Computer-Aided Decision-Making for Formal Relations and Domains of Trust, Distrust, and Mistrust with Cryptographic Applications*

Simon Kramer**, Rajeev Goré, and Eiji Okamoto

[1] University of Luxembourg
simon.kramer@a3.epfl.ch
[2] Australian National University
rajeev.gore@anu.edu.au
[3] University of Tsukuba, Japan
okamoto@risk.tsukuba.ac.jp

Abstract. We propose generic declarative definitions of the concepts of weak and strong *trust relations* between interacting agents, and *trust domains* of trust-related agents in distributed or multi-agent systems. Our definitions yield (1) *transitivity* results for trust relationships, (2) *computational complexity* results for deciding potential and actual trust relationships and membership in trust domains, (3) a positive (negative) *compositionality* result for strong (weak) trust domains, (4) a *computational design pattern* for building up strong trust domains, and (5) a negative *scalability* result for trust domains in general. We instantiate our generic trust concepts in five major cryptographic *applications* of trust, namely: access control, Trusted Third Parties, the Web of Trust, Public-Key Infrastructures, and Identity-Based Cryptography. We also demonstrate that *accountability induces trust.* In particular, accountable access control and cryptographic-key management are trustworthy. Our defining principle for weak and strong trust (domains) is (common) belief in and (common) knowledge of agent correctness, respectively.

Keywords access control; accountability; applied modal logic; CADM; computational trust and trustworthiness; cryptographic-key management; dependable distributed or multi-agent systems; TTP; Web of Trust; PKI.

1 Introduction

The subject matter of this paper is trust in dependable distributed or multi-agent systems. In this section, we introduce the motivation and goal for this matter, as well as the methodology that we employ to meet our goal.

* Cf. [36] for an early technical-report and [35] for a workshop version of this paper.

** This, corresponding, author's contribution was initiated with Grant P 08742 from the Japan Society for the Promotion of Science, and completed with Grant AFR 894328 from the National Research Fund, Luxembourg, cofunded under the Marie-Curie Actions of the European Commission (FP7-COFUND), during an invited stay at the Institute of Mathematical Sciences, Chennai, India.

Table 1. Functionality guarantee of dependable distributed systems

	in spite of	**thanks to**
technology	*incorrect design, natural forces*	*correct design*
agents	*incorrect behaviour:* incorrect use of the technology due to unawareness and/or maliciousness	*correct behaviour:* correct use of the technology

1.1 Motivation

Dependable distributed systems guarantee their functionality both in spite of and thanks to the technologies that implement these systems, and the man and/or machine agents[4] that partake in them (cf. Table 1). The functionality guarantee is conditioned on *naming* (cf. [4, Section 6.4] and [23, Section 19.1]) and *number:*

- on agent identity for the aspects of *anonymity* and *pseudonymity*;
- on a minimal number of correct (or dually, a maximal number of faulty or corrupt) agents for the aspects of *fault tolerance* [44] (classical distributed computation) and *corruption tolerance* [62] (secure multiparty computation).

The notion of *agent correctness* (e.g., as induced by a *policy*) in turn depends on the system itself. For example, correct agent behaviour can include (in various degrees of definitional difficulty):

- algorithm, game-rule, and protocol compliance;
- good book-keeping in audit and control of transactions;
- fairness in the scheduling of legitimately requested services;
- liveness in the sense of absence of crash, e.g., that the agent, say $a \in \mathcal{A}$, has always eventually responded to a ping from some other agent, say $b \in \mathcal{A}$, which can be expressed with a Linear-Temporal Logic [57] formula as follows:

  $$a \mathsf{\ correct\ } b := \overline{\Box}(b \mathsf{\ pings\ } a \rightarrow \Diamond(a \mathsf{\ pongs\ } b)),$$

  where a $\mathsf{correct}$ b macro-defines "a is correct as far as b is concerned", and $\overline{\Box}$ is the "so-far"-modality[5] and $\Diamond$ the "eventually"-modality here;
- access-control compliance [4, Chapter 4], e.g., that a has always been authorised by b and authenticated to access a given resource $r \in \mathcal{R}$, which can be expressed with an LTL formula as follows ($\mathcal{R}$ is finite):

  $$a \mathsf{\ correct\ } b := \bigwedge_{r \in \mathcal{R}} \overline{\Box} \left(\begin{pmatrix} a \mathsf{\ accesses\ } r \wedge \\ \mathsf{canAuthoriseToAccess}(b, a, r) \end{pmatrix} \rightarrow \begin{pmatrix} a \mathsf{\ authenticatedForAccessing\ } r \\ \wedge \mathsf{\ authorisedToAccess}(b, a, r) \end{pmatrix} \right),$$

[4] threads, processes, processor cores and processors, real and virtual machines, users, network nodes, etc.

[5] i.e., the modality "at all times in the past including the present"

where abstract access could be refined w.r.t. concrete *reading* and *writing*, and authorisation w.r.t. *revocation updates*, *delegation indirections*, *security levels* [4, Chapter 8], and *security compartments* [4, Chapter 9];
– absence of cryptographic-key compromise (cf. Section 4);
– etc.

In any case, agent correctness captures the *predictability* of each correct partaking agent that guarantees the functionality of the system to some or even all, correct or faulty, partaking agents. We stress that this predictability is ultimately one of agent behaviour rather than mental attitude. All that ultimately matters is the behavioural *effect* of an attitude, not the attitude itself. (An attitude may have no or no relevant effect.) In sum, system functionality depends on agent correctness, and agents depend on each other via each other's correctness. Whence the *social* need, called *trust*, to know whether or not someone behaves correctly. Note that according to game theory [8, Page 70] this need is also *rational:*

> [...] it isn't rational to trust people without a good reason: [...] trust can't be taken on trust.

That is, we trust when we have at least belief in correctness—in which case we run the risk of mistakingly believing—or *may* trust when we have even knowledge of correctness—in which case we run no such risk. Otherwise, i.e., when we have no such belief (and absence of belief is different from disbelief), we *must* trust (we have no choice). Now at first sight, such knowledge may seem difficult to attain. Yet in our standard understanding of knowledge defined in terms of indistinguishability (an observational equivalence) of system states, an agent a simply attains knowledge about a fact ϕ in a given state s as soon as ϕ also holds in all the states that are indistinguishable from s to a (cf. Section 2). In particular, a need not—but may happen to—have control over the system.

Example 1 (Social Software). The importance of dependable distributed systems for modern society can hardly be overestimated, because *social software* [51], i.e., "the software by which society runs" (e.g., banking and commerce, health care, social networking, voting and governmental administration, etc.) runs on distributed systems. In the age of the Internet, which acts as a generator and amplifier of the virtuality of human relations, trust is crucial [15, 48].

The concept of trust has at least three different aspects, namely trust *relations* and *domains,* and trust *management.* Our intuitions of them are as follows.

Trust relations An agent a trusts an agent b when a believes or even knows that b behaves correctly as far as a is concerned, independently of b's mental attitude. Hence, defining trust based on the correct behaviour of agents is more general than defining trust based on their mental attitude. The reader interested in trust based on mental attitude is referred to [18], which is a substantial study of various trust relations based on belief in and knowledge of mental attitudes.

Trust domains A trust domain is a community of mutually trusting agents with the common belief in or even knowledge of the trust relationships in the community. Informally, a statement ϕ is common belief or knowledge in a community $\mathcal{C}$ when all agents in $\mathcal{C}$ believe or know that ϕ is true (call this new statement ϕ'), all believe or know that ϕ' is true (call this new statement ϕ''), all believe or know that ϕ'' is true (call this new statement ϕ'''), etc. Notice the *mutual awareness* among agents w.r.t. the commonality of their knowledge or belief, which will turn out to be computationally costly (cf. Section 3). More intuitions on common knowledge in distributed systems can be found in [26].

Trust management Trust management is (cf. [54] for a recent survey):

1. the organisation of trust relations into trust domains (compartments), i.e., the sociology
2. the coordination of trust-building actions, i.e., the flow of trust (partial de-hierarchisation and decompartmentation, e.g., by building reputation [55]).

The organisation of trust relations into trust domains requires the ability to *decide* whether or not a given relation is a trust relation, and a given domain is a trust domain. Ideally, this ability appeals to formal definitions and decidability results, in order to support the human brain with *computer-aided decision making (CADM)*, as motivated by the following example.

Example 2 (Group size and the human brain). According to [56], "150 is the cognitive limit to the number of people a human brain can maintain a coherent social relationship with". "More generally, there are several layers of natural human group size that increase with a ratio of approximately three: 5, 15, 50, 150, 500, and 1,500". And "[a]s group sizes grow across these boundaries, they have more externally imposed infrastructure—and more formalized security systems."

The motivation for formal definitions now follows from the assumption that trust is a fundamental element of any coherent social relationship; and the motivation for CADM (requiring decidability results) additionally from the age-old desideratum to extend the cognitive limit of the human brain; cf. [23, Page 214]:

> Trust is the ultimate basis for all dealings that we have with other people. If you don't trust anybody with anything at all, why bother interacting with them?

It turns out that deciding trust relationships can be tractable with the aid of modern computers. However, deciding membership in trust domains, though computationally possible in theory, is computationally intractable in practice, even with the aid of clusters or grids [7], and clouds[6] of (super-)computers. What is worse: not only can we not make use of the promised power of parallel computing for deciding membership in trust domains in general, but also in particular when the candidate domains are the computing complexes themselves!

[6] Cloud computing is the automated outsourcing of IT infrastructure, platforms, and applications for data storage and calculation routines into evolving opaque clouds of anonymous computing units [24].

Example 3 (Cloud Computing). According to [32], in cloud computing, "users are universally required to accept the underlying premise of trust. In fact, some have conjectured that trust is the biggest concern facing cloud computing. Nowhere is the element of trust more apparent than in security, and many believe trust and security to be synonymous." Also according to [47]: "The growing importance of cloud computing makes it increasingly imperative that we grapple with the meaning of trust in the cloud and how the customer, provider, and society in general establish that trust." For more motivation, see [13, 33].

Indeed, the automatic validation of the underlying premise of cloud computing, i.e., trust, is an intractably big concern for the clouds themselves, as indicated above. However, the validation of trust can of course still be a tractably big concern for the relations between the cloud members. Anyway, what remains formally elusive is the *declarative meaning of trust.* As a matter of fact, the vast majority of the research literature on trust focuses on *how* to (operationally) establish and maintain trust of some form (e.g., with protocols, recommendation/reputation systems, reference monitors, trusted computing bases [25], etc.), but without actually defining *what* trust of that form (declaratively) means. And the very few works that do attempt declarative definitions of forms of trust do not provide insights in the tractability of trust domains, or applications to security such as trust in cryptographic-key management (cf. Section 6).

The bottom line is that declaratively defining the meaning of trust and obtaining estimates of the tractability of trust can be difficult. Yet formally defining trust in terms of (declarative) belief in or knowledge of behavioural correctness turns out to be natural, since humans often naturally refer to these or similar notions when informally explaining what they mean by trust. In distributed or multi-agent systems, attaining an *individual* consciousness of trust in terms of belief (weak trust) or knowledge (strong trust) from agent to agent can be computationally *tractable.* Whereas attaining a *collective* consciousness (mutual awareness) of trust in terms of *common* belief or knowledge within a greater domain (e.g., a cluster, cloud, or other collective) of agents is computationally *in*tractable. Computationally speaking, *collective trust does not scale.*

Trust domains should be family-sized, so to speak.

1.2 Goal

Our goal is at least seven-fold, namely:

1. to provide *generic declarative definitions* for trust relations between interacting agents, and trust domains of trust-related agents in distributed systems;
2. to obtain *computational complexity results* for deciding trust relationships and membership in trust domains;
3. to obtain *transitivity* results for trust relations;
4. to obtain *compositionality* and *scalability* results for trust domains;
5. to demonstrate the utility of our framework by showing how easy it is to instantiate it in five difficult cryptographic *applications* of trust, namely:

(a) access control
(b) Trusted Third Parties (TTPs)
(c) the Web of Trust
(d) Public-Key Infrastructures (PKIs)
(e) Identity-Based Cryptography (ID-Based Cryptography);

6. to investigate trust and its *relation to accountability (abusefree auditability),* which intuitively should induce trust;
7. to provide a *computational design pattern* for automatically building up bounded strong trust domains, e.g., in social networks.

Contribution To the best of our knowledge, general formal definitions for trust domains, general complexity results for trust relations as well as trust domains, a correspondence result relating the transitivity of agent correctness with the transitivity of trust relationships, a positive (negative) compositionality result for strong (weak) trust domains, a negative scalability result for trust domains in general, a computational design pattern for building up strong trust domains, and a generic formalisation of trust that can be instantiated in access control, TTPs, the Web of Trust, PKIs, and ID-Based Cryptography and that happens to be induced by (an equally generic) formalisation of accountability are all novel. The resulting (in)tractability insights are of great practical importance for accountable access control and cryptographic-key management (e.g., the design of [inter]national PKIs such as those required by credit-card transactions and ePassports [40]), and could may well be similarly important for computing clouds viewed as trust domains.

1.3 Methodology

Our methodology is to develop our formal definitions for trust relations and trust domains in a generic framework that is a semantically defined, standard modal logic of belief and knowledge (cf. Section 2). In that, we are interested in the *descriptive* (as opposed to deductive) use of an off-the-shelf, general-purpose (as opposed to special-purpose)[7] logic that is as simple as possible and as complex as necessary—both syntactically and semantically as well as computationally. Our *defining principle* for weak and strong trust is belief in and knowledge of agent correctness, respectively. We then derive our complexity results for deciding trust relationships and membership in trust domains by reduction to known results for the complexity of belief and knowledge (cf. Section 3). In spite of the substantial practical significance of our theoretical results, their derivation is quite simple (which increases their value), thanks to our modal logical framework. The difficulty was to find what we believe to be an interesting formal point of view on trust, which happens to be modal. Other points of view have, to the best of our knowledge, not resulted in even one of our contributions.

[7] e.g., the famous BAN-logic, which uses but does not define trust

Table 2. Satisfaction relation

$$
\begin{array}{rl}
(\mathfrak{S},\mathcal{V}),s \models P & \text{:iff } s \in \mathcal{V}(P)\\
(\mathfrak{S},\mathcal{V}),s \models \neg\phi & \text{:iff not } (\mathfrak{S},\mathcal{V}),s \models \phi\\
(\mathfrak{S},\mathcal{V}),s \models \phi\wedge\phi' & \text{:iff } (\mathfrak{S},\mathcal{V}),s \models \phi \text{ and } (\mathfrak{S},\mathcal{V}),s \models \phi'\\
(\mathfrak{S},\mathcal{V}),s \models \mathsf{CB}_{\mathcal{C}}(\phi) & \text{:iff for all } s' \in \mathcal{S}, \text{ if } s \mathrel{\mathrm{D}^{+}_{\mathcal{C}}} s' \text{ then } (\mathfrak{S},\mathcal{V}),s' \models \phi\\
(\mathfrak{S},\mathcal{V}),s \models \mathsf{CK}_{\mathcal{C}}(\phi) & \text{:iff for all } s' \in \mathcal{S}, \text{ if } s \mathrel{\mathrm{E}^{*}_{\mathcal{C}}} s' \text{ then } (\mathfrak{S},\mathcal{V}),s' \models \phi
\end{array}
$$

2 Formal definitions

We develop our formal definitions of trust relations and trust domains in a generic framework that is a semantically defined, standard modal logic of common belief and knowledge. The logic is *parametric* in agent correctness, to be instantiated for each considered distributed system (e.g., Section 1.1 and Sections 4.2–4.4).

Let S designate the considered distributed system (e.g., of sub-systems).

Definition 1 (Framework). *Let*

- $\mathcal{A}$ *designate an arbitrary finite set of unique* agent names[8] *a, b, c, etc.*
- $\mathcal{C} \subseteq \mathcal{A}$ *denote (finite and not necessarily disjoint) communities of agents (referred to by their name);*
- $\mathcal{P} := \{\ a \ \mathsf{correct}\ b \mid a,b \in \mathcal{A}\ \}$ *designate our (finite) set of* atomic propositions P *for referring to agent correctness (finiteness is crucial for complexity);*
- $\mathcal{L} \ni \phi \ ::=\ P \mid \neg\phi \mid \phi\wedge\phi \mid \mathsf{CB}_{\mathcal{C}}(\phi) \mid \mathsf{CK}_{\mathcal{C}}(\phi)$ *designate our modal language of formulae ϕ, with $\mathsf{CB}_{\mathcal{C}}(\phi)$ for "it is common belief in the community $\mathcal{C}$ that ϕ", and $\mathsf{CK}_{\mathcal{C}}(\phi)$ for "it is common knowledge in the community $\mathcal{C}$ that ϕ".*

Then given the set $\mathcal{S}$ (the state space) of system states s *induced by S (e.g., via a reachability or, in modal jargon, temporal accessibility relation)*[9]*, we define the* satisfaction relation $\models$ *of our framework in Table 2. There,*

- *":iff" abbreviates "by definition, if and only if";*
- $(\mathfrak{S},\mathcal{V})$ *designates a (modal)* model *of our framework;*
- $\mathfrak{S} := (\mathcal{S}, \{\mathrm{D}_a\}_{a\in\mathcal{A}}, \{\mathrm{E}_a\}_{a\in\mathcal{A}})$ *designates the (modal)* frame *with appropriate (for the system S)*
  - *serial*[10]*, transitive, and Euclidean*[11] *relations $\mathrm{D}_a \subseteq \mathcal{S}\times\mathcal{S}$ of* doxastic accessibility *(used for defining* belief*),*

[8] i.e., agent names injectively map to agents (names are identifiers here)

[9] For example, suppose that there is a set $\mathcal{S}_i$ of initial states for every system S, T designates the system's reachability or, synonymously, temporal accessibility relation, and T^* designates the reflexive transitive closure of T. Then, $\mathcal{S}$ is induced by S in the sense that $\mathcal{S} := \{\ s \mid \text{there is } s_i \in \mathcal{S}_i \text{ such that } s_i \mathrel{\mathrm{T}^*} s\ \}$.

[10] for all $s \in \mathcal{S}$, there is $s' \in \mathcal{S}$ s.t. $s \mathrel{\mathrm{D}_a} s'$

[11] for all $s, s', s'' \in \mathcal{S}$, if $s \mathrel{\mathrm{D}_a} s'$ and $s \mathrel{\mathrm{D}_a} s''$ then $s' \mathrel{\mathrm{D}_a} s''$

- *equivalence relations* $\mathrm{E}_a \subseteq \mathcal{S} \times \mathcal{S}$ *of* epistemic accessibility *(e.g., state* indistinguishability*, used for defining* knowledge*),*

such that $\mathrm{D}_a \subseteq \mathrm{E}_a$ *for any* $a \in \mathcal{A}$*;*

– $\mathcal{V} : \mathcal{P} \to 2^{\mathcal{S}}$ *designates the* valuation function *(returning for every* $P \in \mathcal{P}$ *the set of states where* P *is true) to be defined according to the appropriate notion of agent correctness for the system* S *(e.g., see Section 4);*

– $\mathrm{D}^+_{\mathcal{C}}$ *designates the transitive closure of* $\bigcup_{a\in\mathcal{C}} \mathrm{D}_a$*;*

– $\mathrm{E}^*_{\mathcal{C}}$ *designates the reflexive transitive closure of* $\bigcup_{a\in\mathcal{C}} \mathrm{E}_a$*.*

Note that defining (common) belief and knowledge abstractly with a serial, transitive, and Euclidean relation, and an equivalence relation, respectively, has emerged as a common practice that gives greater generality over more concrete approaches [46, Section 7.1]: the concrete definitions of the accessibility relations can be freely determined for a given distributed system provided they comply with the prescribed, characteristic properties. Typically, these definitions involve the projection of global states onto agents' local views [22]. For example, let $a \in \mathcal{A}$, and let π_a designate such a projection (function) for a. Then, epistemic accessibility in the sense of state indistinguishability can be defined such that for all $s, s' \in \mathcal{S}$,

$$s \mathrel{\mathrm{E}_a} s' \text{ :iff } \pi_a(s) = \pi_a(s'),$$

which guarantees that E_a is an equivalence relation. Specific applications, might require additional constraints. For example, we might want to stipulate the *a posteriori* constraint that

$$\text{if } s \mathrel{\mathrm{E}_a} s' \text{ then for all } b \in \mathcal{A},\ (\mathfrak{S}, \mathcal{V}), s \models b \mathrel{\mathsf{correct}} a \text{ iff } (\mathfrak{S}, \mathcal{V}), s' \models b \mathrel{\mathsf{correct}} a.$$

Doxastic accessibility $\mathrm{D}_a \subseteq \mathrm{E}_a$ can be defined from E_a by weakening the reflexivity of E_a to seriality as appropriate for the considered application.

Further note the following macro-definitions: $\phi \vee \phi' := \neg(\neg\phi \wedge \neg\phi')$, $\top := a \mathrel{\mathsf{correct}} a \vee \neg(a \mathrel{\mathsf{correct}} a)$, $\bot := \neg\top$, $\phi \to \phi' := \neg\phi \vee \phi'$, $\phi \leftrightarrow \phi' := (\phi \to \phi') \wedge (\phi' \to \phi)$, $\phi \oplus \phi' := \neg(\phi \leftrightarrow \phi')$, $\mathsf{B}_a(\phi) := \mathsf{CB}_{\{a\}}(\phi)$ (for "a believes that ϕ"), $\mathsf{K}_a(\phi) := \mathsf{CK}_{\{a\}}(\phi)$ (for "a knows that ϕ"), $\mathsf{CD}_{\mathcal{C}}(\phi) := \mathsf{CK}_{\mathcal{C}}(\phi) \oplus \mathsf{CK}_{\mathcal{C}}(\neg\phi)$ (for "$\mathcal{C}$ can (epistemically) decide whether or not ϕ"), and $\mathsf{D}_a(\phi) := \mathsf{CD}_{\{a\}}(\phi)$. Likewise we now obtain our declarative definitions of weak and strong notions of trust as mere macro-definitions, i.e., as simple syntactic constructions of semantically defined building blocks (cf. Table 3, conjunction over $\mathcal{C} = \emptyset$ being $\top$). Thereby, the reader is invited not to confuse distrust (using *verb-phrase* negation) with *absence of trust*, e.g., with $\neg\mathsf{B}_a(b \mathrel{\mathsf{correct}} a)$ or $\neg\mathsf{K}_a(b \mathrel{\mathsf{correct}} a)$ (using *sentence* negation). In the sequel of this paper, we shall focus on the positive notions of *trust* and *may-trust* rather than on the other, negative notions, for the ethical reason that the desiderata for well-intended engineering are the positive notions.

Remark 1 (May & Must).

May Here, a synonym for "may" is "can."

Table 3. Weak and strong trust relations and domains

a trusts b := $\mathsf{B}_a(b \text{ correct } a)$		risk of false positive
a distrusts b := $\mathsf{B}_a(\neg(b \text{ correct } a))$		risk of false negative
a mistrusts b := a trusts $b \wedge \neg(b \text{ correct } a)$		wrong trust
a maytrust b := $\mathsf{K}_a(b \text{ correct } a)$		right trust
a musttrust b := $\neg\mathsf{D}_a(b \text{ correct } a)$		no choice
$\mathsf{wTD}(\mathcal{C})$:= $\mathsf{CB}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}} a \text{ trusts } b)$		$\mathcal{C}$ is a weak trust domain
$\mathcal{C}$ trusts $\mathcal{C}'$:= $\mathsf{CB}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b \text{ correct } a \to b \text{ correct } c)) \wedge \mathsf{CB}_{\mathcal{C}}(\mathsf{wTD}(\mathcal{C}'))$		
$\mathcal{C}$ distrusts $\mathcal{C}'$:= $\mathsf{CB}_{\mathcal{C}}(\neg\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b \text{ correct } a \to b \text{ correct } c)) \vee \mathsf{CB}_{\mathcal{C}}(\neg\mathsf{wTD}(\mathcal{C}'))$		
$\mathcal{C}$ mistrusts $\mathcal{C}'$:= $\mathcal{C}$ trusts $\mathcal{C}' \wedge \neg(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b \text{ correct } a \to b \text{ correct } c) \wedge \mathsf{wTD}(\mathcal{C}'))$		
$\mathsf{sTD}(\mathcal{C})$:= $\mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}} a \text{ maytrust } b)$		$\mathcal{C}$ is a strong trust domain
$\mathcal{C}$ maytrust $\mathcal{C}'$:= $\mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b \text{ correct } a \to b \text{ correct } c)) \wedge \mathsf{CK}_{\mathcal{C}}(\mathsf{sTD}(\mathcal{C}'))$		
$\mathcal{C}$ musttrust $\mathcal{C}'$:= $\neg\mathsf{CD}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b \text{ correct } a \to b \text{ correct } c) \wedge \mathsf{sTD}(\mathcal{C}'))$		

Must "Must" must not be confused with "has/have to": "must" refers to an internal obligation—here, the absence of having a choice between trust and distrust, because both are absent in this case, whereas "has/have to" refers to an external obligation; and "must not" refers to an external prohibition, whereas "does/do not have to" refers to the absence of an external obligation.

Remark 2 (May- & Must-trust).

May-trust (trustworthiness) An agent a may trust an agent b in a's sense of correct behaviour (i.e., b is trust*worthy* to a) if and only if b has the will (in case b is willful) and capability (in case there is a need) to control b's behaviour in a's sense.

Must-trust (trustedness) The (system-specific) definition of agent correctness is a necessary condition for may-trust, otherwise we must trust (agent correctness being undefined, we cannot decide its truth [so b would have to be trust*ed* by a]).

Remark 3 (Trust & risk). In the case of [dis]trust, which is belief-based, we run the risk of a wrong (i.e., false positive [negative]) apprehension of agent correctness. In the case of may-trust, we do not run this risk, because may-trust is knowledge-based, which intuitively corresponds to a right apprehension. So belief-based trust is lack of trust, which "[...] is simply a risk, and that can sometimes be handled by standard risk-management techniques such as insurance" [23, Section 13.2.1]. Here we would want to insure beliefs in [in]correctness. As usual, risk quantifies as the cost of an event (here the cost of mistakingly believing in agent correctness) times the probability of the occurrence of the event.

Remark 4 (Trust graphs). Weak and strong trust domains can be represented as *complete* sub-graphs (which when maximal are *cliques*) of graphs of trust and may-trust relationships—bearing common belief and knowledge, respectively. The trust and may-trust graph of a pointed modal model $(\mathfrak{S}, \mathcal{V}), s$ is

$$\langle \mathcal{A}, \{\ (a,b) \in \mathcal{A} \times \mathcal{A} \mid (\mathfrak{S}, \mathcal{V}), s \models a \text{ } \mathsf{relationship} \text{ } b\ \} \rangle,$$

where $\mathsf{relationship} \in \{\mathsf{trusts}, \mathsf{maytrust}\}$. It is conceivable to take trust graphs as accessibility relations of a modal language for speaking *about* trust relations. Indeed, the majority of approaches to trust actually adopts trust relations as *given*, namely as trust graphs. Whereas we here actually *define* and thus explicate trust relations and domains in terms of belief in and knowledge of agent correctness.

We could also define *weak-strong* and *strong-weak* trust domains, i.e., as the common knowledge of weak and the common belief in strong trust relations, respectively. The difference between weak and strong trust is induced by the difference between belief and knowledge, respectively: weak trust possibly is wrong (i.e., mistaken belief), whereas strong trust necessarily is right (i.e., truthful belief). Social network systems furnish evidence for the adequacy of defining trust domains in terms of common knowledge or at least belief. As a matter of fact, the enumeration of "friends" on a member page in such systems constitutes a public announcement to the readers of that page who are logged in, who see all logged-in readers, etc. And it is common knowledge in the community of (dynamic) epistemic logicians that public announcements of (verifiable) elementary facts induce common knowledge within the addressed public (cf. [59] for a public announcement). So, suppose that you are a member of Facebook, and $\mathcal{C}$ designates the set consisting of you and those of your "friends" that are enumerated on your member page. Further, fix the current moment in time, and call it s. (We may talk about time here; see Footnote 9.) Then the formula $\bigwedge_{a,b\in\mathcal{C}} a \text{ } \mathsf{maytrust} \text{ } b \leftrightarrow \mathsf{sTD}(\mathcal{C})$ is true at s (with no outermost $\mathsf{CK}_{\mathcal{C}}$ operator on the left since the aforementioned public announcement implies it). The formula is a (bi-)conditional because we have not fixed the notion of agent correctness for Facebook (*they* should). Note that the trust relations between you and your "friends" are symmetric, because each "friend" had to give their consent for having the privilege of being enumerated as such on your member page in Facebook.

Definition 2 (Truth & Validity). *The formula ϕ is* true *(or* satisfied*) in the model $(\mathfrak{S}, \mathcal{V})$ at the state $s \in \mathcal{S}$:iff $(\mathfrak{S}, \mathcal{V}), s \models \phi$. The formula ϕ is* satisfiable *in the model $(\mathfrak{S}, \mathcal{V})$:iff there is $s \in \mathcal{S}$ such that $(\mathfrak{S}, \mathcal{V}), s \models \phi$. The formula ϕ is* globally true *(or* globally satisfied*) in the model $(\mathfrak{S}, \mathcal{V})$, written $(\mathfrak{S}, \mathcal{V}) \models \phi$, :iff for all $s \in \mathcal{S}$, $(\mathfrak{S}, \mathcal{V}), s \models \phi$. The formula ϕ is* satisfiable *:iff there is a model $(\mathfrak{S}, \mathcal{V})$ and a state $s \in \mathcal{S}$ such that $(\mathfrak{S}, \mathcal{V}), s \models \phi$. The formula ϕ is* valid*, written $\models \phi$, :iff for all models $(\mathfrak{S}, \mathcal{V})$, $(\mathfrak{S}, \mathcal{V}) \models \phi$. (cf. [9])*

Fact 1 (Common belief) *Being defined in terms of a serial, transitive, and Euclidean relation, $\mathsf{CB}_{\mathcal{C}}$ is K45 and $\mathsf{CB}_{\{a\}}$ KD45 for any $a \in \mathcal{C} \subseteq \mathcal{A}$, i.e.:*

K: $\models \mathsf{CB}_{\mathcal{C}}(\phi \to \phi') \to (\mathsf{CB}_{\mathcal{C}}(\phi) \to \mathsf{CB}_{\mathcal{C}}(\phi'))$ *(Kripke's law)*

D: $\models \mathsf{CB}_{\{a\}}(\phi) \rightarrow \neg\mathsf{CB}_{\{a\}}(\neg\phi)$ *(consistency of beliefs, seriality)*
4: $\models \mathsf{CB}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CB}_{\mathcal{C}}(\mathsf{CB}_{\mathcal{C}}(\phi))$ *(positive introspection, transitivity)*
5: $\models \neg\mathsf{CB}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CB}_{\mathcal{C}}(\neg\mathsf{CB}_{\mathcal{C}}(\phi))$ *(negative introspection, Euclideanness)*
N: *if* $\models \phi$ *then* $\models \mathsf{CB}_{\mathcal{C}}(\phi)$ *(necessitation).*

Further, let $\mathsf{EB}_{\mathcal{C}}(\phi) := \bigwedge_{a\in\mathcal{C}} \mathsf{B}_a(\phi)$ *("everybody in* $\mathcal{C}$ *believes that* ϕ*"). Then:*

- $\models \mathsf{CB}_{\mathcal{C}}(\phi) \rightarrow \mathsf{EB}_{\mathcal{C}}(\phi)$
- $\models \mathsf{CB}_{\mathcal{C}}(\phi) \rightarrow \mathsf{EB}_{\mathcal{C}}(\mathsf{CB}_{\mathcal{C}}(\phi))$
- $\models \mathsf{CB}_{\mathcal{C}}(\phi \rightarrow \mathsf{EB}_{\mathcal{C}}(\phi)) \rightarrow (\mathsf{EB}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CB}_{\mathcal{C}}(\phi))$.

For details see [46, Section 7.1].

The difference between belief and knowledge is that belief possibly is wrong (cf. the **D** law), whereas knowledge necessarily is right (cf. the following **T** law).

Fact 2 (Common knowledge) *Being defined in terms of an equivalence relation,* $\mathsf{CK}_{\mathcal{C}}$ *is S5 for any* $\mathcal{C} \subseteq \mathcal{A}$*, i.e.:*

K: $\models \mathsf{CK}_{\mathcal{C}}(\phi \rightarrow \phi') \rightarrow (\mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CK}_{\mathcal{C}}(\phi'))$ *(Kripke's law)*
T: $\models \mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \phi$ *(truth law, reflexivity)*
4: $\models \mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CK}_{\mathcal{C}}(\mathsf{CK}_{\mathcal{C}}(\phi))$ *(positive introspection)*
5: $\models \neg\mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CK}_{\mathcal{C}}(\neg\mathsf{CK}_{\mathcal{C}}(\phi))$ *(negative introspection)*
N: *if* $\models \phi$ *then* $\models \mathsf{CK}_{\mathcal{C}}(\phi)$ *(necessitation).*

Further, let $\mathsf{EK}_{\mathcal{C}}(\phi) := \bigwedge_{a\in\mathcal{C}} \mathsf{K}_a(\phi)$ *("everybody in* $\mathcal{C}$ *knows that* ϕ*"). Then:*

- $\models \mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \mathsf{EK}_{\mathcal{C}}(\mathsf{CK}_{\mathcal{C}}(\phi))$
- $\models \mathsf{CK}_{\mathcal{C}}(\phi \rightarrow \mathsf{EK}_{\mathcal{C}}(\phi)) \rightarrow (\phi \rightarrow \mathsf{CK}_{\mathcal{C}}(\phi))$.

For details see [46, Section 7.1].

Note that depending on the properties of the employed communication lines, common knowledge may have to be pre-established, i.e., off those lines [26].

Fact 3 (Weak-trust consistency) *For all* $a, b \in \mathcal{A}$*,*

$$\models a \ \mathsf{trusts}\ b \rightarrow \neg(a\ \mathsf{distrusts}\ b).$$

Proof. By $\mathbf{D}(\mathsf{CB}_{\{a\}})$.

Fact 4 (Knowledge implies belief) *For all* $\mathcal{C} \subseteq \mathcal{A}$*,* $\models \mathsf{CK}_{\mathcal{C}}(\phi) \rightarrow \mathsf{CB}_{\mathcal{C}}(\phi)$*. In particular when* $\mathcal{C} = \{a\}$*,* $\models \mathsf{K}_a(\phi) \rightarrow \mathsf{B}_a(\phi)$*.*

Proof. By the fact that for all $a \in \mathcal{A}$, $\mathrm{D}_a \subseteq \mathrm{E}_a$ (cf. Definition 1).

The following corollary is immediate.

Corollary 1 (Strong implies weak trust).

1. *For all* $a, b \in \mathcal{A}$*,* $\models a\ \mathsf{maytrust}\ b \rightarrow a\ \mathsf{trusts}\ b$*.*

2. *For all* $\mathcal{C} \subseteq \mathcal{A}$, $\models \mathsf{sTD}(\mathcal{C}) \rightarrow \mathsf{wTD}(\mathcal{C})$.

Trust relations and trust domains can be related as follows.

Proposition 1 (Trust relations & domains). *In trust domains, trust relations are* universal *(i.e., correspond to the Cartesian product on those domains). That is, for all* $a, b \in \mathcal{C}$, $\models \mathsf{wTD}(\mathcal{C}) \rightarrow a\ \mathsf{trusts}\ b$ *and* $\models \mathsf{sTD}(\mathcal{C}) \rightarrow a\ \mathsf{maytrust}\ b$.

Proof. Almost by definition of weak and strong trust relations and domains.

Hence in trust domains, trust relations are equivalence relations. (The universal relation contains all other relations.)

Corollary 2. *In trust domains, trust relations are* $(a, b, c \in \mathcal{C} \subseteq \mathcal{A})$*:*

- reflexive, *i.e.,* $\models \mathsf{wTD}(\mathcal{C}) \rightarrow a\ \mathsf{trusts}\ a$ *and* $\models \mathsf{sTD}(\mathcal{C}) \rightarrow a\ \mathsf{maytrust}\ a$
- symmetric, *i.e.,* $\models \mathsf{wTD}(\mathcal{C}) \rightarrow (a\ \mathsf{trusts}\ b \rightarrow b\ \mathsf{trusts}\ a)$ *and* $\models \mathsf{sTD}(\mathcal{C}) \rightarrow (a\ \mathsf{maytrust}\ b \rightarrow b\ \mathsf{maytrust}\ a)$
- transitive, *i.e.,* $\models \mathsf{wTD}(\mathcal{C}) \rightarrow ((a\ \mathsf{trusts}\ b \wedge b\ \mathsf{trusts}\ c) \rightarrow a\ \mathsf{trusts}\ c)$ *and* $\models \mathsf{sTD}(\mathcal{C}) \rightarrow ((a\ \mathsf{maytrust}\ b \wedge b\ \mathsf{maytrust}\ c) \rightarrow a\ \mathsf{maytrust}\ c)$*).*

A more interesting condition for the transitivity of trust relations than their universality is the existence of an agent (say b) acting as a *trust reference* for the trustworthiness of another agent (say c) to a third agent (say a).

Lemma 1 (Trust reference). *For all* $a, b, c \in \mathcal{A}$:

1. $\models \underbrace{\mathsf{B}_a(c\ \mathsf{correct}\ b \rightarrow c\ \mathsf{correct}\ a)}_{\textit{belief in agent-correctness inclusion}} \rightarrow (\mathsf{B}_a(b\ \mathsf{maytrust}\ c) \rightarrow a\ \mathsf{trusts}\ c)$
2. $\models \underbrace{\mathsf{K}_a(c\ \mathsf{correct}\ b \rightarrow c\ \mathsf{correct}\ a)}_{\textit{knowledge of agent-correctness inclusion}} \rightarrow (\mathsf{K}_a(b\ \mathsf{maytrust}\ c) \rightarrow a\ \mathsf{maytrust}\ c)$.

Proof. By $\mathbf{T}(\mathsf{K}_b)$ and $\mathbf{K}(\mathsf{B}_a)$, and $\mathbf{T}(\mathsf{K}_b)$ and $\mathbf{K}(\mathsf{K}_a)$, respectively.

That is, b acts as a reference of c's trustworthiness to a when a believes or knows that (1) b's notion of correctness about c is included a's, and (2) b may trust c.

Remark 5. Observe that the replacement of $b\ \mathsf{maytrust}\ c$ by the weaker $b\ \mathsf{trusts}\ c$ in Lemma 1 does not yield a validity, due to the absence of the $\mathbf{T}$-law for B_b.

The existence of a trust reference is a sufficient condition (among others) for the *transitivity* of trust relationships—the links in *trust paths* in trust graphs.

Proposition 2 (Trust transitivity).

1. *For all* $a, b, c \in \mathcal{A}$,

$$\models \left(\begin{array}{l}(\underbrace{\mathsf{B}_a(c\ \mathsf{correct}\ b \rightarrow c\ \mathsf{correct}\ a)}_{\textit{belief in agent-correctness inclusion}} \wedge \mathsf{B}_a(b\ \mathsf{maytrust}\ c)) \\ \vee \left(\begin{array}{l}\underbrace{\mathsf{B}_a((c\ \mathsf{correct}\ b \wedge b\ \mathsf{correct}\ a) \rightarrow c\ \mathsf{correct}\ a)}_{\textit{belief in agent-correctness transitivity}} \\ \wedge\ \underbrace{(\mathsf{B}_b(c\ \mathsf{correct}\ b) \rightarrow \mathsf{B}_a(c\ \mathsf{correct}\ b))}_{\textit{inclusion of agent-correctness belief}}\end{array}\right)\end{array}\right) \rightarrow \underbrace{((a\ \mathsf{trusts}\ b \wedge b\ \mathsf{trusts}\ c) \rightarrow a\ \mathsf{trusts}\ c)}_{\textit{trust transitivity}}.$$

2. For all $a, b, c \in \mathcal{A}$,

$$\models \left(\begin{array}{l} (\underbrace{\mathsf{K}_a(c \;\mathsf{correct}\; b \to c \;\mathsf{correct}\; a) \wedge \mathsf{K}_a(b \;\mathsf{maytrust}\; c)}_{\textit{knowledge of agent-correctness inclusion}}) \\ \vee \left(\begin{array}{l} \underbrace{((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a)}_{\textit{agent-correctness transitivity}} \\ \wedge \; \underbrace{(c \;\mathsf{correct}\; a \to \mathsf{K}_a(c \;\mathsf{correct}\; a))}_{\textit{agent-correctness knowledge}} \end{array} \right) \\ \vee \left(\begin{array}{l} \underbrace{\mathsf{K}_a((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a)}_{\textit{knowledge of agent-correctness transitivity}} \\ \wedge \; \underbrace{(\mathsf{K}_b(c \;\mathsf{correct}\; b) \to \mathsf{K}_a(c \;\mathsf{correct}\; b))}_{\textit{inclusion of agent-correctness knowledge}} \end{array} \right) \end{array} \right) \to$$

$$\underbrace{((a \;\mathsf{maytrust}\; b \wedge b \;\mathsf{maytrust}\; c) \to a \;\mathsf{maytrust}\; c)}_{\textit{may-trust transitivity}}.$$

Proof. The sufficiency for trust transitivity of the condition $\mathsf{B}_a(c \;\mathsf{correct}\; b \to c \;\mathsf{correct}\; a) \wedge \mathsf{B}_a(b \;\mathsf{maytrust}\; c)$ and $\mathsf{K}_a(c \;\mathsf{correct}\; b \to c \;\mathsf{correct}\; a) \wedge \mathsf{K}_a(b \;\mathsf{maytrust}\; c)$ is a trivial consequence of Lemma 1; the sufficiency of the condition $\mathsf{B}_a((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a) \wedge (\mathsf{B}_b(c \;\mathsf{correct}\; b) \to \mathsf{B}_a(c \;\mathsf{correct}\; b))$ and $\mathsf{K}_a((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a) \wedge (\mathsf{K}_b(c \;\mathsf{correct}\; b) \to \mathsf{K}_a(c \;\mathsf{correct}\; b))$ follows from $\mathbf{K}(\mathsf{B}_a)$ and $\mathbf{K}(\mathsf{K}_a)$, respectively; and the sufficiency of the condition $((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a) \wedge (c \;\mathsf{correct}\; a \to \mathsf{K}_a(c \;\mathsf{correct}\; a))$ follows from $\mathbf{T}(\mathsf{K}_a)$ and $\mathbf{T}(\mathsf{K}_b)$. (Recall that there is no $\mathbf{T}$-law for belief.)

Note that trust relationships are *not* in general transitive. As a simple counter-example consider that if a trusts b and b trusts c but b's notion of correctness is not included in a's then a need not trust c just because b trusts c.

Theorem 1 (Transitivity correspondence). *For all* $a, b, c \in \mathcal{A}$,

$$\models \underbrace{\begin{pmatrix} (c \;\mathsf{correct}\; b \to \mathsf{K}_b(c \;\mathsf{correct}\; b)) \wedge \\ (b \;\mathsf{correct}\; a \to \mathsf{K}_a(b \;\mathsf{correct}\; a)) \wedge \\ (c \;\mathsf{correct}\; a \to \mathsf{K}_a(c \;\mathsf{correct}\; a)) \end{pmatrix}}_{\textit{agent-correctness knowledge}} \to$$

$$\left(\begin{array}{l} \underbrace{((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a)}_{\textit{agent-correctness transitivity}} \leftrightarrow \\ \underbrace{((a \;\mathsf{maytrust}\; b \wedge b \;\mathsf{maytrust}\; c) \to a \;\mathsf{maytrust}\; c)}_{\textit{may-trust transitivity}} \end{array} \right).$$

Proof. By propositional logic jointly from the simply-to-prove fact that

$$\models \begin{pmatrix} (c \;\mathsf{correct}\; b \to \mathsf{K}_b(c \;\mathsf{correct}\; b)) \wedge \\ (b \;\mathsf{correct}\; a \to \mathsf{K}_a(b \;\mathsf{correct}\; a)) \end{pmatrix} \to$$

$$\begin{pmatrix} ((a \;\mathsf{maytrust}\; b \wedge b \;\mathsf{maytrust}\; c) \to a \;\mathsf{maytrust}\; c) \\ \to \; ((c \;\mathsf{correct}\; b \wedge b \;\mathsf{correct}\; a) \to c \;\mathsf{correct}\; a) \end{pmatrix}$$

and the already-proved fact that $\models (c\ \mathsf{correct}\ a \to \mathsf{K}_a(c\ \mathsf{correct}\ a)) \to ((c\ \mathsf{correct}\ b \land b\ \mathsf{correct}\ a) \to c\ \mathsf{correct}\ a) \to ((a\ \mathsf{maytrust}\ b \land b\ \mathsf{maytrust}\ c) \to a\ \mathsf{maytrust}\ c))$ (cf. Proposition 2.2).

Let us now turn to trust domains.

Proposition 3 (Trust domains).

0. *Emptiness:* $\models \mathsf{wTD}(\emptyset)$ *and* $\models \mathsf{sTD}(\emptyset)$
1. *Decomposition:*
   $\models \mathsf{wTD}(\mathcal{C} \cup \mathcal{C}') \to (\mathsf{wTD}(\mathcal{C}) \land \mathsf{wTD}(\mathcal{C}'))$
   $\models \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}') \to (\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}'))$
2. *Intersection:*
   $\models (\mathsf{wTD}(\mathcal{C}) \land \mathsf{wTD}(\mathcal{C}')) \to \mathsf{wTD}(\mathcal{C} \cap \mathcal{C}')$
   $\models (\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}')) \to \mathsf{sTD}(\mathcal{C} \cap \mathcal{C}')$
3. *Antitonicity:*
   if $\mathcal{C} \subseteq \mathcal{C}'$ *then* $\models \mathsf{wTD}(\mathcal{C}') \to \mathsf{wTD}(\mathcal{C})$ *and* $\models \mathsf{sTD}(\mathcal{C}') \to \mathsf{sTD}(\mathcal{C})$
4. *Mutuality:*
   $\models \mathsf{wTD}(\mathcal{C} \cup \mathcal{C}') \to (\mathcal{C}\ \mathsf{trusts}\ \mathcal{C}' \land \mathcal{C}'\ \mathsf{trusts}\ \mathcal{C})$
   $\models \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}') \to (\mathcal{C}\ \mathsf{maytrust}\ \mathcal{C}' \land \mathcal{C}'\ \mathsf{maytrust}\ \mathcal{C})$

Proof. Straightforward from definitions.

The following theorem provides an insight into trust domains.

Theorem 2 (Strong trust domains). *Merging two strong trust domains is* conditionally compositional *in the sense that if it is common knowledge in their union that they include each other's notions of agent correctness then a necessary and sufficient condition for their merger is that it be common knowledge in their union that each one is a strong trust domain. Formally, for all* $\mathcal{C}, \mathcal{C}' \subseteq \mathcal{A}$,

$$\underbrace{\models \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\left(\begin{array}{l}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a \to b\ \mathsf{correct}\ c)\ \land \\ \bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a \to b\ \mathsf{correct}\ c)\end{array}\right)}_{\text{common knowledge of agent-correctness compatibility}} \to$$
$$\underbrace{(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}')) \leftrightarrow \mathsf{sTD}(\mathcal{C}\cup\mathcal{C}'))}_{\text{trust-domain compositionality}}.$$

Proof. See Appendix A. Lemma 1.2 is crucial for the composability part, which is expressed by the formula $\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}')) \to \mathsf{sTD}(\mathcal{C}\cup\mathcal{C}')$.

Note that there is no analogous result for weak trust domains (cf. Remark 5). That is, merging *weak* trust domains is *non*-compositional in the sense that the result of merging two weak trust domains is not necessarily again a weak trust domain — not even when there is common knowledge (instead of only common belief) in the union of both domains that each domain is a weak trust domain. Nevertheless, we can salvage the following corollary result.

Corollary 3 (Weak and strong trust domains).

$$\models \underbrace{\mathsf{CB}_{\mathcal{C}\cup\mathcal{C}'}\left(\begin{array}{l}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a \to b\ \mathsf{correct}\ c)\ \land \\ \bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a \to b\ \mathsf{correct}\ c)\end{array}\right)}_{\textit{common belief in agent-correctness compatibility}} \to$$
$$\underbrace{(\mathsf{CB}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}')) \to \mathsf{wTD}(\mathcal{C}\cup\mathcal{C}'))}_{\textit{strong- into weak-trust-domain composability}}.$$

Proof. Similarly to the proof in Table 10 of Appendix A, by invoking Lemma 1.1 in lieu of Lemma 1.2.

Theorem 2 yields a simple, though computationally intrinsically costly *design pattern* for recursively building up strong trust domains (cf. Section 5). Again due to the absence of an analogous theorem for weak trust domains, there is no analogous design pattern either. Hence, there really is a strong practical interest in strong trust domains. As a matter of fact, this practical interest is even stronger because checking membership in strong trust domains will turn out to be computationally no more complex (up to a constant) than checking membership in weak trust domains (cf. Section 3).

We continue to define *potential* trust between two agents a, called *potential truster*, and b, called *potential trustee*, and within communities $\mathcal{C}$. The idea is to define *potentiality* as satis*fiability*.

Definition 3 (Potential trust).

- *There is a potential weak (strong) trust relationship between a and b in the system S :iff* a trusts b *(a* $\mathsf{maytrust}$ *b) is satisfiable in the model $(\mathfrak{S}, \mathcal{V})$ induced by S.*
- *The community $\mathcal{C}$ is a potential weak (strong) trust domain in the system S :iff* $\mathsf{wTD}(\mathcal{C})$ *(*$\mathsf{sTD}(\mathcal{C})$*) is satisfiable in the model $(\mathfrak{S}, \mathcal{V})$ induced by S.*

Similarly, we define *actual* trust between two agents a, called *truster*, and b, called *trustee*, and within communities $\mathcal{C}$. The idea is to define (two degrees of) *actuality* as (two degrees of) satis*faction.*

Definition 4 (Actual trust).

- *There is a weak (strong) trust relationship between a and b in the model $(\mathfrak{S}, \mathcal{V})$ at the state $s \in \mathcal{S}$:iff* a trusts b *(a* $\mathsf{maytrust}$ *b) is satisfied in $(\mathfrak{S}, \mathcal{V})$ at s.*
- *There is a weak (strong) trust relationship between a and b in the model $(\mathfrak{S}, \mathcal{V})$:iff* a trusts b *(a* $\mathsf{maytrust}$ *b) is globally satisfied in $(\mathfrak{S}, \mathcal{V})$.*
- *The community $\mathcal{C}$ is a weak (strong) trust domain in the model $(\mathfrak{S}, \mathcal{V})$ at the state $s \in \mathcal{S}$:iff* $\mathsf{wTD}(\mathcal{C})$ *(*$\mathsf{sTD}(\mathcal{C})$*) is satisfied in $(\mathfrak{S}, \mathcal{V})$ at s.*
- *The community $\mathcal{C}$ is a weak (strong) trust domain in the model $(\mathfrak{S}, \mathcal{V})$:iff* $\mathsf{wTD}(\mathcal{C})$ *(*$\mathsf{sTD}(\mathcal{C})$*) is globally satisfied in $(\mathfrak{S}, \mathcal{V})$.*

Since satisfaction implies satisfiability, but not vice versa, actual trust implies potential trust, but not vice versa. For example, if two agents do not even know each other then they can not be in an actual trust relationship. However, they may be in a potential trust relationship: maybe in another system state, their trust potential can become actualised. On the other hand, in a given system two agents may well know each other but not be in a potential trust relationship: the system design might be such that trust between them is impossible.

3 Complexity results

We obtain our complexity results for deciding trust relationships and membership in trust domains by reduction to known results for the complexity of common belief and knowledge (cf. [28] and [27]). As usual for logics, the valuation function (here $\mathcal{V}$) acts as an oracle, which is assumed to decide in a single step whether or not an atomic proposition is true at the current state in the model induced by the considered distributed system S. However, deciding agent correctness is non-trivial and depends on S. For example, in our applications (cf. Section 4), deciding agent correctness can be at least polynomial in the size of the current state, which depending on the system modelling, may contain the history of system events. Another example is the case study in [38], where deterministically deciding agent correctness is quadratic in the size of the current state (containing the history of system events). That is, *state size is system-specific.* For example, when states contain the history of system events, state size can be defined as history length. Hence, we may have to account for the complexity of deciding agent correctness in the complexity of deciding trust relationships and membership in trust domains.

So, suppose that the truth of each atomic proposition can be deterministically decided in a polynomial number $f_S(|s|)$ of steps in the size $|s|$ of the state s of the model $(\mathfrak{S}, \mathcal{V})$ induced by S. The ideal case is when an agent produces its own, formal proof of correctness whenever requested to do so. (Proof-checking is tractable, whereas proof-finding is intractable.) We recall that since potential trust is defined in terms of satisfiability, and actual trust in terms of satisfaction, and since decidability of satisfiability implies decidability of satisfaction, decidability of potential trust implies decidability of actual trust, and complexities of potential trust are upper bounds for complexities of actual trust. Furthermore, satisfiability and validity are inter-solvable (ϕ is valid iff $\neg\phi$ is not satisfiable), and satisfiability complexities yield satisfaction (model-checking) complexities.

Theorem 3. *The computational time complexities of deterministically deciding trust relations is* $\mathcal{O}(f_S(|s|))$ *for potential and actual, weak and strong trust (cf. Table 4).*

Proof. Notice that our definitions of trust relations refer to a finite number (i.e., $|\mathcal{P}|$) of atomic propositions P, and that each definition uses exactly one atomic proposition (e.g., b $\mathsf{correct}$ a) and exactly one modal operator (e.g., B_a or K_a). Now according to [28], the complexity of the satisfiability of formulae $\mathsf{B}_a(\phi)$

Table 4. Computational time complexities

		Trust relations		**Trust domains**	
	degree	*weak* a trusts b	*strong* a maytrust b	*weak* $\mathsf{wTD}(\mathcal{C})$	*strong* $\mathsf{sTD}(\mathcal{C})$
actual	local satisfaction in models $(\mathfrak{S}, \mathcal{V})$ and states s	$\mathcal{O}(f_S(\lvert s\rvert))$		$\mathcal{O}(f_S(\lvert s\rvert) \cdot 2^{\lvert\mathcal{C}\rvert})$	
	global satisfaction in models $(\mathfrak{S}, \mathcal{V})$				
potential	satisfiability in models $(\mathfrak{S}, \mathcal{V})$				

Recall that there is a universal quantification over states s in the definition of global satisfaction, and an existential quantification over states s in the definitions of satisfiability in models.

and $\mathsf{K}_a(\phi)$ in a language with a finite number of atomic propositions and a bounded nesting depth of modal operators B_a and K_a is in (oracle) linear time in the length (here constantly 1) of the formula. Hence, the complexity of the satisfiability of formulae expressing weak and strong trust relations is even in (oracle) constant time, and thus $\mathcal{O}(f_S(|s|))$ without oracle. Yet $\mathcal{O}(f_S(|s|))$ is an absolute lower bound and thus the complexity of *all* trust relationships.

We can learn at least two lessons from these results. The first lesson is that we do have to account for the complexity of deciding agent correctness in the complexity of deciding trust relationships. The second lesson is that, surprisingly, deciding agent correctness is, up to a constant, equionerous to deciding potential and actual as well as weak and strong trust relationships.

Theorem 4. *The computational time complexities of deterministically deciding trust domains is $\mathcal{O}(f_S(|s|) \cdot 2^{|\mathcal{C}|})$ for potential and actual, weak and strong trust (cf. Table 4).*

Proof. According to [27], the complexity of the satisfiability of formulae $\mathsf{CB}_\mathcal{C}(\phi)$ and $\mathsf{CK}_\mathcal{C}(\phi)$ is in (oracle) deterministic single exponential time in the length of the sub-formula ϕ. The intuition is that formulae $\mathsf{CB}_\mathcal{C}(\phi)$ and $\mathsf{CK}_\mathcal{C}(\phi)$ correspond to formulae of infinitely deeply nested operators B_a and K_a with $a \in \mathcal{C}$, respectively, and that in that case, a finite number of atomic propositions does not help. In our case, the length of the conjunctive sub-formula in $\mathsf{wTD}(\mathcal{C})$ and $\mathsf{sTD}(\mathcal{C})$ is polynomial in the size $|\mathcal{C}|$ of the community $\mathcal{C}$. Further, the complexity of each conjunct is $\mathcal{O}(f_S(|s|))$, which we assumed to be polynomial. Yet a single exponential of a polynomial cost is still "only" a single exponential cost. Hence, the complexity of the satisfiability of formulae $\mathsf{wTD}(\mathcal{C})$ and $\mathsf{sTD}(\mathcal{C})$ is deterministic single exponential time in the size of $\mathcal{C}$ times $f_S(|s|)$. That is, the complexity of membership in potential weak and strong trust domains is $\mathcal{O}(f_S(|s|) \cdot 2^{|\mathcal{C}|})$. Finally according to [27], the operators $\mathsf{CB}_\mathcal{C}$ and $\mathsf{CK}_\mathcal{C}$ force satisfying models of a size that is exponential in $|\mathcal{C}|$. Hence, $\mathcal{O}(f_S(|s|) \cdot 2^{|\mathcal{C}|})$ is the complexity of *all*, potential or actual, memberships problems.

Corollary 4. *Computationally, composition of trust domains does not scale.*

Proof. By Theorem 2 and 4.

4 Application to cryptographic-key management

We instantiate our generic trust concepts in five major cryptographic applications of trust, namely: access control (cf. Section 1.1), Trusted Third Parties (TTPs), the Web of Trust, Public-Key Infrastructures (PKIs), and ID-Based Cryptography. For the latter three, we will have to define the valuation function $\mathcal{V}$ on the atomic propositions a $\mathsf{correct}$ b about agent correctness (cf. Definition 1 and Table 3). (We will also have to refine the definition of trust domains in Table 3 for the latter two.) That is, ***each notion of agent correctness is***[12] ***specific to each system rather than general to all systems.*** (Thus, ***trust is system-specific to some extent.***) However, we can define agent correctness *generically* for the Web of Trust and PKIs, with the aid of the following, common auxiliary logic, called *AuxLog*. The logic is a modal fixpoint logic [12] operating on points that are agents $a \in \mathcal{A}$ rather than states $s \in \mathcal{S}$. AuxLog is *parametric* in a binary relation $R \subseteq \mathcal{A} \times \mathcal{A}$ to be fixed separately for the Web of Trust and PKIs, but with the commonality of depending on a fixed state s ($R \in \{\mathrm{DTI}_s, \mathrm{CERT}_s\}$). In other words, we define:

1. trust relations from belief in or even knowledge of agent correctness;
2. agent correctness in the Web of Trust and PKIs with the aid of AuxLog as the existence of reliable designated-trusted-introducer and certification relationships, respectively, whose reliability is grounded in the actual secrecy of the related agents' private keys.

That is, we define trust relations (and domains) from the (common) knowledge of the existence of certain other, application-specific reliable relationships.

Definition 5 (Auxiliary Logic). *Let $\mathcal{X}$ designate a countable set of propositional variables C, and let*

$$\mathcal{L}' \ni \alpha ::= \mathsf{OK} \mid \mathsf{k}_b \mid C \mid \neg\alpha \mid \alpha \wedge \alpha \mid \overline{\Box}\alpha \mid \boldsymbol{\nu} C(\alpha)$$

designate the language $\mathcal{L}'$ of AuxLog, where $b \in \mathcal{A}$ and all free occurrences of C in α of $\boldsymbol{\nu} C(\alpha)$ are assumed to occur within an even number of occurrences of $\neg$ to guarantee the existence of (greatest) fixpoints (expressed by $\boldsymbol{\nu} C(\alpha)$) [12]. Then, given a relation $R \subseteq \mathcal{A} \times \mathcal{A}$, decidable in deterministic constant time, but structurally arbitrary, and an auxiliary interpretation $[\![\cdot]\!] : \mathcal{X} \cup \{\mathsf{OK}, \mathsf{k}_b\} \to 2^{\mathcal{A}}$ partially pre-defined as[13]

$$[\![\mathsf{OK}]\!] := \{\ a \in \mathcal{A} \mid \textit{at most } a \textit{ can access } a\textit{'s private key}\ \}$$
$$[\![\mathsf{k}_b]\!] := \{\ a \in \mathcal{A} \mid \textit{at least } a\textit{'s address is known to } b\ \},$$

[12] like policies, which induce notions of agent correctness, as mentioned in Section 1.1

[13] The phrasing of OK can be made formal provided that a notion of data space $\mathcal{D}$ (including keys) and data derivation for agents is fixed, e.g., à la Dolev-Yao [20]. Then data derivation can be formalised as a relation $\vdash \subseteq 2^{\mathcal{D}} \times \mathcal{D}$ (the first formalisation was in terms of closure operators [52]), and the phrase "at most a can access a's private key" as "for all $b \in \mathcal{A}$, if k is the private key of a and $\mathrm{m}_b(s) \vdash k$ then $b = a$", where $\mathrm{m}_b(s)$ returns the set of data that b generated or received as such in s. Note however that depending on the structure of $\mathcal{D}$, the computability of $\vdash$ may range from polynomial time to undecidability [58].

the interpretation $\|\cdot\|_{[\![\cdot]\!]} : \mathcal{L}' \to 2^{\mathcal{A}}$ *of AuxLog-propositions is as follows:*

$$
\begin{aligned}
\|\mathsf{OK}\|_{[\![\cdot]\!]} &:= [\![\mathsf{OK}]\!] \\
\|\mathsf{k}_b\|_{[\![\cdot]\!]} &:= [\![\mathsf{k}_b]\!] \\
\|C\|_{[\![\cdot]\!]} &:= [\![C]\!] \\
\|\neg\alpha\|_{[\![\cdot]\!]} &:= \mathcal{A} \setminus \|\alpha\|_{[\![\cdot]\!]} \\
\|\alpha \wedge \alpha'\|_{[\![\cdot]\!]} &:= \|\alpha\|_{[\![\cdot]\!]} \cap \|\alpha'\|_{[\![\cdot]\!]} \\
\|\overline{\Box}\alpha\|_{[\![\cdot]\!]} &:= \{\ a \in \mathcal{A} \mid \textit{for all } b \in \mathcal{A}, \textit{ if } b\ R\ a \textit{ then } b \in \|\alpha\|_{[\![\cdot]\!]}\ \} \\
\|\boldsymbol{\nu} C(\alpha)\|_{[\![\cdot]\!]} &:= \bigcup \{\ A \subseteq \mathcal{A} \mid A \subseteq \|\alpha\|_{[\![\cdot]\!]_{[C \mapsto A]}}\ \},
\end{aligned}
$$

where $[\![\cdot]\!]_{[C\mapsto A]}$ *maps* C *to* A *and otherwise agrees with* $[\![\cdot]\!]$.

Further, $\alpha \vee \alpha' := \neg(\neg\alpha \wedge \neg\alpha')$, $\top := \alpha \vee \neg\alpha$, $\bot := \neg\top$, $\alpha \to \alpha' := \neg\alpha \vee \alpha'$, $\alpha \leftrightarrow \alpha' := (\alpha \to \alpha') \wedge (\alpha' \to \alpha)$, $\overline{\Diamond}\alpha := \neg\overline{\Box}(\neg\alpha)$, *and, notably,* $\boldsymbol{\mu} C(\alpha(C)) := \neg\boldsymbol{\nu} C(\neg\alpha(\neg C))$.

Finally, for all $a \in \mathcal{A}$ *and* $\alpha \in \mathcal{L}'$,

$$\langle(\mathcal{A}, R), [\![\cdot]\!]\rangle, a \vDash \alpha \quad \textit{:iff} \quad a \in \|\alpha\|_{[\![\cdot]\!]},$$

and for all $\alpha \in \mathcal{L}'$,

$$\vDash \alpha \quad \textit{:iff} \quad \textit{for all } [\![\cdot]\!] \textit{ and } a \in \mathcal{A},\ \langle(\mathcal{A}, R), [\![\cdot]\!]\rangle, a \vDash \alpha.$$

The reader is invited not to confuse the auxiliary satisfaction relation $\vDash$ of AuxLog with $\models$, the main one from Definition 1. Further note that AuxLog is a member of the family of μ-calculi over the modal system **K**, which is characterised by the laws of propositional logic and the modal laws $\vDash \overline{\Box}(\alpha \to \alpha') \to (\overline{\Box}\alpha \to \overline{\Box}\alpha')$ and "if $\vDash \alpha$ then $\vDash \overline{\Box}\alpha$". The reason is that, as mentioned, $R \subseteq \mathcal{A} \times \mathcal{A}$ is structurally arbitrary. Hence, no more structural properties than those of the modal system **K**, i.e., none, can generally be assumed to hold for $\overline{\Box}$. As a corollary, the model-checking problem, i.e., "Given $a \in \mathcal{A}$ and $\alpha \in \mathcal{L}'$, is it the case that $\langle(\mathcal{A}, R), [\![\cdot]\!]\rangle, a \vDash \alpha$?" is decidable in deterministic polynomial time in the size of α. See [12] for details.

4.1 Trusted Third Parties

The concept of a Trusted Third Party (TTP) is a folklore concept for much of information security, e.g., in PKIs as registration (e.g., the town hall or the local post office, the HR department of a company), certification (e.g., a company, the IT department of a company), and key-escrow authorities [19], for many protocols for authentication and key establishment [11], as well as for secure multiparty computation [62]. Recall that "a secure multiparty computation for function f can be viewed as an implementation of a trusted third party T, which, upon receipt of the input values $x_1, \ldots, x_n$ from parties $P_1, \ldots, P_n$, respectively, produces the output value $y = f(x_1, \ldots, x_n)$. Party T is trusted for (i) providing

Table 5. (Trustworthy) Trusted Third Parties

$\mathsf{wTTP}(c,a,b) := \mathsf{CB}_{\{a,b,c\}}(a \text{ musttrust } c \land b \text{ musttrust } c)$	c is a weak TTP of a and b
$\mathsf{sTTP}(c,a,b) := \mathsf{CK}_{\{a,b,c\}}(a \text{ musttrust } c \land b \text{ musttrust } c)$	c is a strong TTP of a and b
$\mathsf{wtTTP}(c,a,b) := \mathsf{CB}_{\{a,b,c\}}(\mathsf{wTD}(\{c,a\}) \land \mathsf{wTD}(\{c,b\}))$	c is a weakly trustworthy TTP of a and b
$\mathsf{stTTP}(c,a,b) := \mathsf{CK}_{\{a,b,c\}}(\mathsf{sTD}(\{c,a\}) \land \mathsf{sTD}(\{c,b\}))$	c is a strongly trustworthy TTP of a and b

the correct value for y and (ii) [n]ot revealing any further information to parties $P_1, \ldots, P_n$" [60]. In our terminology of agent correctness, the conjunction of Condition (i) and (ii) informally stipulates what it means for T to be correct. Notice that the above definition of secure multiparty computation merely defines the *object* of trust (i.e., agent correctness), but not trust itself (which, in our definition, is belief *in* or even knowledge *of* agent correctness). To the best of our knowledge, the concept of a TTP has never been formally defined, i.e., mathematically analysed into its conceptual constituents. Here, we are able to define the TTP-concept in terms of our trust concepts (cf. Table 5), and instantiate TTPs in the Web of Trust (cf. Section 4.2) and Public-Key Infrastructures (PKIs) (cf. Section 4.3). More precisely, we define the concepts of weak and strong as well as weakly and strongly *trustworthy* TTPs. Trustworthy TTPs are TTPs that may or even must *deserve* the trust of their trusters—and vice versa. Note that the trustworthiness of TTPs (e.g., the certification authorities in a PKI) is absolutely crucial, without which whole security architectures (e.g., an international PKI for ePassports [40]) can break down. Observe that thanks to Theorem 2, the two sides, e.g., $\{c,a\}$ and $\{c,b\}$, in a strongly (but not in a weakly) trustworthy TTP constitute a (strong) trust domain as a whole, i.e., as $\{c,a\} \cup \{c,b\}$ on the sufficient condition $\mathsf{CK}_{\{c,a\}\cup\{c,b\}}(\bigwedge_{x,y\in\{c,a\},z\in\{c,b\}}(y \text{ correct } x \to y \text{ correct } z) \land \bigwedge_{x,y\in\{c,b\},z\in\{c,a\}}(y \text{ correct } x \to y \text{ correct } z))$.

4.2 The Web of Trust

In the (decentralised) Web of Trust, as defined by Philip Zimmermann (cf. [65] and [19]) for PGP in 1992, any agent can independently establish its own domain of trusted correspondents by publicly designating (e.g., on their homepage) so-called *trusted introducers*, who by this very act become commonly known as such. In PGP, the designation of a trusted introducer is implemented as the (publicly exportable) signing of the designated trusted introducer's public key with the designator's private key. Additional assurance can be provided by The Global Internet Trust Register [5]. The role of an agent a's trusted introducer b is to act as a *guarantor* for the trustworthiness of a, and by that, to catalyse the building up of trust relationships between a and those agents c who are only potential (not yet actual) trustees of a but who are (already) actual trustees of b. Notice the importance of distinguishing between potential and actual trust (cf. Definition 3 and 4). Thus, the more guarantors (actual trustees) an agent (as an actual truster) has, the more potential trustees the agent (as a potential truster) has. In the Web of Trust, agents are (socially speaking) trustworthy, or (technically speaking) ***correct if and only if all their designated trusted introducers are, and at most they (the correct agents) can access their (own) private key.*** (Agents with untrustworthy introducers or a corrupt private key are untrustworthy.) Notice the possible *mutuality* in this social notion of agent correctness.

We model the designated-trusted-introducer relationships between agents in system states $s \in \mathcal{S}$ with a family of relations (a kind of data base) $\mathrm{DTI}_s \subseteq \mathcal{A}\times\mathcal{A}$ such that

$$b\,\mathrm{DTI}_s\,a \text{ :iff } b \text{ is a designated trusted introducer of } a \text{ in } s.$$

The valuation function $\mathcal{V}$ on the propositions a $\mathsf{correct}$ b can then be formally defined with the aid of AuxLog as follows:

$$\boxed{\mathcal{V}(a\ \mathsf{correct}\ b) := \{\ s \mid \langle(\mathcal{A}, \mathrm{DTI}_s), \emptyset\rangle, a \vDash \boldsymbol{\nu} C(\mathsf{k}_b \to (\mathsf{OK} \land \overline{\Box}C))\ \},}$$

where $\emptyset$ designates the empty auxiliary interpretation (C is bound!). The greatest-fixpoint assertion $\langle(\mathcal{A}, \mathrm{DTI}_s), \emptyset\rangle, a \vDash \boldsymbol{\nu} C(\mathsf{k}_b \to (\mathsf{OK} \land \overline{\Box}C))$ says that a is in the greatest fixpoint of the interpretation of *the property C such that:*

> if a satisfies C (i.e., a is in the interpretation of C) then a satisfies $\mathsf{k}_b \to (\mathsf{OK} \land \overline{\Box}C)$, which in turn says that if a is known to b then at most a can access a's private key and for all $b \in \mathcal{A}$, if b is a designated trusted introducer of a in the state s then b satisfies C.

Observe that all (=1) free occurrences of C in $\mathsf{k}_b \to (\mathsf{OK} \land \overline{\Box}C)$ of $\boldsymbol{\nu} C(\mathsf{k}_b \to (\mathsf{OK} \land \overline{\Box}C))$ occur within an even (=0) number of occurrences of $\neg$. Hence, our definition is formally well-defined. Further observe that the possible mutuality in our notion of agent correctness corresponds to the co-inductiveness of the greatest fixpoint, which allows direct (*self-designation*) and indirect (*mutual designation*) loops in the designated-trusted-introducer relationships. The interpretation of the corresponding (inductive) least-fixpoint formula would wrongly not allow such loops. Of course, the language of AuxLog allows for other, more complex definitions of agent correctness, e.g., ones disallowing self-designation[14], and/or ones with a more complex notion of being OK. Our present definition is just an inceptive example proposal. The *co*-inductive definition has the following iterative paraphrase from *above* (iterated *de*construction).

> Everybody is correct (the Web of Trust is born in the plenum, so to say); except for the following agents (*ex*clude those which are clearly not OK):
> 0. agents with a corrupt private key (Type 0 agents);

[14] Yet whether or not you declare yourself as trustworthy may well be legally critical (cf. for example such hand-written self-declarations in certain immigration procedures).
Anyway, the disallowance of self-designation could be implemented by introducing a binding operator $\downarrow$ à la hybrid logic [6] into the language of AuxLog, such that $\langle(\mathcal{A}, R), [\![\cdot]\!]\rangle, a \vDash {\downarrow}C(\alpha)$:iff $\langle(\mathcal{A}, R), [\![\cdot]\!]_{[C\mapsto\{a\}]}\rangle, a \vDash \alpha$, and stipulating that $\mathcal{V}(a\ \mathsf{correct}\ b) := \{\ s \mid \langle(\mathcal{A}, \mathrm{DTI}_s), \emptyset\rangle, a \vDash \boldsymbol{\mu} C(\mathsf{k}_b \to (\mathsf{OK} \land \neg{\downarrow}C'(\overline{\Diamond}C') \land \overline{\Box}C))\ \}$.

1. agents with a designated trusted introducer of Type 0 (Type 1 agents);
2. agents with a designated trusted introducer of Type 1 (Type 2 agents);
3. etc.

Clearly, weak or strong trust relations in the Web of Trust must be universal within an agent's domain of correspondents in the sense of Proposition 1: designated trusted introducers are trusted; and they would not act as such, if they did not trust their designator and their designator's other designated trusted introducers, etc. Hence, our trust relations and trust domains from Table 3 as well as our weakly and strongly trustworthy TTPs from Table 5 are fit for the Web of Trust without further adaptation.

4.3 Public-Key Infrastructures

In Public-Key Infrastructures (PKIs), centralised *certificate authorities* (CAs) act as guarantors for the trustworthiness of the public key of their clients by issuing certificates that bind the public key of each client (the legitimate key owner) to the client's (unique) name (cf. [23, Chapters 18–20], [21], [42], [19], and [1–3]). (The legitimacy of a key ownership is warranted by the *registration authority* with which the key owner authenticated in person, i.e., validated the correspondence of her name to her person.) According to [23, Page 284]:

> Key management is the most difficult problem in cryptography, and a PKI system is one of the best tools that we have to solve it with. But everything depends on the security of the PKI, and therefore on the trustworthiness of the CA.

In PKIs, agents are (socially speaking) trustworthy, or (technically speaking) ***correct if and only if all their certified agents are, and at most they (the correct agents) can access their (own) private key.*** (Agents who certify incorrect agents or agents with a corrupt private key are incorrect.) Notice the absence of mutuality in this notion of agent correctness; it is intrinsically unilateral. However, the notion of PKI-trust to be built from this notion of agent correctness will be again bilateral (i.e., mutual, and thus symmetric): the certifying correct agent trusts the certified correct agent, and vice versa.

We model the relationships from certifying agents to certified agents (which may themselves be certifying agents to agents certified by them, etc.) in system states $s \in \mathcal{S}$ with a family of relations (again a kind of data base) $\mathrm{CRT}_s \subseteq \mathcal{A} \times \mathcal{A}$ such that

$$b \; \mathrm{CRT}_s \; a \text{ :iff } b \text{ is certified by } a \text{ in } s,$$

where "b is certified by a in s" means "a has issued a valid certificate for b in s", i.e., a certificate that is non-revoked and non-suspended in s and signed by a with the private key of a. The valuation function $\mathcal{V}$ on the propositions $a \; \mathsf{correct} \; b$ can then be formally defined with the aid of AuxLog as follows:

$$\boxed{\mathcal{V}(a \; \mathsf{correct} \; b) := \{ \; s \mid \langle (\mathcal{A}, \mathrm{CRT}_s), \emptyset \rangle, a \vDash \boldsymbol{\mu} C(\mathsf{k}_b \to (\mathsf{OK} \wedge \overline{\Box} C)) \; \}.}$$

The least-fixpoint assertion $\langle(\mathcal{A}, \mathrm{CRT}_s), \emptyset\rangle, a \vDash \boldsymbol{\mu} C(\mathsf{k}_b \rightarrow (\mathsf{OK} \wedge \overline{\Box} C))$ says that a is in the least fixpoint of the interpretation of *the property* C *such that:*

> if a satisfies $\mathsf{k}_b \rightarrow (\mathsf{OK} \wedge \overline{\Box} C)$ (i.e., a is in the interpretation of $\mathsf{k}_b \rightarrow (\mathsf{OK} \wedge \overline{\Box} C)$)—which in turn says that if a is known to b then at most a can access a's private key and for all $b \in \mathcal{A}$, if b is certified by a in the state s then b satisfies C—then a satisfies C.

Observe that certification is unilateral (i.e., non-mutual, and thus not symmetric) in the sense that certification relationships must not be directly (*self-certification*) nor indirectly (*mutual certification*) looping, which forces a least-fixpoint formulation. The PKI-approach to trustworthiness is thus diametrically opposed to the approach of the Web of Trust. This opposition is reflected first, in the least/greatest fixpoint "duality" of the two paradigms; and second, in the fact that PKIs are based on (ultimately national) *authority* (hierarchical CAs), whereas the Web of Trust is based on (borderless) *peership* (peer-guarantors). (At the international level, it makes sense to link two national [root] CAs via the *bona officia* of a trustworthy trusted third party [a national *bridge CA*], or to organise a group of national CAs as a Web of Trust [*cross* or *mesh certification*] such as the PKIs required by ePassports [40].) Yet again of course, as with the Web of Trust, the language of AuxLog allows for other, more complex definitions of agent correctness, e.g., ones with self-certification for the root-CA[15] (the *trust anchor*), and/or ones with a more complex notion of being OK (e.g., including the possibility of *key escrow*). Again, our present definition is just an inceptive example proposal. The *in*ductive definition has the following iterative paraphrase from *below* (iterated *con*struction).

> Nobody is correct (PKIs are born *ex nihilo*, so to say); except for the following agents (*in*clude those which are clearly OK): agents without a corrupt private key (Type 0 agents), whose certified agents are also of Type 0 (Type 1 agents), whose certified agents are again also of Type 0 (Type 2 agents), etc. (In other words, being of Type 0 is an invariant in the transitive closure of certification relationships.)

Notice the structural difference between this paraphrase for agent correctness in PKIs and the previous one for agent correctness in the Web of Trust. The "duality" is not pure.

As suggested, CAs are commonly organised in a hierarchy, which induces *structured trust domains* in the form of finite trees. Recall that a finite tree is a partially-ordered set (here say $\langle\mathcal{C}, \leq\rangle$) with a bottom (top) element such that for each element in the set, the down-set (up-set) of the element is a finite chain [17]. In PKI trust domains, trust relations are symmetric (up- and downwards the tree branches) and transitive (along the tree branches) but not universal (after all, a tree is a tree and not felt fabric), and the root CA corresponds to

[15] Self-certification for the root-CA could be implemented by introducing an atomic proposition root true at and only at the root-CA agent, and stipulating that $\mathcal{V}(a \text{ correct } a) := \{\ s \mid \langle(\mathcal{A}, \mathrm{CRT}_s), \emptyset\rangle, a \vDash (\mathsf{root} \rightarrow \overline{\Diamond}\mathsf{root}) \wedge \boldsymbol{\mu} C(\mathsf{OK} \wedge \overline{\Box} C)\ \}$.

Table 6. Public-Key Infrastructure trust domains

$\mathsf{wTD}_{\mathsf{PKI}}(\mathcal{C}) := \mathsf{CB}_{\mathcal{C}}(\bigwedge_{a,\,b\,\in\,\mathcal{C}\ \text{and}\ (a \leq b\ \text{or}\ b \leq a)} a\ \mathsf{trusts}\ b)$	$\mathcal{C}$ is a weak PKI trust domain
$\mathsf{sTD}_{\mathsf{PKI}}(\mathcal{C}) := \mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,\,b\,\in\,\mathcal{C}\ \text{and}\ (a \leq b\ \text{or}\ b \leq a)} a\ \mathsf{maytrust}\ b)$	$\mathcal{C}$ is a strong PKI trust domain.

the tree root, the intermediate CA's correspond to the intermediate tree nodes, and the clients to the tree leafs. Hence we can fit our weak and strong trust domains to PKI trust domains $\langle \mathcal{C}, \leq \rangle$ by simply stipulating that the conjunction in the respective definition respect the finite-tree structure $\leq$ of $\mathcal{C}$, and reflect the symmetry and transitivity of the trust relations. And that is all: see Table 6. We can now instantiate our weakly and strongly trustworthy TTPs from Table 5 for PKIs as

$$\mathsf{wtTTP}_{\mathsf{PKI}}(c, a, b) := \mathsf{CB}_{\{a,b,c\}}(\mathsf{wTD}_{\mathsf{PKI}}(\{c, a\}) \wedge \mathsf{wTD}_{\mathsf{PKI}}(\{c, b\}))$$
$$\mathsf{stTTP}_{\mathsf{PKI}}(c, a, b) := \mathsf{CK}_{\{a,b,c\}}(\mathsf{sTD}_{\mathsf{PKI}}(\{c, a\}) \wedge \mathsf{sTD}_{\mathsf{PKI}}(\{c, b\})),$$

respectively, with $\leq := \{(c, a), (c, b)\}$ as (tree) domain structure. Fits for trust domains with other structures (e.g., buses, chains, rings, stars, etc.) can be made by similarly simple stipulations (e.g., for computing clouds, which have not a fixed but a dynamic, i.e., an evolving structure).

In sum, a weak or strong PKI trust domain $\mathcal{C}$ is built from a *certification hierarchy* $\leq$ of certifying (CAs) and certifiable agents $a \in \mathcal{C}$ such that for all states $s \in \mathcal{S}$ there is a (possibly empty) *certification record* $\{\ b \in \mathcal{C} \mid b\ \mathrm{CRT}_s\ a\ \}$. Thereby, the certification hierarchy acts as a constraining skeleton (there is no such skeleton in the Web of Trust; it is free) for the potential and the actual trust relationships in $\mathcal{C}$, and, by that, the respective memberships in the trust domain $\mathcal{C}$ itself. And the certification records act as evidential support for the actuality of the trust relationships and the memberships in the trust domain.

4.4 Identity-Based Cryptography

Identity-Based Cryptography is a variation of Public-Key Cryptography in which the intending sender of a message derives the (public) encryption key from the public identity (e.g., a telephone number, an email address, etc., or a combination thereof) of the intended recipient [31]. In our setting, we abstractly model an agent a's public identity with the symbol 'a'. For the sake of the security of ID-based encryption, an ID-based private key must not be derivable from its corresponding public counterpart without an additional trap-door information. This trap-door information is owned by a central CA (cCA $\in \mathcal{A}$), which therefore can derive the private keys of all its certified agents. (Thus ID-based domains have a star structure: $\leq$ is an n-ary tree of depth 1 with as root cCA and as leaves the n agents certified by cCA.) Hence for ID-Based Cryptography, the definition of an agent being OK (cf. Section 4) must be weakened, e.g.,

$$[\![\mathsf{OK}]\!] := \{\ a \in \mathcal{A} \mid \text{at most } a \text{ and cCA can access } a\text{'s private key}\ \}.$$

Table 7. Generic definition of accountability

$\mathsf{Accountability}$:=	$\mathsf{Abusefreeness} \land \mathsf{Auditability}$
$\mathsf{Abusefreeness}$:=	$\forall a \forall b \Box (a \ \mathsf{correct}\ b \rightarrow \exists/\forall c \mathsf{P}_{(a,c)}(a \ \mathsf{correct}\ b))$
$\mathsf{Auditability}$:=	$\forall a \forall b \Box (\neg(a \ \mathsf{correct}\ b) \rightarrow \exists/\forall c \Diamond \Box \mathsf{P}_{(b,c)}(\neg(a \ \mathsf{correct}\ b)))$

Note that as a consequence, the notion of trust (and thus the value of the trustworthiness of cCA) is weakened! Of course, a restrengthening is possible, e.g., by stipulating that cCA has not *used* the private keys of its certified agents (except possibly for *key escrow*). In sum, the flexibility of ID-Based Cryptography for its users (the certified agents) is paid with a devaluation of the trustworthiness of its provider (cCA).

4.5 Accountable access control and cryptographic-key management

In [38], a generic definition of *accountability* is given, which can be instantiated in one fell swoop with our present notions of agent correctness for access control from Section 1.1 and cryptographic-key management from Sections 4.2–4.4 as displayed in Table 7. There, $\Box$ means "henceforth" and $\Diamond$ "eventually" in the sense of standard temporal logic, and $\mathsf{P}_{(a,b)}$ "a can prove to b that" in the sense of a standard (though even interactive) S4 notion of Gödel-style provability [34]. For accountability, abusefreeness means that correct agents can defend themselves against false accusations of incorrectness, and auditability that incorrect agents can eventually be found out. The connection to trust is that if b knows a's proof of a $\mathsf{correct}$ b then b knows that a is correct as far as b is concerned, and so abusefreeness induces strong trust in the sense of Table 3 (cf. [38] for details).

5 Trust building

Building trust in the sense of building *from absence of trust* (as opposed to *re*building from *dis*trust, cf. 2nd paragraph after Definition 1) is possible if and only if there is at least *potential* trust in the sense of Definition 3. That is, given $a, b \in \mathcal{A}$ and $\mathcal{C} \subseteq \mathcal{A}$, the formulae a trusts b or a $\mathsf{maytrust}$ b, and $\mathsf{wTD}(\mathcal{C})$ or $\mathsf{sTD}(\mathcal{C})$ must at least be *satisfiable in the model induced by the considered system*, in order for a weak or strong trust relationship from a to b to possibly exist, and in order for $\mathcal{C}$ to possibly be a weak or strong trust domain, respectively. Yet thanks to the computability of our notions of trust, a computer can aid us in our decision of whether or not building trust from a current absence of trust in a given system, and between a given pair of agents, or within a given (small) group of agents is actually possible.

When it is indeed possible to actualise a certain potential trust, the next question is how to actually *build up* the trust. Consider that a potential trustee (say a) has at least two non-mutally-exclusive possibilities of earning the trust of a potential truster (say b):

Table 8. The design pattern RD

1. **Input:** a pointed model $(\mathfrak{S}, \mathcal{V}), s$ and $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{A}$ such that

$$(\mathfrak{S}, \mathcal{V}), s \models \mathsf{CK}_{\mathcal{C}_1 \cup \mathcal{C}_2} \begin{pmatrix} \bigwedge_{a,b \in \mathcal{C}_1, c \in \mathcal{C}_2} (b \text{ correct } a \to b \text{ correct } c) \land \\ \bigwedge_{a,b \in \mathcal{C}_2, c \in \mathcal{C}_1} (b \text{ correct } a \to b \text{ correct } c) \end{pmatrix};$$

2. **Divide:** for $i \in \{1, 2\}$ do {
   when $(\mathfrak{S}, \mathcal{V}), s \models \neg \mathsf{sTD}(\mathcal{C}_i)$:
   (a) divide $\mathcal{C}_i$ freely into $\mathcal{C}_{i.1}$ and $\mathcal{C}_{i.2}$;
   (b) $s := \mathrm{RD}((\mathfrak{S}, \mathcal{V}), s, \mathcal{C}_{i.1}, \mathcal{C}_{i.2})$; };
3. **Conquer:** announce to the community $\mathcal{C}_1 \cup \mathcal{C}_2$ that $\mathsf{sTD}(\mathcal{C}_1) \land \mathsf{sTD}(\mathcal{C}_2)$ is true (choose appropriate communication channels and an appropriate protocol), which takes the system from s to some $s' \in \mathcal{S}$ such that s' is reachable (cf. Footnote 9) from s and
$$(\mathfrak{S}, \mathcal{V}), s' \models \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \land \mathsf{sTD}(\mathcal{C}'));$$
4. **Output:** $s' \in \mathcal{S}$;
5. **Effect:** $(\mathfrak{S}, \mathcal{V}), s' \models \mathsf{sTD}(\mathcal{C}_1 \cup \mathcal{C}_2)$.

1. by *behaving correctly* according to the notion of agent correctness a correct b in the system where the computer has found the considered kind of trust to be actually possible. (Behaving correctly can include doing nothing, when the considered notion of agent correctness does not exclude such inactivity.)
2. by *producing evidence* (e.g., a recommendation) for *or proof* (e.g., a signed log file: [16], [38]) of behavioural correctness, whenever requested to do so by the potential truster (here in the role of an auditor).

The potential truster may then decide to trust the potential trustee based on the observation of the potential trustee's behaviour, and/or based on the knowledge of certain data, i.e., evidence or even proof produced by the potential trustee. Depending on the value of the information obtained, a relationship of weak or strong trust is established from the potential truster to the potential trustee, who have now become *de facto* an actual truster and trustee, respectively. This process of building up oriented trust relations can be extended to building up symmetric trust relationships, and trust domains of agents in such relationships. As a matter of fact, Theorem 2 yields a *design pattern*, called RD (cf. Table 8), for building up strong trust domains from a possible absence of trust in a given system via recursive descent. RD relies on the communication abstraction of *public announcement* (cf. 3rd paragraph after Definition 1), which we use as a black-box plug-in. We recall that the effect of a public announcement of a fact is to induce the common knowledge of that fact with the addressed public by minimally changing the epistemic state of each member (cf. [59] for details). Depending on the communication context, the mechanism ranges from trivial (e.g., in a public assembly) to difficult (e.g., with communication asynchrony), possibly requiring implementation as a full-fledged communication protocol. So, although 'announcement' may suggest triviality rather than difficulty, public announcements may well be non-trivial to implement in a given context, which is why we call RD just a *design pattern* rather than a full-fledged algorithm: Building up trust domains from possible absence of trust generally is computationally costly, in particular when made by means of the recursive-descent design pattern RD.

Theorem 5. *The complexity of building up trust domains is exponential in the number of potential members.*

Proof. By the fact that membership in trust domains has to be checked for each potential new member anew, which is exponential in the size of the trust domain being checked in a given model and at a given state (cf. Table 4).

In contrast, it is common knowledge (among humans) that for *destroying* actual trust relationships, and thus also trust domains, a single (side) step is sufficient — metaphorically speaking. And *re*building trust from *dis*trust is difficult. It might require *forgiving* in the sense of forgetting past violations of agent correctness, which would actually reduce rebuilding trust from distrust to building trust from absence of trust in a blank memory—after wiping the bad memories so to say.

6 Related work

There is a huge literature on notions of trust that are not formal in the sense of formal languages and semantics, and also on trust management, which however is not the subject matter of this paper.

We are aware of the following formal work related to ours.

6.1 Trust relations

As explained in the introduction, [18] presents various kinds of trust relations based on belief in and knowledge of mental attitudes, which is less general than belief in or knowledge of agent behaviour.

In [64], trust relationships are defined as four-tuples of a set R of trusters, a set E of trustees, a set C of conditions, and a set P of properties (the actions or attributes of the trustees). Thereby, conditions and properties are fully abstract, i.e., without pre-determined form. According to the authors, "Trust relationship T means that under the condition set C, truster set R trust that trustee set E have the properties in set P.", where the meaning of "trust that" is formally primitive and thus left to interpretation. Given that trust can possibly be wrong, a plausible interpretation of "trust that" could be "believe that" (rather than "know that"). The authors then define operations on trust relationships in terms of set-theoretic operations on the projections of their tuples. We attempt to relate the authors' notion of trust relationship to our notion of weak trust relation (based on belief rather than knowledge) by coercing their definition of T in the following macro-definition of our logic, assuming their P is included in our $\mathcal{P}$:

$$T(C, R, E, P) := (\bigwedge_{c \in C} c) \rightarrow \bigwedge_{r \in R} \mathsf{B}_r(\bigwedge_{e \in E} \bigwedge_{p \in P} p(e)).$$

If the authors agree with this coercion, we have the following definability result:

$$\models T(\{\top\}, \{a\}, \{b\}, \{\cdot \text{ correct } a\}) \leftrightarrow a \text{ trusts } b,$$

where $\cdot$ correct a is an attribute in the authors' sense.

In [61], "trust is a state at which the host believes, expects, or accepts that the effects from the cleint are the positive", although belief is not formal in the sense of modal logic. The authors' idea is that "the host's trust on a client is obtained based on the trust evaluation of the client by the host. When the host trusts a client, the host will believe, expect or accept that the client will do no harm to the host in the given context". Hence, this notion of trust is agent-centric in the sense of being defined in terms of (local) effects at an agent's location. This is a less general notion of trust than ours, which is *systemic* in the sense of being defined in terms of correct agent behaviour within a certain system. Also, we recall again that agent correctness is a standard primitive in the distributed-systems community [44].

In [14], a domain-theoretic model of trust relations between agents is presented. In that model, a given directed trust relation from a truster to a trustee is abstracted as a value reflecting the truster's degree of trust in the trustee. Thus, [14]'s notion of trust is, as opposed to ours, quantitative, but, as opposed to ours, lacks a behavioural (e.g., in terms of agent correctness) and doxastic/epistemic explication (in terms of belief/knowledge). The purpose of the model is the definition of a unique global trust state arising from the trust relations between the agents via trust policies. Complexities for computing portions of that global state are given in [39].

In [45] and [43], axiomatic frameworks for the special purpose of modelling PKIs with the relevant trust relationships are presented. However, the authors do not attempt to actually explicate what trust is. In fact, their respective framework is restricted to the mere *declaration* of the involved trust relationships. The frameworks also ignore that PKIs involve in addition to the trust relations themselves the belief in or even knowledge of these relations.

After the appearance of our work in [36] and [35], appeared [49], in which the author proposes an epistemic semantics of trust relations, but not of trust domains (and without any mathematical result). The author's interpretation of trust is: "If agents know that a target is trustworthy under an interpretation, that agent trusts the target." So the author defines trust in terms of knowledge of trustworthiness, which makes his definition (at least partially) conceptually circular. Whereas we define (dis)trust relations as belief in agent (in)correctness, mistrust relations as mistaken belief in agent correctness, may-trust (trustworthiness) relations as knowledge of agent correctness, and must-trust (trustedness) relations as epistemic undecidability of agent correctness (cf. Table 3). The author then also applies his trust definitions to PKIs and the Web of Trust.

6.2 Trust domains

To the best of our knowledge, the only formal piece of work on trust domains is [63], based on description logic. However, the authors' definition remains a conceptual *modelling,* and moreover one that is limited to PKIs.

6.3 Trust in statements

Four formal notions of trust loosely related to ours are the following. In [29], so-called *trust in belief* and *trust in performance* are formalised in the situation calculus, such that "axioms and theorems can be used to infer whether a thing can be believed". Whereas we define (weak) trust *in agents*, and moreover *in terms of belief* taken off-the-shelf as a standard primitive. In [30], the ideas of trust in belief and trust in performance are taken up again similarly. In [53], trust in a distributed system is captured as the trusted statements about the system, which are added as axioms to the axiomatics of a standard logic of belief. Then in [41], trust in the judgement about the truth of a statement about a distributed system is captured by means of a modal operator $T_{ij}(\phi)$ with the intended meaning "that agent i trusts the judgement of j on the truth of ϕ".

7 Conclusion

Assessment We have delivered simple smooth definitions of weak and strong trust relations and trust domains, and potential and actual trust relationships and membership in trust domains, as well as practical complexity, compositionality, scalability, and transitivity results—and all that in a single standard minimal framework. Moreover, we have provided a design pattern for building up strong trust domains. All our definitions have the advantage of being both declarative *and* computational, as well as parametric in the notion of agent correctness. Thanks to being declarative, our definitions are independent of the manifold manifestations of trust establishment (e.g., via recommendations and/or reputation). They are meaningful in any concrete distributed system with a notion of agent correctness and state space. We recall that agent correctness is a primitive in the distributed-systems community [44], and that state space is forced in a world of digital computers. A surprising insight gained from our computational analysis of trust is that given weak trust, strong trust is for free (up to a constant) from the point of view of complexity theory. A further insight gained from our analysis is that trust can be related to accountability so that accountability induces trust. Finally, we have shown that our trust domains are fit as such for TTPs and the Web of Trust, and that with a minor modification in the form of a constraint, they can be made to fit PKIs, ID-Based Cryptography, and others.

Future work We could unify our notions of weak and strong trust relation (domain) in a notion of *graded trust relation* (*domain*) defined in terms of graded (common) belief instead of plain (common) belief and plain (common) knowledge, respectively [37]. Informally, knowledge is belief with 100% certitude. And with the additional introduction of *temporal* modalities, it becomes possible to study the evolution of the quality and quantity of trust in a given distributed system, by observing the evolution of the grade of each trust relation and trust domain in the system. This could be especially interesting for cryptographic protocols, whose function is, according to [23, Section 13.4, Page 217],

> [...] to minimize the amount of trust required. This is important enough to repeat. The function of cryptographic protocols is to minimize the amount of trust required.

Finally, we could build actual trust-management systems for trust relations and trust domains in our present sense of building trust *from absence of trust* and in a future sense of *re*building trust from *dis*trust.

Acknowledgements The first author thanks R. Ramanujam and S.P. Suresh, and S. Arun-Kumar and S. Prasad, as well as their respective students for delightful discussions on agent correctness and trust.

References

1. Asia PKI consortium. `http://www.apkic.org/`.
2. EuroPKI. `http://www.europki.org/`.
3. Federal PKI. `http://csrc.nist.gov/groups/ST/crypto_apps_infra/pki/index.html`.
4. R. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley, second edition, 2008.
5. R. Anderson, B. Crispo, J.-H. Lee, C. Manifavas, V. Matyas, and F. Petitcolas. *The Global Internet Trust Register*. The MIT Press, 1999.
6. C. Areces and B. ten Cate. *Handbook of Modal Logic*, chapter Hybrid Logics. Volume 3 of Blackburn et al. [10], 2007.
7. A. Arenas, M. Wilson, and B. Matthews. On trust management in grids. In *Proceedings of the ACM Conference on Autonomic Computing and Communication Systems*, 2007.
8. K. Binmore. *Game Theory: A Very Short Introduction*. Oxford University Press, 2007.
9. P. Blackburn and J. van Benthem. *Handbook of Modal Logic*, chapter Modal Logic: A Semantic Perspective. Volume 3 of Blackburn et al. [10], 2007.
10. P. Blackburn, J. van Benthem, and F. Wolter, editors. *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*. Elsevier, 2007.
11. C. Boyd and A. Mathuria. *Protocols for Authentication and Key Establishment*. Springer, 2003.
12. J. Bradfield and C. Stirling. *Handbook of Modal Logic*, chapter Modal Mu-Calculi. Volume 3 of Blackburn et al. [10], 2007.
13. C. Cachin, I. Keidar, and A. Shraer. Trusting the cloud. Technical Column of ACM SIGACT News, June 2009.
14. M. Carbone, M. Nielsen, and V. Sassone. A formal model for trust in dynamic networks. In *Proceedings of the IEEE Conference on Software Engineering and Formal Methods*, 2003.

15. V.G. Cerf. Trust and the Internet. *IEEE Internet Computing*, 14(5), 2010.
16. A. Chuvakin. *Beautiful Security: Leading Security Experts Explain How They Think*, chapter Beautiful Log Handling. In Oram and Viega [50], 2009.
17. B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990 (2002).
18. R. Demolombe. Reasoning about trust: A formal logical framework. volume 2995 of *LNCS*. Springer, 2004.
19. Y.G. Desmedt, editor. *Secure Public Key Infrastructure*. Springer, 2012. To appear.
20. D. Dolev and A. Yao. On the security of public key protocols. *IEEE Transactions on Information Theory*, 29(12), 1983.
21. C. Ellison and B. Schneier. Ten risks of PKI: What you're not being told about public key infrastructure. *Computer Security Journal*, 16(1), 2000.
22. R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
23. N. Ferguson, B. Schneier, and T. Kohno. *Cryptography Engineering: Design Principles and Practical Applications*. Wiley, 2010.
24. European Research Consortium for Informatics and Mathematics. Cloud computing: Platforms, software and applications. *ERCIM News*, 83, 2010.
25. E. Gallery and C.J. Mitchell. Trusted computing: Security and applications. *Cryptologia*, 33(3), 2009.
26. J.Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3), 1990.
27. J.Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3), 1992.
28. J.Y. Halpern and Y. Moses. The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence*, 75(2), 1995.
29. J. Huang and M.S. Fox. An ontology of trust — formal semantics and transitivity. In *Proceedings of the ACM Conference on Electronic Commerce*, 2006.
30. J. Huang and D. Nicol. A calculus of trust and its application to PKI and identity management. In *Proceedings of the ACM Symposium on Identity and Trust on the Internet*, 2009.
31. M. Joye and G. Neven. *Identity-Based Cryptography*. IOS Press, 2009.
32. L.M. Kaufman. Data security in the world of cloud computing. *IEEE Security & Privacy*, 7(4), 2009.
33. K.M. Khan and Q. Malluhi. Establishing trust in cloud computing. *IEEE IT Professional*, 12(5), 2010.
34. S. Kramer. Reducing provability to knowledge in multi-agent systems. In *Proceedings of the LiCS-affiliated Intuitionistic Modal Logics and Applications Workshop*, 2008. `ftp://ftp.research.microsoft.com/pub/tr/TR-2008-90.pdf`.
35. S. Kramer, R. Goré, and E. Okamoto. Formal definitions and complexity results for trust relations and trust domains. In *Proceedings of the ESSLLI-affiliated Workshop on Logics in Security*, 2010.
36. S. Kramer, R. Goré, and E. Okamoto. Formal definitions and complexity results for trust relations and trust domains fit for TTPs, the Web of Trust, PKIs, and ID-Based Cryptography. Logic Column of ACM SIGACT News, March 2010.
37. S. Kramer, C. Palamidessi, R. Segala, A. Turrini, and C. Braun. A quantitative doxastic logic for probabilistic processes and applications to information-hiding. *Journal of Applied Non-Classical Logic*, 19(4), 2009.

38. S. Kramer and A. Rybalchenko. A multi-modal framework for achieving accountability in multi-agent systems. In *Proceedings of the ESSLLI-affiliated Workshop on Logics in Security*, 2010.
39. K. Krukowa and A. Twigg. The complexity of fixed point models of trust in distributed networks. *Theoretical Computer Science*, 389(3), 2007.
40. D. Lekkas and D. Gritzalis. e-Passports as a means towards a globally interoperable public key infrastructure. *Journal of Computer Security*, 18(3), 2010.
41. C.-J. Liau. Belief, information acquisition, and trust in multi-agent systems — a modal logic formulation. *Artificial Intelligence*, 149(1), 2003.
42. A. Lioy, M. Marian, N. Moltchanova, and M. Pala. PKI past, present, and future. *Journal of Information Security*, 5(1), 2006.
43. Ch. Liu, M. Ozols, and T. Cant. An axiomatic basis for reasoning about trust in PKIs. volume 2119 of *LNCS*. Springer, 2001.
44. N.A. Lynch. *Distributed Algorithms*. Morgan Kaufmann Publishers, 1996.
45. U. Maurer. Modelling a public-key infrastructure. volume 1146 of *LNCS*. Springer, 1996.
46. J.-J. Meyer and F. Veltnam. *Handbook of Modal Logic*, chapter Intelligent Agents and Common Sense Reasoning. Volume 3 of Blackburn et al. [10], 2007.
47. B. Michael. In clouds shall we trust? *IEEE Security & Privacy*, 7(5), 2009.
48. K.W. Miller, J. Voas, and P. Laplante. In trust we trust. *IEEE Computer*, 43(10), 2010.
49. T. Muller. Semantics of trust. volume 6561 of *LNCS*. Springer, 2011.
50. A. Oram and J. Viega, editors. *Beautiful Security: Leading Security Experts Explain How They Think*. O'Reilly, 2009.
51. R. Parikh. Social software. *Synthese*, 132, 2002.
52. L.C. Paulson. The inductive approach to verifying cryptographic protocols. *Journal of Computer Security*, 6(1), 1998.
53. P.V. Rangan. An axiomatic basis of trust in distributed systems. In *Proceedings of the IEEE Symposium on Security and Privacy*, 1988.
54. S. Ruohomaa and L. Kutvonen. Trust management survey. volume 3477 of *LNCS*. Springer, 2005.
55. A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2), 2007.
56. B. Schneier. Security, group size, and the human brain. *IEEE Security & Privacy*, 7(4), 2009.
57. Ph. Schnoebelen. The complexity of temporal logic model checking. In *Proceedings of Advances in Modal Logic*, volume 4, 2003.
58. A. Tiu, R. Goré, and J. Dawson. A proof theoretic analysis of intruder theories. *Logical Methods in Computer Science*, 6(3), 2010.
59. H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer, 2007.
60. H.C.A. van Tilborg, editor. *Encyclopedia of Cryptography and Security*, pages 398–400. Springer, 2005.
61. D. Xiu and Z. Liu. A formal definition for trust in distributed systems. volume 3650 of *LNCS*. Springer, 2005.
62. A. Yao. Protocols for secure computations. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, 1982.
63. H. Yu, C. Jin, and H. Che. A description logic for PKI trust domain modeling. In *Proceedings of the IEEE Conference on Information Technology and Applications*, 2005.

64. W. Zhao, V. Varadharajan, and G. Bryan. Analysis and modelling of trust in distributed information systems. volume 3803 of *LNCS*. Springer, 2005.
65. Ph. Zimmermann and J. Callas. *Beautiful Security: Leading Security Experts Explain How They Think*, chapter The Evolution of PGP's Web of Trust. In Oram and Viega [50], 2009.

A A formal proof

In order to simplify our presentation of the proof of Theorem 2, we recall the following standard definition.

Definition 6 (Semantic consequence). *The formula* $\phi' \in \mathcal{L}$ *is* a semantic consequence of $\phi \in \mathcal{L}$, *written* $\phi \Rightarrow \phi'$, *:iff for all models* $(\mathfrak{S}, \mathcal{V})$ *and states* $s \in \mathcal{S}$, *if* $(\mathfrak{S}, \mathcal{V}), s \models \phi$ *then* $(\mathfrak{S}, \mathcal{V}), s \models \phi'$.

Fact 5 $\models \phi \rightarrow \phi'$ *if and only if* $\phi \Rightarrow \phi'$

Proof. By expansion of definitions.

The proof of Theorem 2 is now as follows:

1. $\mathsf{sTD}(\mathcal{C} \cup \mathcal{C}') \Rightarrow \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}'))$ — cf. Table 9
2. $\models \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}') \rightarrow \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}'))$ — 1, Fact 5
3. $\left(\begin{array}{l} \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'} \left(\begin{array}{l} \bigwedge_{a,b \in \mathcal{C}, c \in \mathcal{C}'}(b \text{ correct } a \rightarrow b \text{ correct } c) \wedge \\ \bigwedge_{a,b \in \mathcal{C}', c \in \mathcal{C}}(b \text{ correct } a \rightarrow b \text{ correct } c) \end{array}\right) \wedge \\ \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}')) \end{array}\right) \Rightarrow \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}')$ — cf. Table 10
4. $\models \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'} \left(\begin{array}{l} \bigwedge_{a,b \in \mathcal{C}, c \in \mathcal{C}'}(b \text{ correct } a \rightarrow b \text{ correct } c) \wedge \\ \bigwedge_{a,b \in \mathcal{C}', c \in \mathcal{C}}(b \text{ correct } a \rightarrow b \text{ correct } c) \end{array}\right) \rightarrow (\mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}')) \rightarrow \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}'))$ — 3, Fact 5
5. $\models \mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'} \left(\begin{array}{l} \bigwedge_{a,b \in \mathcal{C}, c \in \mathcal{C}'}(b \text{ correct } a \rightarrow b \text{ correct } c) \wedge \\ \bigwedge_{a,b \in \mathcal{C}', c \in \mathcal{C}}(b \text{ correct } a \rightarrow b \text{ correct } c) \end{array}\right) \rightarrow (\mathsf{CK}_{\mathcal{C} \cup \mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}')) \leftrightarrow \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}'))$ — 2, 4.

Table 9. Decomposability of strong trust domains

$$
\begin{array}{r}
\mathsf{sTD}(\mathcal{C} \cup \mathcal{C}') \Leftrightarrow \\
\text{(by the definition of } \mathsf{sTD}(\mathcal{C} \cup \mathcal{C}')\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}\cup\mathcal{C}'} a \ \mathsf{maytrust}\ b) \Rightarrow \\
\text{(by the definition of } \bigwedge_{a,b\in\mathcal{C}\cup\mathcal{C}'}\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}} a \ \mathsf{maytrust}\ b \wedge \bigwedge_{a,b\in\mathcal{C}'} a \ \mathsf{maytrust}\ b) \Rightarrow \\
\text{(by } \mathbf{4}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}} a \ \mathsf{maytrust}\ b \wedge \bigwedge_{a,b\in\mathcal{C}'} a \ \mathsf{maytrust}\ b)) \Rightarrow \\
\text{(by the distributivity of } \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'} \text{ over } \wedge, \mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}), \text{ and } \mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}} a \ \mathsf{maytrust}\ b) \wedge \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'} a \ \mathsf{maytrust}\ b)) \Rightarrow \\
\text{(by } \models \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\phi) \rightarrow \mathsf{CK}_{\mathcal{C}}(\phi), \mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}), \text{ and } \mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}} a \ \mathsf{maytrust}\ b) \wedge \mathsf{CK}_{\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'} a \ \mathsf{maytrust}\ b)) \Leftrightarrow \\
\text{(by the definition of } \mathsf{sTD}(\mathcal{C}) \text{ and } \mathsf{sTD}(\mathcal{C}')\text{)} \\
\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{sTD}(\mathcal{C}) \wedge \mathsf{sTD}(\mathcal{C}'))
\end{array}
$$

Table 10. Conditional composability of strong trust domains

$$\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\ \wedge\\ \bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\end{pmatrix}\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{sTD}(\mathcal{C})\wedge\mathsf{sTD}(\mathcal{C}'))\Leftrightarrow$$

(by the definition of $\mathsf{sTD}(\mathcal{C})$ and $\mathsf{sTD}(\mathcal{C}')$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\ \wedge\\ \bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\end{pmatrix}\wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\wedge\mathsf{CK}_{\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b))\end{pmatrix}\Leftrightarrow$$

(by the distributivity of $\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}$ over $\wedge$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b))\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b))\end{pmatrix}\Rightarrow$$

(by $\mathbf{T}(\mathsf{CK}_{\mathcal{C}})$, $\mathbf{T}(\mathsf{CK}_{\mathcal{C}'})$, $\mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$, and $\mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Leftrightarrow$$

(by the idempotency of $\wedge$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\end{pmatrix}\wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c))\wedge\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\end{pmatrix}\Leftrightarrow$$

(by the distributivity of $\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}$ over $\wedge$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b\end{pmatrix}\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b\end{pmatrix}\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Rightarrow$$

(by $\mathbf{4}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Rightarrow$$

(by $\models\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\phi)\to\mathsf{CK}_{\mathcal{C}'}(\phi)$, $\models\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\phi)\to\mathsf{CK}_{\mathcal{C}}(\phi)$, $\mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$, and $\mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{CK}_{\mathcal{C}}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Rightarrow$$

(by $\models\mathsf{CK}_{\mathcal{C}'}(\phi)\to\mathsf{EK}_{\mathcal{C}'}(\phi)$, $\models\mathsf{CK}_{\mathcal{C}}(\phi)\to\mathsf{EK}_{\mathcal{C}}(\phi)$, $\mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$, and $\mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{EK}_{\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\mathsf{EK}_{\mathcal{C}}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Leftrightarrow$$

(by the definition of $\mathsf{EK}_{\mathcal{C}'}$ and $\mathsf{EK}_{\mathcal{C}}$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{c\in\mathcal{C}'}\mathsf{K}_c\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{c\in\mathcal{C}}\mathsf{K}_c\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b\end{pmatrix})\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Leftrightarrow$$

(by the distributivity of K_c over $\wedge$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}\mathsf{K}_c(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C},c\in\mathcal{C}'}\mathsf{K}_c(a\ \mathsf{maytrust}\ b)\end{pmatrix}\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}\begin{pmatrix}\bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}\mathsf{K}_c(b\ \mathsf{correct}\ a\to b\ \mathsf{correct}\ c)\\ \wedge\ \bigwedge_{a,b\in\mathcal{C}',c\in\mathcal{C}}\mathsf{K}_c(a\ \mathsf{maytrust}\ b)\end{pmatrix}\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Rightarrow$$

(by Lemma 1, $\mathbf{N}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$, and $\mathbf{K}(\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'})$)

$$\begin{pmatrix}\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{b\in\mathcal{C},c\in\mathcal{C}'}(c\ \mathsf{maytrust}\ b))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}}a\ \mathsf{maytrust}\ b)\ \wedge\\ \mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{b\in\mathcal{C}',c\in\mathcal{C}}(c\ \mathsf{maytrust}\ b))\wedge\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}'}a\ \mathsf{maytrust}\ b)\end{pmatrix}\Leftrightarrow$$

(by the distributivity of $\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}$ over $\wedge$)

$$\mathsf{CK}_{\mathcal{C}\cup\mathcal{C}'}(\bigwedge_{a,b\in\mathcal{C}\cup\mathcal{C}'}a\ \mathsf{maytrust}\ b)\Leftrightarrow$$

(by the definition of $\mathsf{sTD}(\mathcal{C}\cup\mathcal{C}')$)

$$\mathsf{sTD}(\mathcal{C}\cup\mathcal{C}')$$