Trust in Abstract Argumentation

Guido Boella, Serena Villata
Department of Computer Science
University of Torino
Italy
{guido,villata}@di.unito.it

Leendert van der Torre
CSC
University of Luxembourg
Luxembourg
leon.vandertorre@uni.lu

Abstract—**Trust in multiagent systems is used for seeking to minimize the uncertainty in the interactions among the agents. In this paper, we discuss how to use argumentation to reason about trust. Using the methodology of meta-argumentation, first we represent the source of the information from which the argument is constructed in the abstract argumentation framework capturing the fact that *b is attacked* because b is from a particular source s. We show how a source of information can be attacked if it is not evaluated as trustworthy. Second, we provide a fine grained representation of the trust relationships between the information sources in which trust concerns not only the sources but also the single arguments and attack relations the sources propose. Moreover, we represent the evidences in support of the arguments which are put forward by the information sources and agents can express arguments by referring to other agents' arguments. Meta-argumentation allows us not to extend Dung's abstract argumentation framework by introducing trust and to reuse those principles and properties defined for Dung's framework.**

I. Introduction

Trust is a mechanism for managing uncertain information, decision making and dealing with the provenance of information. The result is that trust plays an important role in many research areas of computer science, particularly in the semantic web and multiagent systems where agents interact with other sources. In such interactions, the agents have to reason if they should trust or not the other agents and the extent to which they trust those other agents. The following illustrative example presents an informal argument exchange where several kinds of interactions between arguments and agents are reflected.

- *Witness1: I suspect the guy killed his boss in Rome. (arg a)*
- *Witness1: With a broken car he could not reach the crime scene. (arg b)*
- *Witness2: Witness1 is a compulsive liar. (arg c)*
- *Witness3: I repaired the guy's car at twelve of the crime day. (arg d)*
- *Witness4: I believe that Witness2 is not able to repair that kind of car. (arg e)*
- *Witness5: The guy has another car. (arg f)*
- *Witness6: The guy parked two cars in my underground parking garage three weeks ago. (arg g)*
- *Witness2: Witness6 was on holidays three weeks ago. (arg h)*
- *Witness7: The guy told he killed the boss. (arg i)*
- *Witness3: The guy charges himself to cover up for his wife. (arg l)*

In this informal argument exchange, different kinds of relations can be highlighted between arguments and agents. First, we have that the agents put forward the arguments and the attack relations. We will refer to these assertions by saying that the agents support their arguments and attack relations. Second, the agents can attack the trustworthiness of the other agents. These attacks are always addressed by means of arguments which attack the agent's trustworthiness itself or the trustworthiness of arguments and attack relations supported by this agent. Third, the agents can provide support to the other agents' arguments by putting forward evidences, always under the form of arguments, or by providing arguments which talk about other agents' arguments.

In this paper we argue that argumentation provides a mechanism to reason about trust handling aspects such as the origin of trust and the fine grained trust relationships. The research question addressed in the paper is:

- How to model trust in Dung's argumentation?

This breaks down into the following subquestions:

1)How to represent the information sources and the arguments they support?

2)How to represent an attack to the trustworthiness of the sources of information and a fine grained view of trust relations where trust concerns also single arguments and attacks?

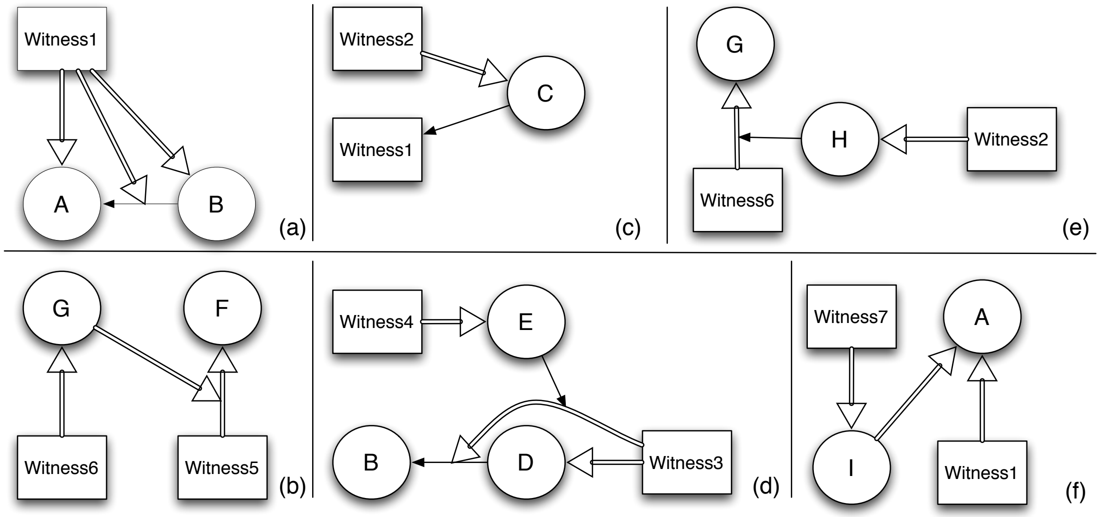

Fig. 1. The patterns involving agents and arguments and the trust relationships.

3)How to represent the evidences provided in support of the arguments?

4)How to model trust when the agents express arguments concerning other agents' arguments?

To answer the research questions we propose to use the methodology of meta-argumentation introduced by Boella et al. [1], [2]. The advantage in using meta-argumentation is that we do not extend Dung's framework [3] in order to introduce trust but we instantiate Dung's theory with meta-arguments. In this way we can reuse all the principles, algorithms and properties already defined for Dung's framework. In meta-argumentation, different entities besides proper arguments are introduced in the meta-level under the form of meta-arguments and the acceptable, *trusted*, meta-arguments are returned. These meta-arguments represent the arguments of the agents and their attack relations. The agents, as sources of arguments and attack relations, are introduced under the form of meta-arguments "*agent i is trustable*".

The research questions ask for patterns where both agents and arguments are composed together and are related to each other by trust relationships. The patterns which emerge from the informal argument exchange are provided in Figure 1 where the common arrows represent the attack relation and the double arrows represent the support relation of the agents to the arguments they built.

We represent the information sources and the arguments they support under the form of meta-arguments. We start with the partial argumentation frameworks of the single agents where the arguments and the attack relations in each agent's mind are provided. At the meta-level, these arguments and attack relations are supported by the agents, introduced in the framework as meta-argument $trust(ag_i)$, by means of another meta-argument Z. In Figure 1.a, the representation of the information sources and the arguments they support is provided. Witness5 supports both arguments a, b and the attack relation between them.

We represent the attacks to agents' trustworthiness and the attacks to the trustworthiness of the single arguments and attack relation by means of attacks at the meta-level. The attacks about trustworthiness are always addressed by means of arguments of the other agents. These arguments can directly attack, in the meta-level, the meta-arguments representing the agents or the meta-arguments representing the other agents' arguments or attack relations. Figure 1.c depicts the attack of Witness2 to the trustworthiness of Witness1. Note that in this case, this agent becomes no more credible in the multiagent system because her credibility has been attacked as a whole. This is not always the case, it may be possible that the agents attack other agents' trustworthiness only concerning a particular argument or attack relation. This is described in Figure 1.d-e where Witness4 and Witness5 attack the trustworthiness of Witness3 and Witness6 respectively only concerning argument g and the attack relation $d \rightarrow b$. In this case, only the attacked argument or attack relation become not acceptable in the framework.

We represent the evidences provided in support of the agents as attacks, in the meta-level, to the Z meta-argument which attacks the argument or attack relation. We introduce evidences in order to have that, even if an agent is considered as not trustable when another agent provides an argument (i.e., an evidence) in support of one of the not trustable agent's argument then this argument becomes accepted. As for the attacks on trustworthiness, evidences are always expressed by means of arguments. In Figure 1.b, the evidence provided by Witness6 in support to the argument of Witness5 is represented.

Finally, we model trust when the agents express arguments concerning other agents' arguments as a kind of support provided by an argument to the reported argument. This is addressed in the meta-level as an attack from the supporting argument to the Z meta-argument attacking the supported argument. arguments about other agents' arguments are represented in Figure 1.f where Witness7 supports by means of his argument i Witness1's argument a.

The paper follows the research questions. Section 2 introduces briefly the methodology of meta-argumentation. In Section 3 we describe how to represent the agents in an argumentation framework and we discuss how to model the patterns defined in Figure 1 where different kinds of attacks to the trustworthiness of the agents are addressed. Related work and conclusions end the paper.

II. Argumentation Theory

A. Abstract Argumentation

Dung's theory [3] is based on a binary *attack* relation among arguments, which are abstract entities whose role is determined only by their relation to other arguments. We restrict ourselves to *finite* argumentation frameworks, i.e., in which the set of arguments is *finite*.

Definition 1 (Argumentation framework AF): An argumentation framework is a tuple $\langle A, \rightarrow \rangle$ where A is a finite set of elements called arguments and $\rightarrow$ is a binary relation called attack defined on $A \times A$.

Definition 2 (Defence): Let $\langle A, \rightarrow \rangle$ be an argumentation framework. Let $\mathcal{S} \subseteq A$. $\mathcal{S}$ defends a if $\forall b \in A$ such that $b \rightarrow a$, $\exists c \in \mathcal{S}$ such that $c \rightarrow b$.

All Dung's semantics are based on the notion of defence. An argumentation framework is a directed graph whose nodes are the arguments and the edges represent the attack relations. A semantics of an argumentation framework consists of a conflict-free set of arguments, i.e., a set of arguments that does not contain an argument attacking another argument in the set.

Definition 3 (Conflict-free CF): Given an argumentation framework $AF = \langle A, \rightarrow \rangle$, a set $S \subseteq A$ is *conflict free*, denoted as *cf*(S), iff $\neg\exists \alpha, \beta \in S$ such that $\alpha \rightarrow \beta$.

We adopt some ideas from Baroni and Giacomin [4]. In particular, an idea we adopt is the use of a function $\mathcal{E}$ that maps argumentation frameworks $\langle A, \rightarrow \rangle$ to its set of extensions, i.e., to a set of sets of arguments. Like [4] we use a function $\mathcal{E}$ mapping an argumentation framework $\langle A, \rightarrow \rangle$ to its set of extensions, i.e., to a set of sets of arguments. However, this function is not formally defined. To be precise, they say: "An extension-based argumentation semantics is defined by specifying the criteria for deriving, for a generic argumentation framework, a set of extensions, where each extension represents a set of arguments considered to be acceptable together. Given a generic argumentation semantics $\mathcal{S}$, the set of extensions prescribed by $\mathcal{S}$ for a given argumentation framework AF is denoted as $\mathcal{E}_{\mathcal{S}}(AF)$." The following definition captures the above informal meaning of the function $\mathcal{E}$. Since Baroni and Giacomin do not give a name to the function $\mathcal{E}$, and it maps argumentation frameworks to the set of accepted arguments, we call $\mathcal{E}$ the *acceptance function*.

Definition 4: Let $\mathcal{U}$ be the universe of arguments. An acceptance function $\mathcal{E} : 2^{\mathcal{U}} \times 2^{\mathcal{U} \times \mathcal{U}} \rightarrow 2^{2^{\mathcal{U}}}$ is a partial function which is defined for each argumentation framework $\langle A, \rightarrow \rangle$ with finite $A \subseteq \mathcal{U}$ and $\rightarrow \subseteq A \times A$, and maps an argumentation framework $\langle A, \rightarrow \rangle$ to sets of subsets of A: $\mathcal{E}(\langle A, \rightarrow \rangle) \subseteq 2^A$.

The following definition summarizes the most widely used acceptability semantics of arguments given in the literature. Which semantics is most appropriate in which circumstances depends on the application domain of the argumentation theory.

Definition 5 (Acceptability semantics): Let $AF = \langle A, \rightarrow \rangle$ be an argumentation framework. Let $\mathcal{S} \subseteq A$. $\mathcal{S}$ defends a if $\forall b \in A$ such that $b \rightarrow a$, $\exists c \in \mathcal{S}$ such that $c \rightarrow b$. Let $D(\mathcal{S}) = \{a \mid \mathcal{S} \text{ defends } a\}$.

- $\mathcal{S} \in \mathcal{E}_{\text{admiss}}(AF)$ iff *cf*$(\mathcal{S})$ and $\mathcal{S} \subseteq D(\mathcal{S})$.
- $\mathcal{S} \in \mathcal{E}_{\text{compl}}(AF)$ iff *cf*$(\mathcal{S})$ and $\mathcal{S} = D(\mathcal{S})$.
- $\mathcal{S} \in \mathcal{E}_{\text{ground}}(AF)$ iff $\mathcal{S}$ is smallest in $\mathcal{E}_{\text{compl}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{pref}}(AF)$ iff $\mathcal{S}$ is maximal in $\mathcal{E}_{\text{admiss}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{ skep-pref}}(AF)$ iff $\mathcal{S} = \cap \mathcal{E}_{\text{pref}}(AF)$.
- $\mathcal{S} \in \mathcal{E}_{\text{stable}}(AF)$ iff *cf*$(\mathcal{S})$ and $\forall b \in A \backslash \mathcal{S}$ $\exists a \in \mathcal{S} : a \rightarrow b$.

Dung's argumentation theory formalizes the reasoning leading to accepted arguments, on the basis of attacks among arguments. In Dung's terminology, it is a theory of argumentation semantics, which relates attack relations among arguments to acceptable arguments. In our terminology, it is a theory of acceptance functions. To *use* Dung's theory, we have to describe the arguments and the attack relation, such that we can use one of the argumentation semantics or acceptance functions to obtain the acceptable arguments. The theory does not assume any structure on the arguments, which are therefore called *abstract* arguments, such that the description of the arguments and the attack relation in Dung's theory is unconstrained, and the theory can be used in many contexts. We call a set of arguments together with an attack relation a *basic* argumentation framework, to distinguish it from the extended argumentation frameworks discussed below. We call this use of the theory, based on an instantiation of abstract arguments, an *instantiation* of Dung's theory.

B. Meta-Argumentation

Consider two politicians arguing about social welfare, using arguments like "employment will go up" or "productivity will go down". Two commentators observing the debate may argue about it, using arguments like "the argument "employment will go up" is accepted by the politicians" or "the politicians accept that the argument "employment will go up" supports the argument that "productivity will go down"." This phenomena of people arguing about other people's arguments is common: lawyers argue about the argumentation of suspects in a courtroom, citizens argue about the argumentation of politicians when making their voting decisions during elections, teachers may argue about the argumentation of their students when evaluating their exams, and parents may argue about their children's argumentation when arguing how to raise their children. We call this arguing about argumentation *meta-argumentation.*

Boella et al. [2] instantiate Dung's theory with meta-arguments, *such that we use Dung's theory to reason about itself.* Wooldridge *et al.* [5] argue that one cannot think of argumentation without thinking of meta-argumentation too. They claim that

> Our key motivation is the following observation: *Argumentation and formal dialogue is necessarily a meta-logical process*. This seems incontrovertible: even the most superficial study of argumentation and formal dialogue indicates that, not only are arguments made about object-level statements, they are also made about arguments. In such cases, an argument is made which refers to another argument. Moreover, there are clearly also cases where the level of referral goes even deeper: where arguments refer to arguments that refer to arguments.

We call this the meta-argumentation viewpoint. In modeling, a viewpoint is associated with a stakeholder with her concerns and gives rise to views on systems. The methodology of meta-argumentation as a way to model argumentation is based on a conceptualization of argumentation using the relation between two theories of argumentation and meta-argumentation.

The motivation of our meta-argumentation methodology comes from the well known and generally accepted observation that Dung's theory of abstract argumentation cannot be used directly when modeling argumentation in many realistic examples, such as multiagent argumentation and dialogues [6], decision making [7], coalition formation [8], combining Toulmin's micro arguments [9], normative reasoning [10], or meta-argumentation. When Dung's theory of abstract argumentation cannot be applied directly, there are two methodologies to model argumentation using the theory, which leads to the dilemma of choosing among these two alternatives.

- **Instantiating abstract arguments.** Starting from a knowledge base, a set of arguments is generated from this base, and the attack relation among the arguments is derived from the structure of the arguments [11].
- **Extending Dung's framework.** Alternatively, the description of argumentation frameworks is extended, for example with preferences among abstract arguments [12], [13], abstract value arguments [14], second- and higher-order attack relations [15], [16], [17], support relations among abstract arguments [18], or priorities among abstract arguments [19].

We assume a fundamental relation about the relation between these two levels: *meta-argumentation has to be able to mirror argumentation*. For example, when politicians argue, the commentators should be able to argue in the same way. For example, if the politicians use as primitives arguments a from a universe of arguments U, together with a mechanism to derive acceptable arguments from relations among the arguments, and the commentators have as primitives meta-arguments ma from a universe of meta-arguments MU together with a mechanism to derive acceptable meta-arguments from relations among the meta-arguments, then the set of arguments must be reflected in the set of meta-arguments, and there must be a relation between the ways acceptable arguments and acceptable meta-arguments are derived.

Meta-argumentation is a particular way to define mappings from argumentation frameworks to extended argumentation frameworks: arguments are interpreted as meta-arguments, of which some are mapped to "argument a is accepted", $acc(a)$, where a is an abstract argument from the extended argumentation framework EAF. The meta-argumentation methodology is summarized in Figure 2.

The function f assigns to each argument a in the EAF, a meta-argument "argument a is accepted" in the basic argumentation framework. We use Dung's acceptance functions $\mathcal{E}$ to find functions $\mathcal{E}'$ between extended argumentation frameworks EAF and the acceptable arguments AA' they return. The accepted arguments of the meta-argumentation framework are a function of the extended argumentation framework $AA = \mathcal{E}'(EAF)$. The transformation function consists of two parts: a function f^{-1} transforms an argumentation framework AF to an extended argumentation framework EAF, and a function g transforms the acceptable arguments of the AF into acceptable arguments of the EAF. Summarizing $\mathcal{E}' = \{(f^{-1}(a), g(b)) \mid (a,b) \in \mathcal{E}\}$ and $AA' = \mathcal{E}'(EAF) = g(AA) = g(\mathcal{E}(AF)) = g(\mathcal{E}(f(EAF)))$.

The first step of our approach is to define the set of extended argumentation frameworks. The second step consists in defining flattening algorithms as a function from this set of EAFs to the set of all basic argumentation frameworks: $f : EAF \rightarrow AF$.

Definition 6 presents the instantiation of a basic argumentation framework as a sequence of partial argumentation frameworks of the agents [20] using meta-argumentation. A sequence of partial argumentation frameworks of the agents $\langle\langle A_1, \rightarrow_1\rangle, \dots, \langle A_n, \rightarrow_n\rangle\rangle$ are sets composed by arguments

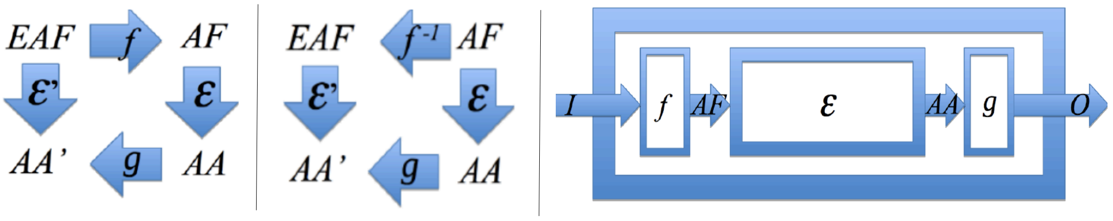

Fig. 2. The meta-argumentation methodology.

NOTATION	MEANING
U	universe of all generated arguments
$A \subset U$	a finite set of arguments
$a, b, c, \ldots \in A$	elements of A
$\rightarrow$	binary relation on A representing attack
MU	universe of all meta-arguments
$accept(a)$	"argument a is acceptable"
MA	a set of meta-arguments
$\longmapsto$	a relation on MA
EAF	an extended AF
$\mathcal{EAF}$	a set of possible EAF
f	function from EAF to AF
AF	a pair of A and $\rightarrow$
$\mathcal{AF}$	a set of possible AF
$\mathcal{E}$	mapping from $\langle A, \rightarrow \rangle$ to sets of subsets of A
g	function from accepted MA to accepted A
X, Y	meta-arguments for attack

TABLE I
META-ARGUMENTATION NOTATION USED IN THIS PAPER.

A_i and a binary attack relation $\rightarrow_i$.

The universe of meta-arguments is $MU = \{acc(a) \mid a \in \mathcal{U}\} \cup \{X_{a,b}, Y_{a,b} \mid a, b \in \mathcal{U}\}$, and the flattening function f is given by $f(EAF) = \langle MA, \longmapsto \rangle$. For a set of arguments $B \subseteq MU$, the unflattening function g is given by $g(B) = \{a \mid acc(a) \in B\}$, and for sets of arguments $AA \subseteq 2^{MU}$, it is given by $g(AA) = \{g(B) \mid B \in AA\}$.

Definition 6: Given an extended argumentation framework $EAF = \langle \langle A_1, \rightarrow_1 \rangle, \ldots, \langle A_n, \rightarrow_n \rangle \rangle$ where for each agent $1 \leq i \leq n$, $A_i \subseteq \mathcal{U}$ is a set of arguments and $\rightarrow_i \subseteq A_i \times A_i$ is a binary relation over A_i, the set of meta-arguments $MA \subseteq MU$ is $\{acc(a) \mid a \in A_1 \cup \ldots \cup A_n\}$ and $\longmapsto \subseteq MA \times MA$ is a binary relation on MA such that: $acc(a) \longmapsto X_{a,b}, X_{a,b} \longmapsto Y_{a,b}, Y_{a,b} \longmapsto acc(b)$ if and only if there is an agent $1 \leq i \leq n$ such that $a, b \in A_i$ and $a \rightarrow_i b$.

The set of acceptable arguments of a meta-argumentation framework $\langle MA, \longmapsto \rangle$ follows from $\mathcal{E}'(EAF) = g(\mathcal{E}(f(EAF)))$. For a given flattening function f, the acceptance function of the extended argumentation theory $\mathcal{E}'$ is defined using the acceptance function of the basic abstract argumentation theory $\mathcal{E}$: an argument of an EAF is acceptable if and only if it is acceptable in the flattened basic AF.

Meta-argumentation has received little attention thus far. On the one hand, Jakobovits and Vermeir [21] present how to use labelings to define what arguments should be accepted or not. All of the labelings and restricted labelings of the argumentation framework, together with their attacks, are represented in the meta-argumentation framework. On the other hand, Cayrol and Lagasquie-Schiex [18] presents a meta-argumentation framework in which are represented two kinds of binary relations between the arguments, the attack relation and the support relation. Similar approaches to meta-argumentation particularly focused on the problem of representing second-order attacks are addressed by Modgil and Bench-Capon [22], Baroni et al. [23], Gabbay [24], [25].

III. MODELLING TRUST IN DUNG'S FRAMEWORK

A number of authors have highlighted that the definition of trust is difficult to pin down precisely, thus in the literature there are numerous different definitions. To pick few of these definitions, [26] define trust as

> a mental state, a complex attitude of an agent x towards another agent y about the behaviour/action a relevant for the goal g

while [27] states that

> trust is the subjective probability by which an individual A expects that another individual B performs a given action on which its welfare depends

The common elements are that there is a consistent degree of uncertainty associated with trust and trust is tied up with the relationships between individuals and particularly it is related to the actions of the individuals and to the effects these actions have on the others. In this paper we does not refer to the actions of the agents but we provide a model for representing the agents' beliefs concerning the trustworthiness of the other agents. We follow the approach proposed by [28] where the influence of trust on the assimilation of acquired information into an agent's belief is considered. [28]'s characteristic axiom is "if agent i believes that agent j has told him the truth of p and he trusts the judgement of j on p, then he will also believe p".

A. Representing the information sources

Let us consider again the informal argument exchange. We have that *Witness2* has a negative opinion about the trustworthiness of *Witness1* while we can infer that all the other witnesses consider *Witness1* a reliable information source since they do not attack him. Agents are introduced in the meta-argumentation framework under the form of meta-arguments "*agent i is trustable*", $trust(i)$, for all the agents i. As in Definition 6, we add meta-arguments "*argument a is accepted*", $acc(a)$, for all arguments in A, and meta-arguments $X_{a,b}$, $Y_{a,b}$ for all arguments a and b such that $a \rightarrow b$. Each argument $a \in A$ in the mind of the agents is put forward, by means of the $Z_{acc(a)}$ meta-argument. This meta-argument attacks the meta-argument $acc(a)$ asking for an evidence in support of argument a. In the simplest case, the $Z_{acc(a)}$ meta-argument is attacked by the meta-argument $trust(i)$ which represents the agent who proposes argument a, as shown in Figure 3. More complex cases of evidences are described in the next sections. For each agent i, if $\rightarrow_i$ contains $a \rightarrow b$, such as if the agent put forward an attack relation, then the meta-argument $trust(i)$ supports the meta-argument $Y_{a,b}$, representing the attack relation, by attacking the meta-argument $Z_{Y_{a,b}}$, as done for the arguments. Also in this case, the attack of the agent to the Z meta-argument is an evidence in support of this attack relation.

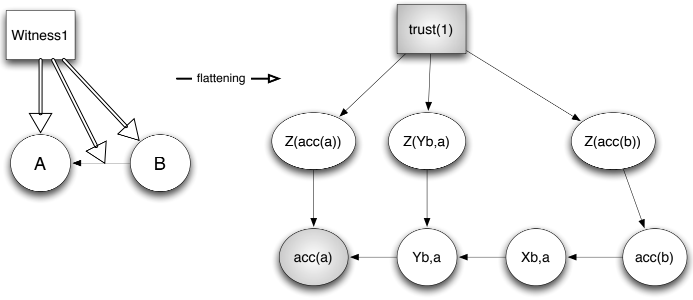

Fig. 3. Introducing the agents in the framework.

We represent the fact that more than one information source sustains the same arguments by let them attacking by means of the $trust(i)$ meta-arguments the same $Z_{acc(a)}$ meta-argument which asks for evidences in support of meta-argument $acc(a)$. An example of multiple support of two agents regarding the same argument is depicted in Figure 4. The same solution is applied to the attack relations where we consider the meta-argument $Y_{a,b}$ instead of meta-argument $acc(a)$.

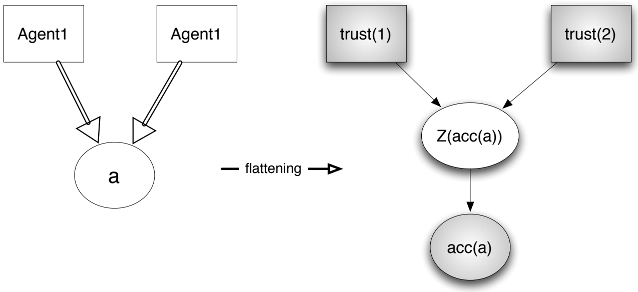

Fig. 4. A multiple support to the same argument.

We extend the EAF proposed in Definition 6 by adding the information sources and second-order attacks, such as attacks from an argument or attack relation to another attack relation. For more details about second-order attacks in meta-argumentation, see [2], [22], [23].

The unflattening function g and the acceptance function $\mathcal{E}'$ are defined as above. In particular, the introduction of the agents in the meta-argumentation framework is defined as follows:

Definition 7: An extended argumentation framework EAF is a tuple $\langle\langle A_1, \rightarrow_1, \rightarrow_1^2\rangle \ldots, \langle A_n, \rightarrow_n, \rightarrow_n^2\rangle\rangle$ where for each agent $1 \leq i \leq n$, $A_i \subseteq \mathcal{U}$ is a set of arguments, $\rightarrow_i$ is a binary relation on $A_i \times A_i$, $\rightarrow_i^2$ is a binary relation on $(A_i \cup \rightarrow_i) \times \rightarrow_i$.

Definition 8: Given an extended argumentation framework $EAF = \langle\langle A_1, \rightarrow_1, \rightarrow_1^2\rangle \ldots, \langle A_n, \rightarrow_n, \rightarrow_n^2\rangle\rangle$, the set of meta-arguments MA is $\{trust(i) \mid 1 \leq i \leq n\} \cup \{acc(a) \mid a \in A_1 \cup \ldots \cup A_n\} \cup \{X_{a,b}, Y_{a,b} \mid a, b \in A_1 \cup \ldots \cup A_n\} \cup \{Z_a \mid a \in A_1 \cup \ldots \cup A_n\}$ and $\longmapsto \subseteq MA \times MA$ is a binary relation on MA such that:

- $acc(a) \longmapsto X_{a,b}, X_{a,b} \longmapsto Y_{a,b}, Y_{a,b} \longmapsto acc(b)$ iff $a, b \in A_i$ and $a \rightarrow_i b$, and
- $trust(i) \longmapsto Z_{acc(a)}, Z_{acc(a)} \longmapsto acc(a)$ iff $a \in A_i$, and
- $trust(i) \longmapsto Z_{Y_{a,b}}, Z_{Y_{a,b}} \longmapsto Y_{a,b}$ iff $a, b \in A_i$ and $a \rightarrow_i b$, and
- $acc(a) \longmapsto X_{a,b\rightarrow c}, X_{a,b\rightarrow c} \longmapsto Y_{a,b\rightarrow c}, Y_{a,b\rightarrow c} \longmapsto Y_{b,c}$ iff $a, b, c \in A_i$ and $a \rightarrow_i^2 (b \rightarrow_i c)$,
- $Y_{a,b} \longmapsto Y_{c,d}$ iff $a, b, c \in A_i$ and $(a \rightarrow_i b) \rightarrow_i^2 (c \rightarrow_i d)$.

Example 1: Let us consider the informal dialogue exchange. We represent the agents in the argumentation framework as shown in Figure 3. Witness1 puts forward two arguments a and b and the attack relation between them. Thanks to the flattening function described in Definition 8, we add the meta-argument $trust(1)$ for representing Witness1 in the framework and we add meta-arguments $acc(a)$ and $acc(b)$ for the arguments of Witness1. The attack relation is represented by means of two meta-arguments $X_{a,b}$ and $Y_{a,b}$ which stay for the inactive and active status of the attack relation $a \rightarrow b$. Witness1 provides an evidence in support to the arguments a and b and the attack relation $a \rightarrow b$ by attacking the respective meta-arguments Z.

B. Representing fine grained trust relationships

In our model, trust is represented as an absence of an attack towards the agents or towards their arguments and attack relations or as the presence of an evidence in support of arguments and attack relations. On the contrary, the distrust relationship is modelled as a lack of evidences in support of the arguments and the attack relations or as an attack relation towards the agents and their arguments and attack relations. The three distrust relationships depicted in Figure 1.c-d-e are of different kind and must be distinguished in the framework in order to reason about trust.

In the informal argument exchange, Witness2 attacks the trustworthiness of Witness1 as a credible witness. In this way, he is attacking each argument and attack relation proposed by Witness1. Witness4, instead, is not arguing against Witness3 but he is arguing against the attack relation $d \rightarrow b$ as proposed by Witness3. Finally, Witness2 reasons about the trustworthiness of Witness6. The untrustworthiness of Witness6 is linked only to the precise argument g. We propose a fine grained view of trust in which the sources of information may be attacked for being unreliable or for being unreliable in sustaining a particular argument or attack relation. Definition 9 presents an extended argumentation framework in which a new relation between arguments is given to represent distrust.

Definition 9: A trust-based extended argumentation framework $TEAF$ is a tuple $\langle\langle A_1, \rightarrow_1, DT_1\rangle, \ldots, \langle A_n, \rightarrow_n, DT_n\rangle\rangle$ where for each agent $1 \leq i \leq n$, $A_i \subseteq \mathcal{U}$ is a set of arguments, $\rightarrow_i \subseteq A_i \times A_i$ is a binary relation and $DT \subseteq A_i \times \vartheta$ is a binary relation such that $\vartheta \in j$ *or* $\vartheta \in A_j$ *or* $\vartheta \in \rightarrow_j$.

The extended argumentation framework $TEAF$ would need new semantics in order to compute what are the accepted arguments. In alternative, we use the meta-argumentation methodology to flatten the $TEAF$ to a meta-argumentation framework where classical Dung's semantics are used to compute the set of acceptable arguments. We define the meta-argumentation framework in the following way where the unflattening function g and the acceptance function $\mathcal{E}'$ are defined as above.

Definition 10: Given a trust-based extended argumentation framework $TEAF = \langle\langle A_1, \rightarrow_1, DT_1\rangle, \ldots, \langle A_n, \rightarrow_n, DT_n\rangle\rangle$, see Definition 9, the set of meta-arguments MA is $\{trust(i) \mid 1 \leq i \leq n\} \cup \{acc(a) \mid a \in A_1 \cup \ldots \cup A_n\} \cup \{X_{a,b}, Y_{a,b} \mid a, b \in A_1 \cup \ldots \cup A_n\} \cup \{Z_a \mid a \in A_1 \cup \ldots \cup A_n\}$ and $\longmapsto \subseteq MA \times MA$ is a binary relation on MA such that:

- $acc(a) \longmapsto X_{a,b}, X_{a,b} \longmapsto Y_{a,b}, Y_{a,b} \longmapsto accept(b)$ iff $a, b \in A_i$ and $a \rightarrow_i b$, and
- $trust(i) \longmapsto X_{trust(i),Z_{acc(a)}}, X_{trust(i),Z_{acc(a)}} \longmapsto Y_{trust(i),Z_{acc(a)}}, Y_{trust(i),Z_{acc(a)}} \longmapsto Z_{acc(a)}, Z_{acc(a)} \longmapsto acc(a)$ iff $a \in A_i$, and
- $trust(i) \longmapsto X_{trust(i),Z_{Y_{a,b}}}, X_{trust(i),Z_{Y_{a,b}}} \longmapsto Y_{trust(i),Z_{Y_{a,b}}}, Y_{trust(i),Z_{Y_{a,b}}} \longmapsto Z_{Y_{a,b}}, Z_{Y_{a,b}} \longmapsto Y_{a,b}$ iff $a, b \in A_i$ and $a \rightarrow_i b$, and
- $trust(i) \longmapsto Z_{acc(a)}, Z_{acc(a)} \longmapsto acc(a), acc(a) \longmapsto X_{acc(a),trust(j)}, X_{acc(a),trust(j)} \longmapsto Y_{acc(a),trust(j)}, Y_{acc(a),trust(j)} \longmapsto trust(j)$ iff $a \in A_i$ and $aT_i trust(j)$, and
- $trust(i) \longmapsto Z_{acc(a)}, Z_{acc(a)} \longmapsto acc(a), acc(a) \longmapsto X_{acc(a),Y_{trust(j),Z_{acc(b)}}}, X_{acc(a),Y_{trust(j),Z_{acc(b)}}} \longmapsto Y_{acc(a),Y_{trust(j),Z_{acc(b)}}}, Y_{acc(a),Y_{trust(j),Z_{acc(b)}}} \longmapsto Y_{trust(j),Z_{acc(b)}}$ iff $a \in A_i, b \in A_j$ and $aT_i b$, and
- $trust(i) \longmapsto Z_{acc(a)}, Z_{acc(a)} \longmapsto acc(a), acc(a) \longmapsto X_{acc(a),Y_{trust(j),Z_{Y_{b,c}}}}, X_{acc(a),Y_{trust(j),Z_{Y_{b,c}}}} \longmapsto Y_{acc(a),Y_{trust(j),Z_{Y_{b,c}}}}, Y_{acc(a),Y_{trust(j),Z_{Y_{b,c}}}} \longmapsto Y_{trust(j),Z_{Y_{b,c}}}$ iff $a \in A_i, b, c \in A_j$ and $aT_i(b \rightarrow_j c)$.

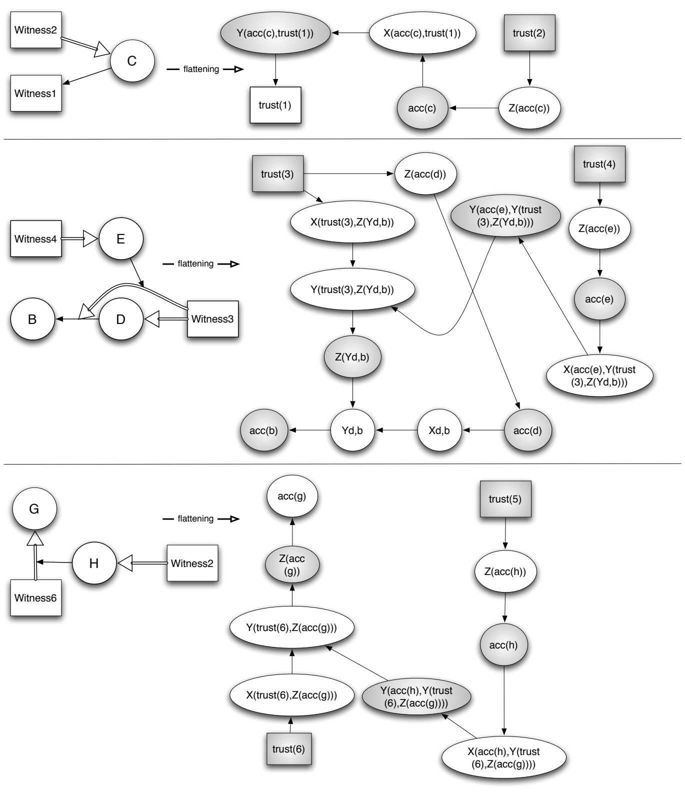

Fig. 5. Fine grained trust in argumentation.

Definition 10 shows how to instantiate an extended argumentation framework composed by a set of arguments, a binary attack relation and a binary distrust relation with meta-arguments. In particular, the last three points model respectively a distrust relationship towards an agent, a distrust relationship towards an argument and a distrust relationship towards an attack relation.

Example 2: In Figure 5 we highlight the three patterns where trust relations between information sources are represented. The first pattern shows that Witness2 attacks the trustworthiness of Witness1 with the argument c. In meta-argumentation, we have that $trust(2)$ proposes

$acc(c)$ by attacking meta-argument $Z_{acc(c)}$ and, with meta-arguments X, Y, it attacks $trust(1)$. This means that if Witness1 is not reliable then each of his arguments and attack relations cannot be acceptable either. If we look at the extension of this pattern, we have that the set of acceptable arguments for the meta-argumentation framework is $\mathcal{E}(f(pattern1)) = \{trust(2), acc(c), Y_{acc(c),trust(1)}\}$. If we consider instead the other two patterns of Figure 5, we have that the attack is directed against a precise element sustained by the other source. On the one hand, Witness4 attacks the attack relation $d \rightarrow b$ proposed by Witness3. This is achieved in meta-argumentation by an attack from meta-argument $acc(e)$, proposed by $trust(4)$, to the attack relation characterized by meta-argument $Y_{d,b}$. The set of acceptable arguments is $\mathcal{E}(f(pattern2)) = \{trust(4), trust(3), acc(d), acc(e), acc(b), Y_{acc(e),Y_{trust(3),Z_{Y_{b,d}}}}, Z_{Y_{d,b}}\}$. Witness3's attack relation $d \rightarrow b$ is evaluated as not reliable for Witness4 and it is not acceptable. On the other hand, Witness2 evaluates unreliable Witness6 concerning argument g. In meta-argumentation, $trust(2)$, by means of meta-argument $acc(h)$, attacks meta-argument $acc(g)$ proposed by $trust(6)$. In this case, the set of acceptable arguments is $\mathcal{E}(f(pattern3)) = \{trust(2), trust(6), acc(h), Y_{acc(h),Y_{trust(6),Z_{acc(g)}}}, Z_{acc(g)}\}$.

C. Representing the evidences supporting arguments

The evidences in favor of the arguments are represented, as discussed before, as a support given by the agents to the arguments at the object level. At the meta-level, this is modeled as an attack relation from meta-argument $trust(i)$ to the Z meta-arguments. However, there are also other cases in which evidences are necessary to support the acceptability of an argument. Let consider the case in which the trustworthiness of an agent is attacked. What does it happen to the arguments put forward by this agent? They become not acceptable. In this case, what is needed to reinstate the acceptability of these arguments is an evidence. This evidence is provided under the form of an argument put forward by another agent.

Definition 8, previously introduced, presents how to instantiate an extended argumentation framework composed by a set of arguments, a binary attack relation, a binary second-order attack relation representing also the information sources. Definition 9, instead, extends Dung's framework with a distrust relation DT. In order to have an extended argumentation framework with both the relations of the EAF of Definition 8 and the $TEAF$ of Definition 10, we define an extended trust-based argumentation framework with an evidence relation $\looparrowright$ which represents the evidences provided in favor of the arguments of the other agents.

Definition 11: A trust-based argumentation framework with evidences $TEAF^2 = \langle\langle A_1, \rightarrow_1, \rightarrow_1^2, \looparrowright_1, T_1\rangle, \ldots, \langle A_n, \rightarrow_n, \rightarrow_n^2, \looparrowright_n, T_n\rangle\rangle$ where $\looparrowright_i$ is a binary relation on $A_i \times A_j$ and the set of meta-arguments MA is $\{trust(i) \mid 1 \leq i \leq n\} \cup \{acc(a) \mid a \in A_1 \cup \ldots \cup A_n\} \cup \{X_{a,b}, Y_{a,b} \mid a, b \in A_1 \cup \ldots \cup A_n\} \cup \{Z_a \mid a \in A_1 \cup \ldots \cup A_n\}$ and $\longmapsto \subseteq MA \times MA$ is a binary relation on MA such that hold the conditions of Definition 8 and Definition 10 and:

- $acc(a) \longmapsto Z_{acc(b)}, Z_{acc(b)} \longmapsto acc(b)$ iff $a, b \in A_i$ and $a \looparrowright_i b$.

Example 3: Let us consider again the informal argument exchange. We have that argument f is "The guy has another car" while argument g by Witness6 is " The guy parked two cars in my underground parking garage three weeks ago". Argument g is an evidence in favor of f. This evidence is expressed in meta-argumentation as an attack from meta-argument $acc(g)$ to the meta-argument $Z_{acc(f)}$ attacking $acc(f)$. This example is described in Figure 6.

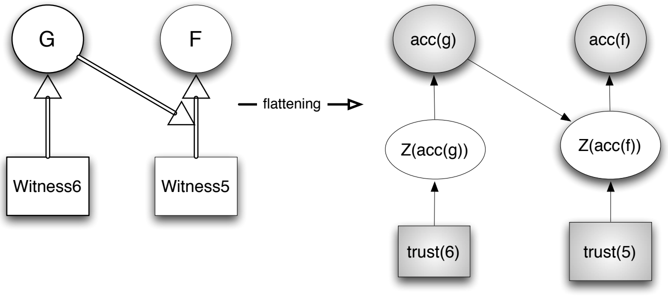

Fig. 6. Introducing evidences in favor of the arguments.

D. Representing arguments about other agents' arguments

The information sources may also express arguments concerning other agents' arguments as in the case of arguments i and a during the informal argument exchange. In this case we have that Witness7 proposes an argument which is based on the argument of another agent, Witness1. Moreover, we have that Witness3 introduces argument l which attacks the support of argument i to argument a. If we add to the extended trust-based argumentation framework $TEAF^2$ a new binary relation $\dashrightarrow \subseteq A_i \times A_i$, we can model the report of other agents' arguments in the following way:

Definition 12: Given an extended trust-based argumentation framework $TEAF^2 = \langle\langle A_1, \rightarrow_1, \rightarrow_1^2, \looparrowright_1, T_1, \dashrightarrow_1\rangle, \ldots, \langle A_n, \rightarrow_n, \rightarrow_n^2, \looparrowright_n, T_n, \dashrightarrow_n\rangle\rangle$, the set of meta-arguments MA is $\{trust(i) \mid 1 \leq i \leq n\} \cup \{acc(a) \mid a \in A_1 \cup \ldots \cup A_n\} \cup \{X_{a,b}, Y_{a,b} \mid a, b \in A_1 \cup \ldots \cup A_n\} \cup \{Z_a \mid a \in A_1 \cup \ldots \cup A_n\} \cup \{dixit_{a,b} \mid a, b \in A_1 \cup \ldots \cup A_n\}$ and $\longmapsto \subseteq MA \times MA$ is a binary relation on MA such that the conditions of Definition 8, Definition 10 and Definition 11 hold and:

- $acc(a) \longmapsto Z_{dixit_{a,b}}, Z_{dixit_{a,b}} \longmapsto dixit_{a,b}, dixit_{a,b} \longmapsto Z_{acc(b)}, Z_{acc(b)} \longmapsto acc(b)$ iff $a, b \in A_i$ and $a \dashrightarrow_i b$ and
- $acc(a) \longmapsto X_{acc(a),dixit_{b,c}}, X_{acc(a),dixit_{b,c}} \longmapsto Y_{acc(a),dixit_{b,c}}, Y_{acc(a),dixit_{b,c}} \longmapsto dixit(b,c)$ iff $a \in A_i, b, c \in A_j$ and $aT_i(b \dashrightarrow_i c)$.

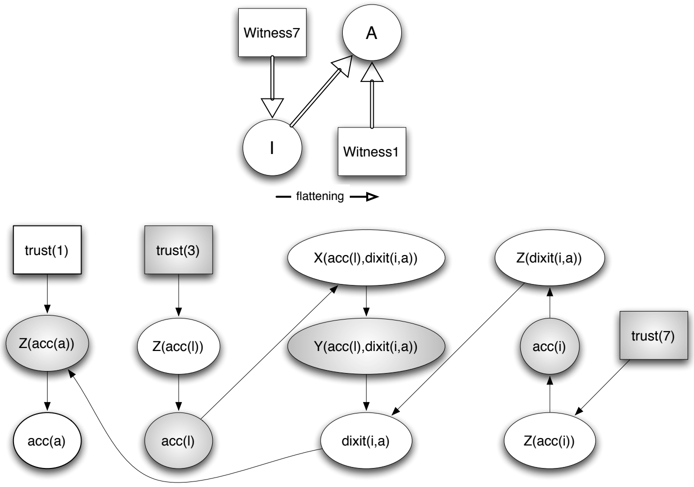

Fig. 7. Introducing arguments about other agents' arguments.

Example 4: Let $TEAF^2$ be defined as $A_1 = \{a, b\}$, $\rightarrow_1 = \{(b, a)\}$, $A_2 = \{c, h\}$, $DT_2 = \{(c, 1), (h, g)\}$, $A_3 = \{b, d, l\}$, $\rightarrow_3 = \{(d, b)\}$, $DT_3 = \{(l, \dashrightarrow_7 (i, a))\}$, $A_4 = \{e\}$, $DT_4 = \{(e, \rightarrow_3 (d, b))\}$, $A_5 = \{a, b, f\}$, $\rightarrow_5^2 = \{(f, (b, a))\}$, $A_6 = \{a, g\}$, $\looparrowright_6 = \{(g, f)\}$, $A_7 = \{i\}$, $\dashrightarrow_7 = \{(i, a)\}$. This extended argumentation framework is the model of the informal argument exchange proposed in the introduction. In Figure 7, we introduce arguments i and l under the form of meta-arguments. We have that argument i reports an argument from the guy which sustains argument a proposed by $trust(1)$. $trust(7)$ proposes meta-argument $acc(i)$ and a *dixit* relation $\dashrightarrow_7 = \{(i, a)\}$ holds between arguments i and a. This is represented in meta-argumentation by the new meta-argument $dixit_{i,a}$ which is sustained, by means of meta-argument $Z_{dixit_{i,a}}$, by $acc(i)$. Witness3 does not agree about the link between arguments i and a because in his opinion the guy admitted to kill the boss only to protect his wife (argument l). This attack has the aim to eliminate the support which argument i gives to argument a. We represent it in meta-argumentation by adding an attack relation from meta-argument $acc(l)$ claimed by $trust(3)$ to meta-argument $dixit_{i,a}$. In this way the attack of $dixit(i, a)$ to $Z_{acc(a)}$ is made ineffective and meta-argument $acc(a)$ would be acceptable only if sustained by some other meta-argument. The set of acceptable arguments of the $TEAF^2$ is $\mathcal{E}'(TEAF^2) = g(\mathcal{E}(f(TEAF^2))) = \{b, c, d, e, f, h, i, l\}$ and the trustable information sources are all the witnesses except Witness1 who has been directly attacked by Witness2.

IV. Related work

Dix et al. [29] present trust as a major issue in MAS applications concerning the research challenges for argumentation. Argumentation has been used in a number of applications to handle trust, for example weighting inferences with measures of trust in the source of the information, and updating these measures as sources prove to be more or less reliable. *Which agents are trustworthy?* This is important for taking decisions and weighing arguments of other agents. Argumentation has already been applied to weigh inferences, e.g. trust in sources given as databases. *Can this theory be extended to define a notion of trust between agents?* We propose to adopt the methodology of meta-argumentation to answer these questions.

Parsons et al. [30] discuss why argumentation has an important role to play in reasoning about trust and highlight what are the mechanisms which need to be investigated through argumentation. The authors claim that a first problem, particularly of abstract approaches such that of [3], is that they cannot express the provenance of trust and they cannot express the fact that b is attacked because b is based on agent s and there is an evidence that s is not trustworthy. In this paper, we propose a methodology which allows us to instantiate Dung's framework with meta-arguments which represent the information sources. Moreover, we show how to express trust relationships between the sources. Another problem highlighted by [30] is the explicit expression of degrees of trust, as adopted by the prevalence of numerical measures of trust in the literature. The authors propose the system TL as possible solution where they introduce in the tuple representing the database a new element which is an ordered sequence of elements from a dictionary and these elements could be associated to numerical measures of trust. In this paper, we do not present an explicit expression of degrees of trust and this is a topic for further research but we present a model where a fine grained view of trust relationships is provided and this allows us to reason about trust in argumentation.

Matt et al. [31] propose an extension to the Dempster-Shafer belief function. The authors allow the evaluator agent to take into account, in addition to the statistical data, a set of justified claims concerning the expected behaviour of the target agent. These claims form the basis of the evaluator's opinions and are formally represented by arguments in abstract argumentation. Two kinds of arguments are defined: forecast arguments and mitigation arguments. Forecast arguments express the trustworthiness or untrustworthiness of the target agent and mitigation arguments attack forecast arguments or other mitigation arguments because of the uncertainties of the validity of forecast arguments. Dempster-Shafer belief function is constructed both from statistical data and from these arguments. Arguments are generated by contracts and a strength function assigns 1 to each unattacked argument and a varying value if it is attacked by a mitigation argument. We propose an approach to the introduction of trust which is more related to modelling and no statistical problems are addressed. We show how to introduce the provenance of the information in the argumentation framework. [31] have focused on the computation of trust by an evaluator of a target in isolation. We propose a model in which all the trust relationships are evaluated together and we do not restrict our model to the contracts.

Stranders et al. [32] propose an approach to trust based on argumentation in which there is a separation between the opponent modeling and decision making. The opponents' behaviour is modeled using possibilistic logic. They start from the work of [33] which supports reasoning under uncertainty with fuzzy logic. The fuzzy generalization of [33]'s approach is combined with a fuzzy rule learner. The paper shows the results based on the ART testbed. In our approach we use Dung's abstract framework and we do not present a decision making approach to trust. We are interested in modeling fine grained trust in argumentation and we do not present experimental results.

Prade [34] presents a bipolar qualitative argumentative modeling of trust where a finite number of levels is assumed in a trust scale and trust and distrust are assessed independently. The author introduces the notion of reputation which is viewed as an input information used by an agent for revising or updating his trust evaluation. Reputation contributes also to provide direct arguments in favour or against a trust evaluation. There are a number of differences between [34] and our approach. First, we does not apply a diagnostic point of view as in [34] but we are interested in a social multiagent perspective. Second, we use a Dung's based approach while in [34] arguments have an abductive form.

An approach related to trust in argumentation is provided by Hunter [35] where the author introduces a logic-based meta-level argumentation framework for evaluating arguments in terms of the appropriateness of their proponents. A further investigation of the relation between trust evaluation and proponents' appropriateness is an interesting direction for future research.

There are two main differences which differentiate our approach to modeling trust in argumentation theory and the works described above. The first one consists in a purely qualitative approach. We do not provide numerical measures of trust as instead it is done by [31], [32] and, with a fuzzy evaluation, by [34]. The second difference consists in a new methodology to introduce trust in argumentation theory from a design perspective. We does not introduce new components in the framework, we just use Dung's argumentation to model itself in such a way to deal with trust.

V. Conclusions

In this paper, a way to model trust in Dung's framework is presented. We answer the research questions using the methodology of meta-argumentation where Dung's framework is used to reason about itself. Meta-argumentation, presented by Boella et al. [1], [2], allows us to introduce the notion of trust without extending Dung's standard argumentation framework, reusing Dung's semantics and properties.

We represent the sources of information in the abstract argumentation framework in order to link the agents to the arguments they construct. We introduce the agents as meta-arguments of the kind "agent i is trustable", $trust(i)$, and each agent is linked to the arguments he proposes by means of meta-arguments Z_x. Meta-arguments $trust(i)$ attack meta-arguments Z_x when x is an argument or an attack relation put forward by agent i. For each agent who sustains argument/attack x, there is an attack from $trust(i)$ to Z_x. In this way, the argumentation framework keeps track of the provenance of the arguments and attack relations and it allows us to represent evidences in the framework. More than one agent can support the same argument. This is expressed in meta-argumentation in the following way. If the agents support directly an argument or attack x, then they both attack meta-argument Z_x. If the agents propose new arguments which sustain other arguments then this is expressed by an attack from meta-argument $acc(a)$ to $Z_{acc(b)}$, where argument a is an evidence of argument b.

The trustworthiness of the agents can be attacked by attacking meta-arguments $trust(i)$ representing the agents in the argumentation framework. The agents, supporting arguments against the trustworthiness of the other agents, attack the reliability of the other agents. Trust is represented as an absence of attacks on the agents' trustworthiness. The agents who are not evaluated as reliable in the framework are those whose meta-argument $trust(i)$ is not in the extension of the meta-argumentation framework.

We present a fine grained view of trust relationships. The agents can express their evaluation on other agents' reliability also concerning single arguments and attack relations proposed by the unreliable agents. We express the evaluation of the untrustworthiness of arguments and attacks by means of attacks to the Y_{Z_x} meta-argument which is used by meta-argument $trust(i)$ to attack meta-argument Z_x.

If the arguments or attack relations evaluated unreliable are not supported by other evidences, such as arguments which attacks the Z_x meta-argument, then they are made unacceptable in the extension of the meta-argumentation framework.

Agents can express reported arguments about arguments expressed by other agents. This is represented in meta-argumentation with meta-argument $dixit_{a,b}$ where argument a reports what is expressed by argument b. Meta-argument $acc(a)$ supports, by means of meta-argument $Z_{dixit(a,b)}$, meta-argument $dixit_{a,b}$ which supports $acc(b)$. In this way, the support of the $dixit$ meta-argument can be attacked by other arguments if the agents believe it to be an unreliable information.

Future research is addressed following different lines. First, we are defining and formally proving a number of properties of our model of trust in argumentation. We are also considering properties, such as the trust transitivity, on which there is not a unique opinion in the literature. Second, we are studying how to express trust revision. As highlighted also by [30], an important aspect in reasoning about trust is the need for a source to be able to revise the trust she has in another source based on experience. Moreover, the notion of reputation may be viewed as another information used by the agents for revising or updating their own trust evaluation.

References

[1]G. Boella, L. van der Torre, and S. Villata, "On the acceptability of meta-arguments," in *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2009*. IEEE, 2009, pp. 259–262.

[2]G. Boella, D. M. Gabbay, L. van der Torre, and S. Villata, "Meta-argumentation modelling i: Methodology and techniques," *Studia Logica*, vol. 93, no. 2-3, pp. 297–355, 2009.

[3]P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games," *Artif. Intell.*, vol. 77, no. 2, pp. 321–357, 1995.

[4]P. Baroni and M. Giacomin, "On principle-based evaluation of extension-based argumentation semantics," *Artif. Intell.*, vol. 171, no. 10-15, pp. 675–700, 2007.

[5]M. Wooldridge, P. McBurney, and S. Parsons, "On the meta-logic of arguments," in *AAMAS*, F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, and M. Wooldridge, Eds. ACM, 2005, pp. 560–567.

[6]T. J. M. Bench-Capon and P. E. Dunne, "Argumentation in artificial intelligence," *Artif. Intell.*, vol. 171, no. 10-15, pp. 619–641, 2007.

[7]A. C. Kakas and P. Moraitis, "Argumentation based decision making for autonomous agents," in *Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS*. ACM, 2003, pp. 883–890.

[8]L. Amgoud, "An argumentation-based model for reasoning about coalition structures," in *ArgMAS*, 2005, pp. 217–228.

[9]S. Toulmin, *The Uses of Argument*. Cambridge University Press, 1958.

[10]K. Atkinson and T. J. M. Bench-Capon, "Legal case-based reasoning as practical reasoning," *Artif. Intell. Law*, vol. 13, no. 1, pp. 93–131, 2005.

[11]H. Prakken, "An abstract framework for argumentation with structured arguments," Utrecht University, Tech. Rep. UU-CS-2009-019, 2009.

[12]L. Amgoud and C. Cayrol, "A reasoning model based on the production of acceptable arguments," *Ann. Math. Artif. Intell.*, vol. 34, no. 1-3, pp. 197–215, 2002.

[13]S. Kaci and L. van der Torre, "Preference-based argumentation: Arguments supporting multiple values," *Int. J. Approx. Reasoning*, vol. 48, no. 3, pp. 730–751, 2008.

[14]T. Bench-Capon, "Persuasion in practical argument using value-based argumentation frameworks," *J. Logic and Computation*, vol. 13, no. 3, pp. 429–448, 2003.

[15]S. Modgil, "An abstract theory of argumentation that accommodates defeasible reasoning about preferences," in *ECSQARU*, 2007, pp. 648–659.

[16]H. Barringer, D. M. Gabbay, and J. Woods, "Temporal dynamics of support and attack networks: From argumentation to zoology," *Mechanizing Mathematical Reasoning*, pp. 59–98, 2005.

[17]S. Modgil, "Reasoning about preferences in argumentation frameworks," *Artif. Intell.*, vol. 173, no. 9-10, pp. 901–934, 2009.

[18]C. Cayrol and M.-C. Lagasquie-Schiex, "On the acceptability of arguments in bipolar argumentation frameworks," in *ECSQARU*, 2005, pp. 378–389.

[19]H. Prakken and G. Sartor, "Argument-based extended logic programming with defeasible priorities," *Applied Non-Classical Logics*, vol. 7, no. 1, pp. 25–75, 1997.

[20]S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis, "On the merging of dung's argumentation systems," *Artif. Intell.*, vol. 171, no. 10-15, pp. 730–753, 2007.

[21]H. Jakobovits and D. Vermeir, "Robust semantics for argumentation frameworks," *J. Log. Comput.*, vol. 9, no. 2, pp. 215–261, 1999.

[22]S. Modgil and T. Bench-Capon, "Integrating object and meta-level value based argumentation," in *COMMA*, vol. 172, 2008, pp. 240–251.

[23]P. Baroni, F. Cerutti, M. Giacomin, and G. Guida, "Encompassing attacks to attacks in abstract argumentation frameworks," in *ECSQARU*, ser. Lecture Notes in Computer Science, C. Sossai and G. Chemello, Eds., vol. 5590. Springer, 2009, pp. 83–94.

[24]D. M. Gabbay, "Fibring argumentation frames," *Studia Logica*, vol. 93, no. 2-3, pp. 231–295, 2009.

[25]——, "Semantics for higher level attacks in extended argumentation frames part 1: Overview," *Studia Logica*, vol. 93, no. 2-3, pp. 357–381, 2009.

[26]C. Castelfranchi and R. Falcone, "Social trust: A cognitive approach," *Trust and Deception in Virtual Societies*, pp. 55–90, 2001.

[27]D. Gambetta, "Can we trust them?" *Trust: Making and breaking cooperative relations*, pp. 213–238, 1990.

[28]C.-J. Liau, "Belief, information acquisition, and trust in multi-agent systems–a modal logic formulation," *Artif. Intell.*, vol. 149, no. 1, pp. 31–60, 2003.

[29]J. Dix, S. Parsons, H. Prakken, and G. R. Simari, "Research challenges for argumentation," *Computer Science - R&D*, vol. 23, no. 1, pp. 27–34, 2009.

[30]S. Parsons, P. McBurney, and E. Sklar, "Reasoning about trust using argumentation: A position paper," in *Seventh International Workshop on Argumentation in Multi-Agent Systems, ArgMAS, In press*, 2010.

[31]P.-A. Matt, M. Morge, and F. Toni, "Combining statistics and arguments to compute trust," in *Ninth International Conference on Autonomous Agents and Multiagent Systems, AAMAS, In press*, 2010.

[32]R. Stranders, M. de Weerdt, and C. Witteveen, "Fuzzy argumentation for trust," in *CLIMA VIII*, ser. Lecture Notes in Computer Science, F. Sadri and K. Satoh, Eds., vol. 5056. Springer, 2007, pp. 214–230.

[33]L. Amgoud and H. Prade, "A possibilistic logic modeling of autonomous agents negotiation," in *EPIA*, ser. Lecture Notes in Computer Science, F. Moura-Pires and S. Abreu, Eds., vol. 2902. Springer, 2003, pp. 360–365.

[34]H. Prade, "A qualitative bipolar argumentative view of trust," in *SUM*, ser. Lecture Notes in Computer Science, H. Prade and V. S. Subrahmanian, Eds., vol. 4772. Springer, 2007, pp. 268–276.

[35]A. Hunter, "Reasoning about the appropriateness of proponents for arguments," in *Proceedings of Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008*, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 89–94.