Coalition Formation via Negotiation in Multiagent Systems with Voluntary Attacks

Guido Boella [a] Dov M. Gabbay [b] Alan Perotti [a] Serena Villata [a]

[a] *University of Turin, Italy*
[b] *King's College London, UK*

Abstract

Argumentation networks are put forward by Dung considering only one kind of attack among arguments. In this paper, we propose to extend Dung's argumentation framework with voluntary attacks in the context of multiagent systems, characterized by the possibility of the attacker to decide whether to attack or not. Enabling voluntary attacks impacts on the acceptability of the arguments in the framework, and therefore it becomes subject of debate between the agents. Agents can negotiate about which subset of voluntary attacks can be raised, and they form coalitions after the negotiation process.

1 Introduction

In this paper, we present a new kind of argumentation framework, called voluntary argumentation framework, which has two kinds of attacks, actual ones and voluntary ones and we show how the agents negotiate about these attacks in order to form coalitions. This framework starts from the observation that, in Dung's argumentation [7], all the attacks are actual ones and the decisions of an agent not to attack another agent's argument even if she has the possibility to do so have not been analyzed. A basic example in law is the difference between an accusation, which is automatically raised and can be represented by an attack in the argumentation framework, and a legal action which is just a threat of suing, that may become an actual sue or nothing at all, depending on the attacker's choice. In this paper, we answer the following research question: *How do agents reason together with the other agents about voluntary attacks?*

We firstly present an extended abstract argumentation framework with voluntary attacks. Secondly, we embed this framework in a multiagent system, partitioning the set of arguments amongst the agents. Each partition of arguments associated to an agent represents the reasons behind her beliefs and actions. Agents propose a number of arguments and try to maximize their acceptability in the multiagent system. The decision of raising or not a voluntary attack corresponds to different kinds of refinements of the abstract framework into a fully instantiated one. Firing a voluntary attack can modify the extensions of the argumentation framework, and therefore it may lead to a contrast between the agents since some of them can see their arguments becoming not acceptable after the addition of the voluntary attack. Agents negotiate about which subset of voluntary attacks shall be fired and they form coalitions based on common interests and strategies or reached compromises.

Consider the scenario of the preparation of an electoral manifesto by an alliance of political parties: some opinions are clearly incompatible and can be modeled with a standard actual attack relation, while some discrepancies are more subtle and agents can be open to negotiate about them. An agent who firmly cares about laicism will attack another agent's proposal to defend the sacred nature of life. The same attacking agent may dislike an argument about the exemption from property taxes for Church-owned properties, but she may decide not to attack that argument if she gains something in return: for instance, getting the second agent to attack one of the first agent's attackers.

This paper is organized as follows. Section 2 provides a background on Dung's theory. In section 3 we introduce voluntary attacks and the relationship between arguments and attack relations and the agents. Section 4 shows how agents negotiate and form coalitions in order to reach a compromise about which voluntary attacks shall be raised. Conclusions end the paper.

2 Argumentation frameworks

Dung's theory [7] is based on a binary *attack* relation among arguments, which are abstract entities whose role is determined by their relation to other arguments. Its structure and its origin may be not known. We restrict ourselves to *finite* argumentation frameworks, i.e., in which the set of arguments is *finite*.

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

We follow Baroni and Giacomin [2] in defining an acceptance function $\mathcal{E}$ that associates with a set of arguments $A \subseteq \mathcal{U}$, a set of arguments produced by a reasoner at a given instant of time, and a binary relation $\rightarrow \subseteq A \times A$, representing the dominance or attack relation among these arguments, the sets of acceptable arguments of A.

Definition 2 (Acceptance function) *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 $A \subseteq \mathcal{U}$ and $\rightarrow \subseteq A \times A$, and which associates with argumentation framework $\langle A, \rightarrow \rangle$ sets of subsets of A: $\mathcal{E}(\langle A, \rightarrow \rangle) \subseteq 2^A$.*

Definition 3 (Conflict free) *Given an argumentation framework $AF = \langle A, \rightarrow \rangle$, a set $C \subseteq A$ is* conflict free*, denoted as cf(C), if and only if there does not exist $\alpha, \beta \in C$ such that $\alpha \rightarrow \beta$.*

The following definition summarizes the most widely used acceptability semantics. Which semantics is most appropriate depends on the application domain of the argumentation theory.

Definition 4 (Acceptability semantics) *Let $AF = \langle A, \rightarrow \rangle$ be an argumentation framework. Let $\mathcal{S} \subseteq A$. $\mathcal{S}$ defends a if for all $b \in \mathcal{A}$ such that $b \rightarrow a$, there exists $c \in \mathcal{S}$ such that $c \rightarrow b$. Let $D(\mathcal{S}) = \{a \mid \mathcal{S}$ defends $a\}$.*

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

A semantics $\mathcal{S}$ satisfies the conflict-free principle if and only if for all AF and for all $E \in \mathcal{E}_{\mathcal{S}}(AF)$: *cf*(E). Caminada [6] show that these semantics can also be described by a three valued argument labeling, where the first two conditions represent conflict free and defense, and the third one represents the so-called reinstatement principle.

Proposition 1 *Let $L^{\langle A, \rightarrow \rangle}_{admiss} : A \rightarrow \{in, out, undecided\}$ be a complete labeling function for semantics satisfying the conflict free principle such that:*

1. *$L^{\langle A, \rightarrow \rangle}_{admiss}(b) = out$ iff $\exists a \rightarrow b : L^{\langle A, \rightarrow \rangle}_{admiss}(a) = in$*
2. *if $L^{\langle A, \rightarrow \rangle}_{admiss}(b) = in$ then $\forall a \rightarrow b : L^{\langle A, \rightarrow \rangle}_{admiss}(a) = out$*

and let $L^{\langle A, \rightarrow \rangle}_{compl} : A \rightarrow \{in, out, undecided\}$ be a complete labeling function such that in addition:

3. *if $\forall a \rightarrow b : L^{\langle A, \rightarrow \rangle}_{compl}(a) = out$ then $L^{\langle A, \rightarrow \rangle}_{compl}(b) = in$.*

Then we have the following results.
$\mathcal{E}_{admiss}(AF) = \{\{a \mid L^{\langle A, \rightarrow \rangle}_{admiss}(a) = in\} \mid \exists L^{\langle A, \rightarrow \rangle}_{admiss}\}$
$\mathcal{E}_{compl}(AF) = \{\{a \mid L^{\langle A, \rightarrow \rangle}_{compl}(a) = in\} \mid \exists L^{\langle A, \rightarrow \rangle}_{compl}\}$

3 Agents negotiation and voluntary attacks

We extend Dung's argumentation framework [7] by adding another kind of binary relation which represents voluntary attacks. Voluntary attacks do not impact on arguments' acceptability in any way, hence the previous definitions of defense, conflict-free and acceptability semantics still apply to our extended argumentation framework. An agent takes into account whether to raise a voluntary attack or not and the voluntary attack becomes either an actual one, impacting on the acceptability of the arguments, or it is eliminated from the framework. The decision of raising a voluntary attack corresponds to a refinement of the arguments involved in the voluntary attack relation. Refining an argument means to refine one of the literals composing it. A literal is refined by assigning to one or more of the variables composing it a constant value. The process of refining the abstract framework adds, in fact, the information about the internal structure of the argument, potentially modifying, as a result, the relations with the other arguments and the nature of framework.

Definition 5 (Extended Argumentation framework) *An extended argumentation framework EAF is a tuple $\langle A, \rightarrow, \twoheadrightarrow \rangle$ where A is a finite set (of arguments), $\rightarrow$ and $\twoheadrightarrow$ are two disjunct binary relations defined on $A \times A$: respectively called attack and voluntary attack relations.*

Let Ag be a finite set of agents. Ag is not part of the EAF but it is related to it via the f_{bel} function.

Definition 6 (f_{bel} function) *Let $f_{bel} : A \rightarrow Ag$, $\forall a \in A\ \exists ag_i \in Ag\ s.t.\ f_{bel}(a) = ag_i$.*

Informally, each agent owns a subset of arguments, and each argument belongs to exactly one agent. Argument j threatens argument k if $j, k \in A$ and $(j, k) \in \twoheadrightarrow$. As mentioned before, voluntary attacks do not impact on the acceptability of arguments, since they are only potential attacks, and can turn into actual attacks or be discarded. Let $(j, k) \in \twoheadrightarrow$ and let $f_{bel}(j) = ag_1$. Agent ag_1 controls the threatening argument j and can decide whether to raise the attack (j, k) or not. In the first case, the voluntary attack becomes an actual one, and the EAF changes as follows: $EAF' =< A, \rightarrow', \twoheadrightarrow' >$, where $\rightarrow' = \rightarrow \cup (j, k)$ and $\twoheadrightarrow' = \twoheadrightarrow \setminus (j, k)$. Otherwise, (j, k) is just removed from $\twoheadrightarrow$. Note that, in the first case, ag_1's decision to fire the attack (j, k) modifies the attack relation $\rightarrow$ and hence the framework's extensions.

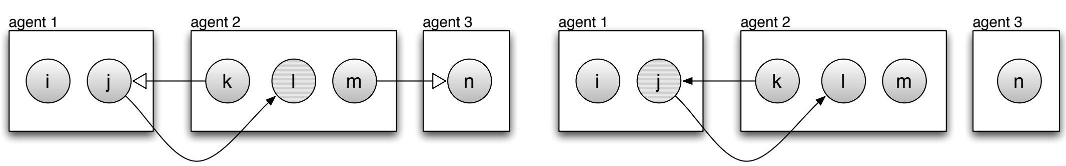

Figure 1: Example of argumentation framework with voluntary attacks and firing choices

The left side of Figure 1 shows an extended argumentation framework $EAF =< \{i, j, k, l, m, n\}, \{(j, l)\}, \{(k, j), (m, n)\} >$ and agents set $Ag = \{1, 2, 3\}$. Arguments are circles, attacks are edges and voluntary attacks are triangle-head arrows. Agents are represented by rectangles. This is a basic framework, but it allows to explain the choice of firing a voluntary attack. Consider agent 2. She owns three arguments in her argument set (k, m, n) and has the chance to raise two voluntary attacks, (k, j) and (m, n). Firing (m, n) is pointless, because it would not change the acceptability of any of agent 2's arguments. But agent 2 may want to raise the (k, j), because doing so she attacks one of her arguments' attacker, thus defending her own argument l. Therefore, if agent 2 chooses to fire (k, j) but not (m, n), the resulting framework is the one depicted in the right side of Figure 1. Note that, after a phase of decision making, the resulting framework contains no voluntary attacks, and is therefore a Dung's argumentation framework. In this example, agent 2 is the only agent who may raise voluntary attacks, and therefore she can make decisions on her own. In a slightly more complex scenario, where different agents can raise voluntary attacks, the social context has to be taken in account, and agents shall communicate and negotiate in order to reach a compromise about which subset of voluntary attacks shall be fired.

3.1 Agents

As stated above, agents know all about the argumentation framework, i.e. their and the others' arguments and attack relations. But each agent holds a specific knowledge about his own arguments, voluntary attacks, and her desire to join coalitions. We introduce three parameters, respectively α, β and γ, in order to characterize agents' preferences on arguments, voluntary attacks and coalitions.

Definition 7 (Agents' parameters)

- *Parameter α_i^j is associated to every argument j owned by ag_i and describes the rating of ag_j.*
- *Parameter $\beta_i^{j,k}$ is defined if $i, j \in A$ and $(j, k) \in \twoheadrightarrow$ and $f_{bel}(j) = ag_i$.*
- *Parameter γ_i is the relative sociality of agent ag_i.*

Parameter α_i^j captures the fact that each agent owns different arguments, but their relevance for her may vary. Parameter $\beta_i^{j,k}$ describes the cost, for agent ag_i, to raise the attack (j, k). Parameter γ_i models the willing to be part of a coalition of agent i. The principal aim of an agent is to maximize her arguments' acceptability. The second aim is to use the least possible effort in voluntary attacks raising and the final aim is to be part of a coalition. There is a sort of triple trade-off among these three goals that each agent aims to. Each agent prefers the arguments with an associated high α parameter and will raise, where useful, lower-cost attacks depending on parameter β; moreover, the agents which desire to be part of a coalition have an high γ parameter. Each agent's sociality γ is relative and has to be weighted with the coalition's importance: an agent will prefer, and therefore join, a powerful team over a weak one, as described in the next section.

Definition 8 (Scenario) *Let $EAF =< A, \rightarrow, \twoheadrightarrow >$ be an extended argumentation framework, we define a scenario $s_{\twoheadrightarrow'}$ as the hypothetical firing of a subset $\twoheadrightarrow'$ of the voluntary attack set $\twoheadrightarrow$, i.e. $s_{\twoheadrightarrow'} =< A, \rightarrow' >$, where $\rightarrow' = \rightarrow \cup \twoheadrightarrow'$.*

The analysis of the scenario is purely potential, because agents have to evaluate the possible dialogue game's outcomes [8] before the dialogue actually happens. So each agent ag_i has three tools for evaluating a scenario $s_{\twoheadrightarrow'}$ such as her utility functions.

Definition 9 (Argument utility function) *Given an agent ag_i and a scenario $s_{\twoheadrightarrow'}$, $ut_i^{arg}(s_{\twoheadrightarrow'}) = \sum_j \alpha_i^j$ s.t. $f_{bel}(j) = ag_i$, $L_{compl}^{\langle A,\rightarrow\rangle}(j) = in$*

Definition 10 (Voluntary attack utility function) *Given an agent ag_i and a scenario $s_{\twoheadrightarrow'}$, $ut_i^{cost}(s_{\twoheadrightarrow'}) = \sum_{(j,k)} \beta_i^{j,k}$ s.t. $f_{bel}(j) = ag_i, (j, k) \in \twoheadrightarrow'$*

Basically, for each scenario $s_{\twoheadrightarrow'}$, $ut_i^{arg}(s_{\twoheadrightarrow'})$ sums up the relevance α_i^j of ag_i's accepted arguments and $ut_i^{cost}(s_{\twoheadrightarrow'})$ sums the costs $\beta_i^{j,k}$ of the voluntary attacks raised by ag_i.

3.2 Coalitions

A coalition is a team of agents who decides to join together in order to mutually protect their interests. Coalitions are not seen as democratic: stronger agents have more influence than weaker ones. But how is the concept of power defined? It is based on the number of accepted arguments each agent owns. In the political scenario, an argument is a proposal of a topic and the acceptance of such point is the defeasible insertion into the electoral manifesto. Discarded topics will not appear in the manifesto, giving no power to the agent who proposes them. For instance, an agent with n arguments, none of them accepted, is weaker than an agent with a single, accepted argument. We define the agent's power as the number of accepted arguments she owns, and the coalition's power as the summation of the agents' powers.

Definition 11 (Power) *Let $pow(ag_i) =\mid \{k : f_{bel}(k) = ag_i, L_{compl}^{\langle A,\rightarrow\rangle}(k) = in\} \mid$. Given $coal_j = \{ag_1, ..., ag_n\}$, $pow(coal_j) = \sum_{ag_w \in coal_j} pow(ag_w)$.*

Note that this is why the γ parameter was introduced as *relative*: on the one hand, the choice for an agent ag_i to join a coalition $coal_j$ depends on ag_i's sociality γ_i, the highest the parameter's value the better chances for the agent to appreciate the idea of joining $coal_j$; on the other hand, the same agent will more likely join a more powerful coalition than a least accredited one. This is captured by the third utility function:

Definition 12 (Coalition utility function) *Let ag_i be an agent and $coal_j$ a coalition. Then $ut_i^{coal}(coal_j) = \gamma_i \times pow(coal_j)$, where $pow(coal_j) = \sum_{ag_w \in coal_j} pow(ag_w)$ and $pow(ag_w) =\mid \{k : f_{bel}(k) = ag_w, L_{compl}^{\langle A,\rightarrow\rangle}(k) = in\} \mid$.*

Furthermore, each agent's relevance in a coalition is defined as her leverage.

Definition 13 (Leverage) *Given an agent* ag_i *and a coalition* $coal_j$ *s.t.* $ag_i \in coal_j$*, let* $lev_i^j = pow(ag_i)/pow(coal_j)$*, where* $pow(ag_i) = \mid \{k : f_{bel}(k) = ag_i, L_{compl}^{\langle A, \rightarrow \rangle}(k) = in\} \mid$ *and* $pow(coal_j) = \sum_{ag_w \in coal_j} pow(ag_w)$.

When a decision has to be made within a coalition, each agent provides her evaluation of the possible outcomes, and this value is weighted by the agent's leverage in the coalition-wide decision.

4 Algorithm and example

In this subsection we describe our algorithm for negotiation and coalition formation. This algorithm takes in input an extended argumentation framework with voluntary attacks and returns a standard, Dung-style one, since every voluntary attack is either raised as an actual one or removed from the framework. The algorithm output can subsequently be subject to modifications and become the next round's input: this is because agents may drop or introduce new arguments or attacks (voluntary or actual). Every round exploits the previous round's coalition structure.

4.1 Phases

Each round consists of three phases which can be iterated until a global agreement is reached. For first, each agent evaluates the possible scenarios and computes their outcome; after that, a coalition phase is faced, when agents decide which proposals shall be made; the third phase is a negotiation dialogue among coalitions about which voluntary attacks shall be fired. The coalition structure is modified throughout the round, according to agreements/disagreements among agents and coalitions.

In the agent phase, each agents evaluates the different scenarios that emerge from the raising of subsets of voluntary attacks. Outcomes are computed considering which agent's arguments would be accepted and which voluntary attacks the agent should fire. Each agent may come out with a different evaluation of the same scenarios. The utility functions' outputs differ from agent to agent, and can favor, for instance, the maximization of accepted arguments over the burden of raising attacks.

In the coalition phase, in each coalition the agents put out their outcomes. Since there is complete information, a normal form game representation can be used and game theory solution concepts (such as Nash Equilibrium) can be used to elect a best choice, which will be the coalition's first proposal in the following dialogue. If the selected scenario has a negative outcome for some agent within the coalition, she has to compare the negativity of the proposal with the value of her membership to the coalition, i.e. coalition utility function versus arguments and attacks utility functions. If sociality prevails, the agent accepts a disadvantageous coalition proposal for the sake of belonging to that coalition; otherwise, she leaves the coalition.

In the framework phase, each coalition has a list of negotiation proposals to put forward. Due to opacity, there's no inter-coalition knowledge about scenarios' outcomes: coalition $coal_i$ does not know anything about what coalition $coal_j$ will propose. Therefore, game theory tools cannot be helpful, and a dialogue phase has to be faced. Coalitions take turns in the dialogue according to their power, strongest coalitions speak first and are therefore advantaged. Proposals are put forward, accepted and refused. Sets of coalitions that reach an agreement can join together in a bigger coalition.

Whenever an agreement is made and voluntary attacks are raised or there is a change in the coalition structure, agents and coalitions need to refine and adjust their proposals, so the algorithm shifts to the coalition phase. When the $\twoheadrightarrow$ set is empty every decision has been made, the resulting framework is in standard Dung-like form, and the round terminates. Subsequent changes to the framework, such as introducing new arguments and voluntary attacks, would trigger another round, whose initial coalition structure would be the previous round's resulting one.

4.2 Moves and dialogue

Since the negotiation phase models an open debate, all messages are broadcasted. We assume that every message is delivered without flaws or relevant delays.

Definition 14 (Move) *A move is a tuple* $< id, id_{ans}, Sa, Rs, \twoheadrightarrow^+, \twoheadrightarrow^-, info >$, *where* $id, id_{ans} \in \mathbb{N}$, $Sa \in \{propose, accept, refuse\}$, $Rs \subseteq CS$, $\twoheadrightarrow^+, \twoheadrightarrow^- \subseteq \twoheadrightarrow$ *s.t.* $\twoheadrightarrow^+ \cap \twoheadrightarrow^- = \emptyset$

In Definition 14 we have that id is the move identifier, id_{ans} is the target of the move; Sa is the speech act, Rs is a subset of the set of coalitions CS to which the move is addressed to; $\twoheadrightarrow^+$ and $\twoheadrightarrow^-$ are two disjunct subsets of $\twoheadrightarrow$ and match the proposed scenario: the attacks in $\twoheadrightarrow^+$ shall be fired, those in $\twoheadrightarrow^-$ shall be removed from the framework; $info$ is an extra field the coalition can use to send more informations. Given a move $m = < id, id_{ans}, Sa, Rs, \twoheadrightarrow^+, \twoheadrightarrow^-, info >$ the function Id (resp. $Target$, $SpeechAct$, $Recipient$, $Voluntary^+$, $Voluntary^-$, $Info$) returns the identifier id of the move (resp. the target move's id, i.e. id_{ans}, the speech act Sa, the recipient set Rs, the sets of voluntary attacks $\twoheadrightarrow^+$ and $\twoheadrightarrow^-$, the $info$ field). Let $Id(m_i) = i$, $SpeechAct(m_0) = propose$, $Target(m_0) = \emptyset$.

The dialogue phase goes through two stages: closed dialogue and open dialogue. In the closed dialogue, after the first move, only selected coalitions are allowed to talk. In the open dialogue stage, making moves is permitted to every coalition. Allowed coalitions take turns in making moves according to their power. The first move has to be a proposal. A generic legal move has a non empty scenario and targets a previous move, proposals are directed to the voluntary attacks' owners, refused proposals can not be repeated.

Definition 15 (Dialogue) *A Dialogue* D *is a non-empty finite sequence of legal moves* $m_0, m_1, ..., m_n$.

Definition 16 (Legal Moves) *A move* m_i *is legal*

- *if* $\forall j$ $s.t.$ $0 \leqslant j < i \exists m_j \in D$, $Target(m_i) < Id(m_i)$, $Voluntary^+(m_i) \cup Voluntary^-(m_i) \neq \emptyset$;
- *if* $SpeechAct(m_i) \in \{propose\}$, *then* $\forall (ar_i, ar_j) \in Voluntary^+(m_i) \cap Voluntary^-(m_i)$ $s.t.$ $f_{bel}(ar_i) = ag_k, ag_k \in Rs$;
- *if* $Target(m_i) = m_j \wedge SpeechAct(m_i) = refuse$, $\nexists m_k$ $s.t.$ $Voluntary^+(m_j) = Voluntary^+(m_k) \wedge Voluntary^-(m_j) = Voluntary^-(m_k)$.

The first negotiation stage succeeds if a set of coalitions agrees upon a proposal. If it happens, the second stage begins; otherwise the dialogue fails. The second stage succeeds if an agreement is reached and fails if all coalitions run out of moves before an agreement is reached. If the second stage succeeds the dialogue result is the second stage's agreement; if it fails the dialogue agreement is the one reached in the first stage. Hence if and only if the first stages succeeds the dialogue succeeds and the argumentation framework is modified according to this agreement. The negotiation ends when there are no more voluntary attacks in the framework.

4.3 Example

Consider the extended argumentation framework in Figure 2. The extended framework is $EAF = \langle \{a, b, c, d, e, f, g, h, i, j, k\}, \{(b, h), (c, b), (c, i)\}, \{(d, c), (j, k)\} \rangle$. The agent set is $\{1, 2, 3, 4, 5, 6\}$, the coalition structure is given from the previous negotiation phase and coalitions are $\{\{1\}\{2\}\{3, 6\}\{4, 5\}\}$. We do not introduce in this example the utility functions due to space constraints.

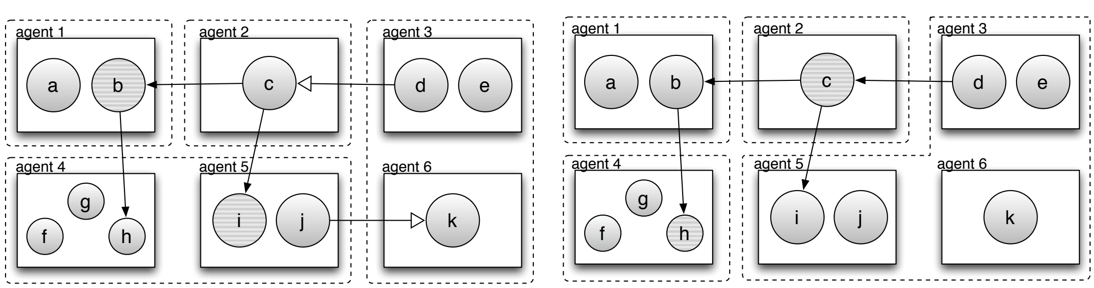

Figure 2: Argumentation framework before and after the algorithm

For the sake of our example, consider agents as different political parties' leaders, and let the arguments be interpreted as follow: *[a]: Keep water services public; [b]: Alternative energy production; [c]: Cut research funding; [d]: More taxes on summer homes; [e]: Quality product seal institution; [f]: Make* *insurance mandatory for each kind of contract; [g]: Capital to host 2015 Auto Show; [h]: Allow 5% deforestation of National Park; [i]: Join European Research Team; [j]: Sign the Atlantic Protocol; [k]: Build an industrial area in the Central Valley.* (b, h) means that by providing energy from alternative sources there will be no need to deforest new areas. (d, c) models that if research funding are cut then the country will not be able to invest on alternate energy production. (c, j) means that cutting down research funding will make the country fall below the community standards needed to join the European Research Team. The voluntary attack (d, c) models the potential attack relation between raising taxes and cutting research funding: money incoming from the newly introduced taxation may be devolved to research grants. The second voluntary attack, (j, k), captures another potential attack relation: argument j proposes to commit to the Atlantic Protocol. There is an optional guideline in the protocol that discourages from creating high-polluting areas, and therefore, if followed, would make the Central Valley industrial area a not accepted argument. Note that in both cases all agents know about the voluntary attacks but the choice is always up to the agent who controls the attack: in the first case it is about how to invest some money, in the second one whether to follow an optional guideline or not.

Now consider a political setting, where the spokesman of six allied political parties is called to prepare the electoral program for the incoming elections. The programs of the single parties are known. First of all, each coalition holds a preliminary reunion in order to prepare a list of proposals to be inserted in the manifesto.

Coalition $\{1\}$ Agent 1 belongs to a singleton coalition, so she can skip the preliminary phase. She knows that agent 2's proposal of cutting research funding leaves no space for alternative energy production to be part of the program, and hopes that $\{3,6\}$ objects to this cut by the taxation proposal, raising the (d, c) voluntary attack.

Coalition $\{2\}$ Agent 2 wants $\{3,6\}$ not to attack his research funding cut argument.

Coalition $\{4,5\}$ Agent 5 wants to propose coalition $\{3,6\}$ to fire its attack against agent 2's argument c, because c attacks her argument i. Agent 4 disagrees, since cutting the research funds defends her argument about deforestation. They find no compromise, so they split up.

Coalition $\{3,6\}$ Agent 3 knows that the money gained from tax raising could support research, but since such argument is not part of her program she decides she will not propose the attack. Agent 6 wants to build an industrial area, and she knows that there is an optional guideline in the Atlantic Protocol which discourages from creating high-polluting areas. So they decide to propose $\{4,5\}$ not to raise the (j, k) attack (i.e. they ask them not to include the optional guideline). None of them would be advantaged from agent 3 firing her (d, c) attack, but they are open to do it as a negotiation's do ut des.

After this phase, the coalitions face a negotiation dialogue. As the most powerful coalition, $\{3,6\}$ gets to talk first. They ask coalition $\{5\}$ not to include the pollution guideline in the Atlantic Protocol ratification. Agent 5 asks them, in exchange, to commit themselves to devolve the money from extra taxes to sponsor research. $\{3,6\}$ accepts and the first stage terminates. Now all the other coalitions can make different proposals. Three singleton coalitions are now allowed to make moves: $\{1\}$,$\{2\}$,$\{4\}$. They do it with respect to their power, so coalition $\{4\}$ is the first one. $\{4\}$ asks $\{3,6\}$ not to devolve money to research funding, but $\{3,6\}$ refuses the proposals because it does not concern the pollution question he cares about. So $\{4\}$ has only one move left: she proposes to both $\{3,6\}$ and $\{5\}$ not to raise their attacks; this is the only possible scenario she can propose, since she does not want $\{3,6\}$ to raise her attack and $\{3,6\}$ does not want $\{5\}$ to fire hers. $\{3,6\}$ accepts, since this is her first choice, but $\{5\}$ refuses because in this case the research funding cut would be approved and joining the European Research Team would be dropped from the manifesto. Therefore, $\{4\}$ has no moves left. It is $\{1\}$ turn now, but she can not put forward new proposals, so she passes, same does $\{2\}$. So coalitions $\{5\}$ and $\{3,6\}$ reach an agreement and join themselves to form a bigger coalition. Voluntary attack (d,c) is raised as (j,k) is removed.

Formally, the dialogue is represented by the following tuples:

- $\{3,6\} : < 0, -, propose, \{\{5\}\}, \{\}, \{(j,k)\}, - >$
- $\{5\} : < 1, 0, propose, \{\{3,6\}\}, \{(d,c)\}, \{(j,k)\}, - >$
- $\{3,6\} : < 2, 1, accept, \{\{5\}\}, \{(d,c)\}, \{(j,k)\}, - >$
- $\{4\} : < 3, 0, propose, \{\{3,6\}\}, \{\}, \{(d,c)\}, - >$

- $\{3,6\} : <\ 4, 3, refuse, \{\{4\}\}, \{\}, \{(d, c)\}, consider(j, k) >$
- $\{4\} : <\ 5, 4, propose, \{\{3, 6\}, \{5\}\}, \{\}, \{(d, c), (j, k)\}, - >$
- $\{3,6\} : <\ 6, 5, accept, \{\{4\}\}, \{\}, \{(d, c), (j, k)\}, - >$
- $\{5\} : <\ 7, 5, refuse, \{\{4\}\}, \{\}, \{(d, c), (j, k)\}, - >$

The resulting framework and coalition structure are represented on the right side of Figure 2.

5 Conclusions

In this paper we present a new kind of argumentation framework in which not only actual attack relations are considered but also potential ones called voluntary attacks. We associate to the agents, composing the multiagent system, a set of arguments, attacks and voluntary attacks. The aim of the agents is to form coalitions for achieving in this way a more important position in the multiagent system. The application of argumentation to coalition formation has been discussed by Amgoud [1], Bulling et al. [5] and Boella et al. [4]. The difference with these approaches is that we do not propose argumentation for doing coalition formation but we propose a new argumentation framework to represent agents' beliefs and we discuss how they can negotiate in order to form coalitions. Bench-Capon et al. [3] propose a framework for the representation and evaluation of arguments in practical reasoning. They assume that rational disagreement is possible, the acceptability of an argument depends in part on the audience to which it is addressed. This idea is a source of inspiration also in our approach where depending on the audience, and on the negotiation this audience carries on, a voluntary attack is fired or not. For more details about the application of argumentation in game theory, see Rahawan and Larson [9] who show the importance of game theory as a tool for analyzing strategic argumentation. Future work addresses the implementation and simulation of the proposed algorithm in order to study its computational properties.

References

[1] Leila Amgoud. An argumentation-based model for reasoning about coalition structures. In Simon Parsons, Nicolas Maudet, Pavlos Moraitis, and Iyad Rahwan, editors, *ArgMAS*, volume 4049 of *Lecture Notes in Computer Science*, pages 217–228. Springer, 2005.

[2] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artif. Intell.*, 171(10-15):675–700, 2007.

[3] Trevor J. M. Bench-Capon, Sylvie Doutre, and Paul E. Dunne. Audiences in argumentation frameworks. *Artif. Intell.*, 171(1):42–71, 2007.

[4] Guido Boella, Leendert van der Torre, and Serena Villata. Social viewpoints for arguing about coalitions. In The Duy Bui, Tuong Vinh Ho, and Quang-Thuy Ha, editors, *PRIMA*, volume 5357 of *Lecture Notes in Computer Science*, pages 66–77. Springer, 2008.

[5] Nils Bulling, Jürgen Dix, and Carlos Iván Chesñevar. Modelling coalitions: Atl + argumentation. In Lin Padgham, David C. Parkes, Jörg Müller, and Simon Parsons, editors, *AAMAS (2)*, pages 681–688. IFAAMAS, 2008.

[6] Martin Caminada. On the issue of reinstatement in argumentation. In *Logics in Artificial Intelligence, 10th European Conference,JELIA 2006*, pages 111–123, 2006.

[7] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[8] Henry Prakken. Coherence and flexibility in dialogue games for argumentation. *J. Log. Comput.*, 15(6):1009–1040, 2005.

[9] I. Rahwan and K. Larson. Argumentation and game theory. *Argumentation in Artificial Intelligence*, pages 321–339, 2009.