Formal Analysis of Dialogues on Infinite Argumentation Frameworks

Francesco Belardinelli  
Laboratoire IBISC  
Université d’Evry, France  
belardinelli@ibisc.fr

Davide Grossi  
Department of Computer Science  
University of Liverpool, UK  
d.grossi@liverpool.ac.uk

Nicolas Maudet  
Sorbonne Universités  
UPMC Univ Paris 06  
CNRS, UMR 7606, LIP6  
F-75005, Paris, France  
nicolas.maudet@lip6.fr

Abstract

The paper analyses multi-agent strategic dialogues on possibly infinite argumentation frameworks. We develop a formal model for representing such dialogues, and introduce $\text{FO}_A$-$\text{ATL}$, a first-order extension of alternating-time logic, for expressing the interplay of strategic and argumentation-theoretic properties. This setting is investigated with respect to the model checking problem, by means of a suitable notion of bisimulation. This notion of bisimulation is also used to shed light on how static properties of argumentation frameworks influence their dynamic behaviour.

1 Introduction

The dialectical and dynamic dimensions of argumentation have been object of scrutiny since the inception of Dung’s abstract argumentation theory (cf. [Dung, 1994; 1995]). However, while the definition and analysis of ‘static’ justifiability criteria (i.e., argumentation semantics [Baroni et al., 2011]) has come to form the bulk of abstract argumentation theory, comparatively little work within Dung’s framework has been dedicated to a systematic study of forms of dynamic and multi-agent interaction. Some research has focused on operationalizations of argumentation semantics via two-player games (see [Modgil and Caminada, 2009] for an overview), while some other has attempted an analysis of strategic behavior in abstract forms of argumentation games (in particular [Procaccia and Rosenschein, 2005; Riveret et al., 2010; Thimm and Garcia, 2010]). This paper pursues further the understanding of multi-agent argumentation over abstract argumentation frameworks (AF) capitalizing on techniques from logic and multi-agent verification.

Contribution of the paper. The paper focuses on the formal analysis of multi-agent strategic interactions on possibly infinite argumentation frameworks. Agents are assumed to exchange arguments from possibly infinite AF. They hold private argumentation frameworks representing their ‘views’ on how arguments attack one another. They interact by taking turns and attacking relevant arguments expanding the framework underlying the interaction. This set up, which we call dynamic argumentation systems (DAS), is general enough to model a wide range of dialogue protocols and games on abstract AF. We analyse this setting formally by introducing $\text{FO}_A$-$\text{ATL}$, a novel first-order extension of the (turn-based) game logic $\text{ATL}$ [Alur et al., 2002]. This logic allows us to specify dynamic properties of strategic interactions in argumentation. The objectives of the paper consist in: (i) the development of techniques to tackle the model-checking problem of $\text{FO}_A$-$\text{ATL}$; (ii) the development of techniques to analyse how static properties of argumentation frameworks influence their dynamic behavior. In other words, we are interested in isolating a class of AF for which their structure allows us to predict their dynamic behaviours. We provide preliminary positive results to both questions.

Technically, the paper models DAS as a special type of infinite-state data-aware systems [Deutsch et al., 2009; Hariri et al., 2013]. This allows us to capitalize on recent results on the formal verification of artifact-centric systems [Belardinelli et al., 2014], thus obtaining truth-preserving bisimulations for $\text{FO}_A$-$\text{ATL}$.

Related work. The paper contributes to several current strands of research in abstract argumentation.

Dynamics of argumentation. How argumentation framework may change by performing operations on their structure has been object of several recent works (e.g., [Baumann, 2012; Bisquert et al., 2013; Booth et al., 2013; Doutre et al., 2014]). However, all mentioned papers assume finite AF, as they rely on the possibility of encoding them into propositional formulas [Besnard and Doutre, 2004]. Neither do they incorporate agency, as they analyse argumentation dynamics from a belief revision paradigm. Our contribution relaxes the finiteness assumption and models agents explicitly as ‘protocols’ dynamically modifying the structure of AF.

Infinite argumentation frameworks. The bulk of known results in abstract argumentation typically pertain to finite AF. However, infinite AF are gaining attention and have been object of several recent contributions [Baroni et al., 2012; 2013; Baumann and Spanring, 2015], which essentially focus on how known results for the finite case generalize to the infinite. Allowing an infinity of arguments is critical in applications where upper bounds on the number of available arguments cannot be established a priori. Our paper contributes to the understanding of infinite AF.

Logics for abstract argumentation. Recently, several formalizations of argumentation theory have been put forward...
(e.g., [Caminada and Gabbay, 2009; Grossi, 2010] for early contributions). These works typically focus on finding logical languages (from modal to many-valued logics) that are sufficiently expressive to represent argumentation semantics. Our focus here is rather to specify the strategic abilities of agents engaging in a dialogue/dispute, such as: ‘the proponent is able to respond to all attacks by maintaining a conflict-free set of arguments’ or ‘the opponent has a strategy to force proponent to run out of arguments’. In this respect the paper is a first contribution to the specification of multi-agent argumentation frameworks by means of temporal logics, as well as to their formal verification (cf. [Lomuscio et al., 2009]).

Outline of the paper. Section 2 introduces the dynamics of (multi-agent) argumentation frameworks and the specification language $\text{FO}_{\text{A}}\text{-ATL}$; then we state the corresponding model checking problem. Section 3 contains the main technical result, namely, bisimilar dynamic argumentation systems (DAS) satisfy the same formulas in $\text{FO}_{\text{A}}\text{-ATL}$. We also investigate the impact of static features of DAS on their dynamics. We conclude in Section 4 and point to future work. For reason of space all proofs are omitted.

2 The Dynamics of AF

In this section we introduce dynamic argumentation systems (DAS). Then, we present $\text{FO}_{\text{A}}\text{-ATL}$, a first-order version of the alternating-time temporal logic ATL [Alur et al., 2002], and state the corresponding model checking problem. We first present the basic terminology to be used in the paper.

2.1 Abstract Argumentation

The theoretical setting is built on abstract argumentation, as introduced in [Dung, 1994; 1995]. In what follows we assume a finite set $A_g = \{a_0, \ldots, a_n\}$ of names for agents.

Definition 1 (Argumentation Framework) Given a set $A_g$ of agent names, a (multi-agent) argumentation framework is a tuple $A = (A, \{\leftarrow_a\} \in A_g)$ s.t. (i) $A$ is a (possibly infinite) set of arguments, and (ii) for every $a \in A_g$, $\leftarrow_a \subseteq A^2$ is an attack relation between arguments.

Notice that Def. 1 allows argumentation frameworks, or AF, that include infinitely many arguments. This choice, while providing greater modelling flexibility, also reflects the fact that upper bounds on the number of arguments available to each agent cannot be easily established in general.

We define $F(A, A_g)$ as the set of all AF built on sets $A$ of arguments and $A_g$ of agent names. Hereafter, we simply write $F(A_g)$ whenever $A$ is clear. Also, given set $A_g = \{p, o\}$ containing the proponent $p$ and opponent $o$, $\overrightarrow{F}(p)$ (resp. $\overrightarrow{F}(o)$) are shorthands for $F(\{p\})$ (resp. $F(\{o\})$), i.e., the set of AF for proponent $p$ (resp. opponent $o$). Finally, we define the (uninduced) attack relation $\leftarrow = \bigcup_{a \in A_g} \leftarrow_a$.

To express relevant properties of AF, we introduce $\text{FO}_{\text{A}},$ a first-order language geared towards expressing properties of attack relations. Let $A_{\varphi}$, for $a \in A_g$, be binary predicate symbols, $P_0, P_1, \ldots$ unary predicate symbols, and $\text{Var}$ a countable set of individual variables. We define first-order formulas $\varphi$ and the set $\text{fr}(\varphi)$ of free variables by mutual recursion.

Definition 2 ($\text{FO}_{\text{A}}$) The formulas $\varphi$ and the set of free variables $\text{fr}(\varphi)$ are defined as follows:

- if $P$ is a predicate symbol and $x \in \text{Var}$ a variable, then $P(x)$ is a formula with $\text{fr}(P(x)) = \{x\}$;
- if $\varphi$ is a formula, then also $\neg \varphi$ is, and $\text{fr}(\neg \varphi) = \text{fr}(\varphi)$;
- if $\varphi$ and $\varphi'$ are formulas with $\text{fr}(\varphi') = \text{fr}(\varphi')$, then also $\varphi \land \varphi'$ is, and $\text{fr}(\varphi \land \varphi') = \text{fr}(\varphi \land \varphi')$;
- if $\varphi$ is a formula with $\text{fr}(\varphi) = \{y\}$, then also $\forall y(A_u(y, x) \rightarrow \varphi)$ and $\forall y \varphi$ are, with $\text{fr}(\forall y(A_u(y, x) \rightarrow \varphi)) = \{x\}$ and $\text{fr}(\forall y \varphi) = \emptyset$.

In the following we use the usual abbreviations $\wedge, \lor, \exists$. Specifically, $\exists y(A_u(y, x) \land \varphi)$ (resp. $\exists y \varphi$) is a shorthand for $\neg\forall y(A_u(y, x) \rightarrow \neg \varphi)$ (resp. $\neg\forall y \neg \varphi$). Notice that every $\text{FO}_{\text{A}}$ formula $\varphi$ has at most one free variable.

Finally, we extend argumentation frameworks with interpretations $\pi$ that assign a subset of $A$ to each predicate symbol $P$, i.e., $\pi(P) \subseteq A$. An interpreted argumentation framework is defined as a couple $(A, \pi)$. We can now introduce the semantics of our first-order language.

Definition 3 (Semantics of $\text{FO}_{\text{A}}$) We define whether an argument $u \in A$ satisfies an $\text{FO}_{\text{A}}$ formula $\varphi$ in an interpreted $\text{AF} (A, \pi)$, or $(A, \pi, u) \models \varphi$, as follows (for clauses (propositional connectives are straightforward, thus omitted):

$(A, \pi, u) \models P(x)$ if $u \in \pi(P)$
$(A, \pi, u) \models \forall y(A_u(y, x) \rightarrow \psi)$ if for every $v \in A$

Provided $u \models \psi$ implies $(A, \pi, v) \models \psi$

$(A, \pi, u) \models \psi$ if for every $v \in A$, $(A, \pi, v) \models \psi$

An $\text{FO}_{\text{A}}$-formula $\varphi$ is true in an interpreted $\text{AF} (A, \pi)$, or $(A, \pi) \models \varphi$, if $(A, \pi, u) \models \varphi$ for every argument $u \in A$; $\varphi$ is valid in $A$, or $A \models \varphi$, if $(A, \pi) \models \varphi$ for every interpretation $\pi$.

From a technical point of view, the language $\text{FO}_{\text{A}}$ is the dyadic fragment of first-order logic with one free variable.\(^1\)

Table 1 shows how this simple logic suffices to formalize several of the key notions from [Dung, 1995] (see also [Grossi, 2010]).

2.2 Dynamic Argumentation Frameworks

In this paper we are interested in analysing dialogues modeled as multi-agent processes over argumentation frameworks. As dialogues progress agents build a ‘shared’ (or ‘public’) argumentation framework which is used to evaluate the available arguments.

Agents and global states

To introduce the dynamics of argumentation frameworks we start with the notion of agent.

Definition 4 (Agent) Given sets $A_g$ of agent names and $A$ of arguments, an agent is a tuple $a = (A, \text{Act}, P_\text{fr})$ where

- $A \in F(a)$ is the agent’s argumentation framework;
- the set Act of actions contains action attack$(x, x')$ to attack argument $x'$ with argument $x$, and action skip;
Definition 5 (Global State) Given a set $A$ of agents, a global state is a couple $(s, a)$ where (i) $s \in \mathcal{F}(A', Ag)$ is an argumentation framework for some $A' \subseteq A$; and (ii) $a \in Ag$.

In Def. 5 all agents are defined on the same set of arguments. However, some literature on argumentation theory suppose that each agent is endowed with a distinct set of arguments (e.g., [Rahwan and Larson, 2011]). We remark that our requirement is not a limitation because, firstly we can always consider the union of the sets of arguments for each agent, and secondly the characterising feature of AF is really the attack relation. Moreover, this assumption simplifies the presentation hereafter. Also, in a state $(s, a)$, $a$ is the active agent, i.e., the agent to act next; it will be omitted whenever clear from the context. We write $G$ to denote the set of all global states. Finally, any set $A' \subseteq A$ can be seen as a global state where, for every $a \in Ag$, $\leftarrow_a$ is empty.

Argument Dynamics

We now introduce the dynamics of argumentation frameworks. In the rest of the paper we suppose that the only two agents are proponent $p$ and opponent $o$. This assumption can be lifted but it crucially simplifies the presentation of the key (conceptual and technical) contributions of the paper.

Definition 6 (DAS) Given set $Ag = \{p, o\}$ of agents defined on the same (possibly infinite) set $A$ of arguments, a dynamic argumentation system is a tuple $P = \langle Ag, I, \tau, \pi \rangle$ where

- $I \subseteq A \times \{o\}$ is the set of initial global states $(s_0, o)$;
- $\tau : G \times (Act_p(A) \cup Act_o(A)) \rightarrow G$ is the transition function, where $\tau((s, a), attack_a((i))$ is defined iff $a = a'$ and $attack_a((i)) \in Pr_a((s))$. Moreover, $(s', a') = \tau((s, a), attack((u, u'))$ iff $a' \neq a$ and $s' = (A', \leftarrow')$

for $A' = A \cup \{u\}$ and $\leftarrow' = \leftarrow \cup \{(u', u)\}$; while $(s', a') = \tau((s, a), skip) \text{ iff } a' \neq a$ and $s' = s$.

- $\pi$ is an interpretation of predicate symbols $P$ as above.

A DAS evolves from an initial state $(s_0, o) \in I$ as specified by the transition function $\tau$, which returns a successor state for the current state and selected action. DAS are turn-based, with each agent taking one action at a time. Hereafter we assume that the opponent always moves first, that is, $I$ is intuitively the set of arguments supported by the proponent. One key feature of DAS is that, since the set $A$ of arguments is infinite in general, they are infinite-state systems normally. Indeed, as the dialogue unfolds, agents can introduce an infinite number of arguments from $A$, thus going through infinitely many states during the system’s execution.

We now fix some notion. The transition relation $\rightarrow_a$ is defined so that $s \rightarrow_a s'$ iff for some action $\alpha_a((i)) \in Act_a((A))$, for $a \in \{o, p\}$, we have that $s \xrightarrow{\alpha_a((i))} s'$, i.e., $\tau((s, a), \alpha_a((i))) = (s', a')$ with $a' \neq a$. A run $\lambda$ from a state $(s, a)$, or $(s, a)$-run, is an infinite sequence $s^0 \rightarrow_a s^1 \rightarrow_a \ldots$, s.t. $s^0 = s$. Notice that, since agents $o$ and $p$ take turns in performing actions, the sequence of agents in an $(s, a)$-run is uniquely determined by $a$, so we omit agents whenever these are clear by the context. Hence, we will often denote a state $(s, a)$ simply as $s$, whenever agent $a$ is understood. For $n, m \in \mathbb{N}$, with $n \leq m$, we define $\lambda(n) = s^n$ and $\lambda[n, m] = s^n, s^{n+1}, \ldots, s^m$. A state $s'$ is reachable from $s$ iff for some $s$-run $\lambda$ and $i \geq 0$, $s' = \lambda(i)$. Let $S$ be the set of states reachable from any initial state $s_0 \in I$. Notice also that the transition relation is serial, as each agent has a skip action enabled at each local state.

DAS, dialogues and games

It is worth comparing DAS briefly with known structures in the literature on dialogue games for abstract argumentation. In the terminology of [Thimm and Garcia, 2010] DAS can be viewed as defining a dialectical game protocol, where at each step an active agent attacks some of the available arguments. DAS can be seen as extensive game forms with infinite horizon (though they are not games, as they do not specify payoffs). They also generalize the structures underpinning the extensive-form argument games of [ Procaccia and Rosenschein, 2005 ] by allowing agents to hold an infinite argumentation framework and to attack any of the available arguments, not only the last one uttered by some other agent. We finally show how standard dialogue games (e.g., for the grounded extension [Dung, 1994] or for credulous admissibility [Vreeswijk and Prakken, 2001]) can be modeled as DAS.

Example 1 In the game for the grounded extension two
agents $o$ and $p$ hold the same private argumentation framework, i.e., $A_o = A_p$. These games are two-player zero-sum perfect information games with possibly infinite horizon. We can view them as DAS. Consider for instance the AF in Figure 1. The DAS corresponding to the game for grounded played on this AF, starting at argument $t_1$ can be defined as follows. Fix $I = \{t_1\}$. For both agents we can define the following protocol: if the current framework contains $t_i$ then attack $t_i$ with $u_i$ or $t_{i+1}$, otherwise skip. This protocol encodes the ‘legal’ moves of $o$ and $p$ according to the game for the grounded extension. The possible runs of this DAS contain all possible sub-graphs of the above AF generated from $t_i$, e.g.: $\{(t_1, \emptyset)\}, \{(t_1, u_1)\}, \{(u_1, t_1)\}$, skip, ...; or $\{(t_1, \emptyset)\}, \{(t_1, t_2)\}, \{(t_2, t_1)\}, \{(t_1, t_{i+1})\}, \{(t_i, t_{i+1})\}, ...$. These runs correspond to possible dialogues in the game for the grounded extension.

Given a DAS $P$ on a (possibly infinite) set $A$ of arguments, we define the corresponding (joint) argumentation framework $A_P = \langle A, \{a \rightarrow (a \in A_P)\rangle$ so that $u \rightarrow\!\!\!\langle u' \rangle$ holds in $A_P$ if $u \rightarrow\!\!\!\langle u' \rangle$ holds in the AF $A_o$ for agent $a \in A$.

Thus, states in $P$ are truthful, yet partial, representations of $A_P$. However, the converse of $(\ast)$ does not hold in general, i.e., $P$ needs not to include all subgraphs of $A_P$ as states. To appreciate this consider the following:

**Example 2** [Example 1 continued] Now endow $o$ with a more restrictive protocol: if the current framework contains $t_i$ then attack $t_i$ with $u_i$, otherwise skip. This protocol makes $o$ play more rationally, selecting arguments to which $p$ cannot reply. The only possible run of this DAS is: $\{(t_1, \emptyset)\}, \{(t_1, u_1)\}, \{(u_1, t_1)\}$, skip, ...

This asymmetry motivates the next definition. An agent is naive if its protocol allows it to move any available attack (cp. this to the more restrictive protocol of $o$ in Example 1).

**Definition 7 (Naive Agent)** An agent $a$ is naive iff for every $A' \in F(A', A)$, attack$(u, u') \in Pr(A')$ iff $u' \in A'$ and $u \rightarrow\!\!\!\langle u' \rangle$ holds in $A_o$.

Intuitively, naive agents perform every possible attack available to them, irrespectively of the strategic behaviour of other agents. A DAS $P$ is naive iff every agent in $A$ is naive. Then, we remark that for naive DAS the converse of $(\ast)$ actually holds.

### 2.3 The Specification Language $FO_A$-ATL

We now introduce a formal language to specify properties of interest of DAS. Arguments in DAS call for the use of first-order logic, whereas strategy operators are needed to account for agents’ behaviour. To the best of our knowledge, no logic combining these two features has ever been studied.

**Definition 8 ($FO_A$-ATL)** The $FO_A$-ATL formulas $\varphi$, with free variables $fr(\varphi)$, are defined as follows:

- if $\psi$ is an $FO_A$-formula, then it is an $FO_A$-ATL formula;
- if $\varphi$ is a formula and $N \subseteq Ag$, then $\neg \varphi$, $(N)X \varphi$, and $(N)G \varphi$ are formulas with $fr(\neg \varphi) = fr((N)X \varphi) = fr((N)G \varphi) = fr(\varphi)$;
- if $\varphi$ and $\varphi'$ are formulas with $fr(\varphi) = fr(\varphi')$ and $N \subseteq Ag$, then $\varphi \rightarrow \varphi'$ and $(N)\varphi U \varphi'$ are formulas with $fr(\varphi \rightarrow \varphi') = fr((N)\varphi U \varphi') = fr(\varphi) = fr(\varphi')$;
- if $\varphi$ is a formula with $fr(\varphi) = \{y\}$, then $\forall y (A_o(y, x) \rightarrow \varphi)$ and $\forall y \varphi$ are formulas with $fr(\forall y (A_o(y, x) \rightarrow \varphi) = \{x\}$ and $fr(\forall y \varphi) = \emptyset$.

The language $FO_A$-ATL is a first-order extension of the alternating-time temporal logic ATL [Alur et al., 2002]. The $FO_A$-ATL formulas $(i)$ $(N)X \varphi$, $(ii)$ $(N)G \varphi$ and $(iii)$ $(N)\varphi U \varphi'$ are read as “the agents in $N$ have a strategy to...” $(ii)$ "...enforce $\varphi$ at the next state”, $(iii)$ "...always enforce $\varphi'$, and $(iii)$ “...enforce $\varphi$ until $\varphi'$”.

To interpret $FO_A$-ATL formulas on DAS we need to introduce the notion of a strategy for a set $N$ of agents. An $N$-strategy is a mapping $f_N : S^+ \rightarrow \bigcup_{a \in N} Act_a(A)$ s.t. $f_N(k \cdot (s, a)) \in Pr_a(s)$ for every $\kappa \in S^+$. Intuitively, a strategy returns an enabled action in $Act_a(A)$ for every non-empty, finite sequence of states in $S^+$. We remark that, according to standard terminology in concurrent game models [Bulling et al., 2010], the agents in DAS have perfect information and perfect recall, that is, their strategies are determined by all information available at each global state, for all states visited up to the current state. Further, the outcome of strategy $f_N$ at state $(s, a)$, or $out((s, a), f_N)$, is the set of all $(s, a)$-runs $\lambda$ s.t. for every $b \in N$, $(\lambda(i+1), b') = \tau((\lambda(i), b), f_N(\lambda[i:1]))$ for all $i \geq 0$. As above, we assume that strategies respect agents’ turns and simply write $out(s, f_N)$, thus omitting agents whenever the latter are clear by the context.

**Definition 9 (Semantics of $FO_A$-ATL)** We define whether an argument $u$ satisfies a formula $\varphi$ at state $s$ in a DAS $P$, or $(P, s, u) \models \varphi$, as follows (clauses for propositional connectives are straightforward and thus omitted):

- $(P, s, u) \models \psi$ if $s, s, u \models \psi$;
- $(P, s, u) \models (N)X \varphi$ if for some $N$-strategy $f_N$ for all $\lambda \in out(s, f_N)$, $(P, \lambda(1), u) \models \varphi$;
- $(P, s, u) \models (N)G \varphi$ if for some $N$-strategy $f_N$ for all $\lambda \in out(s, f_N)$, $i \geq 0$, $(P, \lambda(i), u) \models \varphi$;
- $(P, s, u) \models \neg \varphi$ if for some $N$-strategy $f_N$ for all $\lambda \in out(s, f_N)$, for some $k \geq 0$, $(P, \lambda(k), u) \models \varphi'$ and, for all $j \leq k$, implies $(P, \lambda(j), u) \models \varphi$;
- $(P, s, u) \models \forall y (A_o(y, x) \rightarrow \varphi)$ if for every $v \in s$, $u \rightarrow\!\!\!\langle v \rangle$ implies $(P, s, v) \models \varphi$;
- $(P, s, u) \models \forall y \varphi$ if for every $v \in s$, $(P, s, v) \models \varphi$.

A formula $\varphi$ is true at $s$, or $(P, s) \models \varphi$, if $(P, s, u) \models \varphi$ for every arguments $u \in s$; $\varphi$ is true in $P$, or $P \models \varphi$, if $(P, s_0) \models \varphi$ for all $s_0 \in I$. 

---

864
We illustrate now the expressiveness of FO₄-ATL through some example properties that involve strategic and argumentation-theoretic features.

**Example 3** The following formula states that opponent \( o \) can force proponent \( p \) to run out of moves in the next state:

\[
\langle \langle o \rangle \rangle \exists x \forall y (\nexists y A_p(y, x))
\]  
(1)

This formula is for instance true at argument \( t_1 \) in the DAS of Example 1.

**Example 4** The following formula states that proponent \( p \) has a strategy enforcing the set of arguments in \( P \), which includes the current argument, to be conflict-free (respectively, acceptable, admissible, complete, stable):

\[
P(x) \land \langle \langle p \rangle \rangle G \chi(P)
\]  
(2)

where \( \chi \in \{Cfr, Acc, Adm, Cmp, Stb\} \).

Let us focus for instance on \( P(x) \land \langle \langle p \rangle \rangle G \text{Adm}(P) \). This is a desirable requirement for proponent’s strategies in dialogue games for the grounded extension (recall Example 1) or credulous admissibility [Vreeswijk and Prakken, 2000]. This statement is false of the DAS of Example 1 (at argument \( t_1 \)), where \( o \), by moving \( \text{attack}(u_1, t_1) \), forces \( p \) to run out of moves, no matter what interpretation of \( \pi \) of \( P \) is selected. In fact \( p \) has no strategy to force any set of arguments containing \( t_1 \) (the initial argument) to be admissible. That is: \( P(x) \rightarrow \langle \langle o \rangle \rangle G \lnot \text{Adm}(P) \) or, equivalently, \( P(x) \rightarrow \lnot \langle \langle p \rangle \rangle G \text{Adm}(P) \) for any \( P \).  

3 Verifying Dynamic Argumentation Systems

In Section 2 we introduced DAS as a general model for (abstract) argument-based multi-agent interaction. We tailored an influential game logic in order to provide DAS with the specification language FO₄-ATL. We are then able to define a model checking problem whereby strategic and argumentation-theoretic properties, such as those expressed in (1) and (2) above, can be checked against a given DAS.

**Definition 10 (Model Checking Problem)** Given a DAS \( P \) and an FO₄-ATL-formula \( \varphi \), determine whether for every \( s_0 \in I \), \( (P, s_0, u_0) \models \varphi \) for some argument \( u_0 \).

In case that \( \varphi \) is a sentence with no free variable, the model checking problem reduces to verifying whether \( P \models \varphi \).

Model checking general data-aware systems is known to be undecidable [Deutsch et al., 2007]. In Belardinelli et al., 2012; 2014] the same problem is proved decidable for bounded and uniform systems. Without going into details, we remark that DAS do not normally satisfy these requirements. Therefore new techniques need to be developed.

In this section we introduce a notion of bisimulation to explore under which circumstances two DAS satisfy the same formulas. In particular, we show that bisimilar DAS satisfy the same FO₄-ATL formulas. Then, we investigate how the static properties of AF impact on the dynamics of DAS. This kind of results is key to tackle the model checking problem, as they allow to verify a DAS \( P \) by model checking a bisimilar DAS \( P' \).

**3.1 Bisimulations for Argument Frameworks**

A notion of bisimulation can naturally be defined on AF (cf. [Grossi, 2010]).

**Definition 11 (Static Bisimulation)** Let \( (A, \pi) = \langle A, \{a_r\}_{a \in Ag}, \pi \rangle \) and \( (A', \pi') = \langle A', \{a_r'\}_{a \in Ag}, \pi' \rangle \) be interpreted AF defined on a set \( Ag \) of agents. A static bisimulation is a relation \( S \subseteq A \times A' \) s.t. for \( u, u' \in A \), \( u, u' \in A' \), \( S(u, u') \) implies:

(i) for every predicate symbol \( P \), \( u \in \pi(P) \) iff \( u' \in \pi'(P) \);

(ii) for every \( v \in A \), if \( u \leftarrow_a v \) then for some \( v' \in A' \), \( u' \leftarrow a v' \) and \( S(v, v') \);

(iii) for every \( v' \in A' \), if \( u' \leftarrow a v' \) then for some \( v \in A \), \( u \leftarrow a v \) and \( S(v, v') \).

Two arguments \( u \in A \), \( u' \in A' \) are bisimilar, or \( u \simeq u' \), iff \( S(u, u') \) for some static bisimulation \( S \). Finally, two AF \( A \) and \( A' \) are statically bisimilar (or simply bisimilar) iff (i) for every \( u \in A \), \( u \simeq u' \) for some \( u' \in A' \); and (ii) for every \( u' \in A' \), \( u' \simeq u \) for some \( u \in A \). We denote this as \( A \simeq A' \).

We can now show that bisimilar states satisfy exactly the same FO₄-formulas.

**Lemma 1** Given bisimilar interpreted AF \( (A, \pi) \) and \( (A', \pi') \), and bisimilar arguments \( u \in A \) and \( u' \in A' \), then for every FO₄-formula \( \varphi \),

\[
(A, \pi, u) \models \varphi \iff (A', \pi', u') \models \varphi
\]

As a result, bisimilar AF cannot be distinguished by FO₄-formulas. In the following we explore the conditions under which this applies to DAS as well. In particular, we say that DAS \( P \) and \( P' \) are statically bisimilar iff \( A_P \simeq A_{P'} \) for some static bisimulation that maps initial states into initial states.

3.2 Bisimulations for DAS

We first introduce a notion of dynamic bisimulation, and then explore its properties in the context of DAS. In the rest of the section we let \( P = \langle Ag, I, \pi, \pi \rangle \) and \( P' = \langle Ag, I', \pi', \pi' \rangle \) be two DAS defined on the same set \( Ag \) of agent names. Notice that, albeit agents may have the same name, they might differ as to their argumentation frameworks, actions, or protocols.

**Definition 12 (Dynamic Simulation)** Given DAS \( P \) and \( P' \), a dynamic simulation is a relation \( R \subseteq S \times S' \) s.t. for \( s \in S \), \( s' \in S' \), \( R(s, s') \) implies:

1. \( s \simeq s' \) for some static bisimulation \( S \);
2. for every \( t \in S \), if \( s \rightarrow_a t \) then for some \( t' \in S' \), \( s' \rightarrow_a t' \); and \( t \simeq t' \) for some bisimulation \( S' \subseteq S, \) and \( R(t, t') \).

In Def. 12 we implicitly assume that the simulation relation relates states with turns for the same active agent in \( Ag \). Simulations can then be naturally extended to bisimulations.

**Definition 13 (Dynamic Bisimulation)** A relation \( D \subseteq S \times S' \) is a dynamic bisimulation iff both \( D \) and \( D^{-1} = \{ \{s', s\} \mid D(s, s') \} \) are dynamic simulations.

Two states \( s \in S \) and \( s' \in S' \) are dynamically bisimilar, or \( s \approx s' \), iff \( D(s, s') \) for some bisimulation relation \( D \). It can be shown that \( \approx \) is the largest dynamic bisimulation, and an equivalence relation, on \( S \cup S' \). DAS \( P \) and \( P' \) are dynamically bisimilar, or \( P \approx P' \), iff (i) for every \( s_0 \in I \), \( s_0 \approx s'_0 \).

\[\{5\}] \]
for some $s_0 \in I'$, and (ii) for every $s'_0 \in I'$, there exists $s_0 \approx s'_0$ for some $s_0 \in I$. Notice that arguments $u$ and $u'$ are or are not bisimilar always w.r.t. some states $s \in S$ and $s' \in S'$. We state this explicitly by saying that $u \approx u'$ w.r.t. $s$ and $s'$. In particular, by Def. 13, if $s \approx s'$ and $u \approx u'$ w.r.t. $s$ and $s'$, then $u$ and $u'$ are still bisimilar in all subsequent bisimilar states. This is a key feature for proving preservation of satisfaction for $\text{FO}_A$-ATL formulas.

The first result we prove on bisimulations shows that being statically bisimilar does not imply dynamic bisimilarity, not even in the case of naive agents.

**Lemma 2** Static bisimilarity does not imply dynamic bisimilarity, that is, there exist naive, statically bisimilar $\text{DAS } \mathcal{P}$ and $\mathcal{P}'$ such that $\mathcal{P} \not\approx \mathcal{P}'$.

In Fig. 2 we report $\text{DAS } \mathcal{P}$ and $\mathcal{P}'$, whose underlying AF are statically bisimilar, but that are not dynamically bisimilar.

We now show that dynamically bisimilar states satisfy the same $\text{FO}_A$-ATL formulas.

**Theorem 3** Suppose that $s \approx s'$, and $u \approx u'$ w.r.t. $s$ and $s'$. Then for every $\text{FO}_A$-ATL formula $\phi$, $(\mathcal{P}, s, u) \models \phi$ iff $(\mathcal{P}', s', u') \models \phi$.

As a result, dynamic bisimulations preserve the satisfaction of $\text{FO}_A$-ATL formulas. In particular, we can tackle the problem of model checking a $\text{DAS } \mathcal{P}$ by verifying a bisimilar $\text{DAS } \mathcal{P}'$, and then transferring the result by Theorem 3. In particular, $\text{DAS } \mathcal{P}'$ may exhibit nice structural properties (such as being finite) that can make the verification task feasible.

### 3.3 From Static Properties to Dynamics

In this section we make use of bisimulations to explore how the static properties of $\text{DAS}$ determine their dynamic features. In Lemma 2 we showed that this relationship is not straightforward, not even in the case of naive $\text{DAS}$. However, in some specific cases there is indeed a correspondence. The first result reduces dynamic bisimulations to static bisimulations, together with some constraints on temporal transitions. As usual, we assume that turns are respected.

**Theorem 4** Let $\mathcal{P}$ and $\mathcal{P}'$ be $\text{DAS}$. Suppose that $\mathcal{P}'$ is naive and for every $u \in \mathcal{S}$, $u' \in \mathcal{S}'$, if $s \approx s'$, $u \approx u'$ w.r.t. $s$ and $s'$, and $u \xleftarrow{\alpha} v$ in $\mathcal{P}$ for some $v \in \mathcal{A}$, then $u' \xleftarrow{\alpha} v'$ in $\mathcal{P}'$ for some $v' \in \mathcal{A}'$ and either

1. $v \in s$ and either (i) $v' \in s'$ and $v \approx v'$ w.r.t. $s$ and $s'$, or (ii) $v' \not\in s'$ and for no $w \in v$, $v \xleftarrow{\alpha} w$ in $s$,

2. or $v \not\in s$ and either (i) $v' \not\in s'$, or (ii) $v' \in s'$ and for no $w \in s'$, $v' \xleftarrow{\alpha} w'$ in $s'$.

Then, $D = \{ (s, s') \mid s \approx s' \}$ is a dynamic simulation between $\mathcal{P}$ and $\mathcal{P}'$.

Notice that the assumptions in Theorem 4 imply that $\text{DAS } \mathcal{P}$ and $\mathcal{P}'$ are statically bisimilar. However, they have also to satisfy extra conditions (1) and (2).

The main contribution of Theorem 4 is to show that the notion of dynamic bisimulation can be reduced to static bisimulation, together with some assumptions on the structural properties of $\text{AF } \mathcal{A}_P$ and $\mathcal{A}_{P'}$, in case that we consider naive $\text{DAS}$. Hence, in order to verify $\text{DAS}$ we can simply model check statically bisimilar systems, and then transfer the result by using Theorem 4.

It is of interest to analyse $\text{DAS}$ that actually satisfy the conditions above. For example, it is easy to check that whenever the underlying AF $\mathcal{A}_P$ and $\mathcal{A}_{P'}$ of naive $\text{DAS } \mathcal{P}$ and $\mathcal{P}'$ are directed acyclic graphs (DAG), where every argument is attacked by some other argument, then the conditions in Theorem 4 do indeed hold. So, for this class of naive $\text{DAS}$ structural, static features do determine their dynamic properties.

**Corollary 5** Suppose that $\text{DAS } \mathcal{P}$ and $\mathcal{P}'$ are naive and statically bisimilar, and that $\mathcal{A}_P$ and $\mathcal{A}_{P'}$ are DAG where every argument is attacked by some other argument. Then, $\mathcal{P}$ and $\mathcal{P}'$ are dynamically bisimilar and therefore satisfy the same $\text{FO}_A$-ATL formulas.

### 4 Conclusions and Further Work

In many dialectical situations it is difficult to assume beforehand a bound on the number of arguments that agents have at their disposal. While some properties of such infinite argumentation frameworks have recently been investigated, their implication on debates amongst agents had not been analysed before. In this paper we set up and explored the framework of dynamic argumentation systems through a logic lens. We showed that in general static bisimilarity is not strong enough to capture equivalence of $\text{DAS}$ (in terms of $\text{FO}_A$-ATL formulas they satisfy), and introduced a novel notion of dynamic bisimilarity. For some specific structures (in particular acyclic ones), the static notion remains powerful enough, at least for the naive agents studied here. While in this paper we have focused on such agents (unconstrained in their argumentative moves as long as they are relevant and truthful), an interesting direction of research is to investigate how more restricted protocols would also impact the dynamics of $\text{DAS}$. Another noteworthy feature of our framework is that formulas in $\text{FO}_A$-ATL can express the ability for a group of agents to ensure, for instance, that some arguments get accepted—thus paving the way for a logical analysis of multiparty protocols, see e.g. [Bonzon and Maudet, 2011].
References


Acknowledgements.
The third author benefited from the support of the project AMANDE ANR-13-BS02-0004 of the French National Research Agency (ANR).