Multi-Agent Programming

Brian Logan

School of Computer Science
University of Nottingham

Midlands Graduate School
8th – 12th April 2013
Course Overview

Lecture 1: *Programming agents*
BDI model; PRS and other BDI languages

Lecture 2: *Programming multi-agent systems*
Coordination in MAS; agent communication languages & protocols; programming with obligations and prohibitions

Lecture 3: *Logics for MAS*
LTL, CTL; Rao and Georgeff’s BDI logics; Coalition Logic, ATL

Lecture 4: *Verification of MAS*
A tractable APL and BDI logic: SimpleAPL and PDL-APL
Lecture 3: Logics for MAS
Outline of this lecture

- standard temporal logics, e.g., CTL & ATL
- BDI logics
- Rao and Georgeff’s logic
- multi-agent BDI logics

Background material for this lecture:

Standard Temporal Logics
Computation Tree Logic

- agent programs are just computer programs, so we can use standard modal logics such as CTL and ATL to reason about them
- as an example, we’ll look at using CTL to reason about agents
- CTL stands for Computation Tree Logic
CTL language 1

- $A \bigcirc \phi$ means: on all paths, in the next state, $\phi$ is true
- $E \bigcirc \phi$ means: on some path, in the next state, $\phi$ is true
- in the example below, $t_0$ satisfies $E \bigcirc p$ and does not satisfy $A \bigcirc p$
CTL language 2

- \( A \Diamond \phi \) means: on all paths, eventually \( \phi \) is true
- \( E \Diamond \phi \) means: on some path, eventually \( \phi \) is true
- In the example below, \( t_0 \) satisfies \( E \Diamond q \) and does not satisfy \( A \Diamond p \); it does satisfy \( A \Diamond (p \lor q) \)
CTL language 3

- $A \Box \phi$ means: on all paths, in every state $\phi$ is true
- $E \Box \phi$ means: on some path, in every state $\phi$ is true

In the example below, $t_0$ satisfies $E \Box q$ and does not satisfy $A \Box p$
CTL language 4

- $A(\phi U \psi)$ means: on all paths, eventually $\psi$ holds and in every time point before that, $\phi$ holds
- $E(\phi U \psi)$ means: on some path, eventually $\psi$ holds and in every time point before that, $\phi$ holds
- In the example below, $t_0$ satisfies $E(qUp)$ but not $A(qUp)$ (because $p$ never becomes true on one of the paths)

![Diagram](world w1 → t0: q → t1: p → t2 → ...) and everywhere q but not p}
in addition to the temporal operators, we will have the usual propositional logic: variables $p, q, \ldots$ and boolean connectives $\neg, \land, \lor, \rightarrow, \leftrightarrow$

- a non-redundant set of temporal connectives: $E\bigcirc, E\square, EU$
- defining $A\bigcirc, E\lozenge, A\square, A\lozenge$: exercise
- non-trivial definition: $A(\phi U \psi) = \neg(E(\neg \psi U \neg(\phi \lor \psi)) \lor E\square \neg \psi)$
CTL formal semantics

- the time is branching, discrete and serial (each time point has at least one successor)
- given a world $w$ and a time point $t$, temporal formulas are evaluated as follows:
  - $M, (w, t) \models E \bigcirc \phi$ iff there exists a path $t = t_0, t_1, \ldots, t_n \ldots$ starting in $t$ such that $(w, t_1) \models \phi$
  - $M, (w, t) \models E \Box \phi$ iff there exists a path $t = t_0, t_1, \ldots, t_n \ldots$ starting in $t$ such that for every $t_i$ on the path, $(w, t_i) \models \phi$
  - $M, (w, t) \models E(\phi U \psi)$ iff there exists a path $t = t_0, t_1, \ldots, t_n \ldots$ starting in $t$ such that for some $i \geq 0$, $(w, t_i) \models \psi$, and for all $j$ such that $0 \leq j < i$, $(w, j) \models \phi$. 

Brian Logan

Multi-Agent Programming

MGS 2013
in a similar way, we can use ATL to reason about multi-agent systems
in ATL we can express that a coalition (group) of agents \( A \subseteq N \) has a strategy to enforce a temporal property, whatever the other agents in the system \( N \setminus A \) do

\[
p \in \text{Prop} \mid \neg \phi \mid \phi \rightarrow \psi \mid \langle \langle A \rangle \rangle \bigcirc \phi \mid \langle \langle A \rangle \rangle \Box \phi \mid \langle \langle A \rangle \rangle U(\phi, \psi)
\]

- \( \langle \langle A \rangle \rangle \bigcirc \phi \): the coalition \( A \) can enforce \( \phi \) in the next state
- \( \langle \langle A \rangle \rangle \Box \phi \): the coalition \( A \) can enforce that \( \phi \) always holds
- \( \langle \langle A \rangle \rangle U(\phi, \psi) \): the coalition \( A \) can enforce that \( \phi \) holds until \( \psi \) happens
Limitations of CTL & ATL

- CTL & ATL are well understood, with many high quality verification tools (model checkers etc.)
- however they ignore the propositional attitudes that distinguish agent programming languages
- to use, e.g., CTL to verify agent programs, we need to understand how beliefs, goals etc., are implemented, and verify at the level of the implementation, rather than at the level of beliefs & goals
- often increases the size of the state space (since we include unnecessary implementation detail)
- more importantly, it makes it hard to state and verify ‘BDI properties’, e.g., that beliefs and goals are consistent, or commitment properties
BDI Logics
Logics of agent programs

- we have identified some essential components of an agent programming language:
  - beliefs
  - goals (desires)
  - intentions
  - (possibly) plans
- what should their properties be and how do we go about specifying them?
- for example, what is the language of agent’s beliefs and intentions?
Examples: possible properties of beliefs and goals

- beliefs are consistent
- goals are consistent
- beliefs and goals mutually ‘consistent’
  - the agent does not have a goal to achieve $p$ if it believes that $p$ is already true (only requires beliefs about the current state)
  - the agent does not have a goal to achieve $p$ if it believes that $p$ is impossible to achieve (requires beliefs about the future)
Examples: possible properties of intentions

- Intentions are consistent
- If an agent intends to execute an action, it executes the action unless...
  - If an action is not executable, the agent drops the intention to execute it
  - If an action takes longer than a fixed timeout to complete, the agent drops the intention to execute it
- In general, when should an agent give up an intention?
What do we want the logics for?

- a logic may be useful for specifying and formalising properties of beliefs, desires and intentions
- being able to state properties precisely is useful for concentrating the mind and checking for any ‘side-effects’ of our definitions
- may also allow us to state and (automatically) verify properties of:
  - all programs written in an agent programming language
  - an agent program for all possible inputs (task environments)
  - an agent program for a given input
General shape of the logic

- it should have belief, desire and intention operators/predicates
- it should be extendable to a multi-agent setting (e.g. joint intentions, communication between agents)
- it should be able to formalise dynamics (so it needs to include temporal operators and/or an ability to talk about results of executing actions)
- should be grounded in the agent’s computation in the sense of van der Hoek, Wooldridge, *Towards a Logic of Rational Agency*, Logic Journal of the IGPL, 11 (2) 2003)
Rao & Georgeff’s logic
Rao and Georgeff’s logic

- the logical language has modal operators (for a single agent)
  - $\text{BEL}$ for belief
  - $\text{GOAL}$ for goal (or desire)
  - $\text{INTEND}$ intention
- interpreted using possible worlds models
- each possible world is a branching tree of time points
- the language also contains temporal logic operators to talk about time
- and operators to talk about execution of actions (successful or unsuccessful)
in Rao and Georgeff, each edge in the time tree is ‘labelled’ by an action executed by the agent to bring about the resulting state.

since each time point has one incoming edge, we can just as well label the points rather than the edges.

introduce a (unique) label, which says which event (action) lead to this time point and whether it succeeded or failed ($\text{succ}(e)$ or $\text{fail}(e)$).
Models

- \( \text{BDI}_{\text{CTL}} \) is interpreted over models \( M = (W, T, R, E, B, G, I, L) \) where
  - \( W \) is a non-empty set of possible worlds
  - \( T \) is a non-empty set of time points
  - \( R \) is a serial binary relation on \( T \), such that for each \( w \in W \), \( (T^w, R\,|\,T^w) \) is an infinite tree (where \( T^w \) is the set of time points in \( w \) and \( R\,|\,T^w \) a restriction of \( R \) to \( T^w \))
  - \( E \) is a set of events or primitive actions
  - \( B, G, I \) are accessibility relations (to come)
  - \( L \) is a labelling (valuation function) of time points with propositional variables and events (to come)
Truth definition for propositional variables & events

- $M, (w, t) \models p$ iff $p \in L(t)$
- $M, (w, t) \models succ(e)$ iff $succ(e) \in L(t)$
- $M, (w, t) \models fail(e)$ iff $fail(e) \in L(t)$
- constraint: at most one event label $e$ per time point, and exactly one of $succ(e)$ or $fail(e)$. 
Truth definition for beliefs

- $BEL \phi$ is true in a possible world $w$ at time point $t$ if $\phi$ is true in all belief-accessible worlds $w'$ at $t$
- we assume that if $w'$ is belief-accessible from $(w, t)$ then $t$ exists in $w'$
- accessibility relation for $BEL$: $B \subseteq W \times T \times W$
  - note that this is the same as a binary relation on $W \times T$: $B(((w, t), (w', t)))$ for $B(w, t, w')$
- $B$ is serial, transitive and Euclidean
  \[ (\forall w \forall t \forall v \forall u (B(w, t, v) \land B(w, t, u) \rightarrow B(v, t, u))) \]
- $M, (w, t) \models BEL \phi$ iff for all $w'$ such that $B(w, t, w')$, $M, (w', t) \models \phi$
Truth definition for goals

- \textit{GOAL} \phi is true in a possible world \( w \) at time point \( t \) if \( \phi \) is true in all goal-accessible worlds \( w' \) at point \( t \)
- we assume that if \( w' \) is goal-accessible from \((w, t)\) then \( t \) exists in \( w' \)
- accessibility relation for \textit{GOAL}: \( G \subseteq W \times T \times W \)
- \( G \) is serial
- \( M, (w, t) \models \textit{GOAL} \phi \) iff for all \( w' \) such that \( G(w, t, w') \), \((w', t) \models \phi \)
Truth definition for intentions

Intentions are the same:

- $\text{INTEND} \phi$ is true in a possible world $w$ at time point $t$ if $\phi$ is true in all intention-accessible worlds $w'$ at point $t$
- we assume that if $w'$ is intention-accessible from $(w, t)$ then $t$ exists in $w'$
- accessibility relation for $\text{INTEND}$: $I \subseteq W \times T \times W$
- $I$ is serial
- $(w, t) \models \text{INTEND} \phi$ iff for all $w'$ such that $I(w, t, w'), (w', t) \models \phi$
Goal-accessible worlds are sub-worlds of belief-accessible worlds

- for each belief-accessible world there is a goal-accessible world where things go well
- undesirable paths that exist in the belief-accessible world are pruned

Intention-accessible worlds are sub-worlds of goal-accessible worlds

- intuitively, they contain only those desirable courses of action the agent has committed to
Definition of the subworld relation

- A path (fullpath) in \( w \) is an infinite sequence \( t_0, t_1, \ldots \) of time points in \( w \) such that \( t_0 \) is the root of the time tree in \( w \) and for each pair \( t_i, t_{i+1} \) in the sequence, \( t_{i+1} \) is the child of \( t_i \).

- \( \text{paths}(w) \) is the set of all paths in \( w \).

- \( w \) is a subworld of \( w' \), \( w \sqsubseteq w' \), iff \( \text{paths}(w) \subseteq \text{paths}(w') \).
Example

- suppose $\text{atm}_n$ is an action of extracting $n$ Euros from an ATM
- actions may fail (e.g., if the ATM is out of service)
- an example of a belief accessible world $w_1$, $B(w_0, t, w_1)$:
Example 2

- failure paths are not desirable
- here is a goal-accessible world $w_2$, for which $w_2 \sqsubseteq w_1$ holds:
Example 3

- the agent commits to getting 100 Euro out of the ATM
- here is an intention-accessible world $w_3$, for which $w_3 \sqsubseteq w_2$ holds:
Semantic conditions on $B$, $G$ and $I$ accessibility relations

**CI1** (belief-goal consistency):
\[
\forall w \forall t \forall w'(B(w, t, w') \rightarrow \exists w''(G(w, t, w'') \land w'' \sqsubseteq w'))
\]

**CI2** (goal-intention consistency):
\[
\forall w \forall t \forall w'(G(w, t, w') \rightarrow \exists w''(I(w, t, w'') \land w'' \sqsubseteq w'))
\]
E-formulas

- Let $\phi$ be an $E$-formula (a formula which does not contain positive occurrences of $A$ quantifiers and negative occurrences of $E$ quantifiers outside the scope of $BEL$, $GOAL$, $INTEND$).
- If $M, (w, t) \models \phi$ and $w \sqsubseteq w'$ then $M, (w', t) \models \phi$. 
Properties of beliefs and goals

\[ \text{AI1} \quad \text{GOAL} \phi \rightarrow \text{BEL} \phi \text{ where } \phi \text{ is an } E\text{-formula (if the agent has } \phi \text{ as a goal, then the agent must believe that a path satisfying } \phi \text{ exists in all belief-accessible worlds)} \]

- e.g.: \( \text{GOAL} \ E\diamond p \rightarrow \text{BEL} \ E\diamond p \)
- this is called \text{strong realism}
- valid in BDI_{CTL} because of belief-goal consistency (condition CI1)
Properties of goals and intentions

- $\phi, \text{INTEND } \phi \rightarrow \text{GOAL } \phi$ where $\phi$ is an $E$-formula (the agent only intends desirable paths)
- e.g.: $\text{INTEND } E\Diamond p \rightarrow \text{GOAL } E\Diamond p$
- valid in $\text{BDI}_{\text{CTL}}$ because of goal-intention consistency (condition $\text{CI2}$)
Commitment strategies

Definitions of possible commitment strategies:

- **(blind commitment):**
  \[ \text{INTEND } A \diamond \phi \rightarrow A(\text{INTEND } A \diamond \phi \cup \text{BEL } \phi) \]

- **(single-minded commitment):**
  \[ \text{INTEND } A \diamond \phi \rightarrow A(\text{INTEND } A \diamond \phi \cup \text{BEL } \phi \lor \neg \text{BEL } E \diamond \phi) \]

- **(open-minded commitment):**
  \[ \text{INTEND } A \diamond \phi \rightarrow A(\text{INTEND } A \diamond \phi \cup \text{BEL } \phi \lor \neg \text{GOAL } E \diamond \phi) \]

Sample property:

- For **competent agents** (which satisfy \( \text{BEL } \phi \rightarrow \phi \) for all \( \phi \)), all three commitment strategies result in:
  \[ \text{INTEND } A \diamond \phi \rightarrow A \diamond \phi \]
Multi-Agent BDI Logics
Cohen and Levesque’s logic

- intention is a persistent goal — unlike in Rao and Georgeff’s logic, intention is defined in terms of beliefs, goals and actions

- foundational layer has 4 basic modalities:
  
  - **BEL** (binary, takes an agent and a formula)
  
  - **GOAL**
  
  - **HAPPENS** (which event happens next)
  
  - **DONE** (which event has just occurred)

- each possible world \( w \) is a discrete linear sequence of events, infinitely extended in the past and in the future

- also have time points (integers); events occur *between* time points, so we have something like

  \[ \ldots -1 \ [e_1] \ 0 \ [e_2] \ 1 \ [e_3] \ 2 \ [e_4] \ldots \]
Other approaches


- **KARO**: uses dynamic (PDL) rather than temporal logic as a basis; actions are ‘primary’ (Meyer, van der Hoek, van Linder, “A logical approach to the dynamics of commitments” *Artificial Intelligence*, 113, 1999)

- **BDI-ATL**: substitutes ATL* for CTL* in Rao & Georgeff’s logic, allowing commitment strategies that take account of collaboration among agents (Montagna, Delzanno, Martelli and Mascardi *BDIA TL : An Alternating-Time BDI Logic for Multiagent Systems*, Proc. EUMAS 2005, pp. 214–223)
Problems of classical BDI logics

- the BDI logics we have looked at are ‘classical’ in the sense that they extend existing modal logics with possible worlds semantics
- they have many interesting ideas and can help to formally specify and compare properties of beliefs, desires and intentions, commitment strategies, communication semantics etc.
- however, it is not clear how to implement agents based on these logical specifications
- in particular, what corresponds to belief and goal accessibility relations in the agent programming language / implemented agent?
- this is also a problem for verification of MAS (next lecture)
The next lecture

Verification of MAS