UNIVERSITÉ PARIS DESCARTES
MEMBRE DE USPC Université Sorbonne Paris Cité

Control Argumentation Frameworks

Yannis Dimopoulos, Jean-Guy Mailly, Pavlos Moraitis

LIPADE-TR-N° 4
April 2017

Technical Report
LIPADE
Laboratoire d'Informatique PAris DEscartes

Control Argumentation Frameworks

Yannis Dimopoulos[1], Jean-Guy Mailly[2], Pavlos Moraitis[2]

[1] *Department of Computer Science, University of Cyprus*
[2] *LIPADE, Paris Descartes University, France*

Abstract

Dynamics of argumentation is the family of techniques concerned with the evolution of an argumentation framework (AF), for instance to guarantee that a given set of arguments is accepted. This work proposes *Control Argumentation Frameworks* (*CAF*s), a new approach that generalizes existing techniques by accommodating the possibility of uncertainty in dynamic scenarios. A CAF is able to deal with situations where the exact set of arguments is unknown and subject to evolution, and the existence (or direction) of some attacks is also unknown. It can be used by an agent to ensure that a set of arguments is part of one (or every) extension whatever the actual set of arguments and attacks. A QBF encoding of reasoning with CAFs provides a computational mechanism for determining whether and how this goal can be reached.

Keywords: *Argumentation Dynamics, Uncertainty*

1. Introduction

Argumentation is an important domain in the field of Artificial Intelligence. Accumulated over more than two decades, there is nowadays a vast literature on various aspects of argumentation, such as abstract argumentation frameworks and their semantics (see e.g. [1, 2]), structured argumentation frameworks (see e.g. [3, 4, 5]) and more recently on a particular topic called argumentation dynamics (see e.g. [6, 7, 8, 9, 10]). In this paper we propose a new family of abstract argumentation frameworks, called *control argumentation frameworks,* abbreviated as *CAFs.* A CAF integrates in a unified

Email addresses: yannis@cs.ucy.ac.cy (Yannis Dimopoulos), jean-guy.mailly@parisdescartes.fr (Jean-Guy Mailly), pavlos@mi.parisdescartes.fr (Pavlos Moraitis)

computational framework different notions proposed in the literature on argumentation dynamics, while simultaneously relaxing the basic assumption of complete knowledge, implicit in the majority of past works. The computational methods that are presented in this work are based on Quantified Boolean Formulas (QBFs) [11] solving technology. The aim is to build efficient argumentation systems that can reach certain states (e.g. a set of arguments to be skeptically or credulously accepted that may support goals, decisions, actions, beliefs, etc.) regardless of the different unpredicted threats that they may face when operating in dynamic environments. These threats can be modeled through the different possible changes already studied in the literature that might affect argumentation systems namely addition/removal of arguments, and addition/removal of attacks.

As noted above, in the majority of the previous works, and especially in those proposing computational methods (see e.g. [12, 13]), complete knowledge about the structure of the argumentation theories is assumed. That is, all the arguments of a theory as well as the existence and the direction of the attacks between those arguments are assumed to be known. In reality however, agents need to reason by taking into account aspects of the world that are completely outside their control and may evolve constantly. For instance a decision making/aiding investment banking agent that builds an argument that supports investing in savings accounts which is meaningful when interest rates are high and less so when rates plunge. The problem for an investment agent that aims at generating secure portfolios is that interest rates is a highly uncertain uncontrollable variable. More generally, arguments supporting particular investments decisions depend on uncertain factors such as market fluctuations, expectations, political developments, etc. It is desirable that agents are able to reach conclusions under incomplete information, or even reach long-lasting conclusion, i.e. that remain valid regardless of how the world evolves. This work provides a computational framework that supports reasoning with uncertainty regarding the presence of arguments and the attacks between them. The problem of devising languages that are expressive enough to accommodate such uncertainty has been addressed in some works (see e.g. [14]), without however providing associated computational methods. Indeed, to the best of our knowledge, this is the first work that proposes an argumentation framework handling all possible dynamics under uncertainty, along with efficient computational methods that take advantage of recent progress in methods generalizing the satisfiability problem, namely QBFs.

After a brief description of background knowledge in Section 2, the CAF is presented in Section 3, along with a QBF-based computational method that determines whether a CAF is controllable and how to control it. Some complexity results are given in Section 4, whereas Section 5 concludes the paper.

2. Background

2.1. Argumentation Systems

An *argumentation framework* (AF), as introduced by Dung in [1], is a pair $\langle\mathcal{A},\mathcal{R}\rangle$, where $\mathcal{A}$ is a set of *arguments*, and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an *attack relation*. The relation *a attacks b* is denoted by $a\ \mathcal{R}\ b$ or $(a, b) \in \mathcal{R}$.

In [1], different acceptability semantics were introduced. They are based on two basic concepts: *defence* and *conflict-freeness*. Here we focus on *stable* semantics. Based on the acceptability semantics, we can define the status of any argument, namely *skeptically accepted, credulously accepted* and *rejected arguments*. For space reasons we consider that the reader is familiar with all the above notions.

2.2. Quantified Boolean Formulas

We assume that the reader is familiar with the basics of propositional logic, satisfiability and complexity theory. Quantified boolean formulas (QBFs) are a natural extension of propositional formulas with the universal and existential quantifiers [15]. For instance, the formula $\exists x \forall y (x \vee \neg y) \wedge (\neg x \vee y)$ is satisfied if there is a value for x such that for all values of y the proposition $(x \vee \neg y) \wedge (\neg x \vee y)$ is true. More formally, a "canonical" QBF is of the form $\mathcal{Q}_1 X_1 \mathcal{Q}_2 X_2 \dots \mathcal{Q}_n X_n \Phi$ where Φ is a propositional formula, $\mathcal{Q}_i \in \{\exists, \forall\}$, $\mathcal{Q}_i \neq \mathcal{Q}_{i+1}$, and $X_1, X_2, \dots, X_n$ disjoint sets of propositional variables such that $X_1 \cup X_2 \cup \dots \cup X_n$ coincides with the set of propositional variables of Φ. It is well-known that QBFs span the polynomial hierarchy. E.g., deciding whether the formula $\exists X_1 \forall X_2 \dots \mathcal{Q}_i X_i \Phi$ is true is Σ_i^p-complete, where $\mathcal{Q}_i = \exists$ for odd i, and $\mathcal{Q}_i = \forall$ for even i. The results still hold for Φ in 3CNF. We denote the formula $\exists X_1 \forall X_2 \dots \mathcal{Q}_i X_i \Phi$ by $Q_{i,\exists}$. Finally, a truth assignment on a set of propositional variables $\{x_1, \dots, x_n\}$ is a mapping $M : \{x_1, \dots, x_n\} \rightarrow \{True, False\}$.

3. Control Argumentation Frameworks

On a high level, a CAF is an argumentation framework where arguments are divided in three parts, *factual*, *uncertain* and *control*. The *factual* part of the global theory contains arguments that are based on some (expert) knowledge modeling the regular (w.r.t. some requirements in a specific application) behavior of the system when acting for satisfying an assigned goal (i.e. controlling the ambient temperature of a smart house). The *uncertain* part contains arguments modeling possible changes in the world (w.r.t. the domain of the expertise and the action of the system) that might constitute threats for the regular behavior of the system. These threats are modeled through attacks against arguments in the factual part. Uncertainty is doubly present in this part of the system. Firstly, it captures the lack of information on whether some possible change has really occurred or not. This type of uncertainty is modeled through the presence or absence of the argument representing the change in the theory of this part (i.e. an argument can be "on/off" according to the situation). Secondly, it concerns the lack of information on the type of the threat generated by an occurred change. This type of uncertainty concerns the presence (or not) and the direction of attacks of arguments present in the uncertain part (and representing the occurred changes) against arguments in the factual part. This part simulates all the possible dynamics that may occur (i.e. addition/removal of arguments, addition/removal of attacks) in argumentation systems enhanced with enforcement capabilities. Finally, the *control* part contains arguments that protect the factual part (and therefore the regular behavior of the system w.r.t. attaining its assigned goal) against the threats arising from the uncertain part. More precisely, this part contains arguments that can propose remedial actions against the attacks addressed by arguments in the uncertain part towards arguments in the factual part, but also against arguments in the factual part when the target of the system (i.e. the supporting argument of the goal that has to be skeptically or credulously accepted) is rejected in the current stage of the factual theory.

A notable difference between CAFs and all other "classical" argumentation frameworks (with or without enforcement capabilities) based on a "unique" theory, is that in our framework we don't have to care about the evolution of the factual theory (driving the behavior of the system) as usually happens in classical frameworks (with their unique theory) when the world changes. The reason is that (based on the application requirements) all possible changes/threats (known so far in the domain of the expert and designer of the system) that might occur in an application domain are (or should be) already represented in the uncertain part and the appropriate responses (or remedial actions) in the control part. In classical systems with possibility of enforcement, the theories corresponding to our uncertain and control parts are incrementally merged with the initial theory in a single theory and this merging is on the basis of the system's (unique) theory evolution as the world changes. In CAF systems these theories don't need to merge in a single theory for dealing with the changes in the world. The system has just to recognize which change has occurred (through its uncertain part) at a certain instant, how this change influences the regular behavior (through interaction between factual and uncertain parts) and find the appropriate remedial actions to undertake (through interaction between control, uncertain and factual parts). Thus, the factual theory (corresponding to the initial theory of the classical systems) does not need to evolve (in the sense of classical systems) incrementally for reacting to the world changes when occurred. The system takes into consideration these unpredictable (w.r.t. the factual part) changes through the *interaction* of its three parts. However, the theories of these three parts might evolve during the system's life cycle in a modular way, either automatically through machine learning or through ("off-line") integration of additional knowledge by the designer/expert, if the know-how in the application domain has evolved or the application requirements have changed. This is a main novelty in argumentation frameworks dealing with dynamics. The system ensures its normal behavior when everything goes well, based on the theory in the factual part, and when some unpredicted events/changes arrive (captured through the activation of arguments in the uncertain part), the system is managing to respond for controlling its regular behavior, by using its control part. The designer/expert of a CAF can also take advantage of its modularity w.r.t. maintenance and/or upgrading issues, as he can be focused only on the concerned part of the CAF (i.e. factual, uncertain, control) without caring about the others. More formally a CAF is defined as follows:

Definition 1. *Let $\mathcal{L}$ a language from which we can build arguments and let* $\texttt{Args}(\mathcal{L})$ *be the set which contains all those arguments. A Control Argumentation Framework (CAF)* *is a triple* $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$ *where* $\mathcal{F}$ *is the* factual part, $\mathcal{U}$ *is the* uncertain part *and* $\mathcal{C}$ *is the* control part *of* $\mathcal{CAF}$ *with:*

- $\mathcal{F} = \langle A_F, \rightarrow \rangle$ *where* A_F *is a set of arguments that we know they are present (or active) in the system and* $\rightarrow \subseteq (A_F \cup A_U^+) \times (A_F \cup A_U^+)$ *is an attack relation representing a set of attacks for which we are aware both of their existence and their direction.*

- $\mathcal{U} = \langle A_U, (\rightleftarrows \cup \dashrightarrow) \rangle$ *where* $A_U = A_U^+ \cup A_U^{+/-}$ *is a set of arguments divided in arguments for which we are aware that they are present in the system (*A_U^+*), and arguments for which we are not (*$A_U^{+/-}$*),* $\rightleftarrows \subseteq (((A_U \cup A_F) \times (A_U \cup A_F)) \setminus \rightarrow)$ *is an attack relation representing a set of attacks for which we are aware of their existence but not of their direction and* $\dashrightarrow \subseteq (((A_U \cup A_F) \times (A_U \cup A_F)) \setminus \rightarrow)$ *is an attack relation representing a set of attacks for which we are not aware of their existence but we are aware of their direction, with* $\rightleftarrows \cap \dashrightarrow = \emptyset$*.*

- $\mathcal{C} = \langle A_C, \Rrightarrow \rangle$ *where* A_C *is a set of arguments that the agent can choose to use or not, and* $\Rrightarrow \subseteq (((A_F \cup A_C \cup A_U) \times (A_F \cup A_C \cup A_U)) \setminus (\rightarrow \cup \rightleftarrows \cup \dashrightarrow))$ *is an attack relation.*

$A_F, A_U^+, A_U^{+/-}$ *and* A_C *are disjoint subsets of* $\texttt{Args}(\mathcal{L})$*.*

Before talking about controllability, we need to introduce the notion of completion of a CAF. Intuitively, a completion is a classical AF which is built from the CAF, by choosing one of the possible options for each argument or attack which is concerned by uncertainty.

Definition 2. *Given a CAF* $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$*, a completion of* $\mathcal{CAF}$ *is an AF* $\mathcal{AF} = \langle A, R \rangle$*, s.t.*

- $A = A_F \cup A_C \cup A_U^+ \cup A_{comp}$ *where* $A_{comp} \subseteq A_U^{+/-}$*;*
- *if* $(a, b) \in R$*, then* $(a, b) \in \rightarrow \cup \rightleftarrows \cup \dashrightarrow \cup \Rrightarrow$*;*
- *if* $(a, b) \in \rightarrow$*, then* $(a, b) \in R$*;*
- *if* $(a, b) \in \rightleftarrows$ *and* $a, b \in A$*, then* $(a, b) \in R$ *or* $(b, a) \in R$*;*

- *if* $(a, b) \in \Rrightarrow$ *and* $a, b \in A$*, then* $(a, b) \in R$.

Let us notice that the definition of a completion does not specify anything about the attacks from $\dashrightarrow$, since these attacks may not appear.

Controllability means that we can select a subset $A_{conf} \subseteq A_C$ and the corresponding attacks $\Rrightarrow \cap (A_{conf} \times A_{conf})$ such that whatever the completion of $\mathcal{CAF}$, a given target is always reached. We focus on two kinds of target: credulous acceptance of a set of arguments (this is reminiscent of extension enforcement [10]) and skeptical acceptance of a set of arguments.

Definition 3. *A control configuration of a CAF* $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$ *is a subset* $A_{conf} \subseteq A_C$*. Given a set of arguments* $T \subseteq A_F$ *and a semantics* σ*, we say that* T *is skeptically (resp. credulously) reached by the configuration* A_{conf} *w.r.t.* σ *if* T *is included in every (resp. at least one)* σ*-extension of every completion of* $\mathcal{CAF}' = \langle \mathcal{F}, \mathcal{C}', \mathcal{U} \rangle$*, with* $\mathcal{C}' = \langle A_{conf}, \Rrightarrow \cap ((A_F \cup A_{conf} \cup A_U) \times (A_F \cup A_{conf} \cup A_U)) \rangle$*. We say that* $\mathcal{CAF}$ *is skeptically (resp. credulously)* controllable *w.r.t.* T *and* σ*.*

3.1. Controllability Through Logical Encoding

We propose a method to obtain a control configuration A_{conf} s.t. the set T of arguments is included in all extensions or at least one extension, whatever the evolution of $A_U^{+/-}$ and the actual state of the uncertain attacks. More precisely, our procedure determines if there exists such a configuration, and provides it when it exists. We focus on the stable semantics, extending the encoding from [16].[1]

Let us recall the method to encode the relation between the structure of an AF and its stable extensions. We define first two kinds of propositional variables. Given $\mathcal{AF} = \langle A, R \rangle$,

- $\forall x_i \in A$, acc_{x_i} is a propositional variable representing the acceptance status of the argument x_i;
- $\forall x_i, x_j \in A$, att_{x_i,x_j} is a propositional variable representing the attack from x_i to x_j.

[1] Our method can be adapted with any semantics which can be encoded in propositional logic. Especially, for complete and admissible we can also use encodings from [16].

Φ_{st} is the formula $\Phi_{st} = \bigwedge_{x_i \in A}[acc_{x_i} \Leftrightarrow \bigwedge_{x_j \in A}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j})]$. When the *att*-variables are assigned the truth value corresponding to the attack relation of $\mathcal{AF}$, the models of Φ_{st} (projected on the *acc*-variables) correspond in a bijective way to the stable extensions of $\mathcal{AF}$. Indeed, given $\mathcal{AF} = \langle A, R \rangle$, we define the formula $\Phi_{st}^R = \Phi_{st} \wedge (\bigwedge_{(x_i,x_j) \in R} att_{x_i,x_j}) \wedge (\bigwedge_{(x_i,x_j) \notin R} \neg att_{x_i,x_j})$. Given ω a model of Φ_{st}^R, the set $\{x_i \mid \omega(acc_{x_i}) = \top\}$ is a stable extension of $\mathcal{AF}$. Similarly, for any stable extension ϵ of $\mathcal{AF}$, ω s.t. $\omega(acc_{x_i}) = \top$ iff $x_i \in \epsilon$ is a model of Φ_{st}^R.

Back to the case of CAFs, we see that we cannot directly generalize Φ_{st}^R to obtain an encoding for the stable extensions of the completions of a CAF, since the arguments from A_C which are selected by the agent (*i.e.* the control configuration) are not known in advance. Similarly, the arguments from $A_U^{+/-}$ are not all present in the completion, since they are subject to evolution.

- $\forall x_i \in A_C \cup A_U^{+/-}$, on_{x_i} is a propositional variable which is true when the argument x_i actually appears in the framework.

Thanks to the *on*-variables, we will generalize the formula Φ_{st}^R to consider the fact that an argument x_i has no influence on the extensions when it is not actually in the framework (*i.e.* on_{x_i} is false).

Notation: $\mathbf{A} = A_F \cup A_C \cup A_U$, $\mathbf{R} = \rightarrow \cup \rightleftarrows \cup \dashrightarrow \cup \Rrightarrow$

Now, we can propose an encoding which relates the attack relation and the arguments statuses in $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$:

$$\Phi_{st}(\mathcal{CAF}) = \bigwedge_{x_i \in A_F \cup A_U^+}[acc_{x_i} \Leftrightarrow \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j})] \wedge$$

$$\bigwedge_{x_i \in A_C \cup A_U^{+/-}}[acc_{x_i} \Leftrightarrow (on_{x_i} \wedge \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j}))] \wedge$$

$$\bigwedge_{(x_i,x_j) \in \rightarrow \cup \Rrightarrow} att_{x_i,x_j}$$

$$\bigwedge_{(x_i,x_j) \in \rightleftarrows} att_{x_i,x_j} \vee att_{x_j,x_i}$$

$$\bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}$$

The first line of this definition states in which condition an argument from A_F or A_U^+ is accepted; in this situation it is exactly as in the case of classical AFs: an argument is accepted when all its attackers are rejected. Then, the next line concerns arguments from A_C and $A_U^{+/-}$; since these arguments may not appear in the completion of the CAF, we add the condition that on_{x_i} is true to allow x_i to be accepted. The last lines specify the case in which there is an attack in the completion: attacks from $\rightarrow$ and $\Rrightarrow$ are mandatory, and their direction is known; attacks from $\rightleftarrows$ are mandatory, but the actual direction is not known. We do not give any constraint about $\dashrightarrow$, which is equivalent to the tautological constraint $att_{x_i,x_j} \vee \neg att_{x_i,x_j}$: the attack may appear or not. Finally, we know that attacks which are not in $\mathbf{R}$ do not exist.

Given a set of arguments T, the fact that T must be included in all the stable extensions is represented by:

$$\Phi_{st}^{sk}(\mathcal{CAF}, T) = \Phi_{st}(\mathcal{CAF}) \Rightarrow \bigwedge_{x_i \in T} acc_{x_i}$$

Given a set of arguments T, the fact that T must be included in at least one stable extension is represented by:

$$\Phi_{st}^{cr}(\mathcal{CAF}, T) = \Phi_{st}(\mathcal{CAF}) \wedge \bigwedge_{x_i \in T} acc_{x_i}$$

Now, there is a control configuration s.t. T is skeptically accepted iff the formula

$$\exists\{on_{x_i} \mid x_i \in A_C\}\forall\{on_{x_i} \mid x_i \in A_U^{+/-}\}$$
$$\forall\{att_{x_i,x_j} \mid (x_i, x_j) \in \rightleftarrows \cup \dashrightarrow\}\forall\{acc_{x_i} \mid x_i \in \mathbf{A}\}\Phi_{st}^{sk}(\mathcal{CAF}, T)$$

is valid: this formula specifies that we need to find a valuation of on_{x_i} for $x_i \in A_C$ (*i.e.* a control configuration) such that for every possible valuation of $on_{x_i}, x_i \in A_U^{+/-}$ and every possible valuation of $att_{x_i,x_j}, (x_i, x_j) \in \rightleftarrows \cup \dashrightarrow$ (*i.e.* for every possible completion), each possible assignment of $acc_{x_i}, x_i \in \mathbf{A}$ (*i.e.* each stable extension) implies that the arguments of T are accepted.[2] Therefore, identifying a valuation of the variables $\{on_{x_i} \mid x_i \in A_C\}$ provides a control configuration for the skeptical acceptance of T. For credulous controllability, we should use the following encoding instead:

$$\exists\{on_{x_i} \mid x_i \in A_C\}\forall\{on_{x_i} \mid x_i \in A_U^{+/-}\}$$
$$\forall\{att_{x_i,x_j} \mid (x_i, x_j) \in \rightleftarrows \cup \dashrightarrow\}\exists\{acc_{x_i} \mid x_i \in \mathbf{A}\}\Phi_{st}^{cr}(\mathcal{CAF}, T)$$

[2]We observe that some variables are not quantified. There are implicitely quantified at the first existential level. Since they are not directly concerned by the search of a control configuration, they are removed from the encoding for a matter of simplification.

Let us notice that this time, the acc_{x_i} variables are existentially quantified: T must be implied by at least one stable extension, but not necessarily all of them.

More precisely, to determine whether a CAF is controllable w.r.t. a set of arguments T and the stable semantics, we need to check the validity of one of the previous QBF encodings (depending whether we are interested in skeptical or credulous controllability). To determine the control configuration which corresponds to the controllability, we need to determine the truth assignment of the on_{x_i} variables, for $x_i \in A_C$. The control configuration is given by $A_{conf} = \{x_i \in A_C \mid on_{x_i} \text{ is assigned to } True\}$. Both these tasks can be performed by any modern QBF solver [11].

Example 1. *In a smart home, an intelligent agent is assigned the task of maintaining the temperature over* $25°$ *and the oxygen concentration at a suitable level. The agent is constantly monitoring the temperature and can turn on the heating or the air conditioning (in position "warm") when the temperature drops below* $25°$*. If that equipment is out of order he calls a serviceman for a repair. We assume that air conditioning is considered more energy efficient than heating. The agent knows nevertheless that his user prefers good temperature over high electricity consumption. The agent also monitors oxygen concentration, and may decide to open the house windows when it drops below the desired level. However, opening the windows will decrease the temperature. This contradicts the other goal of keeping the house warm, and the agent will advise the user to get outside for some fresh air.*

The above scenario can be captured in a CAF defined on the following arguments:

- a_1: Keep the ambient temperature t at $t \geq 25°$
- a_2: House temperature $t < 25°$ when the window is open
- a_3: Turn on heating when temperature drops below $25°$
- a_4: Turn on air conditioner (position "warm") when $t < 25°$
- a_5: Open a window when oxygen level is low
- a_6: Advise the user to get outside for fresh air when oxygen level is low

- a_7: Don't turn on heating as it consumes a lot of electricity
- a_8: Good temperature ($t \geq 25°$) is preferred over high electricity consumption
- a_9: Air conditioning is out of order
- a_{10}: Heating is out of order
- a_{11}: Call a serviceman for equipment repair

Our model of the above scenario is based on the following assumptions. 1) the temperature decrease caused by the open windows (a_2) attacks the goal of maintaining the temperature over 25° (a_1); 2) turning on the heating (a_3) and turning on the air conditioning (a_4) are mutually exclusive actions; 3) turning on the heating (a_3) or the air conditioning (a_4) attacks the temperature decrease caused by open windows (a_2); 4) opening the windows when oxygen level is low (a_5) attacks keeping temperature over 25° (a_1). However the lack of oxygen is an abnormal situation inside a house and thus this attack doesn't belong to the knowledge on which the agent is based for making his decisions in usual situations; 5) getting more oxygen by going outside (a_6) attacks (a_5) (i.e. getting more oxygen by staying inside and opening a window) as it avoids to open the window; 6) the decisions based on arguments a_3 and a_7 are logically inconsistent (thus conflicting); however, it is unknown whether a_3 attacks a_7, a_7 attacks a_3 or they mutually attack each other; 7) a_8 attacks argument a_7 as the user prefers to pay a higher electricity bill in order to have a pleasant ambient temperature; 8) argument a_9 attacks argument a_4; 9) argument a_{10} attacks argument a_3; 10) argument a_{11} attacks arguments a_9 and a_{10}.

Based on the above, the scenario can be modeled in a framework $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$, where:

- $\mathcal{F} = \langle A_F, \rightarrow \rangle$ with $A_F = \{a_1, a_2, a_3, a_4\}$ and $\rightarrow = \{(a_2, a_1), (a_4, a_2), (a_3, a_2), (a_4, a_3), (a_3, a_4)\}$
- $\mathcal{U} = \langle A_U, (\rightleftarrows \cup \dashrightarrow) \rangle$ where $A_U = A_U^{+} \cup A_U^{+/-}$ with $A_U^{+} = \emptyset$, $A_U^{+/-} = \{a_5, a_7, a_9, a_{10}\}$, $\rightleftarrows = \{(a_7, a_3), (a_3, a_7)\}$ and $\dashrightarrow = \{(a_5, a_1), (a_9, a_4), (a_{10}, a_3)\}$,
- $\mathcal{C} = \langle A_C, \Rrightarrow \rangle$ with $A_C = \{a_6, a_8, a_{11}\}$ and $\Rrightarrow = \{(a_6, a_5), (a_8, a_7), (a_{11}, a_9), (a_{11}, a_{10})\}$.

The above $\mathcal{CAF}$ is depicted in Figure 1.

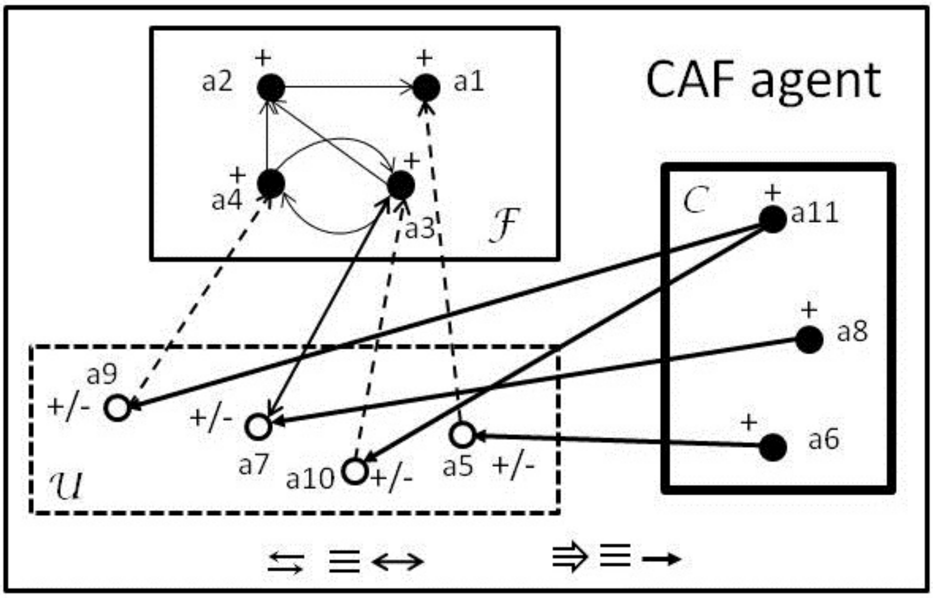

Figure 1: The CAF of the agent

As noted earlier, the target T of the agent is to keep the ambient temperature of the house over 25°. T is satisfied when argument a_1 is skeptically accepted. Indeed, several situations may arise in the course of time, to which the agent can respond accordingly. First, assume for instance, that the user opens a window for some reason (e.g. get fresh air). Then, a_2 is active and attacks a_1. Here, active means that the premisses of an argument are satisfied, and therefore the argument can fire. a_1 is rejected and T fails. Suppose that then the temperature starts decreasing. Arguments a_3 and a_4 are active and attack a_2. a_1 is defended and therefore it is skeptically accepted (i.e. we have two extensions $\{a_{11}, a_8, a_6, a_3, a_1\}$ and $\{a_{11}, a_8, a_6, a_4, a_1\}$). Assume now that an unpredicted event occurs: air conditioning breaks down. In that case argument a_9 is active and attacks a_4. a_4 cannot defend a_1 anymore. Nevertheless, the system is able to react to this event by means of its control part. Indeed, argument a_{11} is active, and the agent will call for an air conditioning repair. Argument a_4 is defended by a_{11}, and defends again a_1. There are again two possible extensions. Suppose now that another event arises before the air conditioning is repaired (assuming that the repair takes a few days). In fact, this time a_1 is only defended by a_3. More specifically, assume that the agent learns about the electricity high consumption of heating from an external information source. In that case argument a_7 is active. As concerns about energy efficiency is not part of the regular behavior of the agent (as defined in theory $\mathcal{F}$), the direction of the attacks between a_7 and a_3 is unknown. Thus, target T might fail if the direction of the attack is from a_7 to a_3. In that case a_1 would be rejected. The above events are unpredicted threats for the system that cannot be normally handled by a classic system where the theory of the agent cannot change (theory $\mathcal{F}$ in our work). The above threats cannot be handled by systems allowing "classic" enforcement either, as they assume full knowledge whereas in the described situations the presence of a_7 (noted as "+/-") and the direction of the attack between a_3 and a_7 are unknown. The same holds for argument a_9 and its attack against argument a_4, and argument a_{11} and its attack against a_9 (this is also the case in our work w.r.t. only $\mathcal{F}$ theory). However, these threats can be addressed by the CAF approach. Indeed, for the threat coming from argument a_7, the system will use argument a_8 (from theory $\mathcal{C}$) for attacking argument a_7 and defending argument a_3. This allows the agent to achieve the target T, as a_1 is defended by a_3 and therefore skeptically accepted again (i.e. there is one extension $\{a_{11}, a_8, a_6, a_3, a_1\}$).

A different situation arises when the agent detects low oxygen concentration. In this case argument a_5 is active and attacks argument a_1. So although a_1 is defended by a_3 and a_4 against the attack of a_2, it is not defended against the attack of a_5. This is also an unpredicted threat that other systems cannot handle (even with classic enforcement), as the presence of argument a_5 (noted as "+/-") and its attack against a_1 are unknown (as also it is the case w.r.t. the $\mathcal{F}$ theory here). This threat can also be handled by CAF through its controller part (i.e. theory $\mathcal{C}$) by argument a_6 which attacks a_5 and defends a_1. Thus, although a_1 is rejected w.r.t. theory $\mathcal{F}$, it is skeptically accepted w.r.t. CAF theory (i.e. we have two extensions $\{a_{11}, a_8, a_6, a_3, a_1\}$ and $\{a_{11}, a_8, a_6, a_4, a_1\}$). Indeed, $T = \{a_1\}$ can be skeptical reached by $A_{conf} = \{a_6, a_8, a_{11}\}$ in $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$. So the target can be achieved regardless of the actual state or evolution of the world.

We illustrate our QBF-based approach with an instantiation of the formula $\Phi_{st}(\mathcal{CAF})$ with the CAF defined previously. Several occurrences of the pattern $att_{x_j,x_i} \Rightarrow \neg acc_{x_j}$ appear in the logical encoding. For a matter of readability, when att_{x_j,x_i} is known to be true, we replace this implication by the fact $\neg acc_{x_j}$. When att_{x_j,x_i} is known to be false, the implication can be removed from the encoding.

$$\Phi_{st}(\mathcal{CAF}) = [acc_{a_1} \Leftrightarrow (\neg acc_{a_2} \wedge (att_{a_5,a_1} \Rightarrow \neg acc_{a_5}))]$$
$$\wedge [acc_{a_2} \Leftrightarrow (\neg acc_{a_3} \wedge \neg acc_{a_4})]$$
$$\wedge [acc_{a_3} \Leftrightarrow (\neg acc_{a_4} \wedge (att_{a_7,a_3} \Rightarrow \neg acc_{a_7}) \wedge (att_{a_{10},a_3} \Rightarrow \neg acc_{a_{10}}))]$$
$$\wedge [acc_{a_4} \Leftrightarrow (\neg acc_{a_3} \wedge (att_{a_9,a_4} \Rightarrow \neg a_9))]$$

$$\wedge [acc_{a_6} \Leftrightarrow on_{a_6}] \wedge [acc_{a_8} \Leftrightarrow on_{a_8}] \wedge [acc_{a_{11}} \Leftrightarrow on_{a_{11}}]$$
$$\wedge [acc_{a_5} \Leftrightarrow (on_{a_5} \wedge \neg acc_{a_6})]$$
$$\wedge [acc_{a_7} \Leftrightarrow (on_{a_7} \wedge \neg acc_{a_8} \wedge (att_{a_3,a_7} \Rightarrow \neg acc_{a_3}))]$$
$$\wedge [acc_{a_9} \Leftrightarrow (on_{a_9} \wedge \neg acc_{a_{11}})]$$
$$\wedge [acc_{a_{10}} \Leftrightarrow (on_{a_{10}} \wedge \neg acc_{a_{11}})]$$

$$\wedge [att_{a_2,a_1} \wedge att_{a_4,a_2} \wedge att_{a_3,a_2} \wedge att_{a_4,a_3}$$
$$\wedge att_{a_3,a_4} \wedge att_{a_6,a_5} \wedge att_{a_8,a_7} \wedge att_{a_{11},a_9} \wedge att_{a_{11},a_{10}}]$$

$$\wedge [att_{a_7,a_3} \vee att_{a_3,a_7}] \wedge \bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}$$

To keep the encoding simple, we do not detail the part concerning absence of attacks, which is summarized in $\bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}$. $\Phi_{st}(\mathcal{CAF})$ is the ground of the encoding. The encoding adapted for the skeptical controllability is $\Phi_{st}^{sk}(\mathcal{CAF}, T) = (\Phi_{st}(\mathcal{CAF}) \Rightarrow \bigwedge_{x_i \in T} acc_{x_i})$. Finally, with the quantifiers, we obtain the following QBF:

$$\exists on_{a_6}, on_{a_8}, on_{a_{11}}, \forall on_{a_5}, on_{a_7}, on_{a_9}, on_{a_{10}}$$
$$\forall att_{a_3,a_7}, att_{a_7,a_3}, att_{a_5,a_1}, att_{a_9,a_4}, att_{a_{10},a_3},$$
$$\forall \{acc_{x_i} \mid x_i \in \mathbf{A}\} \Phi_{st}^{sk}(\mathcal{CAF}, T)$$

Using a QBF solver on this formula gives a valuation of the $on_{a_6}, on_{a_8}, on_{a_{11}}$ variables which corresponds to a control configuration, *i.e.* a subset of A_C such that the target $T = \{a_1\}$ is skeptically accepted.

When CAF is not controllable, *i.e.* there is no subset of C that renders a target argument acceptable for all completions, we may want to seek for a subset of C that achieves that for most of the completions. The recent techniques of [17] on *QBF with soft variables* can be of use here.

Roughly speaking, the idea of soft variables in QBFs is as follows. In a standard QBF $\mathcal{Q}_1 X_1 \mathcal{Q}_2 X_2 \dots \mathcal{Q}_n X_n \Phi$ the quantification level of a variable $x \in X_i$ is i. In other words the quantification level of each variable is fixed. In contrast, the extension presented in [17] allows for soft variables, *i.e.* variables that are not assigned a specific level. Instead, each soft variable is associated with a fixed *set* of allowed quantification levels. Soft variables are prefixed with the symbol $Q_L^{\forall,\exists}$, where L is the set of allowed quantification levels. For instance the formula $F = Q_{\{1,2,3\}}^{\forall,\exists} x \forall y \exists z \Phi$, allows any of the levels 1,2, and 3, giving rise to three possible formulas $F_1 = \exists x \forall y \exists z \Phi$, $F_2 = \forall xy \exists z \Phi$ and $F_3 = \forall y \exists xz \Phi$.

For each allowed quantification level, there is an associated score defined by the user. The optimization problem that arises in this context is to find a level for each soft variable so that the associated propositional formula Φ is satisfied, and the sum of the scores of the soft variables is maximized. If $s(x,l)$ denotes the score assigned to level l for variable x, assume that for the formula F of the previous example $s(x,1) = 3$, $s(x,2) = 2$, $s(x,3) = 1$. Then, the solution to F is a truth assignment that satisfies F_1. If F_1 is unsatisfiable, the solution to F is any satisfying assignment of F_2. Similarly F_3 is the solution if F_2 is unsatisfiable. Intuitively, quantification levels that are more likely to lead to unsatisfiability should be assigned a higher score.

4. Computational Complexity

Now we are interested in the computational complexity of determining whether a CAF is controllable, *i.e.* whether a goal T is skeptically (resp. credulously) reached, for a given configuration. Our encodings lead to obvious upper bound of the complexity: determining whether a goal T is skeptically (resp. credulously) reached belongs to Σ_2^P (resp. Σ_3^P). We provide a lower bound for credulous acceptance.

Definition 4. *The* credulous (resp. skeptical) conclusion problem *is the problem of deciding for a given semantics σ, a Control Argumentation Framework $\mathcal{CAF} = \langle \mathcal{F}, \mathcal{C}, \mathcal{U} \rangle$, and an arguments $q \in A_F$, whether there exists a control configuration A_{conf} such that q is credulously (resp. skeptically) reached by the configuration A_{conf} under semantics σ.*

Proposition 1. *The credulous conclusion problem of a CAF under the stable semantics is Σ_2^p-hard.*

Our framework generalizes existing work. This specific instance of CAFs leads to a lower complexity.

Definition 5. *A Simplified Control Argumentation Framework (SCAF) is a CAF* $\langle F, C, U\rangle$ *such that* $A_U = \emptyset$, $\rightleftarrows = \emptyset$, $\dashrightarrow = \emptyset$, *and is denoted as* $\langle F, C, \emptyset\rangle$.

Note that SCAFs correspond to non-strict normal extension enforcement, since a SCAF $\langle F, C, \emptyset\rangle$ is credulously controllable w.r.t. a set T iff T can be non-strictly enforced in F with a normal expansion [10] by some arguments and attacks from C. So, as a direct consequence of [13], the credulous conclusion problem for a SCAF under the stable semantics is NP-complete.

5. Related Work and Conclusion

In this paper we presented a novel abstract argumentation framework called CAF, integrating in an unified and modular computational framework, all the possible argumentation dynamics considered in the literature, under uncertainty assumption. Among the numerous works on argumentation dynamics, two of them seem more relevant to our approach. The first one is extension enforcement [10]. As said before, there is a correspondence between the credulous conclusion problem for SCAFs and some specific extension enforcement operators. However, even our simplified framework (i.e. SCAFs) is more general than extension enforcement (since it also permits to work with the skeptical conclusion problem). Moreover, since extension enforcement does not consider uncertainty, it cannot be used to tackle situations like our example of smart home. On the opposite, the second one, namely YALLA language [14] seems to be expressive enough to cover any kind of reasoning in abstract argumentation, including reasoning with uncertain attacks or arguments. However, YALLA pays the price of its generality, and we are not aware of any efficient algorithmic approach to handle YALLA-based reasoning. As future works, we plan to extend our complexity study with completeness results, results for the skeptical conclusion problem, and other semantics. In the current state, we search for a solution when a CAF is controllable, *i.e.* there is a configuration which guarantees that the target is reached for any possible completion. However, there are situations where a CAF is not controllable, which leads to an unsatisfiable logical encoding. We will study the related optimization problem, which consists in finding a configuration such that the target is reached in as many completions as possible. As mentioned previously, QBFs with soft variables [17] can be used for that. We are also interested in the development of the structured version of CAFs. Indeed, an agent needs to know the internal structure of an argument to determine whether it is activated or not ($+$ or $-$), depending on which arguments' premises can be deduced from the agent's knowledge base. We do believe, that the computational efficiency of our CAF, while generalizing the possible dynamics through consideration of uncertainty, allowing to handle unpredicted threats in dynamic environments, may be very well suited for building real world applications. Especially, we are interested in implementing self-adaptive systems ensuring real time control tasks in different contexts such as smart homes, surveillance of buildings and streets, personalized self-regulation services for humans, recommendation policies in finance and risk management, etc.

References

[1] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Art. Intel. 77 (1995) 321–357.

[2] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, Knowledge Eng. Review 26 (4) (2011) 365–410.

[3] P. Besnard, A. Hunter, Elements of Argumentation, MIT Press, 2008.
URL `http://mitpress.mit.edu/books/elements-argumentation`

[4] P. M. Dung, R. A. Kowalski, F. Toni, Assumption-based argumentation, in: Argumentation in Artificial Intelligence, 2009, pp. 199–218.

[5] A. C. Kakas, P. Moraitis, Argumentation based decision making for autonomous agents, in: AAMAS'03, 2003, pp. 883–890.

[6] C. Cayrol, F. D. de Saint-Cyr, M. Lagasquie-Schiex, Change in abstract argumentation frameworks: Adding an argument, J. Artif. Intell. Res. (JAIR) 38 (2010) 49–84.

[7] R. Booth, S. Kaci, T. Rienstra, L. van der Torre, A logical theory about dynamics in abstract argumentation, in: SUM'13, 2013, pp. 148–161.

[8] S. Doutre, A. Herzig, L. Perrussel, A dynamic logic framework for abstract argumentation, in: KR'14, 2014.

[9] S. Coste-Marquis, S. Konieczny, J. Mailly, P. Marquis, On the revision of argumentation systems: Minimal change of arguments statuses, in: KR'14, 2014.

[10] R. Baumann, G. Brewka, Expanding argumentation frameworks: Enforcing and monotonicity results, in: COMMA'10, 2010, pp. 75–86.

[11] L. Pulina, The ninth QBF solvers evaluation - preliminary report, in: QBF'16, 2016, pp. 1–13.

[12] S. Coste-Marquis, S. Konieczny, J. Mailly, P. Marquis, Extension enforcement in abstract argumentation as an optimization problem, in: IJCAI'15, 2015, pp. 2876–2882.

[13] J. P. Wallner, A. Niskanen, M. Järvisalo, Complexity results and algorithms for extension enforcement in abstract argumentation., in: AAAI'16, 2016, pp. 1088–1094.

[14] F. D. de Saint-Cyr, P. Bisquert, C. Cayrol, M. Lagasquie-Schiex, Argumentation update in YALLA (yet another logic language for argumentation), Int. J. Approx. Reasoning 75 (2016) 57–92.

[15] H. Kleine Büning, U. Bubeck, Theory of quantified boolean formulas, in: Handbook of Satisfiability, 2009, pp. 735–760.

[16] P. Besnard, S. Doutre, Checking the acceptability of a set of arguments, in: NMR'04, 2004, pp. 59–64.

[17] S. Reimer, M. Sauer, P. Marin, B. Becker, QBF with soft variables, in: AVOCS'14, 2014.