A Formal Derivation of Composite Trust

Tim Muller and Patrick Schweitzer
{tim.muller, patrick.schweitzer}@uni.lu

University of Luxembourg, CSC and SnT

Abstract. Trust appears in asymmetric interactions, where one party (the active party) can easily betray a stakeholder (the passive party). Over the Internet, the amount of information that a passive party can use to determine the integrity of an active party is often limited. The scenario where there is only one passive party and one active party is well studied, and has been solved under some assumptions. We generalize the setting to allow for more parties. In particular, the paper contains a formal derivation of conjunction (and disjunction) of trust opinions.

1 Introduction

Trust has a diverse meaning to different people. Consequently, definitions of trust in the literature vary. A definition of trust can be found in [1]: "First, one trusts another only relatively to a goal, i.e. for something s/he wants to achieve, that s/he desires. (..) Second, trust itself consists of beliefs. Trust is a mental state (..) about the behavior (..) relevant for the result." From that definition, we see that the authors clearly see trust as a mental state, regarding interactions where the result (rather than the intention) is important. In [2] an economical perspective is taken: "Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another." Here, it is clear that intention is relevant, and trust is still about a mental state and a specific interaction. Existing reputation systems must take trustworthiness as an objective property, rather than a mental state, in order for reputation to have meaning. Definitions with a psychological accent often do not depend on interactions, but tie trust to agents, as written in [3]: "Trust in things or people entails the willingness to submit to the risk that they may fail us, with the expectation that they will not (..)." These definitions, as well as informal intuitions share properties that are difficult to characterize. In this paper, we take the view that trust helps us to reason about trust systems (such as reputation systems and recommender systems) on the Internet. This means that trustworthiness, or *integrity*, is taken as an inherent property of the agents. It also means that agents trust each other with respect to interactions within the system. Furthermore, it means that only the result of the interaction matters, not the intention of the agent.

In each of the previously discussed definitions of trust an agent makes an assessment of another agent's future behavior using information they have gathered in the past. Such an assessment is called a *simple trust opinion*. If an agent

makes an assessment about the future behavior of several agents, it is called a *composite trust opinion*. A trust opinion does not only predict the future behavior that is most likely, but also indicates the certainty of the prediction. In this paper, we formalize the aforementioned notion of trust, using trust opinions. To obtain meaningful results, we must specialize the notion of trust.

In an interaction, there are several parties that have an agreement. There is at least one *active party*, who has an opportunity to ignore the agreement, and there is one *passive party*, which cannot affect the outcome of the interaction and may be harmed if active parties ignore the agreement. If an active party adheres to the agreement, we say the active party's behavior is *good behavior*, if he fails to adhere, we say it is *bad behavior*. Since the passive party may be harmed if one of the active parties shows bad behavior, the passive party may form a trust opinion about the active parties before (potentially) interacting. If an agent forms a (composite) trust opinion about (several) agents, he is called the *subject*. The combination of the active parties concerning the potential interaction is called the *target*. To express composite trust opinions we denote the target in propositional logic, where atomic propositions represent good or bad behavior of active parties. To illustrate the use of composite trust, consider the following example.

Example 1. Take an imaginary web service, CLOUD, that offers computational power to users, by CPU-scavenging in a similar fashion to BOINC [4], i.e. CLOUD is a grid. A user that delegates a computation is a client, and a user that offers CPU cycles is a provider. Unlike BOINC, the CLOUD system is a commercial system, where clients pay for computations, and providers get paid for offering computational power.

The identity of the machines in CLOUD is visible, and users can delegate computations to specific (groups of) machines. The infrastructure of CLOUD is very open, which means that malicious users can easily join as a provider. Malicious users may sometimes take shortcuts in the computation, providing wrong results. Furthermore, non-malicious users may prematurely terminate a computation before a result is provided, for example, when the computer shuts down, restarts or drops the network. It may occur that a single computation is delegated to a group of computers working concurrently to reduce latency. It may also occur that a single computation is delegated to more than one (group of) provider(s), to avoid extra latency when one of the (groups of) providers fails.

A client, A, on CLOUD has an instance of an NP-complete problem, and sends the problem to a provider, D, and a copy of the problem to a pair of concurrent providers, B and C. See Figure 1 for a visual representation of the interaction.

In our terminology, clients are passive parties (i.e. potential subjects), and providers are active parties (i.e. potential targets). Good behavior for a provider is delivering a correct result within a specified time frame. Bad behavior for a provider is returning a wrong result, returning it too late, or not at all. Since a client can quickly verify a (positive) solution to an NP-complete problem,

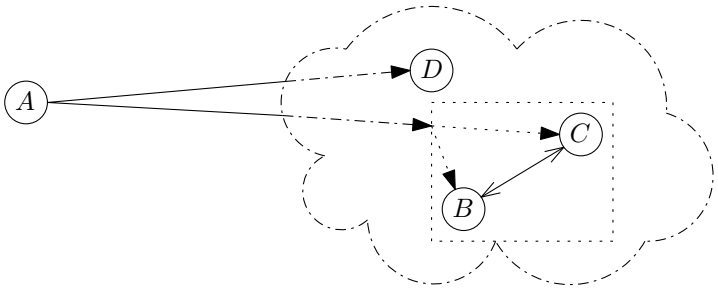

Fig. 1. The two outgoing arrows from A are delegations for a computation. One for D, the other is split and runs concurrently on B and C.

correct and incorrect solutions can easily be distinguished. Hence, it suffices for the subject A to receive at least one correct result within the specified time frame from the target. If either the single provider D or both other providers B and C provide the correct result in time, the whole target's behavior is considered good. We can denote this composite trust opinion as $D \vee (B \wedge C)$.

The subject not only wants to know the probability that the target succeeds, but also the uncertainty. If the probabilities b, c and d of B, C and D succeeding are independent, then one may anticipate that the expected probability of the target succeeding to be $d + b \cdot c - d \cdot b \cdot c$. We formally show the foresight on this trust opinion to be correct in Section 4.

To derive trust opinions, as the ones in the previous example, in a formal way, we need to define the context. We assume to be in an environment where all information comes from interactions between passive and active parties and that active parties operate independently of each other. Furthermore, we assume that each active party has an (hidden) integrity parameter that represents the probability of good behavior in unknown contexts. Therefore, in our framework, a trust opinion is a probability distribution over integrity parameters. From such a probability distribution, one can derive the expected probability of good behavior and the uncertainty of the estimate. In Section 3, we define the context (including trust opinions) formally.

The framework that we operate in is inspired by work in [5] and [6], where the authors independently derived a formal trust model (the *beta model*) of trust opinions. The context of the beta model is very similar to ours, but only allows simple trust opinions, not composite trust opinions. In the beta model, probability distributions known as beta distributions [7] represent trust opinions.

The beta model inspired several very popular extensions, such as Subjective Logic [8], TRAVOS [9] and CertainTrust [10]. Like our trust model, Subjective Logic contains conjunction and disjunction operators, and CertainTrust has been extended to CertainLogic [11] which also contains these operators. Unlike in our trust model, their trust model composite trust opinions are beta distributions. In Section 4, we show that models where composite trust opinions are beta distributions must violate reasonable assumptions

Not all models with conjunction, disjunction and uncertainty are based on the beta model; important examples are Fuzzy logic [12] and Dempster-Shafer theory [13,14]. Trying to apply our results to these models is more difficult, due to some inherently different assumptions and approaches.

There are also trust models that are based on the beta model, which have been extended beyond it. Often, they are the result of tweaking the assumptions of the beta model. In [15], for example, the assertion that behavior is good or bad is generalized into an assertion that a behavior value is selected from a range of values. For CLOUD, this means that incorrect results are distinguished from late results, or lack of results. In [16], the assumption of equal weight to all interactions is dropped. In the case of CLOUD, this implies that interactions that lie further in the past are less relevant than more recent interactions; or that providers behave differently during peak-load periods and off-peak periods.

In Section 2 of this paper, we introduce the necessary concepts from probability theorem. On the basis of these concepts, we formalize the assumptions necessary to reason about trust in our context, in Section 3. We argue that it is useful to have such a collection of simple assumptions from which to derive trust opinions (as opposed to defining how to derive trust opinions directly). That trust operators can be derived from the assumptions is shown in Section 4, where we derive conjunction and disjunction.

2 Preliminaries

The setting of the model is probabilistic in nature. We require the following concepts from probability theory (see e.g. [17,18]).

Definition 1 (σ-algebra, measure, probability measure). *Let Ω be a set of events. A set $\mathcal{F}$ of subsets of Ω is called a σ-algebra if the following three properties hold.*

1. $\emptyset \in \mathcal{F}$.
2. *If* $A \in \mathcal{F}$ *it follows that* $\Omega \setminus A \in \mathcal{F}$.
3. *If* $A_1, A_2, \ldots \subset \mathcal{F}$ *it follows that* $\bigcup_n A_n \in \mathcal{F}$.

Let P be a map from $\mathcal{F} \to \mathbb{R} \cup \{\infty\}$. Then, this map is called a measure if

1. $P(\emptyset) = 0$.
2. $P(A) \geq 0$ *for all* $A \in \mathcal{F}$.
3. *If* $A_1, A_2, \ldots \subset \mathcal{F}$ *such that* $A_k \cap A_l = \emptyset$ *for all* $k \neq l$, *it follows that* $P(\bigcup_n A_n) = \sum_n P(A_n)$.

If P maps to $[0,1]$ and $P(\Omega) = 1$, it is called a probability measure.

The tuple $(\Omega, \mathcal{F})$ from Definition 1 is called a measurable space. The triple $(\Omega, \mathcal{F}, P)$ is called a measure space. If P is additionally a probability measure, the triple is called a probability space.

Definition 2 (Random Variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, \mathcal{E})$ a measurable space. A mapping $X\colon \Omega \to E$ is a random variable, if*

$$\{\omega \in \Omega | X(\omega) \in B\} \in \mathcal{F} \text{ for all } B \in \mathcal{E}.$$

When Ω and E are at most countable, the σ-algebras $\mathcal{F}$ and $\mathcal{E}$ can be assumed to be the power sets over Ω and E, respectively.

In probability theory, the expression $\{\omega \in \Omega | X(\omega) \in B\}$ is often abbreviated to $\{X \in B\}$.

Definition 3 (Probability space of a random variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(E, \mathcal{E})$ a measurable space and $X\colon \Omega \to E$ a random variable. Then $P_X(B) := P(\{X \in B\}), B \in \mathcal{E}$ defines a probability measure P_X on $(E, \mathcal{E})$.*

The expression $P(\{X \in B\})$ is usually shorthanded to $P(X \in B)$.

Definition 4 (Distribution of a random variable). *The probability measure P_X is called the distribution of the random variable X.*

The probability space $(E, \mathcal{E}, P_X)$ is called discrete, if E is at most countable.

Definition 5 (Independence of random variables). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, \ldots, X_n$ be n random variables (over Ω) with values in the measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{1, \ldots, n\}$. The random variables $X_1, \ldots, X_n$ are called independent, if for arbitrary $B_1 \in \mathcal{E}_1, \ldots, B_n \in \mathcal{E}_n$, the events $\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ are independent.*

This definition is equivalent to the following.

$$X_1, \ldots, X_n \text{ indep. } \Leftrightarrow P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P_{X_i}(B_i) \text{ for } B_i \in \mathcal{E}_i.$$

As shorthand notation we write $X \perp\!\!\!\perp Y, Z$ when X, Y, Z are independent.

Definition 6 (Conditional independence of variables). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X, Y, Z be random variables (from Ω) with values in the measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{X, Y, Z\}$. Two random variables X and Y are conditionally independent given the variable Z if*

$$P(X \in A, Y \in B | Z \in C) = P(X \in A | Z \in C) P(Y \in B | Z \in C).$$

for each $A \in \mathcal{E}_X, B \in \mathcal{E}_Y$ and $C \in \mathcal{E}_Z$.

As shorthand we write $P(X, Y|Z) = P(X|Z)P(Y|Z)$, $(X \perp\!\!\!\perp Y)|Z$ or even $X \perp\!\!\!\perp Y|Z$. Note that the definition is equivalent to $P(X|Y, Z) = P(X|Z)$.

Theorem 1 (Law of total probability). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, A and C events and let $B_1, \ldots, B_n$ be a partition in that probability space. Then*

$$P(A|C) = \sum_{i=1}^{n} P(A|B_i, C)P(B_i|C).$$

The law of total probability also holds for continuous random variables X, and Y with positive density functions f_X and f_Y, respectively.

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|X = x) \cdot f_X(x)\,\mathrm{d}x.$$

Theorem 2 (Bayes' law for conditional probabilities). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and B and C events and let $A_1, \ldots, A_n$ be a partition in that probability space. Then*

$$P(A_j|B, C) = \frac{P(B|A_j, C)P(A_j|C)}{P(B|C)} = \frac{P(B|A_j, C)P(A_j|C)}{\sum_{i=1}^{n} P(B|A_i, C)P(A_i|C)}.$$

Note that in this form Bayes' theorem also holds for variables (instead of events). This is true for discrete random variables, continuous random variables as well as a mixture of discrete and continuous random variables. If continuous variables are involved, they need to have a positive density function.

Theorem 3 (Product distribution). *Let X and Y be two independent continuous variables, with positive probability density functions $f(x)$ and $g(x)$. Then $U = X \cdot Y$ is a continuous random variable with probability density function h. Explicitly*

$$h(u) = \int_{-\infty}^{\infty} \frac{1}{|y|} \cdot f(\frac{u}{y}) \cdot g(y)\,\mathrm{d}y.$$

An important distribution we refer to in the next sections is the beta distribution.

Definition 7 (Beta distribution). *A beta distribution is a family of continuous probability distributions in the interval $[0, 1]$, parameterized by two positive parameters, $\alpha, \beta \geq 1$. The probability density function of a beta distribution with parameters α and β is*

$$f_{\mathrm{B}}(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,\mathrm{d}y}.$$

The expression under the fractions is known as the beta function on α and β, and for positive integers α and β, the beta function fulfills $\mathrm{B}(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$.

To quantify information, we define a notion of entropy as in [19].

Definition 8 (Entropy). *The entropy H of a discrete random variable X with possible values $x_1, ..., x_n$ for $n \in \mathbb{N}$ is given by $H(X) = \mathbf{E}(I(X))$, where $\mathbf{E}$ is the expected value and $I(X)$ is the random variable denoting the information content of X. If p denotes the probability mass function of X and $c \in \mathbb{N}$, then the entropy can explicitly be written as*[1]

$$H(X) = \sum_{i=1}^{n} p(x_i)\, I(x_i) = \sum_{i=1}^{n} p(x_i) \log \frac{1}{p(x_i)}.$$

If $p(x_i)$ is equal to 0 for some $i \in \{1, \ldots, n\}$ and $n \in \mathbb{N}$, then $p(x_i) \log(p(x_i)^{-1})$ is taken to be 0.

Entropy can be extended for continuous random variables X ranging from a to b, with probability density function f_X

$$h(X) = \int_a^b f_X(x) \log(\frac{1}{f_X(x)})\, \mathrm{d}x.$$

3 Model assumptions

On a flea market, you see the sellers face to face, see whether they are well organized and you can determine whether they are popular. On an online e-commerce system, such detailed information is not available. You choose the seller based on previous interactions with this sellers, often by including information about interactions that others claim to have had with these sellers. Our trust model is much more relevant for e-commerce systems (and other online services, such as CLOUD), than it is for an actual flea market.

In this section, we formalize the assumptions that we have for trust in a system based on asymmetric interactions (like transactions in e-commerce systems), where expectations are clearly defined. First, we informally introduce our assumptions with motivations, then we formally state the assumptions as relations between random variables.

To reiterate some assertions from the introduction: Interactions are the building blocks in our trust analysis. Interactions are between a passive party (the subject) and active parties (the target). A subject may form a trust opinion about a target, before the subject interacts as passive party with the active parties in the target. The observed behavior of the active party is objectively classified as good (well) or bad (badly). Furthermore, the probability that the active party behaves well is determined by its integrity parameter p. An agent will most likely exhibit non-probabilistic behavior, and will therefore behave well in some situations and badly in others. However, we do not know the correlation between situations and behaviors, nor do we necessarily know the situation. In the light of this, we can view the integrity p as the chance that an agent is in a situation where his behavior is good (or even where behaving well is in his best

[1] In our considerations the base of the logarithm is not important.

interest in some iterative game[2], as in [20]). Lastly, we assume that p neither changes over time nor with respect to the environment. This assumption allows us to treat previous interactions in a mathematically coherent way, since all interactions are equally relevant for the current situation. In the model, an agent will never know the integrity of another agent, but will have an estimate based on these previous interactions.

To formulate the above assumptions in a formal manner, we need to define interactions of agents, integrity parameters of agents, sets of interactions that agents made in the past, and composite targets. To comply to notation used in probability theory (Bernoulli, binomial and beta distributions), we refer to good behavior of the active party as success, S, and bad behavior as failure, F. We are often interested in the previous interactions between a passive party and an active party, which we call an *interaction history* of the passive party about the active party. Furthermore, we take an interaction history to be a pair of natural numbers: the first number as representing the number of successes (good behavior by an active party), the second number as representing the number of failures. Let $\mathbf{A}$ denote the set of agents. The targets $\mathbf{T}$ are defined by $\varphi ::= A \in \mathbf{A} \mid \varphi \wedge \varphi \mid \varphi \vee \varphi$. For $A, B, C \in \mathbf{A}$, $T \in \mathbf{T}$ and a set of events Ω, we define the following random variables.

- $E_T \colon \Omega \to \{\text{S}, \text{F}\}$ is a discrete random variable modeling the outcome of the corresponding interaction with target T.
- $R_T \colon \Omega \to [0, 1]$ is a continuous random variable modeling the (hidden) integrity parameter of target T, defining the probability of success.
- $O_B^A \colon \Omega \to \mathbb{N} \times \mathbb{N}$ is a discrete random variable modeling the interaction history of A about B, representing the past interactions between A as passive party and B as active party.

Recall that a trust opinion is a distribution over the integrity parameter of a target, based on the interaction history about the involved active parties. Hence, if a subject A establishes a trust opinion about a target T, where $B, C, \ldots$ are the active parties in T (denoted $B, C, \ldots \in \text{act}(T)$), the density function looks like $f_{R_T}(x | O_B^A \cap O_C^A \cap \ldots)$. In this setting, the only type of information that is important to the subject, are the interaction histories of this subject. If there are other types of information (interaction histories of others, recommendations, a priori knowledge) available, they can be modeled as additional conditions (on additional random variables).

The definition of the random variables alone does not suffice to compute a query such as $f_{R_{B \wedge C}}(x | O_B^A \cap O_C^A)$. To calculate these trust opinions, we need to provide the dependencies and independencies between the random variables. These (in)dependencies are merely a formal denotation of the assumptions that we have. For a more concise formulation of these (in)dependencies, we introduce

[2] Agents expect to interact multiple times with other agents, and even if betrayal is profitable on the short run, it may be more profitable to conform on the long run.

sets of random variables.

$$\begin{aligned}\mathbb{E} &:= \{E_T : T \in \mathbf{T}\},\\ \mathbb{R} &:= \{R_T : T \in \mathbf{T}\},\\ \mathbb{O} &:= \{O_B^A : A, B \in \mathbf{A}\},\\ \mathbb{W} &:= \mathbb{E} \cup \mathbb{R} \cup \mathbb{O}.\end{aligned}$$

Let $x \in [0,1]$, $n, k \in \mathbb{N}$ and $\lambda\colon \mathbb{N} \to [0,1]$ be a probability distribution. For all $A, B \in \mathbf{A}$ and $S, T \in \mathbf{T}$ we set up the following dependency and independent relations as our assumptions.

D1 R_A is the uniform distribution on $[0,1]$.
If we know nothing about the integrity of A, we assert all values equally likely. For specific applications, statistical data about behaviors of agents may be used to construct an alternative distribution. A suitable distribution has a probability density function that is non-zero on $(0,1)$.

D2 $P(E_T{=}\mathrm{s}|R_T{=}p) = p$.
We assume that the probability of good behavior is determined by the integrity parameter p.

D3 $E_{S\wedge T} = \mathrm{s}$ iff $E_S = \mathrm{s}$ and $E_T = \mathrm{s}$, for $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$.
We define conjunctions of independent targets in such a way that the conjunction succeeds if both targets succeed.

D4 $E_{S\vee T} = \mathrm{s}$ iff $E_S = \mathrm{s}$ or $E_T = \mathrm{s}$, for $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$.
We define disjunctions of independent targets in such a way that the disjunction succeeds if at least one target succeeds.

D5 There exists a function f, with $R_{S\wedge T} = f(R_S, R_T)$, when $\mathrm{act}(S) \cap \mathrm{act}(T) = \emptyset$.
We assert that the integrity of a composite target is determined by the integrity of its active parties.

D6 $P(O_B^A{=}(k, n-k)|R_B{=}x) = \binom{n}{n-k} x^k (1-x)^{n-k} \lambda(n)$.
Assumes that the probability that A and B had an interaction history with size n is $\lambda(n)$, and that each past interaction had success probability x.

I1 For $W \in \mathbb{W}\backslash\{O_B^A\}$, it holds that $O_B^A \perp\!\!\!\perp W|R_B$.
The interaction history is completely determined by its size, and the probability of a success in a single interaction (by Dependency D6).

I2 For $W \in \mathbb{W}\backslash\{E_S : A \in \mathrm{act}(S), E_S \in \mathbb{E}\}$, it holds that $E_A \perp\!\!\!\perp W|R_A$.
The behavior of A is completely determined by its integrity parameter (by Dependency D2).

I3 For $W \in \mathbb{W}\backslash\{R_B\}$, it holds that $R_B \perp\!\!\!\perp W|E_B \cap \bigcap_{C\in\mathbf{A}}\{O_B^C\}$.
The only indicators of the integrity parameter of B, are interactions with it.

Independency I2 can be generalized for composite targets.

Proposition 1. *For all $W \in \mathbb{W}\backslash\{E_S : \mathrm{act}(T) \cap \mathrm{act}(S) \neq \emptyset, R_S \in \mathbb{E}\}$, it holds that $E_T \perp\!\!\!\perp W|R_T$.*

Proof. Apply structural induction. The base case precisely matches Independency I2. For the induction step use, that by definition of $\mathrm{act}(\cdot)$, it holds that $\mathrm{act}(T) \cup \mathrm{act}(T') = \mathrm{act}(T \wedge T') = \mathrm{act}(T \vee T')$. □

Our assumption is that trust adheres to D1-D6 and I1-I3.

A trust opinion of A about T can now be seen as the probability density function given by $f_{R_T}(x|\varphi)$, where φ is a condition that represents all knowledge of A about T, modulo the relations of the random variables. Typically, φ is the intersection of O_C^A, for different agents $C \in \mathbf{A}$. In other words, subjects have interaction histories about a group of active parties. If we restrict φ to merely the interaction history about a single active party, we get a beta distribution representing a simple trust opinion.

Lemma 1. *The simple trust opinion obtained from an interaction history with m successes and n failures is the beta distribution $f_{\mathrm{B}}(x; m+1, n+1)$.*

Proof.

$$
\begin{aligned}
& f_{R_B}(x|O_B^A{=}(m,n)) \\
{=}& \frac{P(O_B^A{=}(m,n)|R_B{=}x) \cdot f_{R_B}(x)}{\int_0^1 P(O_B^A{=}(m,n)|R_B{=}x') \cdot f_{R_B}(x')\,\mathrm{d}x'} \\
{=}& \frac{\binom{m+n}{m} x^m (1-x)^n \lambda(m+n) \cdot f_{R_B}(x)}{\int_0^1 \binom{m+n}{m} (x')^m (1-x')^n \lambda(m+n) \cdot f_{R_B}(x')\,\mathrm{d}x'} \\
{=}& \frac{\binom{m+n}{m} x^m (1-x)^n}{\int_0^1 \binom{m+n}{m} (x')^m (1-x')^n\,\mathrm{d}x'} \\
{=}& f_{\mathrm{B}}(x; m+1, n+1).
\end{aligned}
$$

□

The beta model ([5] and [6]) is based upon the notion that simple trust opinions are beta distribution. We can imagine an operator, trust aggregation, that updates trust opinions by adding more interactions. Formally, if we have a trust opinion X based on interaction history $(x_{\mathrm{S}}, x_{\mathrm{F}})$ and a trust opinion Y based on interaction history $(y_{\mathrm{S}}, y_{\mathrm{F}})$, then the aggregate of X and Y is a trust opinion based on $(x_{\mathrm{S}} + y_{\mathrm{S}}, x_{\mathrm{F}} + y_{\mathrm{F}})$. As such, the beta model inherits the mathematical property that the set of beta distributions is closed under trust aggregation.

Our assumptions regarding simple trust opinions are in line with the beta model, and are in fact sufficient to derive it (as demonstrated in Lemma 1). Hence, those assumptions can be seen as valid for the numerous models based on the beta model [8,9,10]. We extend the assumptions about simple trust opinions, by adding assumptions about composite trust opinions (Dependencies D3, D4 and D5). Under these assumptions, we show in Theorem 5 that composite trust opinions cannot generally be represented as beta distributions.

4 Composite Trust

In Example 1, we introduced the CLOUD grid. An example of a composite target was $D \vee (B \wedge C)$, where B, C and D are providers. The subject, A, has

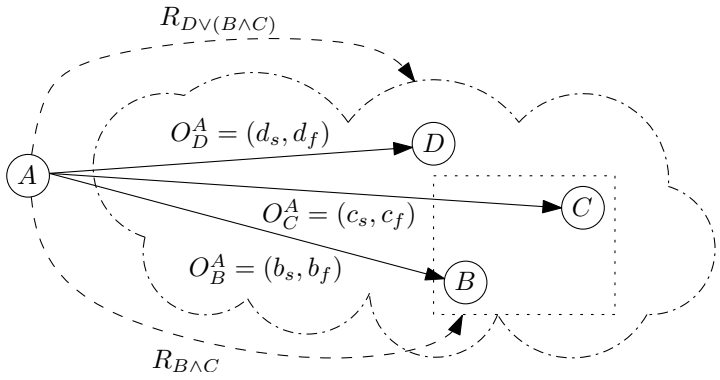

Fig. 2. Solid arrows represent interaction histories. Dashed arrows represent composite trust opinions. Arrows are labeled with the relevant random variables.

a (potentially empty) interaction history about B, C and D. In Example 2, we formally derive the trust opinion of A.

Example 2. The subject wants to form a trust opinion about $D \vee (B \wedge C)$, using only the interaction history of A about active parties B, C and D. The random variables O_B^A, O_C^A and O_D^A represent the interaction history of A about B, C and D. The random variable $R_{D\vee(B\wedge C)}$ represents the (unknown) integrity parameter of the target $D \vee (B \wedge C)$, and the random variable $E_{D\vee(B\wedge C)}$ represents the (unknown) outcomes of the next interaction with the target $D \vee (B \wedge C)$. We are interested not just in the probability of the next outcome of the target is a success ($E_{D\vee(B\wedge C)}$), but also in additional information, i.e. the random variable $R_{D\vee(B\wedge C)}$. Figure 2 depicts the relation between the users and the involved random variables. As stated in Section 3, given failures and successes of past interactions $(b_\mathrm{S}, b_\mathrm{F}, c_\mathrm{S}, c_\mathrm{F}, d_\mathrm{S}, d_\mathrm{F})$, the query for the trust opinion is of the shape $f_{R_{D\vee(B\wedge C)}}(x|O_B^A = (b_\mathrm{S}, b_\mathrm{F}) \cap O_C^A = (c_\mathrm{S}, c_\mathrm{F}) \cap O_D^A = (d_\mathrm{S}, d_\mathrm{F}))$. In other words, the trust opinion represents the probability distribution of a random variable that predicts the probability that the target succeeds.

Whenever a subject wants to compute a composite trust opinion about a target, he chooses the correct conditions and the correct random variable to form a distribution over, as illustrated in Example 2. Therefore, we can assume, without loss of generality, that we are given the term representing the probability distribution, and we want to compute an explicit probability density function.

We are interested in a random variable R_T, where T is not a single agent (unless the subject wants a simple trust opinion). However, we have not provided direct relations between R_T and observation histories O_B^A or integrity parameters of single agents R_A. The only random variable that we can immediately relate R_T to is E_T. For more concise notation, we note the following lemma.

Lemma 2. *If S and T do not share any active parties, then $R_{S\wedge T} = R_S \cdot R_T$.*

Proof. The product $R_S \cdot R_T$ of two random variables is defined as $(R_S \cdot R_T)(\omega) := R_S(\omega) \cdot R_T(\omega)$.

By Dependency D2, it holds that

$$P(E_{S\wedge T} = \mathrm{s} | R_{S\wedge T} = x) = x.$$

And, using Proposition 1 as well as Dependencies D2 and D3 we obtain

$$\begin{aligned}
&P(E_{S\wedge T} | R_S = y \cap R_T = z) \\
=&P(E_S = \mathrm{s} \cap E_T = \mathrm{s} | R_S = y \cap R_T = z) \\
=&P(E_S = \mathrm{s} | E_T = \mathrm{s} \cap R_S = y \cap R_T = z) \cdot P(E_T = \mathrm{s} | R_S = y \cap R_T = z) \\
=&P(E_S = \mathrm{s} | R_S = y) \cdot P(E_T = \mathrm{s} | R_T = z) \\
=&y \cdot z.
\end{aligned}$$

Assume, without loss of generality, that $R_S(\omega) = y$ and $R_T(\omega) = z$. By Dependency D5, there is a function f such that $x = P(E_{S\wedge T} = \mathrm{s} | R_{S\wedge T} = x) = P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = x)$. That implies that $x = f(y, z)$, and thus $P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = f(y, z)) = f(y, z)$. Now, since $R_S(\omega) = y$ and $R_T(\omega) = z$, we have

$$\begin{aligned}
&f(y, z) \\
=&P(E_{S\wedge T} = \mathrm{s} | f(R_S, R_T) = f(y, z)) = f(y, z) \\
=&P(E_{S\wedge T}) \\
=&P(E_{S\wedge T} | R_S = y \cap R_T = z) \\
=&y \cdot z.
\end{aligned}$$

Thus $R_S \cdot R_T = f(R_S, R_T) = R_{S\wedge T}$. □

A similar proof exists for disjunction, using independency over union rather than intersection, yielding $R_{S\vee T} = R_S + R_T - R_S \cdot R_T$

We can derive the probability density function of $R_{S\wedge T}$ under any condition φ.

Theorem 4. *If S and T do not share any active parties, then*

$$f_{R_{S\wedge T}}(x|\varphi) = \int_x^1 \frac{1}{y} \cdot f_{R_S}(\frac{x}{y}|\varphi) \cdot f_{R_T}(y|\varphi)\, \mathrm{d}y.$$

Proof. Apply Theorem 3 and Lemma 2. It suffices to verify the integral bounds. $f_{R_S}(\frac{x}{y}|\varphi) = 0$ for $0 > \frac{x}{y}$ and $1 < \frac{x}{y}$, so we can ignore cases where $y < x$ and $y > 1$. □

The case for disjunction can be calculated in a similar fashion. Theorem 4 is sufficient to derive trust opinions about arbitrary targets (where no active parties appear more than once), given arbitrary interactions with the active parties.

Corollary 1. *For every (finite) target where no active parties appear more than once, an explicit function for the trust opinion can be computed by the subject.*

Proof. Apply structural induction over the shape of the target. The base case (simple trust opinions) is proven in Lemma 1. To prove the induction step, take Theorem 4 as a rewrite rule from left to right. □

In Example 3, we derive an explicit formula for the trust opinion of $B \wedge C$, and look at some of its properties.

Example 3. Assume that the subject, A, wants to establish a trust opinion about the target, $B \wedge C$. In the past, A has interacted as a passive party with B several times; five times B behaved well, and once badly. Furthermore, A has interacted with C, too; four times C behaved well, and twice badly. The trust opinion of A about $B \wedge C$ is $f_{R_{B\wedge C}}(x|O_B^A = (5,1) \cap O_C^A = (4,2))$. Using Theorem 4, the trust opinion can be computed as

$$\int_x^1 \frac{1}{y} \cdot f_{R_B}(\frac{x}{y}|O_B^A = (5,1) \cap O_C^A = (4,2)) \cdot f_{R_C}(y|O_B^A = (5,1) \cap O_C^A = (4,2))\,\mathrm{d}y.$$

By Independency I3, we obtain

$$\int_x^1 \frac{1}{y} \cdot f_{R_B}(\frac{x}{y}|O_B^A = (5,1)) \cdot f_{R_C}(y|O_C^A = (4,2))\,\mathrm{d}y.$$

Which by Lemma 1 is equal to

$$\int_x^1 \frac{\frac{1}{y} \cdot (\frac{x}{y})^6 \cdot (1-\frac{x}{y})^2 \cdot y^5 \cdot (1-y)^3}{\mathrm{B}(5,1) \cdot \mathrm{B}(4,2)}\,\mathrm{d}y.$$

The formula can be formulated without an integral, and instead using some combinatorial functions, so that it reduces to

$$\frac{x^2 \cdot (1-x)^4 \cdot \Gamma(2) \cdot \Gamma(3) \cdot {}_2F_1(2,3;5;\frac{x-1}{x})}{\Gamma(5) \cdot \mathrm{B}(5,1) \cdot \mathrm{B}(4,2)}.$$

where Γ is the gamma function, B the beta function (not to be confused with the beta distribution) and ${}_2F_1$ a hypergeometric distribution. This, in turn, simplifies to

$$2205x^4(1 + 4x - 5x^2 + 2x(2+x)\log(x)).$$

The conjunction operation is depicted graphically in Figure 3. The rightmost distribution is the conjunction of the other two distributions. Recall that the abscissa depicts the integrity parameter of the targets in question. Thus, the more mass is on the right hand side of the graph, the bigger the probability that the target has a high integrity. As we can see, both active parties (B and C) have a relatively high integrity, but their conjunction ($B \wedge C$) does not.

The expected value of the trust opinion about a target is equal to the probability that the target succeeds, computation for $B \wedge C$ yields $\frac{15}{32}$. The expected value for the single agent B to succeed is $\frac{3}{4}$ and for C to succeed is $\frac{5}{8}$. Not coincidentally, the expected value for $B \wedge C$ is the product of that of B and C, namely $\frac{15}{32} = \frac{3}{4} \cdot \frac{5}{8}$.

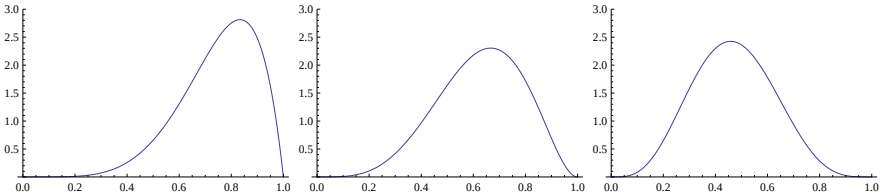

Fig. 3. From left to right: trust opinion about B, about C and about $B \wedge C$.

The entropy of the trust opinion can be calculated by applying Definition 8 to our probability density function. This results in an entropy value of approximately -0.67685 bits. The entropy in the trust opinions for the single agents B and C is -0.86157 bits and -0.62058 bits, respectively. We can see that the amount of information we have about $B \wedge C$ is between the amount of information we have on B and on C. This does not generally hold. If we swap the successes and failures, the entropy for B and C does not change, but the entropy for $B \wedge C$ becomes -2.0811 bits. The reason for the difference in information is obvious, as conjunction is not symmetric with respect to the duality of successes and failures. For conjunction, a failure carries more information, since failures outweigh successes similar to how false overrules true in the logical conjunction.

As we suspected in Section 1, and seen for a specific case in Example 3, the expected behavior of a conjunction of targets, is equal to product of the expected behavior of both targets.

Corollary 2. *If S and T do not share any active parties, then*

$$\mathbf{E}(R_{S \wedge T}) = \mathbf{E}(R_S) \cdot \mathbf{E}(R_T).$$

Proof. Immediate consequence of Lemma 2. □

Although the derivation in Example 3 seems asymmetrical with respect to S and T, commutativity and associativity hold.

Corollary 3. *Conjunctions and disjunctions of independent trust opinions are commutative and associative.*

Proof. Immediate consequence of Lemma 2. □

In Example 3, we have shown a specific composite trust opinion to be

$$f_{R_{B \wedge C}}(x | O_B^A = (5,1) \cap O_C^A = (4,2)) = 2205x^4(1 + 4x - 5x^2 + 2x(2 + x)\log(x)).$$

Now, one can wonder whether there exists a beta distribution with a probability density function of that shape. It is important to realize that if (composite) trust opinions are closed under conjunction (and disjunction), then there must be such a beta distribution.

Theorem 5. *A composite trust opinion need not be representable by a beta distribution.*

Proof. The expression $2205x^4(1 + 4x - 5x^2 + 2x(2 + x)\log(x))$, is a composite trust opinion, but not a polynomial. The probability density function of a beta distribution is always a polynomial (see Definition 7). Hence that composite trust opinion is not based on a beta distribution. □

From Theorem 5, we can conclude that every trust model in which the trust opinions are (isomorphic to) beta models violates one of the assumptions. A famous example is Subjective Logic [8] (binomial, without base rate), other examples include CertainLogic [11]. As the methodology of this paper is inspired by Subjective Logic, Dependencies D1, D2 and D6 are in line with the assumptions in Subjective Logic. Furthermore, the Independencies I1, I2, and I3 are also based on (non-formal formulations in) Subjective Logic. By the pigeon hole principle, Dependency D3 for conjunctions (or Dependency D4 for disjunctions) or Dependency D5 must be violated. Dependency D3 states that $E_{S \wedge T} = \text{S}$ iff $E_S = \text{S}$ and $E_T = \text{S}$ (for independent S and T), and Dependency D5 asserts that the integrity of a composite target is determined by the integrity of the active parties. We believe that these assertions may not be considered erroneous. We do not propose to alter Subjective Logic, as one of the strong points of Subjective Logic is its simple representation (triples with belief, disbelief and uncertainty components), which is isomorphic to beta distributions. And, as proven in Theorem 5, we cannot adhere to all assumptions and have a representation of trust opinions isomorphic to beta distributions.

5 Conclusion

The paper makes several assumptions about the trust domain. The assumptions are designed having interactions over the Internet in mind. There, agents have trust opinions about other agents, and they update their trust opinions when new information becomes available. We argue that a trust opinion is not just an estimated integrity parameter of a target, but that a trust opinion is a probability distribution over the integrity of a target. The advantage is that the subject can derive much more than just the expected value from a probability distribution. Examples of additional key figures that can be deduced from a probability distribution are uncertainty (as entropy), confidence intervals, most probably integrity value (which does not usually equal the expected integrity), error margins (as variance) and the impact of new information (by updating the probability distribution with the new information).

The idea of using probability distributions over an integrity parameter is not new in the trust domain, as it was used in [5] and [6]. The novel idea is to not just use probability distributions over integrity as trust opinions, but to pick $f_X(x|I)$ as the probability distribution, where X is the target of the trust opinion, and I is the information the subject has. To get an explicit formula for the probability density function, we must introduce specific assumptions. The

advantage of deriving the formula from these assumptions is threefold. First, by having explicit assumptions, any criticism on the resulting formula must be reducible to a disagreement about one of the assumptions. Second, if there is a disagreement about the assumptions, one can simply alter the assumption, and look at the implications. In particular, the assumption that all integrity parameters are equally likely for a simple target in the absence of information, is a strong assumption. The assumption can be replaced by asserting a different initial distribution of integrity parameters, and the model does not fundamentally change. Third, extending the formalism with new constructs may be achieved by adding new random variables and assumptions thereon.

An obvious candidate for extending the model is trust chaining, i.e. having the ability to use a recommendation as a (potential) source of information. If we extend the framework with trust chaining, we need to introduce random variables for (possible) recommendations, and introduce assumptions about when agents make honest recommendations and when they make dishonest recommendations. The suggested variants for trust chaining in different models are even more diverse than for conjunction and disjunction, partially due to different implicit assumptions and partially due to different insights [21]. Our approach may help unifying some insights, as well as force the assumptions to be formulated explicitly, thereby mitigating misunderstandings.

In this paper, however, we have applied the approach to composite trust opinions; trust opinions about conjunctions and/or disjunctions of agents. Thus, we have derived an explicit definition of a trust opinion of the shape "Can I trust that both A and B will behave according to agreement?" Of course, more general statements exist, where for "A and B" any propositional formula can be substituted and our result also generalizes to encompass these as well. We have proven some properties about composite trust opinions. First, the trust opinion about a target $S \wedge T$ has the expected value $s \cdot t$, where s and t are the expected values of the trust opinion about S and T. (Similarly, for $S \vee T$, it is $s + t - s \cdot t$.) Second, a composite trust opinion is in general not a beta distribution. Hence, no trust model with elements isomorphic to beta distributions can satisfy all our assumptions.

References

1. Castelfranchi, C., Falcone, R.: Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In: Proceedings of the 3rd International Conference on Multi Agent Systems. ICMAS '98, IEEE Computer Society (1998) 72–79
2. Rousseau, D.: Not so different after all : A cross-discipline view of trust. Academy of Management Review **23**(3) (1998) 393–404
3. Nooteboom, B.: Trust: Forms, Foundations, Functions, Failures and Figures. Edward Elgar (2002)
4. Anderson, D.P.: BOINC: A System for Public-Resource Computing and Storage. In: Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing. GRID '04, Washington, DC, USA, IEEE Computer Society (2004) 4–10

5. Jøsang, A., Ismail, R.: The beta reputation system. In: Proceedings of the 15th Bled Electronic Commerce Conference. Volume 160., Citeseer (2002) 324–337
6. Mui, L., Mohtashemi, M.: A computational model of trust and reputation. In: Proceedings of the 35th Hawaii International Conference on System Science (HICSS). (2002)
7. Johnson, N.L., Kotz, S., Balakrishnan, N.: Beta Distributions. In: Continuous Univariate Distributions. 2 edn. Volume 2. Wiley (1995)
8. Jøsang, A.: Artificial reasoning with subjective logic. In: 2nd Australian Workshop on Commonsense Reasoning. (1997)
9. Teacy, W., Patel, J., Jennings, N., Luck, M.: TRAVOS: Trust and Reputation in the Context of Inaccurate Information Sources. Autonomous Agents and Multi-Agent Systems **12** (2006) 183–198
10. Ries, S.: Certain trust: a trust model for users and agents. In: Proceedings of the 2007 ACM symposium on Applied computing. SAC '07, ACM (2007) 1599–1604
11. Ries, S., Habib, S., Mühlhäuser, M., Varadharajan, V.: Certainlogic: A logic for modeling trust and uncertainty. In: Trust and Trustworthy Computing. Volume 6740 of Lecture Notes in Computer Science. Springer (2011) 254–261
12. George J. Klir, B.Y.: Fuzzy sets and fuzzy logic: Theory and applications. Upper Saddle River, New Jersey : Prentice Hall (1995)
13. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist. **38**(2) (1967) 325–339
14. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
15. Jøsang, A., Haller, J.: Dirichlet Reputation Systems. In: International Conference on Availability, Reliability and Security, Los Alamitos, CA, USA, IEEE Computer Society (2007) 112–119
16. ElSalamouny, E., Sassone, V., Nielsen, M.: HMM-Based Trust Model. In: Formal Aspects in Security and Trust. Volume 5983 of LNCS. Springer (2010) 21–35
17. Billingsley, P.: Probability and measure. 3 edn. Wiley (1995)
18. Gut, A.: Probability: A Graduate Course (Springer Texts in Statistics). Springer (April 2007)
19. McEliece, R.J.: Theory of Information and Coding. 2 edn. Cambridge University Press (2001)
20. McCabe, K.A., Rigdon, M.L., Smith, V.L.: Positive reciprocity and intentions in trust games. Journal of Economic Behavior & Organization **52**(2) (2003) 267 – 275
21. Jøsang, A., Marsh, S., Pope, S.: Exploring different types of trust propagation. In: Trust Management. Volume 3986 of LNCS. Springer (2006) 179–192