The bit transmission problem revisited

Alessio Lomuscio Marek Sergot
Department of Computing
Imperial College of Science, Technology and Medicine
London SW7 2BZ, United Kingdom
{A.Lomuscio,M.Sergot}@doc.ic.ac.uk

ABSTRACT

The design of complex multi-agent systems is increasingly having to confront the possibility that agents may not behave as they are supposed to. In addition to analysing the properties that hold if protocols are followed correctly, it is also necessary to predict, test, and verify the properties that would hold if these protocols were to be violated. We illustrate how the formal machinery of deontic interpreted systems can be applied to the analysis of such problems by considering three variations of the bit transmission problem.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: DAI—Multiagent systems

General Terms

Theory

1. INTRODUCTION

The design of complex multi-agent systems is increasingly having to confront the possibility that agents may not behave as they are supposed to. In e-commerce, in security, in automatic negotiation, in any application where agents are programmed by different parties with competing interests, it is unrealistic to assume that all agents will behave according to some given protocol or standard of behaviour. In addition to analysing the properties that hold if protocols are followed correctly, it is also necessary to predict, test, and verify the properties that would hold if these protocols were to be violated, and to test the effectiveness of introducing proposed enforcement mechanisms.

Consider an online auction mechanism modelled and implemented as a multi-agent system, where agents play the parts of auctioneers and bidders. At each round the protocol of the system gives the opportunity to a selected number of agents to bid for some goods, according to the specific auction protocol being employed. Suppose that an agent bids an amount which it is unable to fulfil (either by mistake, or perhaps with the sole intention of raising the bid so that an opponent will have to pay a higher price). If that bid is made in the prescribed manner then it would still count as a valid bid — even if it were to be regarded as illegal, anti-social, unethical to make a bid in those circumstances. Undesirable behaviour of this kind will often have adverse effects on the system as a whole. Perhaps the agent will be forced to de-commit from the commitment it entered upon by bidding, as a result of which the seller will (at least temporarily) lose the deal, which in turn may have consequences on other deals it has entered upon with other agents. At the very least, the resulting disputes will have to be resolved at a cost.

One way out of such problems is to attempt to devise tighter controls in the protocols, such as stipulating that a bid by an agent is valid only if made with previously cleared funds. The opportunities for inventing such controls are limited, however, and in any case have associated costs too (in the example, the costs of the clearing mechanism and the consequently diminished liquidity of the market). Moreover forcing agents of an open multi-agent system to behave according to a rigid set of rules is seen as increasingly unfeasible because of the inherent distribution of the system. One alternative, having specified what is correct, desirable, or acceptable behaviour of the agents within a given context, is to strive to handle and reason about violations of these norms within the system. The required formalisms not only have to be capable of handling *local violations*, but must also be capable of representing the usual attributes ascribed to agents in a multi-agent setting, such as knowledge, intentions, and goals (see e.g., [Woo00]).

These are inherently complex issues, and we do not expect to have methods able to give comprehensive answers to all of them in the near future. One approach being explored by the authors of this note as a step in this direction is an attempt to extend interpreted systems [FHMV95], a mainstream and well-developed semantics for reasoning about knowledge in multi-agent systems, with a deontic component. Deontic interpreted systems [LS01a, LS01b] are designed to represent correct and incorrect functioning behaviour of agents as well as their epistemic properties.

The specific aims of this paper are to illustrate how deontic interpreted systems may be applied to model and reason about violations in a simple example, the bit-transmission problem [FHMV95].

AAMAS'02, July 15-19, 2002, Bologna, Italy.

2. THE BIT TRANSMISSION PROBLEM

The bit-transmission problem [FHMV95] involves two agents, a *sender* S, and a *receiver* R, communicating over a possibly faulty communication channel. S wants to communicate some information — the value of a bit for the sake of the example — to R. We would like to design a protocol that accomplishes this objective while minimising the use of the communication channel.

The interesting scenario arises when the channel is working correctly at certain times while failing at others. There are several ways in which this can be approached, for instance by stipulating that the channel delivers messages with a fixed probability $P > 0$ at any given round. In this paper we do not make any of these assumptions; instead, we analyse the most general case with respect to a most simple protocol. The protocol is as follows. S immediately starts sending the bit to R, and continues to do so until it receives an acknowledgement from R. R does nothing until it receives the bit; from then on it sends acknowledgements of receipt to S. S stops sending the bit to R when it receives an acknowledgement. Note that R will continue sending acknowledgments even after S has received its acknowledgement. Intuitively S will know for sure that the bit has been received by R when it gets an acknowledgement from R. R, on the other hand, will never be able to know whether its acknowledgement has been received since S does not answer the acknowledgement.

The bit-transmission problem can be analysed using the formalism of interpreted systems. To do this we follow the approach taken by Halpern and colleagues in [FHMV95].

There are three active components in the scenario: a sender, a receiver, and a communication channel. In line with the spirit of the formalism of interpreted systems, it is convenient to see sender and receiver as agents, and the communication channel as the environment. Each of these can be modelled by considering their local states, their actions, and their protocols. We refer to [FHMV95] for more details.

In this way we can generate the set of reachable global states that the protocols of the agents and the transition function of the system generates. This, together with an evaluation for a suitable set of propositional atoms, permits us to represent and check properties of the system directly on the semantical models. For example, by employing standard temporal connectors on the runs as constructed above one can check that:

$$IS_{\mathrm{b}} \not\models \mathbf{recbit} \rightarrow \Diamond \mathbf{recack}$$

which represents the intrinsic unreliability of the channel.

Irrespective of the analysis of the dynamic properties of the system, there is one interesting static epistemic property that is worth observing. By ascribing knowledge to the agents using the standard [FHMV95] approach, it can be checked that

$$IS_{\mathrm{b}} \models \mathbf{recack} \rightarrow K_S\big(K_R\,(\mathbf{bit} = \mathbf{0}) \vee K_R\,(\mathbf{bit} = \mathbf{1})\big)$$

So, if an ack is received by S, then S is sure that R knows the value of the bit. Intuitively this represents the fact that although the channel is potentially faulty, if messages do manage to travel back and forth the protocol is strong enough to eliminate any uncertainty in the communication problem.

3. VIOLATION OF SPECIFICATIONS

In the previous section we have assumed that both agents and environment follow the protocols. Let us now assume the possibility of R being faulty. The possibility that S is faulty, and other combinations of faulty R, S and E, can be treated in similar fashion. Intuitively this may compromise the communication protocol.

To analyse this, we can use the machinery of deontic interpreted systems [LS01a, LS01b]. Because of space limitations we are unable to report the details and encourage the interested reader to refer to [LS02]. It is argued there that this scenario can be analysed by describing the two agents in terms of red and green local states (representing correct and faulty functioning behaviour), relate these to the functioning of the local protocols, and define an appropriate transition function. By computing the resulting set of reachable states one can check that the mechanism of acknowledgements that is no longer reliable. Indeed we find in our model IS'_{b} that:

$$IS'_{\mathrm{b}} \not\models \mathbf{recack} \rightarrow K_S\big(K_R\,(\mathbf{bit} = \mathbf{0}) \vee K_R\,(\mathbf{bit} = \mathbf{1})\big),$$

where K is again the modal operator of knowledge defined on local states as in [FHMV95].

What is more interesting though, is that a particular form of knowledge also still holds. Intuitively if S could make the assumption of R's correct functioning behaviour, then, upon receipt of an acknowledgement, it would then make sense for it to know that R does know the value of the bit. To model this intuition we use the operator $\widehat{K}_i^j$ "knowledge under the assumption of correct behaviour" as presented in [LS01a]. This describes the knowledge agent i has if attention is restricted to states in which j is performing as intended. We refer to [LS01b] for more details. This operator is of particular interest in this circumstance because it captures precisely our intuition about the example. Indeed we have:

$$IS'_{\mathrm{b}} \models \mathbf{recack} \rightarrow \widehat{K}_S^R\big(K_R\,(\mathbf{bit} = \mathbf{0}) \vee K_R\,(\mathbf{bit} = \mathbf{1})\big)$$

To summarise: we have modified the scenario of the bit-transmission problem, relaxing slightly the assumptions of correct functioning behaviour that hold true for it. In particular, we have allowed for the possibility that one of the agents, the receiver R, may violate the protocol by performing an action it should not perform according to the protocol. We have seen that some key properties of the system then no longer hold. In particular under the assumptions we have studied, S will never know whether or not R knows the value of the bit; at best S can know this only under the assumption of correct functioning behaviour for R. This is an intuitive result that was validated syntactically, and semantically on the model.

Acknowledgements. This work was supported by the EU-funded FET project *ALFEBIITE* (IST-1999-10298).

4. REFERENCES

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge.* MIT Press, 1995.

A. Lomuscio and M. Sergot. Extending interpreting systems with some deontic concepts. In *Proceedings of TARK 2001*, Morgan Kauffman.

A. Lomuscio and M. Sergot. On multi-agent systems specification via deontic logic. In J.-J Meyer, editor, *Proceedings of ATAL 2001.* Springer Verlag, July 2001.

A. Lomuscio and M. Sergot. Violation, error recovery, and enforcement in the bit transmission problem. In *Proceedings of DEON'02*, London, May 2002.

M. Wooldridge. *Reasoning about rational agents.* MIT Press, July 2000.