BDIO_{CTI}
Obligations and the Specification of Agent Behavior

Jan Broersen
ICS, Utrecht University
The Netherlands
broersen@cs.uu.nl

Mehdi Dastani
ICS, Utrecht University
The Netherlands
mehdi@cs.uu.nl

Leendert van der Torre
CWI Amsterdam
The Netherlands
torre@cwi.nl

1 Introduction

Recently several agent architectures have been proposed that incorporate obligations. However, agent specification or verification languages that take obligations into account have received less attention. Our research question is how properties involving obligations can be specified or verified in an extension of Rao and Georgeff's BDICTL- In Section 2 we extend BDICTL with so-called Standard Deontic Logic, and in Section 3 and 4 we introduce various single agent and multiagent properties.

2 BDIO_{CTI}

We use an equivalent reformulation of Rao and Georgeff's formalism 11991] presented by Schild [2000], which we extend with Standard Deontic Logic (SDL, Von Wright 11951]). We only consider the semantics.

Definition 1 (Syntax BDIOCTL) *Assume n agents. The admis sable formulae of* BDI_{CTL} *are categorized into two classes, state formulae and path formulae.*

S1 Each primitive proposition is a state formula.

S2 If α *and* β *are state formulae, then so are* $\alpha \wedge \beta$ *and* $\neg\alpha$.

S3 If α *is a path formula,* $E\alpha$ *and* $A\alpha$ *are state formulae.*

S4 If α *is a state formula and* $1 \leq i \leq n$, *then* $B_i(\alpha), D_i(\alpha), I_i(\alpha), O_i(\alpha)$ *are state formulae as well.*

P If α *and* β *are state formulae, then* $X\alpha$ *and* $\alpha U \beta$ *are path formulae.*

The semantics of BDICTL involves two dimensions. The truth of a formula depends on both the world *w* and the temporal state s. A pair *(w,s)* is called a situation in which BDICTL formulae are evaluated.

Definition 2 (Situation structure BDIO_{CTL}) *Assume* n *agents. A structure* $M = \langle \Delta, \mathcal{R}, \mathcal{B}_1, \mathcal{D}_1, \mathcal{I}_1, \mathcal{O}_1, \ldots, \mathcal{O}_n, L\rangle$ *forms a situation structure if* Δ *is a set of situations,* $\mathcal{R} \subseteq \Delta \times \Delta$ *is a binary relation such that* $w = w'$ *whenever* $\langle w, s\rangle \mathcal{R} \langle w', s'\rangle$, $Z_i \subseteq \Delta \times \Delta$ *for* $Z \in \{\mathcal{B}, \mathcal{D}, \mathcal{I}, \mathcal{O}\}$ *and* $1 \leq i \leq n$ *are binary relations such that* $s = s'$ *whenever* $\langle w, s\rangle Z_i \langle w', s'\rangle$, *and* L *an interpretation function that assigns a particular set of situations to each primitive proposition.* $L(p)$ *contains all those situations in which* p *holds.*

A speciality of CTL is that some formulae - called path formulae- are not interpreted relative to a particular situation. What is relevant here are full paths. The reference to *M* is omitted whenever it is understood.

Definition 3 (Semantics BDIO_{CTL}) *Assume* n *agents. A full path in situation structure* M *is a sequence* $\chi = \delta_0, \delta_1, \delta_2, \ldots$ *such that for every* $i \geq 0$, δ_i *is an element of* Δ *and* $\delta_i \mathcal{R} \delta_{i+1}$, *and if* χ *is finite with* δ_n *its final situation, then there is no situation* δ_{n+1} *in* Δ *such that* $\delta_n \mathcal{R} \delta_{n+1}$. *We say that a full path starts at* δ *iff* $\delta_0 = \delta$. *If* $\chi = \delta_0, \delta_1, \delta_2, \ldots$ *is a full path in* M, *then we denote* δ_i *by* χ^i $(i > 0)$.

Let M *be a situation structure,* δ *a situation, and* χ *a full path. The semantic relation* $\models$ *for* BDIO_{CTL} *is then defined as follows:*

S1 $\delta \models p$ *iff* $\delta \in L(p)$ *and* p *is a primitive proposition*

S2 $\delta \models \alpha \wedge \beta$ *iff* $\delta \models \alpha$ *and* $\delta \models \beta$

$\delta \models \neg\alpha$ *iff* $\delta \models \alpha$ *does not hold*

S3 $\delta \models E\alpha$ *iff* $\exists$ *full path* χ *in* M *starting at* δ *s.t.* $\chi \models \alpha$

$\delta \models A\alpha$ *iff* $\forall$ *full path* χ *in* M *starting at* δ, $\chi \models \alpha$

S4 $\delta \models B_i(\alpha)$ *iff for every* $\delta' \in \Delta$ *such that* $\delta \mathcal{B}_i \delta'$, $\delta' \models \alpha$

$\delta \models D_i(\alpha)$ *iff for every* $\delta' \in \Delta$ *such that* $\delta \mathcal{D}_i \delta'$, $\delta' \models \alpha$

$\delta \models I_i(\alpha)$ *iff for every* $\delta' \in \Delta$ *such that* $\delta \mathcal{I}_i \delta'$, $\delta' \models \alpha$

$\delta \models O_i(\alpha)$ *iff for every* $\delta' \in \Delta$ *such that* $\delta \mathcal{O}_i \delta'$, $\delta' \models \alpha$

P $\chi \models X\alpha$ *iff* $\chi^1 \models \alpha$

$\chi \models \alpha U \beta$ *iff there is at least one* $i \geq 0$ *such that* $\chi^i \models \beta$ *and for all* $j (0 \leq j < i)$, $\chi^j \models \alpha$

As usual 'globally α*' (on a path) is defined as* $G(\alpha) = \alpha U \perp$ *and 'at some future moment* α *holds' as* $F(\alpha) = \neg G(\neg\alpha)$.

We do not discuss the correspondence between modal properties and conditions on binary relations $\mathcal{B}_i$, $\mathcal{D}_i$, $\mathcal{I}_i$, and $\mathcal{O}_i$, but discuss possible specification properties expressible as modal formulae.

Rao and Georgeff discuss realism properties

$$B_i p \rightarrow D_i p \qquad D_i p \rightarrow \neg B_i \neg p$$
$$D_i EFp \rightarrow B_i EFp$$

and commitment strategies, e.g.:

$$I_i AFp \rightarrow A(I_i AFp U B_i p)$$
$$I_i AFp \rightarrow A(I_i AFp U (B_i p \vee \neg B_i EFp))$$

3 Single agent properties

The following properties characterize the relation between mental attitudes of a single agent.

3.1 Regimentation

A main question when developing a normative system is whether the norms can be violated or not, i.e. whether the norms are soft or hard constraints. In the latter case, the norms are said to be regimented. Regimented norms correspond to preventative control systems in computer security. For example, in the metro in Paris it is not possible to travel without a ticket, because there is a preventative control system, whereas it is possible to travel without a ticket on the French trains, because there is a detective control system. Norm regimentation for agent *i* is characterized by:

$$O_i p \rightarrow p$$

Strong and weak epistemic norm regimentation are:

$$O_i p \rightarrow B_i p \qquad O_i p \rightarrow \neg B_i \neg p$$

And intentional norm regimentation is:

$$O_i p \rightarrow I_i p \qquad O_i p \rightarrow \neg I_i \neg p$$

A variant of the latter regimentation is conditional to a conflict between agent i's internal and external motivations. For example, if an agent is obliged to work but desires to go to the beach, then it intends to go to work. Or at least it does not intend to go to the beach. An agent is called strongly or weakly respectful if:

$$(O_i p \wedge D_i \neg p) \rightarrow I_i p \qquad (O_i p \wedge D_i \neg p) \rightarrow \neg I_i \neg p$$

Moreover, a respectful agent can internalize its obligations in the sense that they turn into desires, or at least it cannot decide to violate the obligation. For example, if an agent is obliged to work then it also desires to work, or at least it cannot desire not to work.

$$O_i p \rightarrow D_i p \qquad O_i p \rightarrow \neg D_i \neg p$$

Instead of respectful, agents may also be egocentric, which can be characterized by similar properties:

$$D_i p \rightarrow O_i p \qquad D_i p \rightarrow \neg O_i \neg p$$

3.2 Persistence

Obligations typically persist until a deadline, e.g. deliver the goods before noon, or they persist forever, e.g. don't kill. We denote a deadline obligation by *Oi(p, d),* where achievement of the proposition *d* is the deadline for the obligation to achieve p. A deadline obligation *Oi(p,d)* persists until it is fulfilled or becomes obsolete because the deadline is reached:

$$O_i(p, d) =_{def} A(O_i p U(p \vee d))$$

A deadline obligation *0{(p,p),* for which the only deadline is the achievement of the obligation itself, is called an 'achievement obligation'. We may characterize that *Oip* is an achievement obligation by:

$$O_i p \rightarrow A(O_i p U p)$$

Alternatively, we may characterize that *Oip* persists forever, i.e. that it is a 'maintenance obligation', by:

$$O_i p \rightarrow AGO_i p$$

4 Multi-agent Obligations

In a multi-agent setting agents interact with each other, thereby creating obligations. For example, in an electronic market where agents interact to buy and sell goods, sending a confirmation to buy an item creates the obligation of payment by the buyer and the obligation of shipment of the item by the seller. Social systems may be designed in which obligations are related to the mental attitudes of other agents. For example, there may be communities in which agent *i* may adopt all the obligations of agent *j.*

$$B_i(O_j(\alpha)) \rightarrow O_i(\alpha)$$

The following property characterizes that agent *i* adopts the desires of agent *j* as its obligations. For example, if agent *j* desires to eat then agent *i* is obliged to see to it that he gets something to eat.

$$B_i(D_j(\alpha)) \rightarrow O_i(\alpha)$$

In a master slave relationship, the intentions of the master become the obligations of the slave.

$$B_i(I_j(\alpha)) \rightarrow O_i(\alpha)$$

Agent *i* is a dictator if for every other agent *j* it holds that:

$$I_i(O_j(\alpha)) \rightarrow O_j(\alpha)$$

Finally, a particular kind of dictator *i* is one whose desires immediately turn into obligations of another agent j.

$$D_i(O_j(\alpha)) \rightarrow O_j(\alpha)$$

5 Summary

In this paper we have introduced the BDIOCTL logic; a combination of Rao and Georgeff's BDICTL formalism and standard deontic logic SDL. We have defined several specification and verification properties in this logic. The formalization of other properties is the subject of further research. The option most discussed in deontic logic is whether violations of norms can trigger new obligations, i.e. whether there is contrary-to-duty reasoning. For example, it is often assumed that the legal code does not contain contrary-to-duty norms. Properties related to contrary-to-do reasoning are therefore of particular interest.

References

[Rao and Georgeff, 1991] A.S. Rao and M.R Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91),* pages 473-484. Morgan Kaufmann Publishers, 1991.

[Schild, 2000] K. Schild. On the relationship between BD1-logics and standard logics of concurrency. *Autonomous agents and multi-agent systems,* 3:259-283, 2000.

[Wright, 1951] G.H. von Wright. Deontic logic. *Mind,* 60:1-15, 1951.