Trust on Beliefs: Source, Time and Expertise

Victor S. Melo Alison R. Panisson Rafael H. Bordini

Pontifical Catholic University of Rio Grande do Sul (PUCRS)
Postgraduate Programme in Computer Science – School of Informatics (FACIN)
Porto Alegre – RS – Brazil
{victor.melo.001,alison.panisson}@acad.pucrs.br,r.bordini@pucrs.br

Abstract

Trust is an important mechanism that describes how credible is the relation between agents in a multi-agent system. In this work, we extend the idea of trust to beliefs of agents, combining not only the provenance of information but also the outdated of such information. The resulting approach allows the agent generate different trust values for beliefs, depending on which meta-information is more important for that particular application, the trust in the source or how recent the information is. For this end, we describe some profiles of agents with different characteristics, combining the trust on the source and the outdated of information. Furthermore, we discuss how patterns of reasoning like argumentation schemes can play an important role in our approach, considering the expertise of the source of information.

1 Introduction

From the dictionary, *trust* means "belief that someone or something is reliable, good, honest, effective, etc." [TRU]. This definition fits well with the context of multi-agent systems, where much work has been devoted to that topic [PAH$^+$12, PSM12, PTS$^+$11, PSM13, TCM$^+$11]. The principal focus of trust in multi-agent systems is to describe the relation between agents about how credible agents appear to be to each other in such system.

However, as can be observed in the definition from the dictionary, trust is a term broadly applicable, and it can be applicable to agents, information, objects (such as vehicles, electronics, etc.), among others. In this work, we focus on the different sources of information available to agents in a multi-agent system and the trust on each of those sources. Furthermore, it is very important for dynamic environments, i.e., environments that are constantly changing, like the ones where multi-agent systems commonly are situated, the consideration of the time that information is stored/received; because of the constant changes on the environment, the information becomes outdated very quickly.

Therefore, in addition to the different sources of information available to agents in multi-agent systems, we consider how outdated is the information. To this end, we introduce some profiles for agents, which differ in the weights that are attributed to each of those criteria discussed in this work, i.e., meta-information about beliefs available in multi-agent systems.

The main contributions of this paper are: (i) we discuss various meta-information available in multi-agent systems, which are very useful when the agent has conflicting beliefs, allowing it to decide in which one to

In: J. Zhang, R. Cohen, and M. Sensoy (eds.): Proceedings of the 18th International Workshop on Trust in Agent Societies, Singapore, 09-MAY-2016, published at http://ceur-ws.org

believe; (ii) the meta-information considered in this work is inspired from practical platforms to develop multi-agent systems, which makes our work attractive in practical terms; and (iii) we introduce some interesting agent profiles, which are based on various criteria applicable to the meta-information considered. These profiles are interesting for different application domains where different meta-information may have different weights, as discussed in this work.

The remainder of this paper is structured as follows. We first describe the background of our work, including some interesting features from agent-oriented programming languages for our approach, and trust in multi-agent systems. Next, in Sections 3, 4 and 5, we discuss the application of trust values for beliefs, the possibilities of using time as meta-information, and the trust values combined to meta information such as time, respectively. In Section 6, we show how patterns of reasoning (named argumentation schemes) can play an interesting role in our approach, considering the expertise of the source of information. After that we discuss some related work and, finally, we conclude the paper with some final remarks.

2 Background

2.1 Agent-Oriented Programming Languages

There are many agent-oriented programming languages, such as Jason, Jadex, Jack, AgentFactory, 2APL, GOAL, Golog, and MetateM, as discussed in [BDDS09], each one with different characteristics. In this work, we choose Jason [BHW07]. Jason extends AgentSpeak(L), an abstract logic-based agent-oriented programming language introduced by Rao [Rao96], which is one of the best-known languages inspired by the BDI (*Beliefs-Desires-Intentions*) architecture, one of the most studied architectures for cognitive agents.

Jason has some interesting characteristics for our approach and, in this section, we describe some of such features.

- **Strong negation:** Strong negation helps the modelling of systems where uncertainty cannot be avoided, allowing the representation of things that the agent believes to be true, believes to be false, and things that the agent is ignorant about. Therefore, an agent is able to believe that, for example, a particular block is blue, represented by the predicate `blue(block)`, or that the block is not blue, represented by the predicate ¬`blue(block)`[1]. Furthermore, Jason allows agents to have both information in its belief base, with different annotation, indicating different sources, time-steps, etc. as described below;

- **Belief annotations:** One interesting characteristic of Jason is that it automatically generates annotations for all beliefs in the belief base about the source from where the belief was obtained (sensing the environment, communication with another agent, or a mental note created by the agent itself). An annotation has the following format: $blue(block)[source(john)]$, stating that the source of the belief that the block is blue is agent john. In addition to the automatic annotation of the source, the programmer can treat the events of receiving/perceiving any information, including annotation of time and any other meta-information he wants to store.

- **Speech-act based communication:** Jason uses performatives based on speech acts in its communication language, and formal semantics has been given for the changes in mental attitudes caused by the performatives available in the Jason extension of AgentSpeak. The performatives available in Jason can be easily extended, and their effects over the agent mental state can also be customised. Among such customisations, it is possible to add the annotations mentioned above.

There are other interesting characteristics in Jason, as a series of functions of the interpreter that are customisable, more details can be found in [BHW07].

2.2 Trust in Multi-agent Systems

Trust is a useful mechanism for decentralised systems, where autonomous entities deal with uncertain information and have to decide what to believe [PAH+12, PTS+11, TCM+11]. In trust-based approaches, agents can use the level of trust associated with the sources of contradictory information in order to decide about which one to believe. There are many different approaches to trust in the literature [PAH+12, PSM12, PTS+11, PSM13, TCM+11, CFP03], but here we will build our definitions mostly based on the concepts presented in [PTS+11,

[1] Where, we use the ¬ symbol for representing strong negation.

TCM[+]11]. First, in this section, we describe trust simply as a relation between agents, while in Section 3 we expand it, associating trust values for beliefs, which represent how much an agent trusts in that belief based on the sources which it have for it.

Considering trust as a relation between agents and following the definition presented in [TCM[+]11], a *trust relation* can be formalised as:

$$\tau \subseteq Ags \times Ags$$

where the existence of the relation indicates that an agent assigns some level of trust to another agent. For example, $\tau(Ag_i, Ag_j)$ means that agent Ag_i has at least some trust on agent Ag_j. It is important to realise that this is not a symmetric relation, so if $\tau(Ag_i, Ag_j)$ holds, this does not imply that $\tau(Ag_j, Ag_i)$ holds too.

A *trust network* is a directed graph representing a trust relation. It can be defined as:

$$\Gamma = \langle Ags, \tau \rangle$$

where Ags is the set of nodes in the graph, representing the agents of the trust network, and τ is the set of edges, where each edge is a pairwise trust relation between agents of Ags. An example of a trust network can be seen in Figure 1.

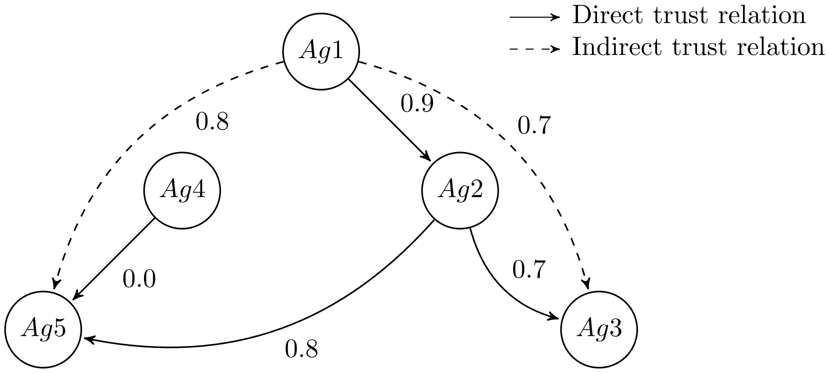

Figure 1: Trust Network Example

In order to measure trust, we follow the definition given in [PTS[+]11, TCM[+]11] where a function tr with the following signature:

$$tr : Ags \times Ags \mapsto \mathbb{R}$$

is used. It returns a value between 0 and 1, representing how much an agent trusts another. However, differently from [PTS[+]11, TCM[+]11], we define the relation between tr and τ as:

$$\begin{array}{lcl} tr(Ag_i, Ag_j) \geq 0 & \Leftrightarrow & (Ag_i, Ag_j) \in \tau \\ tr(Ag_i, Ag_j) = \texttt{null} & \Leftrightarrow & (Ag_i, Ag_j) \notin \tau \end{array}$$

so, in our definition, a trust level can in fact be zero, represented by $tr(Ag_i, Ag_j) = 0$, which means that Ag_i does not trust Ag_j. This is different from cases where Ag_i has no trust value assigned to Ag_j, represented by $tr(Ag_i, Ag_j) = \texttt{null}$. We use $(Ag_i, Ag_j) \notin \tau$ to denote that Ag_i has no acquaintance with Ag_j, i.e., it is not able to assess how trustworthy Ag_j is. Both cases can be seen in Figure 1, where there we have $tr(Ag4, Ag5) = 0$ and $tr(Ag1, Ag4) \notin \tau$.

Trust is a transitive relation, so an agent Ag_i can trust Ag_j directly or indirectly. Direct trust occurs when agent Ag_i directly assigns a trust value to Ag_j. Indirect trust occurs when, continuing the previous example, Ag_j trusts another agent Ag_k: in this case we could say that Ag_i indirectly trusts Ag_k.

We say there is a *path* between agents Ag_i and Ag_j if it is possible to create sequence of nodes of length n, $n \geq 1$:

$$\langle Ag_0, Ag_1, Ag_2, \ldots, Ag_{n-1}, Ag_n \rangle$$

so that

$$\tau(Ag_0, Ag_1), \tau(Ag_1, Ag_2), \ldots, \tau(Ag_{n-1}, Ag_n)$$

with $Ag_0 = Ag_i$ and $Ag_n = Ag_j$. In order to measure the trust from one particular path from Ag_i to Ag_j we need to use an operator to consider all the direct trust values in that path. Following the idea proposed in [PTS+11], a general operator $\otimes^{tr}$ can be defined as follows:

$$tr(Ag_i, Ag_j) = tr(Ag_0, Ag_1) \otimes^{tr} ... \otimes^{tr} tr(Ag_{n-1}, Ag_n)$$

which will define the trust value that Ag_i has on Ag_j according to the path $Ag_0, \ldots, Ag_n$ from Ag_i to Ag_j, constructed as defined above. If it happens that there are m different paths between Ag_i and Ag_j, a first possible path having a trust value of $tr(Ag_i, Ag_j)^1$ and the mth having $tr(Ag_i, Ag_j)^m$, following [PTS+11] we can define a generic operator $\oplus^{tr}$ so that:

$$tr(Ag_i, Ag_j) = tr(Ag_i, Ag_j)^1 \oplus^{tr} \ldots \oplus^{tr} tr(Ag_i, Ag_j)^m$$

For simplicity, in this paper we use those generic operators instantiated as:

- The trust of a path operator $\otimes^{tr}$ is the minimum trust value along the path. That is, it is defined as:

  $$tr(Ag_i, Ag_j) = min\{tr(Ag_0, Ag_1), \ldots, tr(Ag_{n-1}, Ag_n)\}$$

  given a path $Ag_0, \ldots, Ag_n$ from Ag_i to Ag_j as defined above.

- The $\oplus^{tr}$ over trust paths is defined as:

  $$tr(Ag_i, Ag_j) = max\{tr(Ag_i, Ag_j)^1, \ldots, tr(Ag_i, Ag_j)^m\}$$

  where m is the number of different possible paths between Ag_i and Ag_j.

In practical terms, the trust framework makes the agent explicitly aware of how much the other agents in that multi-agent system are trustworthy, and this info available for the agent by means of predicates such as `trust(ag1,0.8)`, meaning that the trust value that this agent places on `ag1` is `0.8`.

3 Trust on Beliefs

In this section, we introduce an way to calculate trust applied to beliefs, which is based on the trust value applied to the sources of these beliefs. We consider not only other agents as sources of information, but also perception of the environment, artifacts, and "mental notes" (beliefs created by an agent itself). For trust values for information received from other agents, we assume that these values are explicitly asserted in the belief base of agents (but calculated dynamically) based on the approach presented in the previous section. For trust values of information perceived from the environment, these values depend on the application domain, where, for example, multiple sensors could have varying degrees of trustworthiness.

For the purpose of a running example, we use the following trust values:

Source	Trust Value
ag1	0.3
ag2	0.4
ag3	0.5
ag4	0.8
self	1.0
percept_1	0.9
percept_2	0.6

Table 1: Values of Trust on Individual Sources of Information

Therefore, we expand trust to be a relation between an agent and the possible sources of information. So function $tr(Ag_i, Ag_j)$ that returns the trust level of Ag_i on Ag_j is generalised to:

$$tr(Ag_i, s_j)$$

where s_j represents one of the sources of information for agent Ag_i. This way, an agent Ag_i has a trust level on other kinds of sources, percepts or mental notes. This is interesting for cases when, using a similar example to the one presented in [AB14], an agent Ag_i has a sensor s_t which is known to have an accuracy of 80%. This way, the trust Ag_i has on s_t is defined as $tr(Ag_i, s_t) = 0.8$, associating the known percentage of accuracy with the trust value on s_t.

Further, the trust value of a particular sensor could be learned from experience, which seems more appropriate to the concept of trust used in this work.

It is important to emphasise that function *tr* returns the value of trust an agent has on some source. Now we can define a trust value associated to beliefs using function *tr*.

As a belief φ of an agent Ag_i can come from multiple sources, in order to know how much Ag_i trusts φ, we must consider the *tr* value associated with each source of φ for Ag_i. For this, we introduce the function trb_i below:

$$trb_i : \varphi \rightarrow \mathbb{R}$$

where $trb_i(\varphi)$ returns the trust value that Ag_i has on belief φ based on the trust level Ag_i has on the sources that asserted information φ. The operation that calculates $trb_i(\varphi)$ varies according to agent profiles, corresponding to different attitudes towards one's sources of information.

We introduce two agent profiles for calculating trust values over beliefs. They both may be interesting in different domains, depending on whether we are interested in credulous or sceptical agents.

Definition 1 (Credulous Agent) *A credulous agent considers only the most trustworthy source of information, and does not look for an overall social value.*

The formula used by a credulous agent to consider the most trusted source is:

$$trb_i(\varphi) = max\{tr(Ag_i, s_1), ..., tr(Ag_i, s_n)\}$$

where $\{s_1, ..., s_n\}$ is the set of sources that informed φ to Ag_i.

Definition 2 (Sceptical Agent) *A sceptical agent considers the number of sources from which it has received the information, and the trust value of each such source, in order to have some form of social trust value.*

A sceptical agent considers the quantity of sources that the information φ comes from. Therefore, we use a formula that sums the trust values of sources that information φ has been received from by Ag_i, determining a social trust value as follows:

$$trb_i(\varphi) = \frac{\sum\limits_{s \in S_\varphi^+} tr(Ag_i, s)}{|S_\varphi^+| + |S_\varphi^-|}$$

where $S_\varphi^+ = \{s_1, ..., s_n\}$ is the set of n different sources of φ and S_φ^- is the set of sources for $\overline{\varphi}$.

For example, considering an agent Ag_i with the trust values presented in Table 1, if Ag_i receives an information φ from a set of sources $S_\varphi^+ = \{Ag1, Ag2, Ag3\}$ and receives $\overline{\varphi}$ from $S_\varphi^- = \{Ag4\}$, then:

- A credulous agent will consider only the maximum trust values in S_φ^+ and S_φ^-, then it will get $trb_i(\varphi) = 0.5$ and $trb_i(\overline{\varphi}) = 0.8$.
- A sceptical agent will consider all the various sources. In particular, it will get $trb_i(\varphi) = \frac{0.3+0.4+0.5}{4} = 0.3$ and $trb_i(\overline{\varphi}) = \frac{0.8}{4} = 0.2$.

As another example, when Ag_i receives an information φ where the sources of φ are $S_\varphi^+ = \{percept_1\}$ and receives $\overline{\varphi}$ with sources $S_\varphi^- = \{Ag2, Ag3, Ag4\}$, then:

- A credulous agent will have $trb_i(\varphi) = 0.9$, and $trb_i(\overline{\varphi}) = 0.8$, having greater trust in φ than in its negation.
- A sceptical agent however will have $trb_i(\varphi) = \frac{0.9}{4} = 0.2$ and $trb_i(\overline{\varphi}) = \frac{0.4+0.5+0.8}{4} = 0.4$, preferring to believe $\overline{\varphi}$ instead.

There are cases when, for an information φ received by an agent Ag_i, $trb_i(\varphi)$ is equal to $trb_i(\overline{\varphi})$. For a *credulous* agent, it is easy to note that this occurs when the maximum trust value $tr(Ag_i, s_v)$, for a source $s_v \in S^+_\varphi$ equals the maximum trust value $tr(Ag_i, s_w)$ for a source $s_w \in S^-_\varphi$.

Differently, for sceptical agents, this occurs when $\sum_{s \in S^+_\varphi} tr(Ag_i, s)$ is equal to $\sum_{s \in S^-_\varphi} tr(Ag_i, s)$.

For these cases, we can consider other meta-information such as the time the beliefs were acquired (e.g., giving preference to more recent information) in order to decide what to believe. In next sections, we describe some possibilities for such extra criteria.

3.1 Expanding Trust Assessment

In Section 2.2, we defined the $\oplus^{tr}$ operator, which calculates the trust level that an agent Ag_i has on another agent Ag_j when there are n paths between them, as the maximum operator. However, considering the agent profiles, such as the *credulous* and *sceptical* profiles presented, a *sceptical* agent could consider the number n of paths between the two agents Ag_i and Ag_j to calculate $tr(Ag_i, Ag_j)$.

For example, consider two agents, Ag_i and Ag_j, where, using the *max* operator, we have $tr(Ag_i, Ag_j) = 0.6$, while $np(Ag_i, Ag_j) = 1$ where np is a function returning the number of all different paths between Ag_i and Ag_j in the trust network. This way we have $tr(Ag_i, Ag_j) = tr(Ag_i, Ag_j)^1$. Now consider another agent Ag_k, where $tr(Ag_i, Ag_k) = 0.6$ is defined using the *max* operator too, but $np(Ag_i, Ag_k) = 4$. Then we have $tr(Ag_i, Ag_k) = tr(Ag_i, Ag_k)^1 \oplus^{tr} ... \oplus^{tr} tr(Ag_i, Ag_k)^4$. This way, it is possible for Ag_i to consider Ag_k as more reliable than Ag_j taking into consideration the various different paths between them.

In organisational-based multi-agent systems/societies such as [HSB07], agents could assign trust values to *groups* of agents (or entire organisations) in the multi-agent system, in order to avoid some undesired bias. For example, consider an agent Ag receiving information from a set of agents $Ags = \{Ag_1, Ag_2, \ldots, Ag_n\}$, all belonging to the same organisation B. Then, it might be misleading if Ag considers those agents from Ags as different sources, thereby giving a social weight to that information, as those agents could spread some untrue information of interest to organisation B, or be biased in any way simply for belonging to the same organisation. For those cases, there may be a way for agent Ag to keep the information that the agents in Ags represent the same organisation, and this way Ag may not consider them as different sources. We will investigate that in future work.

More importantly, in future work we aim to combine our platform with existing trust systems (such as [KSS13]) so that we use realistic forms of update of levels of trust in sources of information while agents interact with each other and the environment, building on the extensive work done in the area of trust and reputation in multi-agent systems [PSM13].

4 Using Time

As it was presented in the last section, even using trust the agent can have contradictory information with the same trust values, and for these cases, other meta-information are needed. An example of such meta-information is *time*. There can exist scenarios where the time a piece of information was acquired can be even more important than the trust on its sources, and this will be explored below, with the definition of two different profiles, which may be interesting in different domains.

The first thing to consider is the way that an agent can maintain the information about time. As it was presented before, in Jason [BHW07], a belief can be annotated with each source that informed it. Besides, annotations can also be easily used for recording the time that the belief was informed by each source.

Considering that an environment is constantly changing, the search for the most recent information is often related to the search for the most accurate information. Sometimes it can be even possible that an agent Ag_j informs φ to Ag_i, and some time later, Ag_j informs $\overline{\varphi}$. Considering that the only source of φ and $\overline{\varphi}$ is Ag_j, using time, Ag_i can easily decide for $\overline{\varphi}$, as it is the most recent belief informed and the trust level of the source Ag_j is the same. This way, it is possible to realise that there exists a timeline of acquired beliefs. For example, consider the discrete time structure of Table 2 representing the instants of time at which beliefs are acquired by an agent Ag_i:

This way, Ag_i acquired two beliefs, being $\overline{\varphi}$ the latest acquired belief. Considering that the trust level of Ag_1 and Ag_2 are the same, at time 1 and 2, Ag_i will believe in φ, while at time 3 and 4, Ag_i will believe in $\overline{\varphi}$.

The function that returns the most recent time that a belief φ was received by a source s_j is $time(\varphi, s_j)$. This way, considering the discrete representation of time in table 2, for the belief φ acquired by Ag_i from agent Ag_1, it is used $time(\varphi, Ag_1) = 1$.

Time	time 1	time 2	time 3	time 4
Belief	φ		$\overline{\varphi}$	
Source	Ag_1		Ag_2	

Table 2: Discrete time representation

Note that an information φ can be received from multiple sources at multiple times. For example, for an agent Ag_i, if Ag_1 inform φ at time 1, and Ag_2 inform φ at time 3. This way, it is interesting to realise that each source will be annotated with its own time, for example, as it is in Jason, the belief that a block is blue could be:

$$blue(block)[source([ag1, ag2]), time([t1, t3])].$$

Considering the discrete time structure represented before, we can define as outdated time a value acquired from the difference between the actual time and the time that an information was informed by a source. This difference, for an information received by an agent Ag_i, we call T_i.

Definition 3 *Considering an agent Ag_i, $T_i(\varphi, s_j)$ is the difference between the actual time and the time that φ was acquired from source s_j.*

The formula of T_i could be, considering a belief φ received by an agent Ag_i from a source s_j, and a variable *now* which represents the actual time:

$$T_i(\varphi, s_j) = now - time(\varphi, s_j)$$

For a more complex example, it is interesting to note that a belief can come from different sources. Consider $S(\varphi) = \{s_1, ..., s_n\}$ and $S(\overline{\varphi}) = \{s_1, ..., s_m\}$ as the sets of sources for the information φ and $\overline{\varphi}$, respectively. In this case, an agent could compare a source $s_{v1} \in S(\varphi)$ with another source $s_{w1} \in S(\overline{\varphi})$, where s_{v1} and s_{w1} are the sources with the minimal T_i function value from their sets. If T_i is the same for s_{v1} and s_{w1}, then another source $s_{v2} \in S(\varphi)$ and $s_{w2} \in S(\overline{\varphi})$ can be selected, where the time annotated in s_{v2} is the most recent in the set $S(\varphi) \setminus \{s_{v1}\}$ and the time of s_{w2} is the most recent in $S(\overline{\varphi}) \setminus \{s_{w1}\}$. This could be applied recursively until it finds some $s_{vi} \in S(\varphi)$ greater or lower than an $s_{wi} \in S(\overline{\varphi})$, or if there are no more sources for φ or $\overline{\varphi}$ to compare, the agent would remain uncertain about φ.

5 Trust and Time

To combine trust and time, it is important to realise how one affects the other. Here trust is the focus, with the time being used to set the trust level adequately, prioritising the most recent information. This way, the older an information is, the less trusted it should be, for the reasons presented in the previous section. So we have:

Definition 4 (Outdated Information) *Considering a belief φ acquired by an agent Ag_i from a source s_j, the greater the $T_i(\varphi, s_j)$ value is, the lower the trust level for this belief φ should be.*

Note that naturally the trust level of φ will decrease as time passes, unless some sources keep informing φ by an amount that equalises the φ trust loss.

Considering an agent Ag_i, we can define a function $trs_i(\varphi, s_j)$ that returns the trust of φ considering just the information received by the source s_j at time $time(\varphi, s_j)$. Considering $S(\varphi) = \{s_1, s_2, ..., s_n\}$ the set of sources that informed φ, so Ag_i may have a different *trs* value for φ associated with each source. Here we use a generic operator $\odot^{trs}$ to relate the trust of the source with the time since the belief was informed by it:

$$trs_i(\varphi, s_j) = tr_i(Ag_i, s_j) \odot^{trs} T_i(\varphi, s_j)$$

Now, another function can be defined. Considering the same set of sources $S(\varphi)$, the function *trt* has just φ as parameter. As in Jason each belief has annotated all the sources that informed this belief, it is not needed to pass as parameter the set of sources of φ. We use a generic operator $\ominus^{trt}$, making a relation between all the *trs* assigned to each source. So we have:

$$trt_i(\varphi) = trs_i(\varphi, s_1) \ominus^{trt} ... \ominus^{trt} trs_i(\varphi, s_n)$$

Now there is a generic operator to relate the trust of the sources with the time that they informed a belief.

We defined two agent profiles in section 3, credulous and sceptical, where each one calculates $trb_i(\varphi)$ according to it owns attitude. Now we define other profiles for calculating $trt_i(\varphi)$, again, they both may be interesting in different domains, depending on whether we are interested in meticulous or conservative agents. It is interesting to note that an agent may be credulous meticulous, sceptical meticulous, credulous conservative or sceptical conservative.

Definition 5 (Conservative Agent) *A conservative agent uses time just when the trust on conflict beliefs are the same.*

For contradictory beliefs, a conservative agent will calculates the $trb_i(\varphi)$ and $trb_i(\overline{\varphi})$. The way that trb_i will be calculated depends on if the agent is credulous or sceptical too. If the trust values of the beliefs are the same, it will calculate the $trt_i(\varphi)$ and $trt_i(\overline{\varphi})$ to determinate the trust of each one considering the time they were informed.

Definition 6 (Meticulous Agent) *A meticulous agent uses trust just when the time of the conflict beliefs are the same.*

Differently from conservative agents, a meticulous agent Ag_i will first calculate the $trt_i(\varphi)$ and $trt_i(\overline{\varphi})$, and if it is the same, Ag_i will consider just the trust, ignoring the time that they were acquired, calculating $trb_i(\varphi)$ and $trb_i(\overline{\varphi})$.

As example, consider an agent Ag_i and the beliefs acquired according to the timeline of table 3:

Time	time 1	time 2	time 3	time 4
Belief	φ	φ		$\overline{\varphi}$
Source	Ag_1	Ag_2		Ag_3

Table 3: Time discrete representation

And consider the trust that Ag_i has in each source according to the table 4:

Source	**Trust Value**
Ag_1	0.6
Ag_2	0.4
Ag_3	0.5

Table 4: Values of Trust on Individual Sources of Information

Considering, for simplicity, that the $\odot^{trs}$ is a fraction operator, then we have $trs_i(\varphi, s_j) = \frac{tr_i(Ag_i, s_j)}{T_i(\varphi, s_j)}$. This definition keeps the idea that how bigger T_i is, less trusted a belief should be. Now, consider the $\ominus^{trt}$ operator as a max operator, then we have $trt_i(\varphi) = max\{trs_i(\varphi, s_1), ..., trs_i(\varphi, s_n)\}$, for a set of sources $\{s_1, ..., s_n\}$. Then, consider that Ag_i is a credulous agent and that the actual time in the timeline presented is time 5, so:

- A credulous conservative agent will consider $trb_i(\varphi) = 0.6$ and $trb_i(\overline{\varphi}) = 0.5$, opting to believe in φ.
- A credulous meticulous agent will consider $trs_i(\varphi, Ag_1) = \frac{0.6}{5-1} = 0.15$, and $trs_i(\varphi, Ag_2) = \frac{0.4}{5-2} = 0.13$. Then, we have $trt_i(\varphi) = max\{0.15, 0.13\} = 0.15$. And for $\overline{\varphi}$, we have $trs_i(\overline{\varphi}, Ag_3) = \frac{0.5}{5-4} = 0.5$. This way, $trt_i(\overline{\varphi}) = max\{0.5\} = 0.5$. So, a meticulous agent will believe in $\overline{\varphi}$, as $trt_i(\varphi) = 0.15$ and $trt_i(\overline{\varphi}) = 0.5$.

As it was presented, the trust of a belief depends on the trust of its sources. A natural approach is, when a belief φ is shown to be true or false, the trust of the sources of φ changes, increasing in case of true or decreasing otherwise. Considering time, we can improve this idea. The older an information is, the more time it had to change in the environment. Thus, there can be cases when an information φ is acquired, it is true, but after some time, φ becomes false. Thus, some of φ sources might not have informed something false, but as it had time to change in the environment, it became false. Thus, the idea is, considering an agent Ag_i, for a belief φ received by a source s_j, the longer $T_i(\varphi, s_j)$ is, the less trust s_j will lose in case of φ shows itself to be false to Ag_i.

6 Considering the Expertise of a Source

Another interesting criteria to combine, or even to generate trust values for beliefs, is to consider the expertise of the source in regards to specific kinds of information. For example, when a friend tells you that it is going to rain today, and you watch on television that it is going to be a sunny day, although you have more trust in your friend, it is reasonable to consider that the weatherperson is an expert in that subject (i.e., weather) and it is more reasonable to assume that it will be a sunny day.

A way to consider the expertise of the source is to use patterns of reasoning, for example, the so-called argumentation schemes [WRM08]. In particular, regarding the expertise of the source, Walton [Wal96] introduces the argumentation scheme called *Argument from position to know*, described below[2]:

> **Major Premise**: Source a is in a position to know about things in a certain subject domain S containing proposition A.
> **Minor Premise**: a asserts that A (in domain S) is true (or false).
> **Conclusion**: A is true (or false).

The associated critical questions (CQs) for the *Argument from position to know* are:

- **CQ1**: Is a in a position to know whether A is true (or false)?
- **CQ2**: Is a an honest (trustworthy, reliable) source?
- **CQ3**: Did a assert that A is true (or false)?

Therefore, the pattern of reasoning can be analysed in a dialectical way, where its conclusion is evaluated through the critical questions, due to its defeasible nature, and if the pattern of reasoning is valid, the trust value of that information can be incremented, considering that it comes from an expert source (i.e., someone in a position to know), and there are no reasons to doubt that.

Of course, as we are dealing with a value-based framework, it is necessary to attribute some kind of values for expert sources, including how much critical questions are correctly answered. Again, these values could depend on the application domain, where safety-critical applications such as the ones related to health could give greater consideration to the expertise of the source. For example, it is more reasonable to consider the opinion of a doctor who is expert/specialised on the particular health problem than the opinion of a general practitioner.

On the other hand, in some domains the source expertise may not be so important, for example, in our previous example about the weather: the consequences of taking or not the umbrella are not as strong as in the case of a wrong diagnosis of a serious illness.

We can observe that the argumentation scheme of position to know itself considers how much the source is trustworthy. Further, it considers whether the source is in a position to know such subject, and if it was that same source that provided such information directly.

In order to exemplify these ideas, imagine the following scenario related to the stock market: an agent, named ag_2, has informed (to agent ag_1) that an expert, named ex_1, said that a particular kind of stock, named st_1, has a great investment potential. Further, consider the following trust values related to the argumentation scheme for the ag_1:

Source/Belief	Trust Value
ag_2	0.6
ex_1	0.8
$expert(ex_1, st_1)$	0.9

Table 5: Values of Trust and Beliefs.

Considering the profiles introduced in Section 3, the *credulous* and *sceptical* agents consider the trust value for the information of st_1 having great investment potential, $great\_invest\_potential(st1)$, as 0.6, because there is only one source for $great\_invest\_potential(st1)$, named ag_2, with trust value of 0.6.

However, as we argued, there are some application domains in which it is reasonable to consider the expertise of the source, giving extra weight to such information when the agent believe the source is an expert. With this idea in mind, we introduce the following profiles based on the argumentation scheme (reasoning pattern) described.

[2]For simplicity, we use the more general argumentation scheme from position to know instead of the argumentation scheme for expert opinion, which is a subtype of the argument from position to know [WRM08, p. 14].

Definition 7 (Suspicious Agent) *A* suspicious *agent considers only the trust value for the source who provided the information to it, and ignores the trust on the original source who provided the information to that agent who informed the suspicious agent.*

In our scenario, the trust value of $great\_invest\_potential(st1)$ for a *suspicious* agent is 0.6. However, the agent that received that information is able to ask directly for the source of that information, in this case the expert ex_1, about the investment and, when receiving the information $great\_invest\_potential(st1)$ from ex_1, the trust value becomes to 0.8.

As observed, for a suspicious agent, even when receiving the information directly from the source, it aggregates the trust it has over the source, and not over the expertise of the source. To consider the expertise rather than the trust over the source can be very useful in some application domains, mainly due to the fact that trust values can be learned from experience, while the expertise of that particular source could be acquired from a reliable newspaper, web-page, etc.

Definition 8 (Expertise-recogniser Agent) *An* expertise-recogniser *agent considers the trust value of the information based on how much the source is an expert in that subject.*

Considering our scenario, the trust value for $great\_invest\_potential(st1)$ becomes to 0.9.

7 Related Work

Tang et al., in [TCM+11], combine argumentation and trust, taking into account trust on information used in argument inferences. That work is based on work presented by Parsons et al. [PTS+11], which proposes a formal model for combining trust and argumentation, aiming to find relationships between these areas.

Our work differs from [PTS+11, TCM+11] in some points. We introduce an approach for computing trust values for beliefs that differs from [PTS+11, TCM+11], where trust on a piece of information is assumed to be more directly available to argumentation. Different from those approaches, we allow for different sources for the same information (which is often the case in Jason agents) a propose ways to combine them into a single trust value for that information. We also define some agent profiles to facilitate the development of agents that require different *social* perspectives on the trust values of multiple sources; this is not considered in [PTS+11, TCM+11] either.

Another difference from [PTS+11, TCM+11] is that we consider other meta-information available in the multi-agent system (e.g., time annotation), which is inspired by agent-oriented programming languages that have such meta-information easily available.

Alechina et al., in [ABH+06], introduce a well-motivated and efficient algorithm for belief revision for AgentSpeak. The authors do not use trust or reputation in that work though. Therefore, the main point where our work differs from [ABH+06] is in the use of trust. We also argue that our approach could be used in order to improve the belief process presented in [ABH+06], as the trust values and the reasoning pattern introduced by us could play an important role in belief revision process too.

Amgoud and Ben-Naim, in [ABN15], propose a new family of argumentation-based logics (built on top of Tarskian logic) for handling inconsistency. An interesting aspect in [ABN15] is that the work defines an approach in which the arguments are evaluated using a "ranking semantics", which orders the arguments from the most acceptable to the least acceptable ones. The authors argue that, with a total order of arguments, the conclusions that are drawn are ranked with regards to *plausibility*. Although [ABN15] does not use trust, the proposed approach provides ordered arguments thus avoiding undecided conflicts. Our approach follows the same principles, but considering different criteria for ranking the information which agents have available in theirs belief bases, considering different meta-information available in agent-oriented languages. To use such meta information in argumentation-based approaches is part of our ongoing work [MPB16].

Parsons et al., in [PAH+12], identify ten different patterns of argumentation, called schemes, through which an individual/agent can acquire trust on another. Using a set of critical questions, the authors show a way to capture the defeasibility inherent in argumentation schemes and are able to assess whether an argument is good or fallacious. Our approach differs from [PAH+12] in that we are not interested in agents arguing about the trust they have on each other. We are interested in using such trust values and combining them with other meta-information available in the multi-agent system in order to use trust to resolve conflicts between beliefs, what might be interesting in domains where it might be important for the agents to not be undecided about some information.

Similarly we used the reasoning pattern based in an argumentation schemes, we argue that argumentation schemes from [PAH+12] could be used in order to evolve the trust values of different sources. In our future work we intend to investigate such integration.

Biga and Casali, in [AB14], present an extension of Jason, called G-Jason, to allow the creation of more flexible agents to reason about uncertainty, representing belief degrees and grades using the annotation feature provided by Jason. The authors define `degOfCert(X)`, where `X` is a value between 0 and 1, as a value associated with certainty of a belief and `planRelevance(LabelDegree)` as a value associated with plans, where the *LabelDegree* value is based on its context and its triggering event's *degOfCert* level. Our approach differs from [AB14] in that we use the notion of trust on agents and sensors in order to infer a level of certainty on beliefs. Further, we consider other meta-information, and we define profiles that combine different uses of such information, which is not considered in [AB14] .

8 Final Remarks

In this work, we described how different sources of information available to an agent in a multi-agent system and the trust of each of those sources can be combined to generate trust values for beliefs. Further, considering multi-agent shared environments, which are constantly changing, we combine also the time that information was stored/received by the agent, allowing agents to take into consideration how outdated a piece of information is.

In additional, we discussed how dialectical pattern of reasoning (i.e., argumentation scheme) can play an interesting role in our approach. We showed that argumentation schemes could guide agents in order to consider the expertise of the source instead of only the trust the agent has over that source. This idea seems interesting, considering that the trust values normally are learned from the experience, while the expertise of a particular source can be acquired from, also, reliable sources of information like the specification of the multi-agent system.

Finally, considering the different weights given to the meta-information discussed in this work, we introduce some agent profiles that can be useful for different application domains, as discussed in our brief examples.

In our future work, we intend to evaluate the impact of each profile introduced in this work for different kinds of applications, keeping an open mind for new (or even *middle ground*) profiles. This will allow us to identify the best profiles for each application domain, depending on the overall behaviour desired for the multi-agent system.

Furthermore, we also intend to combine the different profiles introduced in this work, as well as new profiles we will investigate, with practical argumentation-based reasoning mechanisms such as [PMVB14], where trust relations among agents may play an important role in decisions over conflicting arguments. Some initial investigation in this direction can be found [MPB16].

References

[AB14] Ana Calasi Adrián Biga. G-jason: An extension of jason to engineer agents capable to reason under uncertainty. In *G-JASON: An Extension of JASON to Engineer Agents Capable to Reason under Uncertainty*, 2014.

[ABH+06] Natasha Alechina, Rafael H Bordini, Jomi F Hübner, Mark Jago, and Brian Logan. Automating belief revision for agentspeak. In *Declarative Agent Languages and Technologies IV*, pages 61–77. Springer, 2006.

[ABN15] Leila Amgoud and Jonathan Ben-Naim. Argumentation-based ranking logics. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1511–1519. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[BDDS09] Rafael H. Bordini, Mehdi Dastani, Jrgen Dix, and Amal El Fallah Seghrouchni. *Multi-Agent Programming: Languages, Tools and Applications.* Springer Publishing Company, Incorporated, 1st edition, 2009.

[BHW07] Rafael H. Bordini, Jomi Fred Hübner, and Michael Wooldridge. *Programming Multi-Agent Systems in AgentSpeak using Jason (Wiley Series in Agent Technology).* John Wiley & Sons, 2007.

[CFP03] Cristiano Castelfranchi, Rino Falcone, and Giovanni Pezzulo. Trust in information sources as a source for trust: a fuzzy approach. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 89–96. ACM, 2003.

[HSB07] Jomi F Hubner, Jaime S Sichman, and Olivier Boissier. Developing organised multiagent systems using the moise+ model: programming issues at the system and agent levels. *International Journal of Agent-Oriented Software Engineering*, 1(3-4):370–395, 2007.

[KSS13] Andrew Koster, W. Marco Schorlemmer, and Jordi Sabater-Mir. Opening the black box of trust: reasoning about trust models in a BDI agent. *J. Log. Comput.*, 23(1):25–58, 2013.

[MPB16] Victor S. Melo, Alison R. Panisson, and Rafael H. Bordini. Argumentation-based reasoning using preferences over sources of information. In *fifteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2016.

[PAH+12] Simon Parsons, Katie Atkinson, Karen Haigh, Karl Levitt, Peter McBurneye Jeff Rowe, Munindar P Singh, and Elizabeth Sklar. Argument schemes for reasoning about trust. *Computational Models of Argument: Proceedings of COMMA 2012*, 245:430, 2012.

[PMVB14] Alison R. Panisson, Felipe Meneguzzi, Renata Vieira, and Rafael H. Bordini. An Approach for Argumentation-based Reasoning Using Defeasible Logic in Multi-Agent Programming Languages. In *11th International Workshop on Argumentation in Multiagent Systems*, 2014.

[PSM12] Simon Parsons, Elizabeth Sklar, and Peter McBurney. Using argumentation to reason with and about trust. In *Argumentation in multi-agent systems*, pages 194–212. Springer, 2012.

[PSM13] Isaac Pinyol and Jordi Sabater-Mir. Computational trust and reputation models for open multi-agent systems: a review. *Artificial Intelligence Review*, 40(1):1–25, 2013.

[PTS+11] Simon Parsons, Yuqing Tang, Elizabeth Sklar, Peter McBurney, and Kai Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 879–886. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[Rao96] Anand S. Rao. AgentSpeak(L): BDI agents speak out in a logical computable language. In *Proceedings of the 7th European workshop on Modelling autonomous agents in a multi-agent world : agents breaking away: agents breaking away*, MAAMAW '96, pages 42–55, Secaucus, NJ, USA, 1996. Springer-Verlag New York, Inc.

[TCM+11] Yuqing Tang, Kai Cai, Peter McBurney, Elizabeth Sklar, and Simon Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, page 38, 2011.

[TRU] Trust definition. `http://www.merriam-webster.com/dictionary/trust`. Accessed: 2016-02-11.

[Wal96] Douglas Walton. *Argumentation schemes for presumptive reasoning.* Routledge, 1996.

[WRM08] D. Walton, C. Reed, and F. Macagno. *Argumentation Schemes.* Cambridge University Press, 2008.