The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Control Argumentation Frameworks

Yannis Dimopoulos
Department of Computer Science
University of Cyprus
yannis@cs.ucy.ac.cy

Jean-Guy Mailly
LIPADE
Paris Descartes University, France
jean-guy.mailly@parisdescartes.fr

Pavlos Moraitis
LIPADE
Paris Descartes University, France
pavlos@mi.parisdescartes.fr

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, namely normal extension enforcement, 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. We also provide some results concerning soundness and completeness of the proposed encoding as well as complexity issues.

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.* (Dung 1995; Baroni, Caminada, and Giacomin 2011)), structured argumentation frameworks (see *e.g.* (Besnard and Hunter 2008; Dung, Kowalski, and Toni 2009; Kakas and Moraitis 2003)) and more recently on a particular topic called argumentation dynamics (see *e.g.* (Cayrol, de Saint-Cyr, and Lagasquie-Schiex 2010; Booth et al. 2013; Doutre, Herzig, and Perrussel 2014; Coste-Marquis et al. 2014; Baumann and Brewka 2010)). 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 computational framework different notions proposed in the literature on argumentation dynamics, while simultaneously relaxes 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) (Pulina 2016) solving technology. The aim is to build effective 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. So, while most works on argumentation dynamics consist in defining methods which modify an argumentation framework, the aim of CAFs is different. Indeed, the goal of our work is to define a framework which can resists to any future change that could occur, while maintaining some desired properties of the system.

As noted above, in the majority of the previous works, and especially in those proposing computational methods (see *e.g.* (Coste-Marquis et al. 2015; Wallner, Niskanen, and Järvisalo 2016)), complete knowledge about the structure of the argumentation theories is assumed. That is, all the arguments of a theory as well as the existence (or not) 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 but less so when rates plunge. The problem for an investment agent that aims at generating secure portfolios is that interest rates are 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.* (Dupin de Saint-Cyr et al. 2016)), 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.

The second section of the paper presents background knowledge. The third section describes the CAF formalism. Then, we present a QBF-based computational method that determines whether a CAF is controllable and how to control it. Some complexity results are given, and finally the last sections describe relevant related work and interesting research tracks.[1]

Background

Argumentation Systems

An *argumentation framework* (AF), as introduced in (Dung 1995), is a pair $\mathcal{AF} = \langle A, R \rangle$, where A is a set of *arguments*, and $R \subseteq A \times A$ is an *attack relation*. The relation a *attacks* b is denoted by $(a, b) \in R$.

In (Dung 1995), different acceptability semantics were introduced. They are based on two basic concepts: *defence* and *conflict-freeness*. Here we focus on *stable* semantics: a set of arguments $S \subseteq A$ is a stable extension of $\mathcal{AF} = \langle A, R \rangle$ iff 1) S is conflict-free (*i.e.* $\forall a_i, a_j \in S$, $(a_i, a_j) \notin R$) and 2) $\forall a_j \in A \setminus S$, $\exists a_i \in S$ s.t. $(a_i, a_j) \in R$. Let us notice that our approach can be adapted to any extension-based semantics. Based on the acceptability semantics, we can define the status of any argument, namely *skeptically accepted* (belonging to each extension), *credulously accepted* (belonging to some extension) and *rejected arguments* (belonging to no extension). For more details about argumentation semantics, we refer the reader to (Dung 1995; Baroni, Caminada, and Giacomin 2011).

Quantified Boolean Formulas

We assume that the reader is familiar with the basics of propositional logic, satisfiability and complexity theory. Otherwise, see (Biere et al. 2009; Arora and Barak 2009) for an overview. Quantified Boolean Formulas (QBFs) are a natural extension of propositional formulas with the universal and existential quantifiers (Kleine Büning and Bubeck 2009). 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. For instance, 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 $V = \{x_1, \dots, x_n\}$ is a mapping $\omega : V \to \{0, 1\}$.

[1]For space reasons, proofs are omitted.

Control Argumentation Frameworks

This section defines the control argumentation framework. On a high level, a CAF is an argumentation framework where arguments are divided in three parts, *fixed*, *uncertain* and *control*. The fixed part of the global theory describes the background (basic) knowledge of an agent that allows him to solve problems when acting in a static environment. It is static (i.e. in the sense that it does not change over time) and neither the agent itself nor the environment can have an influence on it (i.e. only the user can change it).

The $uncertain$ part is a part that the agent cannot control either. It captures the changes that may occur in the environment where the agent is acting and the context dependant information. This part is controlled by the environment in the sense that changes in the environment are represented in its theory. It can change over time reflecting the dynamics of the environment. It is to some extent the sensory input of the system. When the agent has some goal about the fixed part of the CAF (*e.g.* some arguments supporting an action/decision should be skeptically accepted), then this uncertain part might constitute threats for the goal of the agent. These threats are modeled through attacks against arguments in the fixed 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 fixed part. This part captures 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 is controlled solely by the agent, which can decide which arguments are actually used in the system or not. The environment has no influence on this part. These arguments can "protect" the agent's 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 fixed part, but also against arguments in the fixed 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 fixed theory.

One of the positive features of this partitioning is that our framework is modular. Indeed, if the system has to be updated, only a part of it is concerned with this update. For instance, the agent becomes aware of the existence of new threats, through the update of the uncertain part. No other part is affected. When new remedial actions are made available to the agent, the control part is concerned with this update, but not the other two parts. In the classical AFs any change modeled through the dynamics is integrated in the (updated) basic knowledge of the agent and from that moment onwards it is part of the knowledge of the agent. However, some of these changes should not probably be integrated in a permanent way in the background knowledge of an agent as they could not be coherent with the agent's goals satisfaction and therefore having a negative impact on the action of the agent when trying to optimize his goals. For instance why in a negotiation context integrating permanently in the basic theory of a seller agent implementing his selling policy the attacks of a buyer agent representing his preferences? In our framework uncertain knowledge of a negotiating agent about his opponent (i.e. which arguments he does have in his theory and which arguments he doesn't) as well as exchanged arguments and attacks (expressing preferences) during a negotiation dialogue, would be represented in the uncertain part. So in classical AFs the basic theory is updated permanently when the environment changes irrespectively of the nature of the dynamics, while in our framework only the parts that are concerned with these dynamics are updated. So, modularity makes it easier to maintain the system and protects the satisfaction of its goals.

Formally, a CAF is defined as follows:

Definition 1. *Let $\mathcal{L}$ be 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* fixed 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 belong to 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 is a set of arguments for which we are not sure that they belong to the system, $\rightleftarrows \subseteq (((A_U \cup A_F) \times (A_U \cup A_F)) \setminus \rightarrow)$ 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 $\Rrightarrow \subseteq \{(a_i, a_j) \mid a_i, a_j \in (A_F \cup A_C \cup A_U)$ and $a_i \in A_C$ or $a_j \in A_C\} \setminus (\rightarrow \cup \rightleftarrows \cup \dashrightarrow))$ is an attack relation.*

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

$\mathcal{F}$ is the fixed part of the CAF, *i.e.* the part of the system which cannot be influenced either by the agent or by the environment. $\mathcal{U}$ represents the possible changes of the environment and the context dependant information. This can be seen as threats against a goal related to the fixed part. $\mathcal{C}$ represents everything which can be decided by the agent; this part is seen as the remedial actions to protect the goal.

Intuitively, $\langle A_F, \rightarrow \cap (A_F \times A_F) \rangle$ is a classical AF where everything is known. The originality of our approach lies in the other arguments and attacks. For instance, there is an attack $(a_i, a_j) \in \rightleftarrows$ or $(a_i, a_j) \in \dashrightarrow$, with $a_i, a_j \in A_F$, when we are sure that two arguments exist but we are not sure of the direction or the existence of attacks between them (*e.g.* due to a lack of information about the context).
An attack $(a_i, a_j) \in \rightarrow$, with $a_i \in A_U$ and $a_j \in A_F$, represents a situation where it is uncertain whether a_i is present in the system (*e.g.* some of its premises could be false at the current time), but *if a_i is in the system*, then it certainly attacks a_j.

Example 1. *We define a CAF $\mathcal{CAF} = \langle F, C, U \rangle$ as follows:*

- *the fixed part is $\mathcal{F} = \langle \{a_1, a_2, a_3, a_4, a_5\}, \{(a_2, a_1), (a_3, a_1), (a_4, a_2), (a_4, a_3)\} \rangle$;*
- *the uncertain part is $\mathcal{U} = \langle \{a_6\}, \rightleftarrows \cup \dashrightarrow \rangle$, with $\rightleftarrows = \{(a_6, a_4), (a_4, a_6)\}$, and $\dashrightarrow = \{(a_5, a_1)\}$;*
- *the control part is $\mathcal{C} = \langle \{a_7, a_8, a_9\}, \{(a_7, a_5), (a_8, a_5), (a_8, a_6), (a_9, a_6)\} \rangle$.*

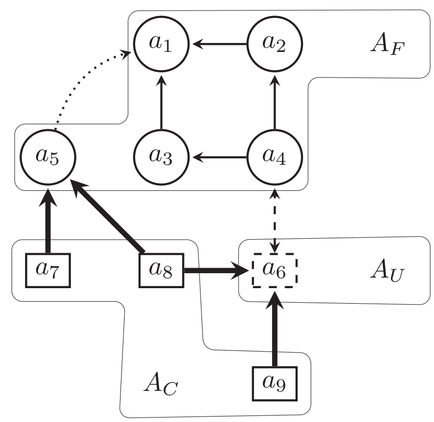


Figure 1: The CAF $\mathcal{CAF}$

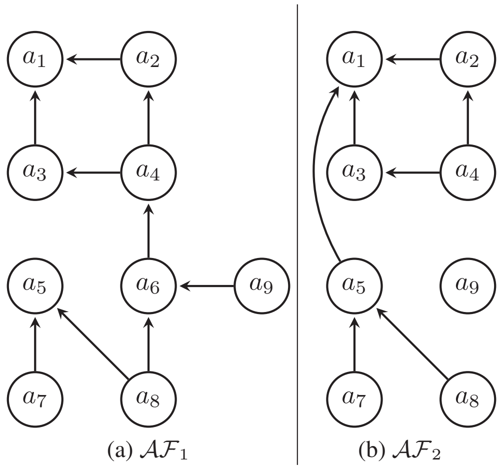
Before talking about controllability, we need to introduce the notion of completion of a CAF. Intuitively, a completion is a classical AF which is built from the CAF, by choosing one of the possible options for each uncertain argument or attack.

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

- *$A = A_F \cup A_C \cup A_{comp}$ 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.

Example 2 (Example 1 cont.)**.** *We describe two possible completions of $\mathcal{CAF}$. First, we consider a completion $\mathcal{AF}_1$ where the attack (a_5, a_1) is not included, while the argument a_6 (with the attack (a_6, a_4)) is included. Another possible completion is $\mathcal{AF}_2$, where a_6 is not included (so, neither the* *attacks related to it) while the attack (a_5, a_1) is included. The completions are "classical" AFs, where no distinction is made between the different kinds of arguments and attacks.*



(a) $\mathcal{AF}_1$ (b) $\mathcal{AF}_2$

Figure 2: Two possible completions of $\mathcal{CAF}$

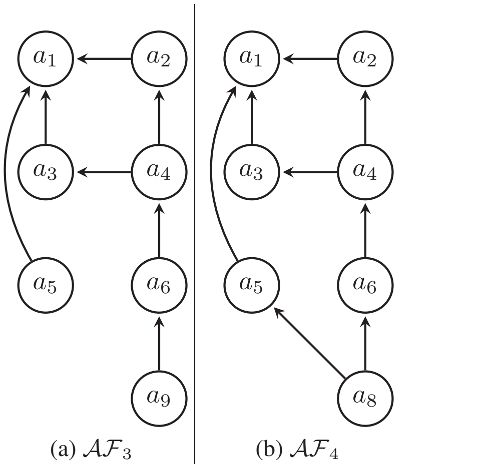
Controllability means that we can select a subset $A_{conf} \subseteq A_C$ and the corresponding attacks $\{(a_i, a_j) \in \Rightarrow \mid a_i, a_j \in (A_F \cup A_U \cup A_{conf})\}$ such that whatever the completion of $\mathcal{CAF}$, a given target is always reached. We focus on two kinds of targets: credulous acceptance of a set of arguments (this is reminiscent of extension enforcement (Baumann and Brewka 2010)) and skeptical acceptance of a set of arguments.

Definition 3. *A control configuration of a CAF $\mathcal{CAF} = \langle F, C, U \rangle$ 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 $\mathcal{CAF}' = \langle F, C', U \rangle$, with $\mathcal{C}' = \langle A_{conf}, \{(a_i, a_j) \in \Rightarrow \mid a_i, a_j \in (A_F \cup A_U \cup A_{conf})\} \rangle$. We say that $\mathcal{CAF}$ is skeptically (resp. credulously)* controllable *w.r.t. T and σ.*

Example 3 (Example 1 cont.)**.** *We suppose that the target is $T = \{a_1\}$. When we consider the control configuration $A'_{conf} = \{a_9\}$, T is neither skeptically nor credulously reached by A'_{conf} w.r.t. the stable semantics. Indeed, for the configured CAF $\mathcal{CAF}'$ where $\mathcal{C}$ is reduced to A'_{conf}, we can exhibit the completion $\mathcal{AF}_3$ where a_1 does not belong to a stable extension. On the contrary, the control configuration $A''_{conf} = \{a_8\}$ yields a configured CAF $\mathcal{CAF}''$ such that every completion accepts a_1 both credulously and skeptically (see $\mathcal{AF}_4$ for an example of such a completion). So $\mathcal{CAF}$ is credulously (and skeptically) controllable w.r.t. $\{a_1\}$ and the stable semantics.*

Controllability Through Logical Encoding

We propose a method to obtain a control configuration A_{conf} s.t. a set T of arguments is included in all extensions



(a) $\mathcal{AF}_3$ (b) $\mathcal{AF}_4$

Figure 3: Completions of configured CAFs

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. For illustrating our approach, we focus on the stable semantics, extending the encoding from (Besnard and Doutre 2004). However, our method is generic, and can be adapted to any semantics. Especially, for complete and admissible we can also use propositional encodings from (Besnard and Doutre 2004).

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.

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) are in one-to-one correspondence with the the stable extensions of $\mathcal{AF}$. Indeed, given $\mathcal{AF} = \langle A, R \rangle$, we define the formula $\Phi^R_{st} = \Phi_{st} \wedge (\bigwedge_{(x_i,x_j) \in R} att_{x_i,x_j}) \wedge (\bigwedge_{(x_i,x_j) \notin R} \neg att_{x_i,x_j})$. Given ω a model of Φ^R_{st}, 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^R_{st}.

Back to the case of CAFs, we see that we cannot directly generalize Φ^R_{st} 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 x_i currently 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 currently 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 $\mathcal{CAF} = \langle F, C, U \rangle$:

$$\begin{gathered}\Phi_{st}(\mathcal{CAF}) = \bigwedge_{x_i \in A_F}[acc_{x_i} \Leftrightarrow \\ \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j})] \wedge \\ \bigwedge_{x_i \in A_C \cup A_U}[acc_{x_i} \Leftrightarrow (on_{x_i} \wedge \\ \bigwedge_{x_j \in \mathbf{A}}(att_{x_j,x_i} \Rightarrow \neg acc_{x_j}))] \wedge \\ \bigwedge_{(x_i,x_j) \in \rightarrow \cup \Rrightarrow} att_{x_i,x_j} \\ \bigwedge_{(x_i,x_j) \in \rightleftarrows} att_{x_i,x_j} \vee att_{x_j,x_i} \\ \bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}\end{gathered}$$

The first two lines of this definition states in which condition an argument from A_F 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 two lines 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. So, any model of $\Phi_{st}(\mathcal{CAF})$ represents a configuration of $\mathcal{CAF}$ (given by the variables $on_{x_i}, x_i \in A_C$), its completions (given by the att_{x_i,x_j} and $on_{x_i}, x_i \in A_U$ variables), and one of its stable extensions (given by the acc_{x_i} variables).

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

$$\begin{aligned}&\exists\{on_{x_i} \mid x_i \in A_C\}\forall\{on_{x_i} \mid x_i \in A_U\} \\ &\forall\{att_{x_i,x_j} \mid (x_i,x_j) \in \dashrightarrow\}\Psi_{\forall}^{\rightleftarrows}(\Phi_{st}^{sk}(\mathcal{CAF}, T))\end{aligned} \qquad (1)$$

is valid, with $\Psi_{\mathcal{Q}}^{\rightleftarrows}$ defined, for $\mathcal{Q} \in \{\forall, \exists\}$, by

$$\begin{aligned}\Psi_{\mathcal{Q}}^{\rightleftarrows}(\Phi) = \bigwedge_{(x_i,x_j) \in \leftrightarrows}[&((\neg att_{x_i,x_j} \wedge att_{x_j,x_i}) \wedge \Psi_{\mathcal{Q}}(\Phi)) \\ &\vee((att_{x_i,x_j} \wedge \neg att_{x_j,x_i}) \wedge \Psi_{\mathcal{Q}}(\Phi)) \\ &\vee((att_{x_i,x_j} \wedge att_{x_j,x_i}) \wedge \Psi_{\mathcal{Q}}(\Phi))]\end{aligned}$$

and $\Psi_{\mathcal{Q}}(\Phi) = \mathcal{Q}\{acc_{x_i} \mid x_i \in \mathbf{A}\}\Phi$.

This formula follows the definition of skeptical controllability: which consists in determining whether there is a control configuration (given by the existential quantification over $on_{x_i}, x_i \in A_C$) such that for every completion, every extension satisfies the target. However, for the universal quantifications, we cannot make a trivial use of Boolean quantifiers to quantify over the set of every possible completion. Indeed, if we use $\forall\{att_{x_i,x_j} \mid (x_i,x_j) \in \leftrightarrows\}$, the formula should be true for any assignment of the Boolean variables att_{x_i,x_j}, including the case when both att_{x_i,x_j} and att_{x_j,x_i} are false. This specific assignment would be conflicting with the clause $att_{x_i,x_j} \vee att_{x_j,x_i}$ from $\Phi_{st}(\mathcal{CAF})$. See *e.g.* in example 1 the case of argument a_6 and the attacks it is involved: att_{a_6,a_4} and att_{a_4,a_6} cannot be false at the same time. For this reason, the quantification over the possible completions is made by the combination of classical Boolean quantification over the variables $on_{x_i}, x_i \in A_U$ and $att_{x_i,x_j}, (x_i,x_j) \in \dashrightarrow$, and a weaker form of quantification implemented by $\Psi_{\mathcal{Q}}^{\leftrightarrows}$. This function allows to quantify over all the valuations of $att_{x_i,x_j}, (x_i,x_j) \in \leftrightarrows$, except those which do not correspond to completions (and would violate $att_{x_i,x_j} \vee att_{x_j,x_i}$). Then, for each completion, the universal quantification over the acc_{x_i} variables allows to verify that the target T is contained in every extension.

For credulous controllability, we use the following encoding instead:

$$\begin{aligned}&\exists\{on_{x_i} \mid x_i \in A_C\}\forall\{on_{x_i} \mid x_i \in A_U\} \\ &\forall\{att_{x_i,x_j} \mid (x_i,x_j) \in \dashrightarrow\}\Psi_{\exists}^{\rightleftarrows}(\Phi_{st}^{cr}(\mathcal{CAF}, T))\end{aligned} \qquad (2)$$

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 $A_{conf} = \{x_i \in A_C \mid on_{x_i}$ is assigned to $1\}$. Both these tasks can be performed by any modern QBF solver (Pulina 2016). The method is formally described in Algorithm 1. We suppose that $\mathrm{QBFTruth}$ is a sound and complete procedure which determines whether a QBF is true or not; $\mathrm{QBFModel}$ is a sound and complete procedure which returns, when it exists, a consistent assignment of the Boolean variables which are quantified at the first existential level.

Example 4 (Example 1 cont.)**.** *We illustrate our QBF-based approach with an instantiation of the formula $\Phi_{st}(\mathcal{CAF})$ with $\mathcal{CAF}$ as 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*

Algorithm 1 CAFControl

Require: $\mathcal{CAF} = \langle F, C, U \rangle, T \subseteq A_F, x \in \{sk, cr\}$
$QBF_{sk}(\mathcal{CAF}, T)$ is the formula defined by (1)
$QBF_{cr}(\mathcal{CAF}, T)$ is the formula defined by (2)
if $\mathrm{QBFTruth}(QBF_x(\mathcal{CAF}, T))$ **then**
  $A_{conf} = \{x_i \in A_C \mid on_{x_i}$ is assigned 1 in $\mathrm{QBFModel}(QBF_x(\mathcal{CAF}, T))\}$
  return A_{conf}
else
  return $\perp$
end if

att_{x_j,x_i} is known to be false, the implication can be removed from the encoding.

$$\Phi_{st}(\mathcal{CAF}) = [acc_{a_1} \Leftrightarrow (\neg acc_{a_2} \wedge \neg acc_{a_3} \wedge (att_{a_5,a_1} \Rightarrow \neg acc_{a_5}))] \wedge [acc_{a_2} \Leftrightarrow \neg acc_{a_4}] \wedge [acc_{a_3} \Leftrightarrow \neg acc_{a_4}] \wedge [acc_{a_4} \Leftrightarrow (att_{a_6,a_4} \Rightarrow \neg a_6)]$$

$$\wedge [acc_{a_7} \Leftrightarrow on_{a_7}] \wedge [acc_{a_8} \Leftrightarrow on_{a_8}] \wedge [acc_{a_9} \Leftrightarrow on_{a_9}] \wedge [acc_{a_5} \Leftrightarrow (\neg acc_{a_7} \wedge \neg acc_{a_8})] \wedge [acc_{a_6} \Leftrightarrow (on_{a_6} \wedge \neg acc_{a_8} \wedge \neg acc_{a_9} \wedge (att_{a_4,a_6} \Rightarrow \neg acc_{a_4}))]$$

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

$$\wedge [att_{a_6,a_4} \vee att_{a_4,a_6}] \wedge \bigwedge_{(x_i,x_j) \notin \mathbf{R}} \neg att_{x_i,x_j}$$

To keep the encoding simple, we do not provide details on the part concerning the 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 base of the encoding. The encoding adapted for the skeptical controllability is $\Phi^{sk}_{st}(\mathcal{CAF}, T) = (\Phi_{st}(\mathcal{CAF}) \Rightarrow \bigwedge_{x_i \in T} acc_{x_i})$. Finally, with the quantifiers, we obtain the following QBF:

$$\exists on_{a_7}, on_{a_8}, on_{a_9}, \forall on_{a_6} \forall att_{a_5,a_1} [((\neg att_{x_4,x_6} \wedge att_{x_6,x_4}) \wedge \Psi) \vee ((att_{x_4,x_6} \wedge \neg att_{x_6,x_4}) \wedge \Psi) \vee ((att_{x_4,x_6} \wedge att_{x_6,x_4}) \wedge \Psi)]$$

with $\Psi = \forall \{acc_{x_i} \mid x_i \in \mathbf{A}\} \Phi^{sk}_{st}(\mathcal{CAF}, T)$. It is possible to compute CAFControl($\mathcal{CAF}, T, sk$): using a QBF solver on this formula gives a valuation of the $on_{a_7}, on_{a_8}, on_{a_9}$ variables which corresponds to a control configuration, i.e. *a subset of A_C such that the target $T = \{a_1\}$ is skeptically accepted. Here, the solutions are the control configurations which contain either a_8, or a_7 and a_9 together.*

Now we prove that our procedure for determining a control configuration is sound and complete.

Proposition 1. *Given a CAF $\mathcal{CAF} = \langle F, C, U \rangle$ and a target $T \subseteq A_F$,*
1. if CAFControl($\mathcal{CAF}, T, sk$) $= A_{conf}$, then $\mathcal{CAF}$ is skeptically controllable w.r.t. T and the stable semantics, and T is skeptically reached by A_{conf} w.r.t. the stable semantics;
2. if CAFControl($\mathcal{CAF}, T, cr$) $= A_{conf}$, then $\mathcal{CAF}$ is credulously controllable w.r.t. T and the stable semantics, and T is credulously reached by A_{conf} w.r.t. the stable semantics.

Proposition 2. *Given a CAF $\mathcal{CAF} = \langle F, C, U \rangle$ and a target $T \subseteq A_F$,*
1. if $\mathcal{CAF}$ is skeptically controllable w.r.t. T and the stable semantics, then CAFControl($\mathcal{CAF}, T, sk$) $= A_{conf}$ s.t. T is skeptically reached by A_{conf} w.r.t. the stable semantics;
2. if $\mathcal{CAF}$ is credulously controllable w.r.t. T and the stable semantics, then CAFControl($\mathcal{CAF}, T, cr$) $= A_{conf}$ s.t. T is credulously reached by A_{conf} w.r.t. the stable semantics.

When a 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 (Reimer et al. 2014) 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 (Reimer et al. 2014) 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 $Q^{\forall,\exists}_L$, where L is the set of allowed quantification levels. For instance the formula $F = Q^{\forall,\exists}_{\{1,2,3\}} x \forall y \exists z \Phi$, 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.

Computational Complexity

We now investigate 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^P_2 (resp. Σ^P_3). 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 $\mathcal{CAF} = \langle F, C, U \rangle$, and an argument $q \in A_F$, 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 3. *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 (Baumann and Brewka 2010) by some arguments and attacks from C. So, as a direct consequence of (Wallner, Niskanen, and Järvisalo 2016), the credulous conclusion problem for a SCAF under the stable semantics is NP-complete.

Proposition 4. *The credulous conclusion problem for a Simplified Control Argumentation Framework under the stable semantics is* NP*-complete.*

We also obtain a partial result for the skeptical conclusion problem.

Proposition 5. *The skeptical conclusion problem for a Simplified Control Argumentation Framework under the stable semantics is* NP*-hard.*

Related Work

Many works on argumentation dynamics or uncertainty in argumentation have been proposed in recent years. We describe here the more relevant approaches.

First, Partial Argumentation Frameworks (PAFs) have been proposed in (Coste-Marquis et al. 2007). Such a PAF is similar to classical Dung's AFs, with an additional relation between arguments representing ignorance. Formally, a PAF can be seen as a CAF with $A_C = A_U = \emptyset$, and $\Rrightarrow = \leftrightarrows = \emptyset$, while the ignorance relation of the PAF corresponds to our $\dashrightarrow$ relation. (Coste-Marquis et al. 2007) uses PAFs as a tool to merge several AFs. Later, (Baumeister, Neugebauer, and Rothe 2015) studies PAFs (renamed as Attack-Incomplete AFs). The question studied in this paper is to determine the complexity of the verification problem for several argumentation semantics, *i.e* given a set of arguments S and a PAF, is S an extension of every (or some) completion of the PAF. (Baumeister, Rothe, and Schadrack 2015) handle the same question for a variant of the framework, named Argument-Incomplete AF. This kind of AFs are built on two different sets of arguments, corresponding respectively to A_F and A_U in CAFs. For all these frameworks (Coste-Marquis et al. 2007; Baumeister, Neugebauer, and Rothe 2015; Baumeister, Rothe, and Schadrack 2015), we observe that CAFs propose a more general setting to represent uncertainty in argumentative scenarios. Moreover, none of these works is concerned with argumentation dynamics nor proposes a computational method.

(Boella et al. 2011) introduces a concept of potential attacks, which is related to the uncertain attacks relation $\dashrightarrow$. However, the difference is that potential attacks are discussed by agents, which decide to "accept" them (which means to move them from $\dashrightarrow$ to $\rightarrow$) or "reject" them (which means to remove them from the framework) to obtain finally a classical AF which optimizes some criterion.

Uncertainty in argumentation has also been studied in the context of probabilities (Hunter 2014). This supposes that the agent has some richer information than what is required in the definition of CAFs. It seems reasonable to consider that this kind of numeric information is not always available for the agent. However, using probabilities in argumentation dynamics could be a possible extension of our work and is kept for future work.

There is a strong relation between our contribution and extension enforcement (Baumann and Brewka 2010). 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 where the agent knowledge is incomplete. On the opposite, the YALLA language (Dupin de Saint-Cyr et al. 2016) 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: we are not aware of any efficient algorithmic approach to handle YALLA-based reasoning.

There is a connection with input/output modules (Baroni et al. 2014). However, the idea of input/output is to study the relations between sub-frameworks, *i.e.* how the arguments' status in a sub-framework influences other sub-frameworks. These modules are not related to the enforcement of a set of arguments, and do not incorporate uncertainty.

Finally, the acronym CAF has already been used in some work related to argumentation dynamics (see e.g. (Liao, Jin, and Koons 2011)) where it stands for Conditioned Argumentation Frameworks. In this work, the authors propose a division of an argumentation theory (*i.e.* affected, unaffected and conditioning parts) w.r.t. an update. However this division is completely different to our partitioning. Moreover, this work assumes complete information concerning the addition/deletion of arguments and attacks.

Conclusions and Future Work

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. We have proposed the notion of controllability of a CAF, *i.e.* the fact that a CAF can anticipate all the possible threats against a given goal related to arguments acceptance.

As future work, we will study a negative counterpart of the skeptical and credulous conclusion problems, *i.e.* determining whether it is possible to control a CAF such that some arguments are rejected. 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 (Reimer et al. 2014) can be used for that. We will identify concrete applications of CAFs in order to extract benchmarks from these applications, and run our (soft) QBF-based algorithms on these benchmarks.

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, depending on which arguments' premises can be deduced from the agent's background knowledge. 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

Arora, S., and Barak, B. 2009. *Computational Complexity - A Modern Approach*. Cambridge University Press.

Baroni, P.; Boella, G.; Cerutti, F.; Giacomin, M.; van der Torre, L. W. N.; and Villata, S. 2014. On the input/output behavior of argumentation frameworks. *Artif. Intell.* 217:144–197.

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An introduction to argumentation semantics. *Knowledge Eng. Review* 26(4):365–410.

Baumann, R., and Brewka, G. 2010. Expanding argumentation frameworks: Enforcing and monotonicity results. In *COMMA'10*, 75–86.

Baumeister, D.; Neugebauer, D.; and Rothe, J. 2015. Verification in attack-incomplete argumentation frameworks. In *ADT'15*, 341–358.

Baumeister, D.; Rothe, J.; and Schadrack, H. 2015. Verification in argument-incomplete argumentation frameworks. In *ADT'15*, 359–376.

Besnard, P., and Doutre, S. 2004. Checking the acceptability of a set of arguments. In *NMR'04*, 59–64.

Besnard, P., and Hunter, A. 2008. *Elements of Argumentation*. MIT Press.

Biere, A.; Heule, M.; van Maaren, H.; and Walsh, T., eds. 2009. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Boella, G.; Gabbay, D. M.; Perotti, A.; van der Torre, L. W. N.; and Villata, S. 2011. Argumentative agents negotiating on potential attacks. In *KES-AMSTA'11*, 280–290.

Booth, R.; Kaci, S.; Rienstra, T.; and van der Torre, L. 2013. A logical theory about dynamics in abstract argumentation. In *SUM'13*, 148–161.

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

Coste-Marquis, S.; Devred, C.; Konieczny, S.; Lagasquie-Schiex, M.; and Marquis, P. 2007. On the merging of dung's argumentation systems. *Artif. Intell.* 171(10-15):730–753.

Coste-Marquis, S.; Konieczny, S.; Mailly, J.; and Marquis, P. 2014. On the revision of argumentation systems: Minimal change of arguments statuses. In *KR'14*.

Coste-Marquis, S.; Konieczny, S.; Mailly, J.; and Marquis, P. 2015. Extension enforcement in abstract argumentation as an optimization problem. In *IJCAI'15*, 2876–2882.

Doutre, S.; Herzig, A.; and Perrussel, L. 2014. A dynamic logic framework for abstract argumentation. In *KR'14*.

Dung, P. M.; Kowalski, R. A.; and Toni, F. 2009. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*. 199–218.

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

Dupin de Saint-Cyr, F.; Bisquert, P.; Cayrol, C.; and Lagasquie-Schiex, M. 2016. Argumentation update in YALLA (yet another logic language for argumentation). *Int. J. Approx. Reasoning* 75:57–92.

Hunter, A. 2014. Probabilistic qualification of attack in abstract argumentation. *Int. J. Approx. Reasoning* 55(2):607–638.

Kakas, A. C., and Moraitis, P. 2003. Argumentation based decision making for autonomous agents. In *AAMAS'03*, 883–890.

Kleine Büning, H., and Bubeck, U. 2009. Theory of quantified boolean formulas. In *Handbook of Satisfiability*. 735–760.

Liao, B. S.; Jin, L.; and Koons, R. C. 2011. Dynamics of argumentation systems: A division-based method. *Artif. Intell.* 175(11):1790–1814.

Pulina, L. 2016. The ninth QBF solvers evaluation - preliminary report. In *QBF'16*, 1–13.

Reimer, S.; Sauer, M.; Marin, P.; and Becker, B. 2014. QBF with soft variables. In *AVOCS'14*.

Wallner, J. P.; Niskanen, A.; and Järvisalo, M. 2016. Complexity results and algorithms for extension enforcement in abstract argumentation. In *AAAI'16*, 1088–1094.