
- In cases where the state (of the truth of P) is known to be dynamic, a lack of observations over a given period of time increases uncertainty.
- Before any measurements, the trustor A has a prior trust a^A regarding the trust statement. The weight W^A of a^A in determining trust is also subjective. The inputs to the trust computation are $p^A_{P,i}$, $n^A_{P,i}$, a^A and W^A.
- The output of the computational trust model after the i^{th} observation is a trust value $tv^A_{P,i}$, a real number in [0,1]. It is dynamically adjusted with every piece of new evidence (based on observations on new measurements).

The model we propose satisfying these properties is a Bayesian trust model, i.e. trust is considered as a subjective probability and computed based on Bayesian inference. Some rules and parameters are up to the subjective choice of the trustor, and we provide examples for those.

In this paper, the trust statements of interest are those that can not be partitioned into simpler trust statements with the available information. Combining trust, or combining evidence coming from different sources is out of the scope of this paper.

III. Background: Bayesian inference

In this paper, a trust value is seen as subjective probability and is computed by making use of Bayesian inference. In this section, we briefly describe what is relevant for computational trust but will not be going too much into the details of Bayesian probability theory as treated in [12]. In Bayesian probability theory, the posterior probability of a random event is the conditional probability after taking the evidence into account. Trust value in favor of one of the outcomes of a binomial event can be seen as the expected value of the posterior probability distribution for this binomial event given a prior trust.

Consider an experiment with 2 outcomes, success S and failure F. Let the probability of success be $p(S) = \theta \in [0, 1]$, which also can be treated as a random variable with a certain distribution. The parameter θ needs to be estimated based on evidence D, a sequence of trials. According to Bayes' rule, the posterior distribution can be obtained from the prior distribution $p(\theta)$ and the likelihood via the equation [12]:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta')p(\theta')d\theta'} \tag{1}$$

where $p(\theta|D)$ stands for the likelihood. Sometimes, prior and posterior distributions are of the same type. This is called conjugacy between the likelihood and the prior distribution. For a binomial event, the likelihood function is the binomial likelihood function $\mathcal{L}(\theta|n, k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$, where n and k denote the numbers of trials and successes, respectively. The function $\mathcal{L}(\theta|n, k)$ and a beta prior $Beta(\alpha, \beta)$ are conjugates with the posterior $Beta(k+\alpha, \beta+n-k)$. A beta distribution is therefore convenient, as it gives intuition about how the likelihood function affects a prior distribution. $Beta(\alpha, \beta)$ is

a continuous distribution on $[0, 1]$ of a random variable Z, whose probability density function (PDF) is:

$$f(z; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} z^{\alpha-1}(1-z)^{\beta-1}$$

where α and β are two positive shape parameters and $B(\alpha, \beta)$ is a constant to ensure the PDF integrates to 1. The expected value of the beta distribution with random variable Z and parameters α, β is known to be:

$$E[Z] = \int_0^1 z f(z; \alpha, \beta) dz = \frac{\alpha}{\alpha + \beta} \tag{2}$$

Hence the expected value of $Beta(k+\alpha, \beta+n-k)$ is $\frac{k+\alpha}{n+\alpha+\beta}$.

For a positive constant γ, for which $\alpha' = \gamma\alpha$, $\beta' = \gamma\beta$, the expected value of the Beta distribution with shape parameters (α', β') will also be $\frac{\alpha}{\alpha+\beta}$. For larger values of γ, the distribution will be narrower (with less variance around $\frac{\alpha}{\alpha+\beta}$).

IV. Trust Literature

Trust has been the subject of many scientific studies not only in computer science, but also in different disciplines such as social sciences [6] and philosophy [3].

Trust can be understood as "*an attitude of an agent who believes that another agent has a given property*" [7]. Thus it is not a property of just the trustee, but of the trustor and the trustee together. A widely used definition of Gambetta is [8]: "*Trust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects its own action*". The perception of trust as subjective probability is common, which is also the standing point here: we see computational trust as a data driven input-output mechanism, which provides a prediction or expectation of subjective probability. Some of the other approaches are fuzzy approaches, and multi-criteria decision making based approaches, and these are compared in [1].

There have been several attempts to formalise trust as a computational concept. Some iexamples are [2], [18], [19], [22], and [25]. Audun Jøsang studied trust in great detail through the years ([14], [15], [16], [17]) via subjective logic, which he founded, and collected his work in a recent book [18]. Different trust formalization methods serve different purposes. Trust is a psychological phenomenon mostly and it is subjective. Hence, it is hard to map it to physical process and analyze objectively, and come up with a natural benchmark for formal trust models [18]. The inputs to a trust model can be data coming from measurements, experiment outputs, experiences, referrals or a collection of ratings. For example, in social sciences, trust is measured via rating systems or scales [24]. A similar approach can be taken, e.g., for trust assessment in social networks [23].

We do not study general trust, i.e. trust has a scope determined by a predicate and an associated context in this paper. General trust and situational trust are distinguished also in the literature. For example, [22] introduces a top-down approach where situational trust is obtained from general trust, but weighted according to utility and importance factors. Our proposed approach is to collect the relevant evidence for a defined trust scope, which is bottom-up.

Evindence continuity, i.e. not being limited to a set of discrete values of evidence was studied in [14], [15], [25].

Trust is not static, hence a computed trust value should be revised based on new evidence. The relationship between evidence and time is critical and not all available evidence may be equally relevant at all times. This is described as *aging of evidence* in [25], as *forgetting* in [14], as *longetivity* in [15], [16] and was reflected to trust computation in [4]. In this paper we give a step by step guide to computational trust and relate the update rules to the nature of the predicate; we distinguish the cases where aging of evidence is relevant from the cases where evidence does not age.

Augmenting trust with machine learning is a trending approach (see [28], [30]) for example in social networks or distributed systems. In [20], neural networks are proposed as a method to compute trust in multi-agent systems, and this is studied in [21]. In peer to peer systems, Bayesian network based trust management models can be used, and in [9], a distributed learning method is used to model the behaviors of peers based on their behavior history. There are recent studies about computational trust in the field of trust in social networks [10], [26] and in the Internet of Things (IoT) [11]. Trust in ubiquitous computing is often studied via a model called CertainTrust [25]. In [13], behavioral computer science is considered as the domain to apply computational trust where Artificial Intelligence (AI), IoT and behavioral sciences meet. With the advances in IoT and AI, there is more interaction among intelligent machines or between humans and intelligent machines. Therefore, trust across machines, trust of humans to machines and vice versa are of potential interest.

V. A Formalization of the Proposed Computational Trust Model

The proposed computational trust model is depicted in Figure 1. This model is an extension of existing Bayesian trust models ([18], [25], [27], etc.). This section describes the steps towards trust value computation.

A. Step 1: Definition of a Trust Statement

The trust statement P is either true or false and its subject is the trustee T. The predicate $\wp$ narrows down the scope of trust and the relevant domain knowledge, and is associated with a time horizon δ of validity for trust. This time horizon follows the time instance t_i of last observation $o^A_{P,i}$ and is given by

$$\delta = [t_i + \Delta_1, t_i + \Delta_2]$$

where Δ_1 and Δ_2 are non-negative real numbers ($\Delta_2 \geq \Delta_1$). When Δ_2 is greater than zero we say that the trust statement P *looks to the future*. If both Δ_1 and Δ_2 are zero we say that P *looks to the present*.

We distinguish the trust statements where the nature of the phenomenon remains constant (see example 1 in Section VI)

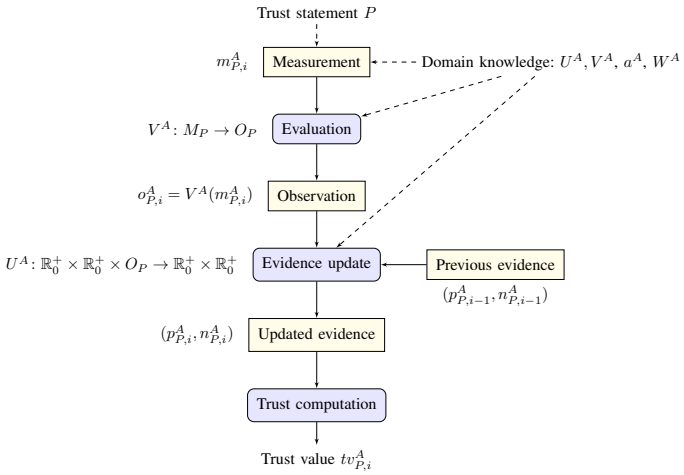


Fig. 1: Recipe for trust value computation

from the cases where it may change over time (see example 2 in Section VI). In the former (a type-1 trust statement), observations throughout the history are equally relevant. In the latter (a type-2 trust statement), the chronological order in the observation history (memory) is important and contributions of individual observations to the evidence may vary. For example, observations may become less relevant as they get older, emphasizing the impact of the most recent observations on the overall evidence. Different trustors can interpret the domain knowledge differently to determine their use of memory in computing trust accordingly. Typical examples of type-2 trust statements are regarding time series data whose values need to remain inside (or outside) critical intervals.

The trust statement can refer to a phenomenon that can be directly observed or to something that has to be inferred (such as occupancy detection without being able to see the room using, e.g. temperature sensors). In the former, the trust statement refers to the future, otherwise the problem becomes trivial. In the latter, the trust statement can refer to the present or to the future (both cases being non-trivial), inferring from relevant observations.

B. Step 2: Evaluating Measurements

The next step in the trust computation is an evaluation of measurements; i.e. a mapping from measurements $(m^A_{P,i})_{i=1}^N$ to observations $(o^A_{P,i})_{i=1}^N$. For type-2 predicates where the nature of the phenomenon is changing it is typical to schedule measurements at time intervals of equal length. However, in practice some of these measurements may be missing (e.g. due to measurement failures), as well as the corresponding observations. Missing observations should lead to reduced (positive and negative) evidence and contribute to uncertainty. For example, we may have a high level of trust that Dr. Yang is an excellent surgeon now, but we should not maintain the same level of trust over several years if we have no observations in between.

The value $o^A_{P,i}$ quantifies success for the i^{th} observation with the mapping $V^A\colon M_P \to O_P$ where M_P is the measurement space and $O_P \subseteq [0,1] \cup \{-1\}$ is the observation space. Here $o^A_{P,i} = 1$ means absolute success and $o^A_{P,i} = 0$ means absolute failure for that observation. Any value of $o^A_{P,i}$ between 0 and 1 refers to a partial success. For some measurements, for example coin tosses, the measurements are only absolute successes or failures. For some others, for example in case of movie ratings, there is partial success. For practicality, range is extended to include $\{-1\}$, representing a missing observation. Missing the i^{th} observation gives $o^A_{P,i} = -1$.

Considering a trust statement P and its negation $\neg P$, when the same set of measurements are taken into account ($m^A_{P,i} = m^A_{\neg P,i}$), there is symmetry around positive and negative evidence. For a successfully performed measurement the following holds:

$$o_{P,i} = 1 - o_{\neg P,i}$$

For example consider the trust statement $\neg P$ that "The machine is *not* working properly". The measurement $m^A_{P,i}$ is the warmth of the machine where a high temperature is a sign of malfunctioning. One can say if the machine is sizzling, the high temperature measurement leads to a success observation for $\neg P$, i.e. $o_{\neg P,i} = 1$. On the other hand, for the statement P that "The machine is working properly", the same observation is considered a failure observation, i.e. $o_{P,i} = (1 - o_{\neg P,i}) = 0$.

C. Step 3: Collecting and Updating Evidence

Observations $(o^A_{P,i})_{i=1}^N$ are mapped to positive evidence $p^A_{P,i} \in \mathbb{R}_0^+$, and negative evidence $n^A_{P,i} \in \mathbb{R}_0^+$. Their values are updated using a trustor-specific update rule (a.k.a. update function) U^A after each observation.

In [4], different types of update rules were discussed, for discrete and continuous observations and a list of requirements were also given. Here, we underline the fact that the requirements depend on the nature of the phenomenon we are making a trust assessment over. For type-1 trust statements, all the evidence is equally relevant. For type-2 trust statements, more recent observations will typically be more relevant.

The update rule $U^A\colon \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times O_P \to \mathbb{R}_0^+ \times \mathbb{R}_0^+$ updates evidence taking the observations into account. It is designed to make sure that more relevant observations contribute more to the computed evidence (so evidence is not necessarily monotonically increasing). The update rule can also incorporate the fact that there are missing observations or initial lack of knowledge into trust computation.

$$(p^A_{P,i}, n^A_{P,i}) = U^A(p^A_{P,i-1}, n^A_{P,i-1}, o^A_{P,i})$$

$$(p^A_{P,0}, n^A_{P,0}) = (0,0), \quad p^A_{P,i} \geq 0, \quad n^A_{P,i} \geq 0 \, \forall i$$

Update rules are subjective, i.e. they vary across trustors and their computational trust models. The following is a list of some cases where different update rules are relevant.

1) Updates for type-1 trust statement: The update rule should satisfy the following:

- Every observation is equally important and evidence does not wear out. That is, the times or the order of the observations do not matter.

- Missing observations do not imply additional uncertainty.
- Positive and negative evidence are monotonically increasing and can not be negative.

An update rule capturing this is as follows:

$$p^A_{P,i} = \begin{cases} p^A_{P,i-1} & o^A_{P,i} = -1 \\ p^A_{P,i-1} + o^A_{P,i} & \text{otherwise} \end{cases}, \quad n^A_{P,i} = \begin{cases} n^A_{P,i-1} & o^A_{P,i} = -1 \\ n^A_{P,i-1} + (1 - o^A_{P,i}) & \text{otherwise} \end{cases} \tag{3}$$

In case of discrete output, the total evidence is the number of recorded measurements, and positive evidence is the number of success outcomes among these. An example trust statement for a given fixed coin is "The next toss of the coin will yield heads", for which it is fair to assume that the nature (i.e. the bias) of the coin does not change over time.

2) Update for type-2 trust statements: The update rule should satisfy the following:

- The order of observations matters.
- More recent observations should have more impact on the (relevant) evidence.
- Missing observations imply additional uncertainty and shall reduce the positive evidence and the negative evidence at the same rate.
- Positive and negative evidence are not necessarily monotonically increasing and can not be negative.

Various update rules can capture these properties. This flexibility is discussed in [4] where a few examples of update rules are provided for discrete observations and continuous observations. The following update rule from [4] for type-2 trust statements can be seen as a generalization of the update rule given in (3) for type-1 trust statements. Let ζ, γ, ψ be constants in $[0, 1]$.

$$p^A_{P,i} = \begin{cases} \zeta \cdot p^A_{P,i-1} & o^A_{P,i} = -1 \\ \psi^{1-o^A_{P,i}} \cdot p^A_{P,i-1} + o^A_{P,i} & \text{otherwise} \end{cases}, \quad n^A_{P,i} = \begin{cases} \zeta \cdot n^A_{P,i-1} & o^A_{P,i} = -1 \\ \gamma^{o^A_{P,i}} \cdot n^A_{P,i-1} + (1 - o^A_{P,i}) & \text{otherwise} \end{cases} \tag{4}$$

A similar but simpler update rule that scales evidence so that more recent evidence weighs more is given in [16].

3) Uncertainty: A lack of evidence implies uncertainty. The trustor uses the domain knowledge to determine its (subjective) way to build evidence. Regardless of how they build evidence, what remains constant across trustors is that increasing the amount of positive and negative evidence at hand by having more observations shall increase the trustor's confidence in its trust computation. We use Josang's definition to quantify uncertainty $u^A_{P,i} \in (0, 1]$:[1]

$$u^A_{P,i} = \frac{W}{W + p^A_{P,i} + n^A_{P,i}}$$

[1] W is determined by the two shape parameters of the prior beta distribution, which are positive. For more information on Bayesian statistics see [12].

Obviously there is lack of evidence at the beginning when there are no observations. This leads to absolute uncertainty $p^A_{P,0} = 0$, $n^A_{P,0} = 0$, $u^A_{P,0} = 1$. If the evidence taken by the trustor from individual observations is low (i.e. if a lot of observations are needed to gather reasonable evidence as far as trustor A is concerned), then the weight W^A would play a significant role in trust computation for a long time. On the other extreme, if each observation gives a lot of evidence (as far as this trustor is concerned), W^A's impact on the trust value computation is diminished after a small number of observations.

Failure to make new observations can also increase uncertainty, more so in a system where the most recent observations matter the most, or the nature of the phenomenon in the trust statement is changing over time (like well-being of a patient). This is reflected to uncertainty via the evidence update, since uncertainty is negatively correlated with the total evidence.

Confidence is tied to uncertainty: High uncertainty implies lack of confidence. The relation between uncertainty $u^A_{P,i}$ and confidence value $cv^A_{P,i}$ is defined as follows [4]:

$$cv^A_{P,i} = 1 - u^A_{P,i}$$

D. Step 4: Trust Computation

Bayesian inference as described in Section III can be used to calculate trust over time, under the light of new evidence.

More precisely, in equation (1), $p(\theta)$ represents the distribution for random variable $z = p(S)$. For any prior trust value a (initial trust without any evidence, also called the *base rate*), a beta distribution $Beta(\alpha, \beta)$ with expected value a^A can be taken as the prior distribution so that, the conjugacy between beta distribution and binomial likelihood can be used to calculate the expected value of the posterior distribution i.e. trust. The parameter W^A, namely weight relating a^A with (α, β) as in $\alpha = a^A \cdot W^A$ and $\beta = (1 - a^A) \cdot W^A$, determines the strength of the prior trust. When W^A is chosen larger, the corresponding prior (beta) distribution will be narrower, which can be interpreted as a stronger prior trust. When the prior trust is strong, more evidence is necessary to change it. In other words, for large values of W^A (hence of α and β), if k observations out of n point to positive evidence for the truth of P, it will take longer for $\frac{k+\alpha}{n+\alpha+\beta}$, the expected value of posterior distribution (computed trust) to deviate from a^A.

In [18], Bayesian inference is used to update the prior probability with evidence to calculate a "projected probability". Projected probability is defined over what we call an *opinion*, the four-tuple (b, d, u, a) - belief, disbelief, uncertainty, and prior trust, where $b + d + u = 1$. In [18] projected probability for an opinion is defined as $b + u \cdot a$ and it is assumed that projected probability is equal to the expected value of the posterior distribution. This implies the following bijection between (p_P, n_P), and (b, d, u, a) where $p_P = k$, $n_P = n - k$:

$$b = \frac{k}{n + \alpha + \beta} = \frac{p_P}{p_P + n_P + W}$$

$$u = \frac{\alpha + \beta}{n + \alpha + \beta} = \frac{W}{p_P + n_P + W}$$

These equations indicate that uncertainty u as defined by Jøsang decreases as more evidence is collected. The value of u is always positive, since W is tied to positive shape parameters α, β. In [4], the approach described above is followed, where opinions are generated based on positive and negative evidence and trust is computed via the formula $b + u \cdot a$:

$$tv^A_{P,i} = \frac{p^A_{P,i} + a^A \cdot W^A}{p^A_{P,i} + n^A_{P,i} + W^A} \tag{5}$$

In this paper, we use equation (5) to quantify trust as well. However, there are some drawbacks of the approach in (5) as is. When there is lack of observations, the evidence should be decreased if the trust statement is of type-2. This is done for example in [4] by scaling positive and negative evidence. Hence, the total evidence heavily depends on relevant (more recent) evidence. That implies, when there is no new evidence for long enough, the positive and negative evidence will essentially approach zero, and the trust value will approach back to a^A. However, most likely the initial trust value a^A is not relevant anymore. Moreover, if a^A is relatively large, that might even imply an increase in trust in the absence of new evidence in some cases, which is unacceptable. In [4], confidence ($cv^A_{P,i} = 1 - u^A_{P,i}$) is used in combination with the computed trust in the decision making process. This approach in practice solves the problem of giving too much weight to the prior trust when there is lack of evidence.

An additional approach that we propose in this paper is to adjust the trust model to repeatedly update the prior trust, for example by taking the computed trust to be the prior trust for future observations. Note that every time the prior trust is updated the evidence needs to be reset to zero in order to prevent double counting. Using dynamic base rates for reputation systems was earlier studied in [29] without going into trust computation. One challenge then is to determine how often the prior trust should be updated. For simplicity W can be kept constant across updates of the prior. This is illustrated in Example 2 in Section VI.

The answer to the question "how much evidence or how many observations are enough?" can be considered as a result of domain knowledge in combination with the subjective approach in evidence building. The answer mainly depends on W^A and the amount of evidence collected from observations, hence on how the impact of the prior trust decays as more observations are made. This can be quantified by careful choices of W and evidence weights and can be used together with an appropriate threshold setting in decision making.

VI. Evaluation of the Proposed Trust Formalization on Examples

Example 1. P_1: "The next toss of the given (fixed) coin will yield heads."

Consider a trustor A who wants to compute trust to P_1. For this trust statement and the binary domain $\mathbb{X} = \{x, \overline{x}\}$, assume that x corresponds to heads and $\overline{x}$ corresponds to tails. Context and time are irrelevant here.

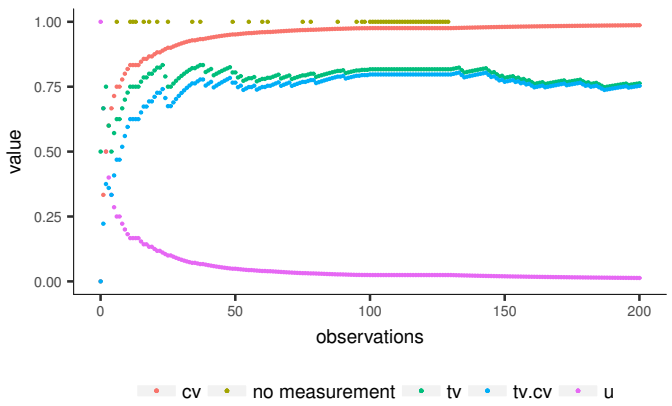


Fig. 2: Behaviors of trust, uncertainty, confidence, and trust times confidence over observation index.

Note that this is a type-1 trust statement and measurements are direct experiments in the form of coin tosses. Observations in this case are the same as measurements (evaluations are implicit). For evidence update, an update rule as in equation (3) can be used. That is, evidence update is just updating the total number of positive and negative observations so far, and the trust computation is as in equation (5).

The trustor can assume nothing is known about the coin initially. Thus any value in $[0, 1]$ can be considered equally likely (i.e. a uniform distribution) before measurements start. That implies, at the beginning he has a prior trust $a^A = 0.5$ with $W^A = 2$. If the trustor has a stronger bias for a certain value of a^A, he can choose a larger W^A and a beta distribution with shape parameters $\alpha = a^A \cdot W^A, \beta = (1 - a^A) \cdot W^A$.

In our simulation, an unfair coin whose probability of heads is 0.7 is chosen and some noise is added to the data (D_i denotes the set of first i observations); some tosses are missed by the trustor. In Figure 2, the behaviors of computed trust, uncertainty and confidence over time are shown. When there is no measurement, trust, uncertainty or confidence remain constant. The figure also shows the trust-confidence product, which can be an important metric in decision making when combined with thresholding. In Figure 3 the corresponding beta distributions at the beginning, after 30, 100 and 200 observations are shown. Starting from a prior distribution over $\theta = p(S)$ that is uniform with $E(p(\theta)) = 0.5$, increasing number of observations implies for $tv^A_{P,i} = E(p(\theta|D_i))$ to be closer to 0.7 and with smaller variance around the mean (i.e. less uncertainty and more confidence).

Example 2. P_2: "A given room is occupied."

Consider a room where we measure the current CO_2 level with a sensor frequently. There is domain knowledge that as people generate CO_2 someone's presence in the room shall lead to higher levels of CO_2. For low levels of CO_2, however, a decrease or increase is not indicative, since noise created by external sources may play a dominant role.

The evaluation could be based on levels of CO_2 combined

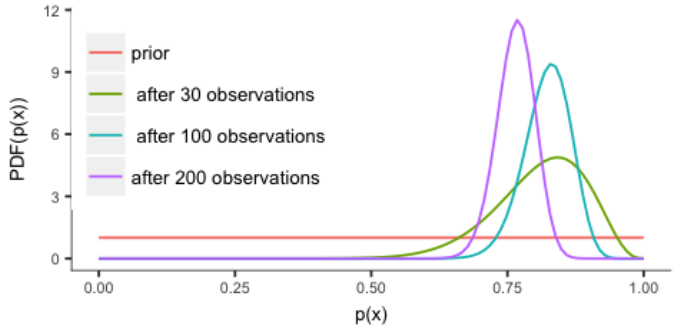


Fig. 3: The four graphs correspond to the uniform prior distribution and the updated distributions after 30, 100 and 200 observations in Example 1.

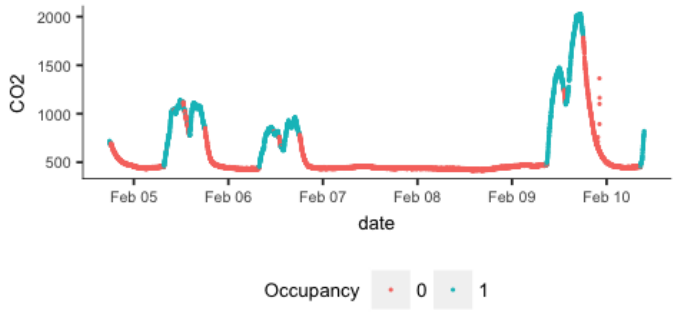


Fig. 4: CO_2 level (in parts per million by volume - ppm) versus occupancy over a period of 6 days in Example 2.

with its trend (behavior) as given by the derivative of the measured CO_2 levels: increasing levels of CO_2 support the truth of the predicate. Hence, $o^A_{P,i}$ can be a real number in $[0, 1]$ tied to the CO_2 trend: close to or equal to 0 when CO_2 level is low and/or when CO_2 level is decreasing.

The data used in this is a set of occupancy detection data from the UCI Machine Learning Repository [5]. It consists of multivariate time series data consisting of CO_2, temperature, humidity, and light measurements with a period of one minute[2]. The data are meant as input for binary classification of occupancy in a room. The ground truth data (labels corresponding to data samples) are available thanks to annotation of camera recordings (photos taken one minute apart). The occupancy parameter stands for the ground truth, where "1" stands for occupancy, "0" stands for no occupancy. Figure 6 shows how CO_2 behaves over time. The graph is colored based on ground truth occupancy. The following rule is used to obtain observation values from measurements:

$$o^A_{P_2,i} = \begin{cases} 0 & \text{if } m^A_{P_2,i} \leq 500, \\ 0 & \text{if } m^A_{P_2,i} > 500 \text{ and } m^A_{P_2,i} < m^A_{P_2,i-1} \\ -1 & \text{no measurement captured} \\ 1 & \text{otherwise} \end{cases}$$

[2]It consists of three data subsets of the same type and we employ the data subset named *training data*.

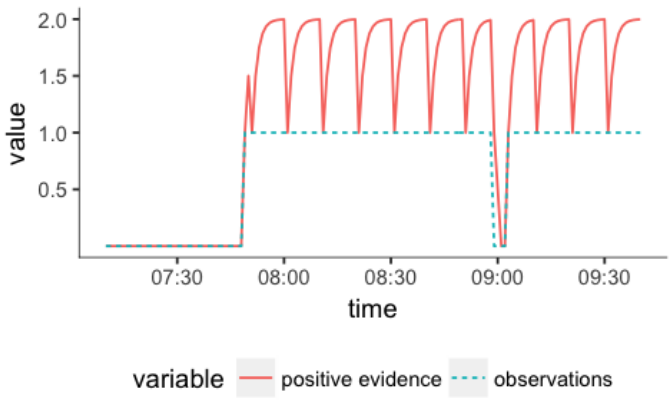


Fig. 5: Observations, (positive) evidence and trust over time in Example 2, showing only a few hours of data for readability.

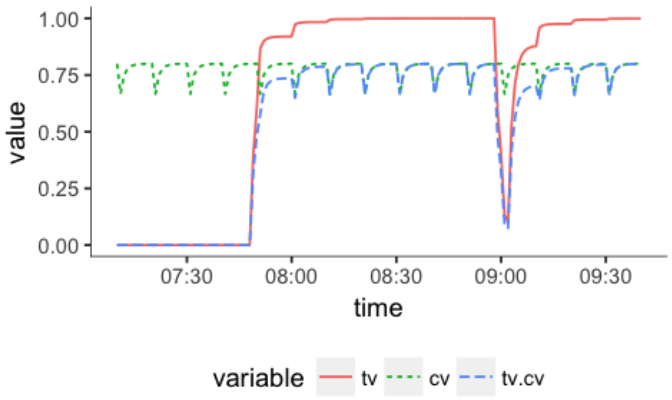


Fig. 6: Trust and confidence values over time in Example 2.

In this experiment we employ the proposed dynamic prior trust. After a certain number of observations (a window length L), the positive evidence is reset to zero, and the trust value computed after the previous observation is taken as the new prior trust. We use the update rule in (6), which is very similar to (4), and choose $L = 10$, $\zeta = \psi = \gamma = 0.5$. Initially, the prior trust is set to 0.5.

$$p^A_{P,i} = \begin{cases} \zeta \cdot p^A_{P,i-1} & o^A_{P,i} = -1 \\ o^A_{P,i} & i \equiv 1 (mod\, 10) \\ \psi \cdot p^A_{P,i-1} + o^A_{P,i} & \text{otherwise} \end{cases},$$
$$n^A_{P,i} = \begin{cases} \zeta \cdot n^A_{P,i-1} & o^A_{P,i} = -1 \\ 1 - o^A_{P,i} & i \equiv 1 (mod\, 10) \\ \gamma^{o^A_{P,i}} \cdot n^A_{P,i-1} + (1 - o^A_{P,i}) & \text{otherwise} \end{cases} \quad (6)$$

Typically, computed trust and confidence values are inputs to a decision making task, which can, e.g., compare $tv \cdot cv$ to a threshold. We used 0.6 as the threshold value. In our decision making $tv \cdot cv$ values greater than 0.6 indicate occupancy in the room. Comparing to the ground truth with our trust based occupancy detection on CO_2 data gives 89% accuracy, 98% precision and 52% recall rate.

VII. Conclusion and Future Work

In this paper, we provided a formalization of computational trust. It is an extension of existing Bayesian trust models (e.g. [18], [25], [27]). Our first contributions are in the evaluation and update steps, where measurements are evaluated as observations, and observations are used to update evidence in accordance with the expected behavior of the trust curve encoded into update rules. Using different update rules gives freedom to reflect the anticipated behavior to the resulting trust curve. We provided a few example update rules. We also proposed a dynamic prior trust (base rate) such that in the absence of evidence, trust is not based on the initial prior trust, which is potentially not relevant anymore. Moreover, we listed the properties of a computational trust model and steps needed to be taken for trust computation. For a given trust statement, a trust value can be computed and updated over time by employing these steps and the domain knowledge (relevant measurements, a prior belief when past is relevant, an understanding of how the trust curve should behave).

Computed trust is usually an input to a decision task. Our ultimate goal is to position trust in decision making, that is to come up with a decision making framework.

Multiple trust statements can be relevant to a decision to be taken. In that case, most likely a combined trust value is needed. For example, the well-being of a system may require trust assessments of its parts and an interpretation of how these are related. Moreover, for a given trust statement different measurements could be relevant and there may be a need to combine the evidence from these. How these combinations should be performed is an area of potential research.

Another potential contribution for the future is in the learning of the update rules, for example, via machine learning. At the moment, the proposed model does not adjust itself for improving the update rules and these are static. Ideally, the trustor shall provide an initial update rule according to an expected trust behavior (a trust curve over time) and the model shall try to improve it over time.

Acknowledgment

This work is a result of the MANTIS project funded by H2020 ECSEL under grant agreement No 662189.

References

[1] Ashtiani, M. and Azgomi, M.A., 2016. A hesitant fuzzy model of computational trust considering hesitancy, vagueness and uncertainty. Applied Soft Computing, 42, pp.18-37.

[2] Abdul-Rahman, A., 2005. A framework for decentralised trust reasoning (Doctoral dissertation, University of London).

[3] Buchak, L., 2014. Belief, credence, and norms. Philosophical Studies, 169(2), pp.285-311.

[4] Bui, V., Verhoeven, R. and Lukkien, J., 2014, June. Evaluating trustworthiness through monitoring: The foot, the horse and the elephant. In International Conference on Trust and Trustworthy Computing (pp. 188-205). Springer International Publishing.

[5] Candanedo, L.M. and Feldheim, V., 2016. Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models. Energy and Buildings, 112, pp.28-39.

[6] Coleman, J.S. and Coleman, J.S., 1994. Foundations of social theory. Harvard university press.

[7] Demolombe, R., 2001. To trust information sources: a proposal for a modal logical framework. In Trust and deception in virtual societies (pp. 111-124). Springer Netherlands.

[8] Gambetta, D., 2000. Can we trust trust. Trust: Making and breaking cooperative relations, 13, pp.213-237.

[9] Ghaderzadeh, A., Kargahi, M. and Reshadi, M., 2017. DisTriB: Distributed trust management model based on gossip learning and Bayesian networks in collaborative computing systems. Journal of Advances in Computer Research, 8(4), pp.37-57.

[10] Golbeck, J.A., 2005. Computing and applying trust in web-based social networks (Doctoral dissertation).

[11] Guo, J., Chen, R. and Tsai, J.J., 2017. A survey of trust computation models for service management in internet of things systems. Computer Communications, 97, pp.1-14.

[12] Hoff, P.D., 2009. A first course in Bayesian statistical methods. Springer Science & Business Media.

[13] Johansen, C., Pedersen, T. and Jøsang, A., 2016, July. Towards Behavioural Computer Science. In IFIP International Conference on Trust Management (pp. 154-163). Springer International Publishing.

[14] Jøsang, A. and Ismail, R., 2002, June. The beta reputation system. In Proceedings of the 15th bled electronic commerce conference (Vol. 5, pp. 2502-2511).

[15] Jøsang, A., Luo, X. and Chen, X., 2008, June. Continuous ratings in discrete bayesian reputation systems. In IFIP International Conference on Trust Management (pp. 151-166). Springer US.

[16] Jøsang, A., Hird, S. and Faccer, E., 2003, May. Simulating the effect of reputation systems on e-markets. In International Conference on Trust Management (pp. 179-194). Springer Berlin Heidelberg.

[17] Jøsang, A. and Quattrociocchi, W., 2009, August. Advanced Features in Bayesian Reputation Systems. In TrustBus (Vol. 5695, pp. 105-114).

[18] Jøsang, A., 2016. Subjective Logic: A Formalism for Reasoning Under Uncertainty. Springer.

[19] Krukow, K., 2006. Towards a theory of trust for the global ubiquitous computer. University of Aarhus, PhD Thesis.

[20] Lu, G., Lu, J., Yao, S. and Yip, Y.J., 2009. A review on computational trust models for multi-agent systems. The open information science journal, 2, pp.18-25.

[21] Lu, G., 2011. Neural Trust Model for Multi-agent Systems (Doctoral dissertation, University of Huddersfield).

[22] Marsh, S.P., 1994. Formalising trust as a computational concept.

[23] Mui, L., 2002. Computational models of trust and reputation: Agents, evolutionary games, and social networks (Doctoral dissertation, Massachusetts Institute of Technology).

[24] Rempel, J.K., Holmes, J.G. and Zanna, M.P., 1985. Trust in close relationships. Journal of personality and social psychology, 49(1), p.95.

[25] Ries, S., 2009. Trust in ubiquitous computing (Doctoral dissertation, Technische Universität).

[26] Sherchan, W., Nepal, S. and Paris, C., 2013. A survey of trust in social networks. ACM Computing Surveys (CSUR), 45(4), p.47.

[27] Teacy, W.T.L., Patel, J., Jennings, N.R. and Luck, M., 2006. Travos: Trust and reputation in the context of inaccurate information sources. Autonomous Agents and Multi-Agent Systems, 12(2), pp.183-198.

[28] Traverso, G., Cordero, C.G., Nojoumian, M., Azarderakhsh, R., Demirel, D., Habib, S.M. and Buchmann, J., 2017. Evidence-Based Trust Mechanism Using Clustering Algorithms for Distributed Storage Systems. IACR Cryptology ePrint Archive.

[29] Whitby, A., Jøsang, A. and Indulska, J., 2004, July. Filtering out unfair ratings in bayesian reputation systems. In Proc. 7th Int. Workshop on Trust in Agent Societies (Vol. 6, pp. 106-117).

[30] Zhao, K. and Pan, L., 2014, September. A machine learning based trust evaluation framework for online social networks. In Trust, Security and Privacy in Computing and Communications (TrustCom), 2014 IEEE 13th International Conference on (pp. 69-74). IEEE.