Agent-based Abstractions for Verifying Alternating-time Temporal Logic with Imperfect Information

Francesco Belardinelli
Laboratoire IBISC, UEVE – IRIT Toulouse
belardinelli@ibisc.fr

Alessio Lomuscio
Department of Computing
Imperial College London
a.lomuscio@imperial.ac.uk

ABSTRACT

We introduce a 3-valued abstraction technique for alternating-time temporal logic (ATL) in contexts of imperfect information. We introduce under- and over-approximations of an agent’s strategic abilities in terms of must- and may-strategies and provide a 3-valued semantics for ATL based on agents in interpreted systems (IS). We define a relation of simulation between the agents and prove that it preserves defined truth values of ATL formulas. Finally, we introduce a notion of abstraction on IS and show that it simulates the concrete interpreted system. Under this setting we present a procedure that enables the direct construction of a finite abstraction from an infinite-state system.

1. INTRODUCTION

In recent years logic-based formalisms for representing and reasoning about strategic abilities, both individual and coalitional, have been a thriving area of research in formal methods for multi-agent systems [8, 6, 20]. Several multi-modal logics have been introduced to provide a formal account of complex strategic reasoning and behaviours for single agents and groups, including alternating-time temporal logic (ATL), coalition logic, and strategy logic [1, 10, 31, 32].

A key issue concerning these logics for strategic reasoning regards the development of techniques for verifying multi-agent systems (MAS) with respect to properties expressed in these rich formal languages. This endeavour is problematic as the typical verification problems for several of these formalisms, including model checking, are computationally harder than those for the temporal logics (e.g., CTL, LTL) that they subsume [15, 35].

Related to formal verification, a crucial distinction operated in game-theoretic contexts is whether players have perfect or imperfect information about other players and the environment they are interacting in and with. Originally, most of the logics for strategies mentioned above have been introduced in contexts of perfect information [1, 10], partly because this setting exhibits better computational properties. However, for many applications of interest, including autonomous agents, distributed computing, and economic theory, perfect information is either unrealistic as a working hypothesis or unattainable [16]. Imperfect information is known to make the verification task computationally more costly. As an example, while verifying ATL under perfect information is polynomial, the corresponding problem for imperfect information is \( \Delta^P_2 \)-complete [22]. When perfect recall is assumed, the problem goes from PTIME-complete to undecidable [15]. Thus, for logics of strategies to be adopted as specification languages in contexts of imperfect information, efficient verification tools and techniques, supporting model checking for practical cases of interest, should be developed.

In this paper we aim at contributing towards this long-term objective. Specifically, we introduce a novel 3-valued abstraction technique for the verification of ATL specifications with respect to (possibly infinite) MAS with imperfect information. Taking inspiration from [5, 19, 3, 4], we introduce under- and over-approximations of an agent’s strategic abilities in terms of must- and may-strategies, then provide a 3-valued semantics for ATL based on agents in interpreted systems (IS). Next, we define a relation of simulation between agents and prove that it preserves defined truth values of ATL formulas. Finally, we introduce a notion of abstraction on IS and show that it simulates the corresponding concrete interpreted system. We illustrate the formal machinery by means of a toy example based on the Train Gate Controller [21], and conclude by discussing applications to the verification of strategic behaviours of agents in multi-agent systems.

Related work. The literature on abstraction-based techniques applied to the model checking problem has grown steadily in the past two decades [11, 12]. Here we focus on the contributions most closely related to the present work.

Multi-valued interpretations of modal logics has since long appeared in the literature [17, 18]. This approach has also been applied to the verification of temporal and epistemic logics [26, 27, 28], including ATL* under imperfect information [23]. In this line formulas are interpreted on some designated algebraic structure (possibly infinite) and modal operators correspond to operations on the values in the structure.

Here we also adopt a multi-valued semantics, 3-valued specifically, but our approach is essentially different from the references above, as it is based on the definition of under- and over-approximations of transition systems [5, 19]. Indeed, the main inspiration for this paper comes from [3, 34], which put forward 3-valued abstraction techniques for CTL and the alternating \( \mu \)-calculus (\( \mu \textit{PC} \)), assuming perfect information nonetheless. Abstractions are shown to preserve
defined truth values of formulas in the relevant logic, then
a notion of pre-order between abstractions, akin to simula-
tion, is introduced. A crucial difference here is that we con-
sider imperfect information. Indeed, in $\mathsf{APC}$ with perfect
information, the ATL operators $\langle \Gamma \rangle U$, $\langle \Gamma \rangle G$, and $\langle \Gamma \rangle F$ are definable in terms of the ‘next’ modality $\langle \Gamma \rangle X$ and the
fixed-point operators $\mu$ and $\nu$. Hence, the technique in [3]
deals directly only with the next fragment of $\mathsf{APC}$, then ap-
plies standard procedures for calculating fixed-points. How-
ever, this method fails when imperfect information is taken
into account, as ATL operators can no longer be expressed
via fixed points [7, 14]. Thus, original techniques have to be
developed. Moreover, the semantics here proposed is agent-
based, and so is the simulation relation and abstraction tech-
nique we develop, again differently from [3].

Recently, other 3-valued abstraction techniques have ap-
ppeared [29, 30, 4], which differ from our account as regards
the 3-valued semantics for ATL. Notably, here we introduce
original notions of may- and must-strategies and prove that,
differently from [30], our 3-valued semantics conservatively
extends the standard 2-valued semantics for ATL.

Scheme of the paper. The rest of the paper is struc-
tured as follows. In Section 2 we present ATL and pro-
vide it with a semantics in terms of agent-based interpreted
systems, suitable for MAS representation. In Section 3 we
introduce a 3-valued semantics for ATL, based on general
agents and general IS. Then, we show that, differently from
[30], the 3-valued semantics conservatively extends the 2-
valued one. In Section 4 we introduce simulation relations
between agents and between IS, and prove that these indeed
preserve the interpretation of ATL formulas (Theorem 1).
Then, in Section 5 we define agent-based abstractions and
prove that they simulate the original MAS (Theorem 2).
We illustrate the formal machinery with an infinite version
of the Train Gate Controller scenario.

2. INTERPRETED SYSTEMS WITH IMPER-
FECT INFORMATION

In this section we introduce the formal machinery that
will be used throughout the paper. Hereafter we assume a
set $\mathcal{Ag} = \{1, \ldots, m\}$ of indexes for agents and a set $\mathsf{AP}$
of atomic propositions. Given a set $U$, $\overline{U}$ denotates its com-
plement (w.r.t some $V \supseteq U$). Also, we denote the $i$-th element
of a tuple $v$ as either $v_i$ or $v.i$.

Definition 1 (Agent). Given a set $\mathcal{Ag}$ of agent in-
dices, an agent is a tuple $i = (L, \mathsf{Act}, P, t)$ such that

- $L$ is the (possibly infinite) set of local states;
- $\mathsf{Act}$ is the (finite) set of individual actions;
- $P : L \to 2^{\mathsf{Act}}$ is the protocol function;
- $t : L \times \mathsf{Act} \to L$ is the local transition function, where
  $\mathsf{Act} = \mathsf{Act}_1 \times \cdots \times \mathsf{Act}_{|\mathcal{Ag}|}$ is the set of joint actions,
  s.t. for $l \in L$, $a \in \mathsf{Act}$, $t(l, a)$ is defined iff $a_i \in P(l)$.

Intuitively, an agent $i$ is situated in some local state $l \in L$,
which represents the information she has about the whole
system, and she can perform some action $a \in \mathsf{Act}$, accord-
ing to protocol $P$. Performing a joint action brings about
a change in the state of the agent, according to transition
function $t$. Hereafter we often identify an agent index $i$
with the corresponding agent, the context will disambiguate.
Also, we assume w.l.o.g. that for every local state $l \in L$,
$P(l) \neq \emptyset$ by considering a null action $\text{skip}$, enabled in every
local state, such that $t(l, \text{skip}) = l$ whenever $a_i = \text{skip}$. Hence,
the protocol $P$ is a function from $L$ to $2^{\mathsf{Act}} \setminus \emptyset$.

Given a set $\mathcal{Ag}$ of agents, a global state $s \in G$ is defined
as a tuple $(l_1, \ldots, l_{|\mathcal{Ag}|})$ of local states, one for each agent in
$\mathcal{Ag}$. Notice that an agent’s protocol and transition function
depend only on its local state, which might contain strictly
less information than the global state $s$. In this sense agents
have imperfect information about the system. This is in
marked contrast with [3], where agents are assumed to have
perfect information.

To describe formally the execution of a multi-agent sys-
tem, we introduce the notion of interpreted system.

Definition 2 (IS). An interpreted system is a tuple
$M = (\mathcal{Ag}, I, T, \Pi)$ such that

- every $i \in \mathcal{Ag}$ is an agent;
- $I \subseteq G$ is the set of (global) initial states;
- $T : G \times \mathsf{Act} \to G$ is the global transition function
  such that $s' = T(s, a)$ iff for all $i \in \mathcal{Ag}$, $s'_i = t_i(s, a)$;
- $\Pi : G \times \mathsf{Act} \to \{tt, ff\}$ is the labelling function.

An interpreted system describes the interactions of a group
$\mathcal{Ag}$ of agents, starting from some initial state in $I$, accord-
ing to the transition function $T$. Atomic propositions are
assigned values true ($tt$) or false ($ff$). Notice that the global
transition function $T$ is defined on global state $s$ for joint
action $a$ iff $a_i \in P_i(s)$ for every $i \in \mathcal{Ag}$. Finally, we intro-
duce the set $S \subseteq G$ of global states reachable from set $I$ of
initial states, through the transition function $T$.

To reason about the strategic abilities of agents in inter-
preted systems, we make use of alternating-time temporal
logic (ATL).

Definition 3 (ATL). Formulas $\varphi$ in ATL are defined
by the following BNF, for $q \in \mathsf{AP}$ and $\Gamma \subseteq \mathcal{Ag}$:

$\varphi ::= q | \neg \varphi | \varphi \land \varphi | \langle \Gamma \rangle X \varphi | \langle \Gamma \rangle (\varphi U \varphi) | \langle \Gamma \rangle G \varphi$

As customary, a formula $\langle \Gamma \rangle \Phi$ is read as ‘the agents in
coalition $\Gamma$ have a (collective) strategy to achieve $\Phi$’. The
meaning of ATL operators ‘next’ $X$, ‘until’ $U$, and ‘always’ $G$
is standard. For a set $\Gamma \subseteq \mathcal{Ag}$ of agents, we define $\Gamma$-formulas
as the formulas where $\Gamma$ is the only coalition appearing in
ATL modalities.

Since the behaviour of agents in interpreted systems de-
pends only on their local state, to provide an interpretation
of ATL formulas on IS it is appropriate to consider a notion
of strategy that takes into account local states only.

Definition 4 (Uniform Strategy). A (uniform, mem-
oryless) strategy for agent $i \in \mathcal{Ag}$ is a function $f_i : L_i \to
\mathsf{Act}_i$ such that for every local state $l_i \in L_i$, $f_i(l_i) \in P_i(l_i)$.

Intuitively, Def. 4 prescribes that in all global states $s, s'$,
in which the local state of agent $i$ is the same, i.e., $s_i = s'_i$, a
uniform strategy $f_i$ returns the same action $f_i(s_i) = f_i(s'_i)$,
in line with the standard account on uniform strategies [24].
Notice that we focus on memoryless (position) strategy,
for which only the current local state is relevant to select
the action. The results below can in principle be extended
to memoryfull (perfect recall) strategies, but the technical details are more cumbersome. We leave this extension for future work.

Given an IS $M$, a path $p$ is any infinite sequence $s_1s_2\ldots$ of global states, in which $p^i$ denotes the $i$-th element $s_i$. Further, for a set $F_1 = \{f_1 | i \in \Gamma\}$ of strategies, a path $p$ is $F_1$-compatible iff for every $j \geq 1$, $p^{j+1} = T(p^j, a)$ holds for some joint action $a \in ACT$ such that for every $i \in \Gamma$, $a_i = f_i(p^j, i)$. Let $out(s, F_1)$ be set of all $F_1$-compatible paths starting from $s$.

We can now assign a meaning to ATL formulas on IS.

**Definition 5 (Satisfaction).** The 2-valued satisfaction relation $\models^2$ for an IS $M$, state $s \in S$, and ATL formula $\phi$ is defined as follows (clauses for propositional connectives are immediate and thus omitted):

- $(M, s) \models^2 q$ iff $I_1(s, q) = tt$
- $(M, s) \models^2 \langle \Gamma \rangle X \phi$ iff for some $F_1$, for all $p \in out(s, F_1)$, $(M, p^2) \models^2 \phi$
- $(M, s) \models^2 \langle \Gamma \rangle G \phi$ iff for some $F_1$, for all $p \in out(s, F_1)$, for all $k \geq 1$, $(M, p^k) \models^2 \phi$
- $(M, s) \models^2 \langle \Gamma \rangle \phi U \psi'$ iff for some $F_1$, for all $p \in out(s, F_1)$, for some $k \geq 1$, $(M, p^k) \models^2 \phi'$ and for all $j$, $1 \leq j < k \Rightarrow (M, p^j) \models^2 \psi$

A formula $\phi$ is true in an IS $M$, or $M \models^2 \phi$, iff for all initial states $s \in I$, $(M, s) \models^2 \phi$.

**Definition 6 (Model Checking).** Given an IS $M$ and an ATL formula $\phi$, the model checking problem amounts to determining whether $M \models^2 \phi$.

Observe that the semantics provided in Def. 5 is in line with similar proposals for interpreting ATL in contexts of imperfect information [24, 6]. In particular, the strategies generating paths are uniform, i.e., they only take into account the local state of agents and provide the same action in indistinguishable states. Further, here we adopt the objective interpretation of ATL with imperfect information, as opposed to the subjective interpretation whereby ATL operators are evaluated at state $s$ against paths $p \in out(s', F_1)$, for all states $s'$ indistinguishable from $s$ for the agents in $\Gamma$, i.e., $out(s, F_1) = \bigcup_{s' \equiv s} out(s', F_1)$ [33]. Hereafter we deal primarily with the objective interpretation and discuss briefly the necessary modification concerning the subjective semantics.

The features of the above semantics entail that well-known principles, which are valid under perfect information, fail under imperfect information. Notably, the fixed-point characterizations of ATL operators $\langle \Gamma \rangle G$, $\langle \Gamma \rangle F$, and $\langle \Gamma \rangle U$ do not hold (see [20] for counterexamples), thus making the model checking procedures based on them unusable and the verification task strictly more complex. In particular, ATL with imperfect information is not subsumed by the alternating $\mu$-calculus [7, 14], as it is the case for perfect information, and therefore the techniques developed in [3] do not apply to the present setting. An original methodology will be developed in the next section.

We conclude this formal presentation with a toy example, whose unique purpose it to illustrate the machinery above.

### 2.1 The Train Gate Controller Scenario

We discuss a variant of the Train Gate Controller (TGC) scenario [21]. In the version we consider two trains, $T_0$ and $T_1$, try to access a tunnel, whose entrance is managed by controller $C$. Controller $C$ allows only one train in the tunnel at any time and keeps track of how many times each train has accessed the tunnel by using counters $c_0$ and $c_1$. In case the trains are both requesting access, controller $C$ grants it to the train with smaller count. If $c_0 = c_1$, then the choice is non-deterministic. By doing so, controller $C$ tries to keep a fair access to the tunnel.

The formal specification for each train can be given as detailed in Def. 1, as follows:

**Definition 7 (Train).** For $i \in \{0, 1\}$, each train $T_i = (L_i, Act_i, P_i, t_i)$ is defined as

- $L_i = \{\text{away, wait, tunnel}\}$
- $Act_i = \{\text{approach, enter, leave, skip}\}$
- $P_i(\text{away}) = \{\text{approach, skip}\}$
- $P_i(\text{wait}) = \{\text{enter}\}$
- $P_i(\text{tunnel}) = \{\text{leave}\}$

- $t_i(\text{away}, a) = \text{wait if } a_i = \text{approach}$
- $t_i(\text{away}, a) = \text{away if } a_i = \text{skip}$
- $t_i(\text{wait}, a) = \text{tunnel if } a_i = \text{enter and } c_C = \text{enter}$
- $t_i(\text{tunnel}, a) = \text{away if } a_i = \text{leave and } c_C = \text{leave}$

By Def. 7 each train tries to access the tunnel in an execution loop. In particular, the entering and exiting from the tunnel are synchronised with controller $C$.

To model controller $C$ we make use of four variables: $\text{status}$, with values green $g$ and red $r$, counters $c_0$ and $c_1$ with range 1, and variable $q$ with values 0, 1, and 2. Hereafter we use primed variables to denote their values at the next state. Variables not explicitly mentioned remain unchanged.

**Definition 8 (Controller).** Controller $C = (L_C, Act_C, P_C, t_C)$ is defined as

- $L_C$ is the set of tuples $(\text{status}, c_0, c_1, q)$ s.t. $\text{Rand}(\text{status}) = (g, r)$, $\text{Rand}(c_0) = \text{Rand}(c_1) = N$, and $\text{Rand}(q) = \{0, 1, 2\}$
- $Act_C = \{\text{enter}_0, \text{leave}_0, \text{enter}_1, \text{leave}_1\}$
- for every $l \in L_C$ and $i \in \{0, 1\}$, $P_C(l) = \{\text{enter}_i\}$ if status = $g$ and $q = 2$ and $c_i < c_{i-1}$, $P_C(l) = \{\text{enter}_0, \text{enter}_1\}$ if either status = $g$ and $q \neq 2$, or status = $g$ and $q = 2$ and $c_0 = c_1$, $P_C(l) = \{\text{leave}_0, \text{leave}_1\}$ if status = $r$
- $t_C(l, a) = l'$ holds whenever status = $g$, $a_C = \text{enter}_i$, $a_i = \text{enter, status'} = r$, $c_i' = c_i + 1$, and $q' = q - 1$; or status = $r$, $a_C = \text{leave}_i$, $a_i = \text{leave, status'} = g$; or status = $g$, $a_i = \text{approach, and } q' = q + 1$.

By Def. 8, controller $C$ manages the entrance to the tunnel with semaphore $\text{status}$, which is synchronised with the entering and exiting of trains. Also, counters $c_0$ and $c_1$ keep track of how many times each train has accessed the tunnel. Finally, variables $q$ (for queue) registers the number of trains in state $\text{wait}$. Notice that, since counters $c_0$ and $c_1$ take values in $N$, controller $C$ has infinitely many local states.

We now introduce the IS corresponding to this TGC scenario. In the following we consider a set $AP$ containing atoms $c_0 < c_1$, $c_0 > c_1$, $c_0 = c_1$, in_tunnel$_0$, and in_tunnel$_1$, to be interpreted below.
Definition 9 (IS). The IS for the Train Gate Controller scenario is the tuple $M_{\text{TGC}} = (\text{Ag}, I, T, \Pi)$ such that

- $\text{Ag} = \{T_0, T_1, C\}$;
- $I$ contains only the tuple (away, away, (g, 0, 0, 0)), i.e., both trains are away at the beginning, the semaphore is green, both counters are equal to 0, and no train is waiting;
- $T$ is the composition of the local transition functions $t_i$, for $i \in \text{Ag}$, as per Def. 2;
- the labelling function $\Pi$ is such that a comparison $c_0 \times c_1$ is true at state $s$ iff the corresponding relation holds between the counters. Also, $\Pi(s, \text{in}_\text{tunnel}) = \text{tt}$ iff $s_i = \text{tunnel}.$

We remark that, since controller $C$ has infinitely many local states, the IS $M_{\text{TGC}}$ is an infinite-state system.

Then, we can check, for instance, that in state $s = (\text{wait}, \text{wait}, (g, 0, 0, 0))$, for enabled joint action $a = (\text{enter}, \text{enter}, \text{enter}_r)$, we have transition $T(s, a) = (\text{tunnel}, \text{wait}, (r, 1, 0, 1)).$ Therefore, we can already check that in state $s$ controller $C$ can (has a strategy to) let train $T_0$ enter the tunnel, thus making $c_0$ greater than $c_1$:

$$(M, s) \models \langle C \rangle X(c_0 > c_1 \land \text{in}_\text{tunnel}_0)$$

Moreover, in ATL we can specify various behaviours and strategic abilities of controller $C$ and trains $T_0$ and $T_1$. As an example, we might state that, whenever the counter for train $T_0$ is smaller, she has a strategy to eventually make the counters equal.

$$AG((c_0 < c_1) \rightarrow \langle T_0 \rangle F(c_0 = c_1))$$

where $AG$ is shorthand for $\langle \emptyset \rangle G$.

Further, we can ask whether train $T_0$ has a strategy to engage the tunnel “infinitely often”:

$$\langle T_0 \rangle G(\langle T_0 \rangle F \text{ in}_\text{tunnel}_0)$$

Finally, we can express that, whenever the counters are equal, train $T_0$ and controller $C$ have a strategy so that at the next step train $T_1$ has a strategy to make the counters equal:

$$AG((c_0 = c_1) \rightarrow \langle C, T_0 \rangle X(T_1) X(c_0 = c_1))$$

Intuitively, formula (2) is true by the strategy whereby train $T_0$ keeps on engaging the tunnel; while the same strategy is not sufficient to make formula (1) true. Indeed, (1) is false in general. Also, (3) is true whenever controller $C$ grants $T_0$’s request, and then $T_1$ engage the tunnel.

In this TGC scenario we checked formulas (1)-(3) manually. However, we aim at developing an automated verification procedure capable of dealing with infinite-state systems, such as the IS $M_{\text{TGC}}$, for which standard model checking techniques cannot be immediately applied.

3. 3-VALUED SEMANTICS FOR ATL WITH IMPERFECT INFORMATION

This section is devoted to introducing a generalisation of the notion of agent in Def. 1 in terms of over- and under-approximations of their strategic abilities. Then, we present a 3-valued semantics for ATL, and show that this conservatively extends the 2-valued version in Section 2.

Definition 10 (Generalised Agent). A (generalised) agent is a tuple $i = (L, Act, P_{\text{may}}, P_{\text{must}}, t_{\text{may}}, t_{\text{must}})$ such that

- sets $L$ of local states and $Act$ of actions are given as in Def. 1;
- $P_{\text{may}}$ and $P_{\text{must}}$ are protocol functions from $L$ to $2^{Act}$;
- $t_{\text{may}}$ and $t_{\text{must}}$ are local transition relations defined on $L \times ACT \times L$ such that, for $x \in \{\text{may}, \text{must}\}$, $l \in L,$ and $a \in ACT$, transition $t_x(l, a, l')$ is defined for some $l' \in L$ iff $a_i \in P^x(l).$

By Def. 10 generalised agents have the same components of standard agents, but differently from Def. 1, we now distinguish between $\text{may}$ and $\text{must}$ protocols and transitions. This distinction can be understood in terms of approximations of the agents’ abilities. Intuitively, the $\text{may}$ protocol and transitions represent an over-approximation of these abilities, while $\text{must}$ components can be seen as an under-approximation. Here the distinction between $\text{must}$ and $\text{may}$ components and related terminology derive from the literature [5, 19]. The intuitive meaning will become apparent in Section 5, where we define agent abstractions. Moreover, the standard agents in Def. 1 are the limit case in which under- and over-approximations coincide, or formally, $p_{\text{may}} = p_{\text{must}}$ and $t_{\text{may}} = t_{\text{must}}$ is a function.

Hereafter we simply refer to agents, the context will distinguish between generalised and standard agents. Further, we introduce a generalisation of the interpreted systems in Def. 2, in which atoms can be assigned a third truth value uu for ‘undefined’. In what follows, for $x = \text{may}$ (resp. $\text{must}$), $\tau = \text{must}$ (resp. $\text{may}$).

Definition 11 (Generalised IS). A (generalised) interpreted system is a tuple $M = (Ag, I, T, \Pi)$ such that

- $Ag$ and $I$ are given as in Def. 2;
- $T : G \times ACT \rightarrow G$ is the global transition function such that $s' = T(s, a)$ iff for all $i \in Ag$, either $s'_i = t_{\text{may}}(s_i, a)$ or $s'_i = t_{\text{must}}(s_i, a)$;
- $\Pi : G \times AX \rightarrow \{\text{tt}, \text{ff}, \text{uu}\}$ is the labelling function.

The undefined value uu can be interpreted in various ways, for instance, unknown, unspecified, or inconsistent, depending on the application in hand. We do not discuss this matter further, as it is not relevant for our technical contribution. We say that the truth value $\tau$ is defined whenever $\tau \neq \text{uu}$. If every agent in $Ag$ is standard and the truth value of every atom is defined, then we say that the IS is standard as well, and we are back to Def. 2.

Given an IS $M$ and a coalition $\Gamma \subseteq Ag$ of agents, we define an indexed global transition relation $T_{\Gamma} \subseteq G \times ACT \times G$ so that $T_{\Gamma}(s, a, s')$ holds iff (i) for all $i \in \Gamma$, $t_{\text{must}}(s_i, a, s'_i)$; and (ii) for all $i \in \Gamma$, $t_{\text{may}}(s_i, a, s'_i)$. Intuitively, the transition relation $T_{\Gamma}$ exhibits a conservative stance on the strategic abilities of coalition $\Gamma$ (by considering the under-approximation $t_{\text{must}}$), and an optimistic view of the abilities of adversarial $\Gamma$. Finally, we introduce the set $S \subseteq G$ of global states reachable from set $I$ of initial states, through the transition relation $T$.

Next, as it was the case for $\text{must}$ and $\text{may}$ protocols and transitions, we introduce also $\text{must}$- and $\text{may}$-strategies.
Definition 12 (Uniform x-Strategy). For \( x \in \{\text{may}, \text{must}\} \), a (uniform, memoryless) x-strategy for agent \( i \in \text{Ag} \) is a function \( f^{x}_{i} : L_{i} \rightarrow \text{Act}_{i} \), such that for every local state \( l \in L_{i} \), \( f^{x}_{i}(l) \in \text{P}^{x}_{i}(l) \).

Here we distinguish between may and must strategies to over- and under-approximate the strategic abilities of agents. Again, the distinction collapse in the case of standard agents. Moreover, these notions can be extended to the case of memoryful (perfect recall) strategies; again, we leave this for future work.

For a set \( \text{P}^{\text{must}}_{i} = \{ f_{i}^{\text{must}} \mid i \in \Gamma \} \) of strategies, a path \( p \) is \( \text{P}^{\text{must}}_{i} \)-compatible if for every \( j \geq 1 \), \( T_{i}(p^{j}, a, p^{j+1}) \) holds for some joint action \( a \in \text{ACT} \) such that for every \( i \in \Gamma \), \( a.i = f_{i}^{\text{must}}(p^{j}.i) \). Symmetrically, a path is \( \text{P}^{\text{may}}_{i} \)-compatible if for every \( j \geq 1 \), \( T_{i}(p^{j}, a, p^{j+1}) \) holds for some joint action \( a \in \text{ACT} \) such that for every \( i \in \Gamma \), \( a.i = f_{i}^{\text{may}}(p^{j}.i) \). For \( x \in \{ \text{may, must} \} \), let \( \text{out}(s, \text{P}^{x}_{i}) \) be the set of all \( \text{P}^{x}_{i} \)-compatible paths starting from \( s \).

Definition 13 (Satisfaction). The 3-valued satisfaction relation \( \models^{3} \) for an IS \( M \), states \( s \in S \), and ATL formula \( \phi \) is defined as in Fig. 1. In particular, in all other cases the value of \( \phi \) is undefined (uu).

We observe that, in the clauses for ATL operators, must-strategies are used to make formulas true, while may-strategies are used to falsify them. Further, we can introduce a 3-valued, subjective interpretation of ATL by defining in Def. 13 the set \( \text{out}(s, \text{P}^{x}_{i}) \) of outcomes as \( \bigcup_{i \in \Gamma \cdotp s \in S} \text{out}(s, \text{P}^{x}_{i}) \), for \( x \in \{ \text{may, must} \} \). Finally, \( (M, s) \models^{3} \phi \) is tt (resp. ff) iff for all (resp. some) \( s \in I, (M, s) \models^{3} \phi \) is tt (resp. ff). Otherwise, \( (M, s) \models^{3} \phi \) is uu.

We conclude the presentation of the 3-valued semantics with the following result, which shows that for standard IS it coincides with the 2-valued version. This is in contrast with recent proposals in this area [30, 29].

Proposition 1. In every standard IS \( M \), for every state \( s \in S \) and ATL formula \( \phi \), the truth value \( (M, s) \models^{3} \phi \) is always defined and

\[
\begin{align*}
(M, s) \models^{3} \phi & = \text{tt} \text{ iff } (M, s) \models^{2} \phi \\
(M, s) \models^{3} \phi & = \text{ff} \text{ iff } (M, s) \models^{2} \phi
\end{align*}
\]

Proof. The proof is by induction on \( \phi \). If \( \phi \) is an atom \( p \), then \( (M, s) \models^{3} p \) is defined as \( \Pi(s, p) \). Further, \( (M, s) \models^{3} \neg p \) is tt if \( \Pi(s, \neg p) = \text{tt} \), iff \( (M, s) \models^{2} \neg p \). The case for \( (M, s) \models^{3} X \psi \) is similar. The cases for propositional connectives are immediate by the induction hypothesis. As for \( \phi = (\Gamma')X\psi \), since for every \( i \in \text{Ag} \), \( P^{\text{may}} = P^{\text{must}} \) and \( l^{\text{may}} = l^{\text{must}} \) defined by the distinction between may- and must-strategies collapse. Hence, \( (M, s) \models^{3} (\Gamma')X\psi \) is tt iff for some strategy \( F_{i} \), for all \( p \in \text{out}(s, F_{i}) \), \( (M, p^{i}) \models^{3} \psi \) is tt. By induction hypothesis we obtain that for some \( F_{i} \), for all \( p \in \text{out}(s, F_{i}) \), \( (M, p^{i}) \models^{2} \psi \), that is, \( (M, s) \models^{3} \phi \). The case for \( (M, s) \models^{3} (\Gamma')X\psi \) is similar. To prove that the truth value of \( \phi \) is defined, suppose that \( (M, s) \models^{3} (\Gamma')X\psi \) is not tt. That is, for every strategy \( F_{i} \), for some \( p \in \text{out}(s, F_{i}) \), \( (M, p^{i}) \models^{3} \psi \) is not tt. By induction hypothesis, the truth value of \( \psi \) is defined at each \( p^{i} \), therefore \( (M, p^{i}) \models^{3} \psi \) is ff. As a consequence of the satisfaction clause for falsehood, \( (M, s) \models^{3} \phi \) is ff. The case for \( (M, s) \models^{3} (\Gamma')X\phi \) is symmetric. The proofs for operators \( G \) and \( U \) are similar and therefore omitted.
and $\preceq^\text{may}_H$ are partial orders, i.e., reflexive and transitive relations, whenever $H$ is such. Further, since agents in Def. 1 are a particular case of generalised agents (for which may- and must- protocols and transitions coincide), Def. 15 defines a relation of simulation for (non-generalised) agents as well. Hereafter, given a set $\Gamma \subseteq \text{Ag}$ of agents, we use $\text{Ag}^\L_\Gamma = \{\langle i, \tau \rangle \mid i \preceq^\text{may}_J \tau, i \in \Gamma \} \cup \{\langle i, \tau \rangle \mid i \preceq^\text{may}_J \tau, i \in \Gamma \}$ to refer to the set of must- and may-simulating agents $i'$, exactly one for each $i \in \text{Ag}$. Finally, a global state $\Gamma'$ defined on $\text{Ag}^\L_\Gamma$ simulates $s$ on $\Gamma$ (w.r.t. $H$), or $s <^\text{may}_H s'$, if (i) for every $i \in \Gamma$, $s <^H s_i$; (ii) for every $i \in \Gamma$, $s_i <^H s_i$

**DEFINITION 16 (IS SIMULATION).** Given a set $\Gamma \subseteq \text{Ag}$ of agents, an IS $M' = \langle \text{Ag}^\L_\Gamma, \Gamma', \Pi' \rangle$ $\Gamma$-simulates IS $M = \langle \text{Ag}, I, T, \Pi \rangle$ (w.r.t. $H$), or $M \preceq^H M'$, iff

1. $\text{Ag}^\L_\Gamma$ defined as above is the set of simulations for agents in $\text{Ag}$;
2. for every $s \in I$, $s <^H s'$ for some $s' \in I'$;
3. for every $s \in S$, $s' \in S'$, if $s <^H s'$ and $\Pi'(s', p) = t$, for $t \in \{tt, ff\}$, then $\Pi(s, p) = t$.

By Def. 16, if $M'$ $\Gamma$-simulates $M$ then every initial state in $M$ is $\Gamma$-simulated by some initial state in $M'$, and defined truth values of atoms are preserved from $M'$ to $M$. Clearly, IS simulations are also partial orders, provided that $H$ is. We observe that simulations for interpreted systems are indexed to sets of agents. Indeed this is normally the case for alternating simulations, as studied for instance in [2, 3].

The main result of this section shows that IS simulations preserve defined truth values of ATL formulas. To prove this, we need the following auxiliary lemma.

**LEMMA 1.** If $s <^H s'$ then

1. for every strategy $F^\text{must}_{i'}$, there exists strategy $F^\text{must}_i$ such that for all $p \in \text{out}(s, F^\text{must}_{i'})$, there exists $p' \in \text{out}(s', F^\text{must}_i)$ such that $p^k <^H p'^k$ for every $k \geq 1$;
2. for every strategy $F^\text{may}_{i'}$, there exists strategy $F^\text{may}_i$ such that for all $p' \in \text{out}(s', F^\text{may}_{i'})$, there exists $p \in \text{out}(s, F^\text{may}_i)$ such that $p^k <^H p'^k$ for every $k \geq 1$.

**PROOF.** As regards (1), suppose that $s <^H s'$ and let $F^\text{must}_{i'}$ be a strategy for primed coalition $\Gamma'$. We inductively define $F^\text{must}_i$ on the length $n$ of paths, and prove that it satisfies the statement of the lemma. For $n = 1$, for $i' \in \Gamma'$, consider $f^\text{must}(s') \in F^\text{pre}_i\text{must}(s')$ and $a_i \in F^\text{pre}_i\text{must}(s')$ such that $H_i(a_i, f^\text{must}(s'))$: the existence of such $a_i$ is guaranteed by Def. 14.2 and 15.2. We set $f^\text{pre}_i\text{must}(s') = a_i$ for $i \in I$. Further, if $T_i(s, a, p^2)$ for a extending $F^\text{pre}_i\text{must}(s')$, then (i) for all $i \in \Gamma$, $t^\text{pre}_i(s_i, a, p^2)$; and (ii) for all $i \in \Gamma$, $t^\text{pre}_i(s_i, a, p^2)$. Then, by definition of local simulation, if $t^\text{pre}_i(s_i, a, p^2)$ for $i \in I$, then for some $i' \in I$, $t^\text{pre}_i(s_i', a', p')$ for $a'$ extending $F^\text{pre}_i\text{must}(s_i')$ in particular. Moreover, $s <^H s'$ implies $s_i <^H s'_{i'}$ for every $i' \in \Gamma$. Again, by definition of local simulation, if $t^\text{pre}_i(s_i, a, p^2)$ for $i \in I$, then for some $s_i' \in S_i$, $t^\text{pre}_i(s_i', a', p')$ for $a'$ extending $F^\text{pre}_i\text{must}(s_i')$. Intuitively, Lemma 1 says that in $\Gamma$-similar states, a must-simulation $i' \in \Gamma'$ simulates may-strategies of $i' \in \Gamma$, while her must-strategies are simulated by $i$. This can appear counterintuitive, but notice that in order to simulate may-strategies an agent must be capable of simulating must-transitions. Symmetrically for may-simulations and must-strategies.

By Lemma 1 we can prove our main preservation result.

**THEOREM 1.** If $M <^H M'$, $s <^H s'$ and $\tau \in \{tt, ff\}$, then for every $\Gamma$-formula $\phi$, $(M', s') \models_3 \phi \Rightarrow \tau \text{ implies } (M, s) \models_3 \phi \Rightarrow \tau$.

**PROOF.** The proof is by induction on the structure of $\phi$. The base case for atoms follows by Def. 16.3 of IS simulation.


Figure 1: The 3-valued satisfaction relation $\models_3$ for an IS $M$, state $s \in S$, and ATL formula $\phi$. 
By Theorem 1 we immediately obtain the following corollary.

**Corollary 1.** If $M \succeq^\Gamma M'$, then for every $\Gamma$-formula $\phi$,

$$(M' \models^3 \phi) = tt \quad \text{implies} \quad (M \models^3 \phi) = tt$$

By Corollary 1 a positive answer to the model checking problem for simulating IS $M'$ entails a positive answer for simulated IS $M$ as well. On the other hand, if $\phi$ is false in $M'$, nothing can be derived about $M$.

We conclude this section by briefly discussing the modifications necessary to account for the subjective interpretation. Specifically, we require that $s \preceq^\Gamma s'$ implies (i) for every $v \in S$, if $s_i = v_i$ for some $i \in \Gamma$, then $s'_i = v'_i$ for some $i' \in S'$ such that $v \preceq^\Gamma v'$; and (ii) for every $v' \in S'$, if $s'_i = v'_i$ for some $i' \in \Gamma$, then $s_i = v_i$ for some $v \in S$ such that $v \preceq^\Gamma v'$. We state without proof that this modification is sufficient to prove Lemma 1, and then Theorem 1, for the subjective interpretation. The proofs follow similar lines of reasoning.

In the next section we apply the results above to derive finite, 3-valued abstractions of (possibly infinite) interpreted systems that preserve the defined truth value of ATL formulas.

## 5. ABSTRACTION

In this section we define an agent-based notion of abstraction for interpreted systems, indexed to a set $\Gamma$ of agents. Then, we prove that abstractions $\Gamma$-simulate the original, concrete IS. Hence, they can be used in the verification procedure to make specific model checking instances more amenable.

To begin with, for every agent $i \in Ag$, let $\approx_i$ be an equivalence relation on $L_i$, and $\llbracket [l] \rrbracket = \{l' \in L_i | l', l \approx_i l \}$ be the equivalence class of $l$ according to $\approx_i$. We first introduce abstractions for agents.

**Definition 17 (Abstract Agent).** Given an agent $i = (L_i, Act_i, P_i^{may}, P_i^{must}, t_i^{may}, t_i^{must})$ in $\Gamma$, the abstract agent $i' = (L', Act', P_i'^{may}, P_i'^{must}, t_i'^{may}, t_i'^{must})$ is defined such that

1. $L' = \{\llbracket [l] \rrbracket | l \in L\}$;
2. $Act' = Act$;
3. $t_i'^{may}(l', a, l'_2)$ iff for every $l_1 \in l'_1$, $t_i^{may}(l_1, a, l_2)$ for some $l_2 \in l'_2$; and $t_i'^{must}(l', a, l'_2)$ iff for some $l_1 \in l'_1$, $l_2 \in l'_2$, $t_i^{must}(l_1, a, l_2)$;
4. for every $l' \in L'$, $P_i'^{may}(l') = \{a' \in Act | t_i'^{may}(l', a', l'_2) \}$ for some $l_2 \in l'_2$ and $P_i'^{must}(l') = \bigcup_{l' \in L'} P_i'^{must}(l)$.

On the other hand, if $i \in \Gamma$, then $i'$ is defined such that (1) and (2) holds and

3'. $t_i'^{may}(l'_1, a, l'_2)$ iff for some $l_1 \in l'_1$, $l_2 \in l'_2$, $t_i^{may}(l_1, a, l_2)$; and $t_i'^{must}(l'_1, a, l'_2)$ iff for every $l_1 \in l'_1$, $t_i^{must}(l_1, a, l_2)$ for some $l_2 \in l'_2$;
4'. for every $l' \in L'$, $P_i'^{may}(l') = \bigcup_{l' \in L'} P_i'^{may}(l)$ and $P_i'^{must}(l') = \{a' \in Act | t_i'^{must}(l', a', l'_2) \}$ for some $l_2 \in l'_2$.

By Def. 17 the local states of abstract agent $i'$ are the equivalence classes of local states for $i$; while the set of actions is the same. For $i \in \Gamma$, the $may$ (resp. $must$) protocol in an abstract local state $l'$ includes all actions that are enabled in some (resp. all) concrete states $l \in l'$; while a $may$ (resp. $must$) transition holds between abstract local states $l_1'$ and $l_2'$ iff from some (resp. all) concrete states $l_1 \in l_1'$, there is a transition to some $l_2 \in l_2'$.

We can now introduce the notion of abstract interpreted system.

**Definition 18 (Abstract IS).** Given an IS $M = \langle Ag, I, T, \Pi \rangle$ and a set $\Gamma$ of agents, the abstract IS $M^\Gamma = \langle Ag^\Gamma, I^\Gamma, T^\Gamma, \Pi^\Gamma \rangle$ is defined as

- $Ag^\Gamma$ is the set of abstractions of agents in $Ag$;
- $I^\Gamma = \{\llbracket [l] \rrbracket, \ldots, \llbracket [l_{Ag}] \rrbracket | [l_i, \ldots, l_{Ag}] \in I\}$;
- for $\tau \in \{tt, ff\}$ and $p \in AP$, $\Pi^\Gamma(\llbracket [l] \rrbracket, \ldots, \llbracket [l_{Ag}] \rrbracket, p) = \tau$ iff for all $l_i' \in \llbracket [l] \rrbracket$, $\Pi(\llbracket [l_i] \rrbracket, \ldots, \llbracket [l_{Ag}] \rrbracket, p) = \tau$; otherwise, $\Pi^\Gamma(\llbracket [l] \rrbracket, \ldots, \llbracket [l_{Ag}] \rrbracket, p) = uu$.
By Def. 18 an atom $p$ receives value true (resp. false) in an abstract state $s'$ iff it receives such value in all corresponding concrete states $s$. Since the equivalence relations $\equiv_i$ are not assumed to respect the assignment $\Pi$, some atoms might become undefined in some abstract states. Thus, the abstraction of a standard IS is not necessarily standard itself. However, here the aim is to trade definiteness with a smaller state space in $M^\Delta$ in comparison to $M$.

We can now prove the main result of this section.

**Theorem 2.** The abstraction $M^\Delta \Gamma$-simulates the interpreted system $M$.

**Proof.** By Lemma 2, every abstract agent $i' \in Ag^\Delta$ simulates agent $i \in Ag$. Further, for every $s \in I$, $s \equiv^\Delta s'$ for $s' = ([s_1], \ldots, [s_{|Ag|}]$. Finally, for $s \in M$, $s' \in M'$, if $s \equiv^\Delta s'$ and $\Pi'(s', p) = \tau$, then for every $l_i \in s'_i$, $\Pi((l_1, \ldots, l_{|Ag|}), p) = \tau$. In particular, $\Pi((s_1, \ldots, s_{|Ag|}), p) = \tau$.

As an immediate consequence of Theorem 1 and 2, we obtain the following result.

**Corollary 2.** If $M^\Delta$ is the abstraction of IS $M$, $s \in s'$, and $\tau \in \{|t, ff\}$, then for every $\Gamma$-formula $\phi$,

$$((M', s') \models^\Delta \phi) = \tau \implies ((M, s) \models^\Delta \phi) = \tau$$

By Corollary 2, under specific conditions, we can transfer the (defined) verification result from abstraction $M'$ to concrete IS $M$. This result is of particular interest in all cases where $M$ is infinite, while $M'$ is finite, as in the following toy example. We conclude by stating that, with minor modifications, the abstraction procedure above can be applied to the subjective semantics as well.

### 5.1 The Abstract Train Gate Controller

Here we show how we can define a 3-valued abstraction for the infinite IS $MTGC$ in Section 2.1 that has the interesting feature of being finite. Specifically, by borrowing ideas from predicate abstraction for IS [30], we say that local states $l$ and $l'$ for agent $i$ are equivalent iff they both satisfy the same local predicates. Then, notice that the only infinite component of the TGC scenario is the controller $C$ itself, while both trains are finite. Further, in specifications (1) and (2) only train $T_0$ appears in ATL modalities, and therefore we set $\Gamma = \{T_0\}$. Now notice that each agent must- and may-simulates herself and both $T_0$ and $T_1$ are finite. Hence, we only need to define an abstract controller $C^A$ that may-simulates $C$. To do so, we observe that the relevant predicates, appearing in the protocols and specifications, concern the values of $c_0$ and $c_1$ being either equal ($c_0 = c_1$) or one greater than the other ($c_0 < c_1$ and $c_0 > c_1$). Then, we introduce abstraction $C^A$ for controller $C$ in Def. 8, according to Def. 17, as follows:

**Definition 19.** (Abstraction $C^A$). The abstract controller $C^A = (L^C_A, \text{Act}^C_A, P^A_C, t^A_C)$ is given as:

- $L^C_A$ is the set of tuples $\langle \text{status}, \text{comp}, q \rangle$ such that $q$ and $\text{comp}$ are defined as for controller $C$, while $\text{Ran}(\text{comp}) = \{c_0 < c_1, c_0 > c_1, c_0 = c_1\}$
- The actions are the same as for controller $C$: $\text{Act}^C_A = \text{Act}_C$
- The may- and must-protocol coincide: for every $l \in L^C_A$ and $i \in \{0, 1\}$, $P^A_C(l) = \{\text{enter}_i\}$ if status = $g$, $q = 2$ and $\text{comp} = \{c_0 < c_{i-1}\}$;
- $P^A_C(l) = \{\text{enter}_0, \text{enter}_1\}$ if either status = $g$ and $q \neq 2$, or status = $g$, $q = 2$ and $\text{comp} = \{c_0 = c_1\}$;
- $P^A_C(l) = \{\text{leave}_0, \text{leave}_1\}$ if $\text{status} = r$;

- we have that for abstractions of standard agents we have $t'^\text{must} \subseteq t'^\text{may}$. Hence, for every must-transition hereafter there is a corresponding may-transition (omitted):

$$A^\text{may}((g, c_0 < c_1, q), (enter, a_1, enter_0)), (r, c_0 < c_1, q - 1))$$
$$A^\text{may}((g, c_0 < c_1, q), (enter, a_1, enter_0)), (r, c_0 = c_1, q - 1))$$
$$A^\text{may}((g, c_0 > c_1, q), (a_0, enter, enter_1)), (r, c_0 > c_1, q - 1))$$
$$A^\text{may}((g, c_0 > c_1, q), (a_0, enter, enter_1)), (r, c_0 < c_1, q - 1))$$
$$A^\text{may}((r, \text{comp}, q), (\text{leave}, a_1, \text{leave}_0), (g, \text{comp}, q))$$
$$A^\text{may}((r, \text{comp}, q), (\text{leave}, a_1, \text{leave}_0), (g, \text{comp}, q))$$
$$A^\text{may}((g, \text{comp}, q), (\text{enter}, a_0, a_0), (c, \text{comp}, q + 1))$$
$$A^\text{may}((g, \text{comp}, q), (\text{enter}, a_0, a_0), (c, \text{comp}, q + 1))$$

The key remark here is that, while the original IS $MTGC$ has infinitely many states, its abstraction $M^\Delta_{TG}$ is finite with $\mid L_0 \times L_1 \times L^2_{\Delta} \mid = 162$ global states. Hence, we can model check formulas (1)-(3) in Section 2.1, for instance, and then, in case we obtain a defined answer, transfer the result on the original IS $MTGC$ by means of Corollary 2.

### 6. CONCLUSIONS

In this paper we have introduced a 3-valued abstraction technique for infinite-state imperfect-information MAS specified via ATL. As discussed in the paper, and as it is to be expected in 3-valued abstraction, the technique cannot always resolve the value of a specification. To address this issue, in future work we plan to investigate refinement techniques [11] in contexts of imperfect information. Further, we plan to adapt the technique to support more powerful logics for strategic reasoning under incomplete information (e.g. Strategy Logic [9]). Finally, we envisage to implement the developed procedure in a model checking tool for the verification of strategic behaviours of agents.

Acknowledgements. This research was funded by EP-SRC under grant EP/100520X. F. Belardinelli acknowledges the support of the ANR JCJC Project SVeDaS (ANR-16-CE40-0021).

### REFERENCES


