Approximate Verification of Strategic Abilities under Imperfect Information Using Local Models

Damian Kurpiewski, Wojciech Jamroga, Yan Kim

21/08/2025

Institute of Computer Science
Polish Academy of Sciences

Faculty of Mathematics and Computer Science
Nicolaus Copernicus University in Toruń

Motivation

- Verification of strategic ability under imperfect information is **challenging**
  - Complexity ranges from **NP**-complete to undecidable
  - Traditional fixpoint equivalences **fail** in imperfect information setting

Motivation

- Verification of strategic ability under imperfect information is **challenging**
  - Complexity ranges from **NP**-complete to undecidable
  - Traditional fixpoint equivalences **fail** in imperfect information setting
- Existing fixpoint approximation (Jamroga et al., 2019):
  - Operates on **global model** of the system
  - Still suffers from **state/transition-space explosion**

Motivation

- Verification of strategic ability under imperfect information is **challenging**
  - Complexity ranges from **NP**-complete to undecidable
  - Traditional fixpoint equivalences **fail** in imperfect information setting
- Existing fixpoint approximation (Jamroga et al., 2019):
  - Operates on **global model** of the system
  - Still suffers from **state/transition-space explosion**
- **Key insight**: For asynchronous MAS, we can leverage modular representation
  - Use **local models** instead of global model
  - Achieve **exponential reduction** in model size

Previous Work and Our Contribution

Previous Approach (Jamroga et al., 2019)

- Translation of $\mathbf{ATL}_{\mathrm{ir}}$ to alternating epistemic μ-calculus
- Provides a **lower bound** for verification
- Still operates on **global model** (suffers from state-space explosion)

Previous Work and Our Contribution

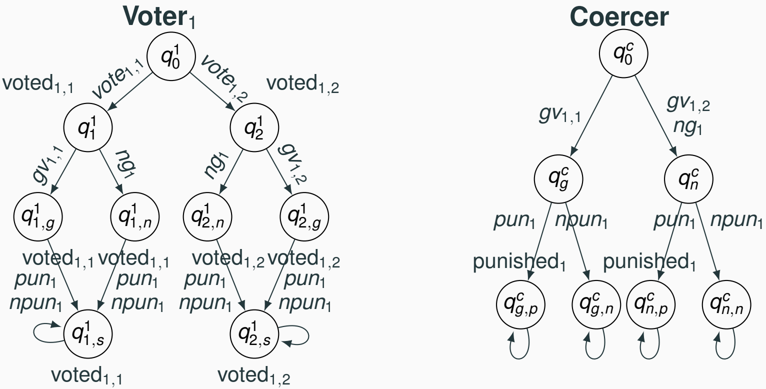
Previous Approach (Jamroga et al., 2019)

- Translation of $\textbf{ATL}_{\mathrm{ir}}$ to alternating epistemic μ-calculus
- Provides a **lower bound** for verification
- Still operates on **global model** (suffers from state-space explosion)

Our New Approach

- Leverage **modular representation** of asynchronous MAS
- Perform fixpoint computation on **local model(s)**
- Key observation: epistemic classes in global model $\leftrightarrow$ local states
- **Exponentially smaller** model for verification

Asynchronous MAS: Voting Example



- ASV_n^k: n voters, k candidates, 1 coercer
- Voter chooses candidate & whether to share receipt
- Coercer chooses to punish or not
- Events shared between agents must be executed synchronously

Formal Background

Asynchronous MAS (AMAS)

- n agents $\mathcal{A} = \{1, \ldots, n\}$
- Each agent i has local states L_i, events Evt_i, repertoire function Roc_i
- Global state: $(l_1, \ldots, l_n) \in L_1 \times \cdots \times L_n$

Formal Background

Asynchronous MAS (AMAS)

- n agents $\mathcal{A} = \{1, \ldots, n\}$
- Each agent i has local states L_i, events Evt_i, repertoire function Roc_i
- Global state: $(l_1, \ldots, l_n) \in L_1 \times \cdots \times L_n$

Strategic Ability (ATL$_{\mathrm{ir}}$)

- $\langle\langle i \rangle\rangle \mathrm{F}\phi$: agent i has a strategy to eventually achieve ϕ
- $\langle\langle i \rangle\rangle \mathrm{G}\phi$: agent i has a strategy to always maintain ϕ
- Imperfect information: strategies based on **local states**

Local Approximation Model

Definition (Local Approximating Model)
For agent i, $M_i = (L_i, Evt_i, Roc_i, PV_i, Tapp_i)$ where:

- L_i: local states of agent i
- $Tapp_i$: transition relation capturing essential behavior
  - $(l, \epsilon, l) \in Tapp_i$ if global model has ϵ-loop at l
  - $(l, \tau, l) \in Tapp_i$ if global model has livelock cycle at l
  - $(l, \alpha, l') \in Tapp_i$ if global model has path from l to l' via α

Key insight: The local model captures all relevant behavior for agent i's strategic abilities.

Fixpoint Approximation on Local Models

Translation of sATL $_{ir}$ formulas

1. $tr_L(\langle\langle i \rangle\rangle \mathrm{F}\phi) = \mu Z.(\phi \vee \langle i \rangle Z)$
2. $tr_L(\langle\langle i \rangle\rangle \mathrm{G}\phi) = \nu Z.(\phi \wedge \langle i \rangle Z)$
3. $tr_L(\langle\langle i \rangle\rangle \psi \,\mathrm{U}\, \phi) = \mu Z.(\phi \vee (\psi \wedge \langle i \rangle Z))$

Fixpoint Approximation on Local Models

Translation of sATL $_{ir}$ formulas

1. $tr_L(\langle\langle i \rangle\rangle \mathrm{F} \phi) = \mu Z.(\phi \vee \langle i \rangle Z)$
2. $tr_L(\langle\langle i \rangle\rangle \mathrm{G} \phi) = \nu Z.(\phi \wedge \langle i \rangle Z)$
3. $tr_L(\langle\langle i \rangle\rangle \psi \,\mathrm{U}\, \phi) = \mu Z.(\phi \vee (\psi \wedge \langle i \rangle Z))$

Verification Procedure

- Generate local approximating model M_i for agent i
- Verify $tr_L(\phi)$ on M_i using standard fixpoint algorithm
- If true in M_i, then ϕ is true in the global model (lower approximation)

Experimental Setup

Benchmarks

- **ASV**: Asynchronous Simple Voting protocol
- **ASV+R**: ASV with revoting capability

Verified formula

$$\phi_1 = \langle\langle \mathit{Voter}_1 \rangle\rangle \mathrm{F}(\mathit{vote}_{1,1} \wedge \neg \mathit{give}_1)$$

Voter 1 can vote for candidate 1 without sharing receipt

Implementation

- Local model generation: UPPAAL
- Verification: STV model checker
- Comparison: local approximation vs. global model

Experimental Results: ASV Protocol

#V	Model generation (s)			Verification (s)		
	Global	Approx.	Optimized	Global	Approx.	Result
2	0.04	6.60	6.54	$<$0.01	$<$0.01	TRUE
3	0.10	6.62	6.60	0.29	$<$0.01	TRUE
4	1.22	6.93	6.91	30.15	$<$0.01	TRUE
5	35.80	8.71	8.70	2659	$<$0.01	TRUE
6	1206	36.95	29.42	timeout	$<$0.01	TRUE
7	timeout	282.48	280.62	-	$<$0.01	TRUE
8	timeout	5539	4046	-	$<$0.01	TRUE

Key observation: Verification time for approximated model is **constant** regardless of number of voters.

Experimental Results: ASV with Revoting

#V	Model generation (s)			Verification (s)		
	Global	Approx.	Optimized	Global	Approx.	Result
2	0.82	19.43	19.27	8.20	$<$0.01	TRUE
3	131.61	26.44	19.28	timeout	$<$0.01	TRUE
4	timeout	524.93	19.25	-	$<$0.01	TRUE
5	timeout	-	19.34	-	$<$0.01	TRUE
6	timeout	-	19.40	-	$<$0.01	TRUE
7	timeout	-	19.41	-	$<$0.01	TRUE
8	timeout	-	19.43	-	$<$0.01	TRUE
9	timeout	-	19.44	-	$<$0.01	TRUE

Key observation: Optimized model generation time grows **linearly** with number of voters.

Conclusions and Future Work

Conclusions

- Proposed new fixpoint approximation using **local models** instead of global model
- Proved correctness: if formula holds in local model, it holds in global model
- Achieved **exponential speedup** in verification time
- Model generation can be optimized to grow **linearly** with system size (for some models)

Limitations and Future Work

- Current approach: observable goals, individual strategies only
- Future: general non-observable properties
- Future: proper coalitions (not just single agents)
- Future: nested strategic reasoning