Competent agents and customising protocols

Ulle Endriss, Wenjin Lu, Nicolas Maudet, & Kostas Stathis
Imperial College London
City Univerisity London

ue@doc.ic.ac.uk
{lue,stathis}@soi.city.ac.uk
maudet@lamsade.dauphine.fr

Funded by the EU as part of the SOCS project
(Societies Of ComputeeS)

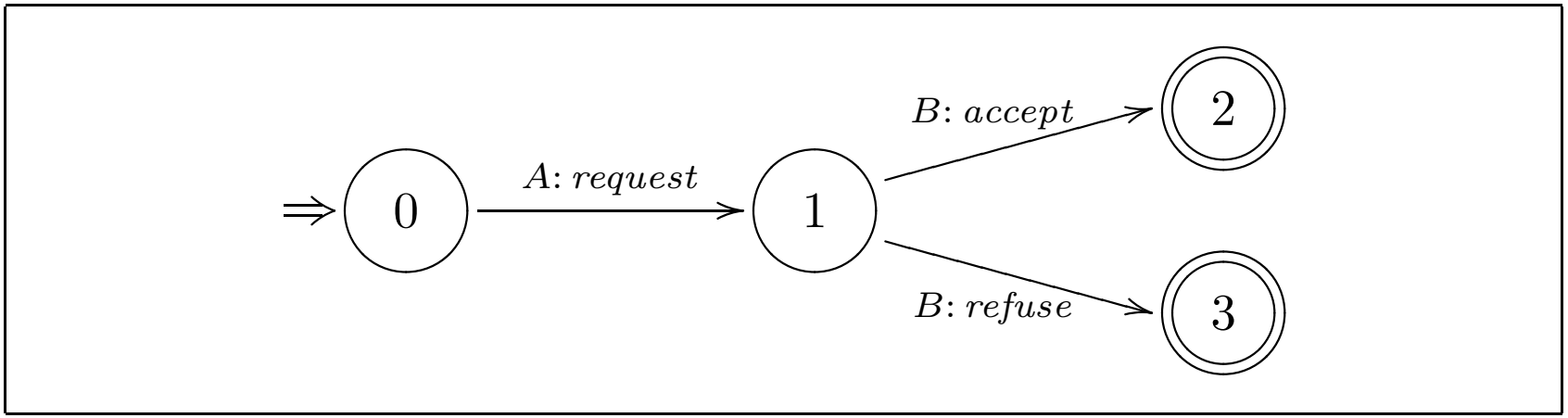
Introduction

- private strategies *vs.* public protocols;
- communication protocols and strategies cannot be assumed to always match perfectly;
- issues of conformance and *competence*;
- automatically checking *a priori* conformance and competence using logic-based methods.

Plan

- Introduction
- Protocols for logic-based agents
- Conformance
- Competence as exhaustive conformance
- Competence as reachability
- Customising protocols
- Conclusion and future work

Protocols for logic-based agents (i)



$$\underbrace{\begin{array}{lcl} START(T) & \Rightarrow & request(T{+}1) \\ accept(T) & \Rightarrow & STOP(T{+}1) \\ refuse(T) & \Rightarrow & STOP(T{+}1) \end{array}}_{\mathcal{P}_A} \qquad \underbrace{\begin{array}{lcl} request(T) & \Rightarrow & accept(T{+}1) \vee \\ & & refuse(T{+}1) \end{array}}_{\mathcal{P}_B}$$

Protocols for logic-based agents (ii)

Shallow protocols —automata where it is possible to determine the next state of the dialogue on the sole basis of the previous event.

Agents' strategies —a set of integrity constraints of the following form:

$$\underset{\textit{move received}}{P(T)} \quad \wedge \quad \underset{\textit{conditions}}{Cond} \quad \Rightarrow \quad \underset{\textit{move uttered}}{P'(T{+}1)}$$

Response space —abstraction of the communication strategy: removing all conditions and collecting the consequents whom antecedents are the same into a single disjunction.

Conformance

Levels of Conformance

- an agent is *weakly conformant* to a protocol $\mathcal{P}$ iff it never utters an illegal dialogue move (with respect to $\mathcal{P}$).
- an agent is *exhaustively conformant* to a protocol $\mathcal{P}$ iff it does utter a legal dialogue move whenever required to do so by $\mathcal{P}$.

Checking weak conformance —a simple sufficient criterion (based on agents' response spaces) can be exhibited, see Endriss *et al.* [IJCAI03]

Competence as exhaustive conformance (i)

An agent A with communicating strategy $\mathcal{S}$, and a protocol $\mathcal{P}$

Checking exhaustive conformance —more difficult than weak conformance because involves inspecting agents' knowledge bases

Criteria —A is exhaustively conformant if weakly conformant +
• $\mathrm{COND}_{\mathcal{S}}(P)$ is a logical theorem for every expected input P
or
• $\mathrm{COND}_{\mathcal{S}}(P)$ is a logical consequence of the agent's knowledge base for every expected input P

($\mathrm{COND}_{\mathcal{S}}(P)$: disjunction of all the private conditions that appear in $\mathcal{S}$ in a constraint together with the trigger $P(T)$)

Competence as exhaustive conformance (ii)

Competence = exhaustive conformance? —not a sufficient criterion

Consider an agent with the following response space

$$\mathcal{S}^* = \{\mathit{request}(T) \Rightarrow \mathit{refuse}(T{+}1)\}$$

Even if this agent was indeed exhaustively conformant, it would intuitively not be competent in the sense that it could never reach state 2 (see Fig.)

Competence as reachability (i)

Joint competence —for some protocol $\mathcal{P}$, two agents have the joint competence to reach P' from P iff they have the ability to generate a sequence of dialogue moves that are legal with respect to $\mathcal{P}$ and that include P' once P has been uttered.

Flattened responce spaces —propositional representation of the response spaces obtained by removing reference to time.

Checking joint competence — two agents with flattened response spaces $\mathcal{S}_A^*$ and $\mathcal{S}_B^*$ that are exhaustively conformant to protocol $\mathcal{P}$ have the competence to reach P' from P iff $\mathcal{S}_A^* \cup \mathcal{S}_B^* \cup \{P\} \models P'$

Competence as reachability (ii)

Models and dialogues —minimal models of $\mathcal{S}_A^* \cup \mathcal{S}_B^* \cup \{START\}$

- models including $STOP$ ($\Rightarrow$ terminating dialogues),
- models not including $STOP$ ($\Rightarrow$ potentially infinite loops).

Loops — we can then distinguish *good* and *bad* loops.

- *good* loops: at least one path to a final final state is still open.
- *bad* loops: agents have no other choice than to repeat the same sequence of utterances

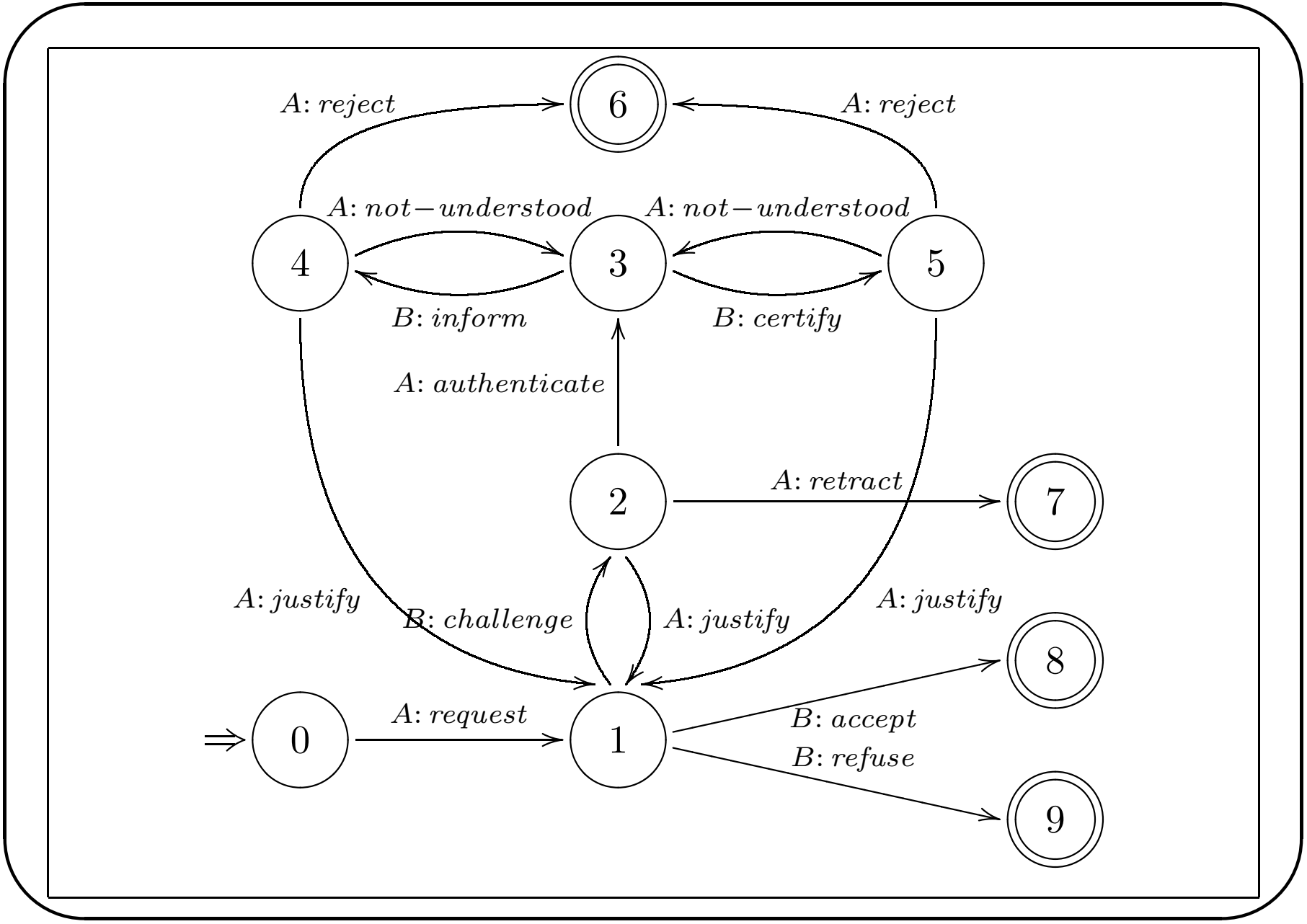
If for every P in the model of a loop, $\mathcal{S}_A^* \cup \mathcal{S}_B^* \cup P$ still has got at least a minimal model including STOP, then this is a good loop, otherwise this is a bad loop.

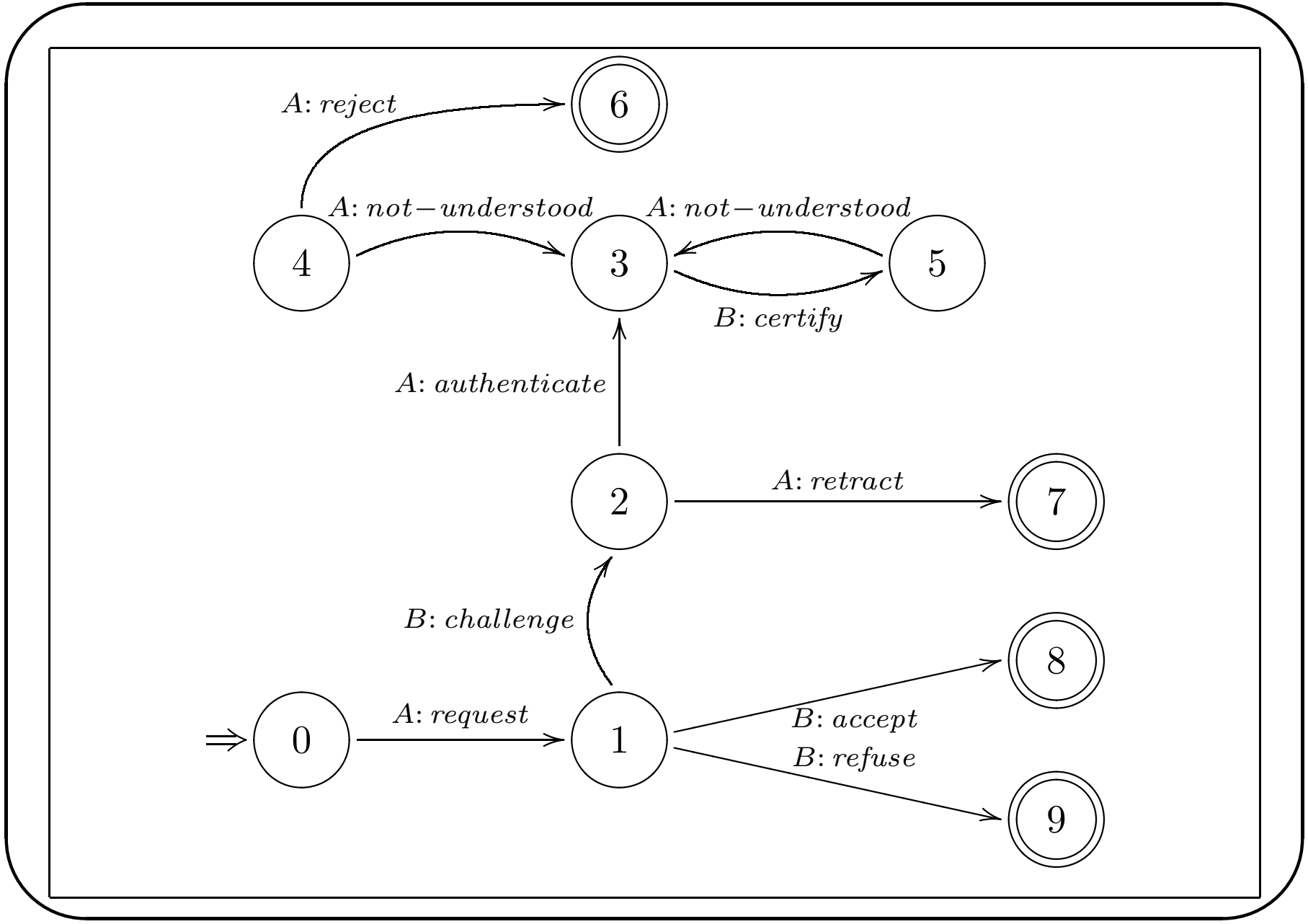
Customising protocols (i)

Problem —a protocol can turn out to be problematic because it is used by agents that are not fully competent. In this case, it can be useful for the designer of the application to *customise* his protocol, in order to avoid these undesirable situations.

Customisation process —

- identify the bad loops resulting from the agents' response spaces,
- remove all occurences in the protocol of those moves of the bad loops that do not allow to reach termination.


A: reject
6
A: reject
A: not−understood
A: not−understood
4
3
5
B: inform
B: certify
A: authenticate
2
A: retract
7
A: justify
B: challenge
A: justify
A: justify
8
0
A: request
1
B: accept
B: refuse
9


A: reject
6
A: not−understood
A: not−understood
4
3
5
B: certify
A: authenticate
2
A: retract
7
B: challenge
8
0
A: request
1
B: accept
B: refuse
9

Customising protocols (ii)

identify bad loops

$$\{ \quad \{START, request, accept, STOP\}$$
$$\{START, request, refuse, STOP\}$$
$$\{START, request, challenge, retract, STOP\}$$
$$\{\underline{START, request, challenge, authenticate, certify, not\text{-}understood}\}\}$$

delete moves

$$\{START, request, challenge, \cancel{authenticate}, \cancel{certify}, \cancel{not\text{-}understood}\}$$

Conclusion and Future work

- we have argued that competence cannot be reduced to conformance (even exhaustive)
- we have put forward a notion of competence as reachability
- we have shown preliminary results that allow to automatically check competence for a class of logic-based agents
- we have shown how to use these results to customise protocols
- we would like to investigate whether this notion could be of any help to design *fair* protocols