The Size of Message Set Needed for the Optimal Communication Policy

Tatsuya Kasai, Hayato Kobayashi, and Ayumi Shinohara

Graduate School of Information Sciences, Tohoku University, Japan

The 7th European Workshop on Multi-Agent Systems (EUMAS 2009)
Ayia Napa, Cyprus
Dec17-18, 2009

Background

- Multi-agent coordination with communication.

> Main objective :
> To find the optimal *action policy* δ^A and *communication policy* δ^M

- We are interested in an approach based on autonomous learning.

- Definition of policies for agent i in our proposed methods

	Signal Learning (SL) [Kasai+ 08]	**Signal Learning with Messages (SLM)** [Kasai+ AAMAS09]
Action Policy	$\delta_i^A : \Omega_i \times \boldsymbol{M}_i^{receive} \rightarrow A_i$	
Communi-cation Policy	$\delta_i^M : \Omega_i \rightarrow M_i$	$\delta_i^M : \Omega_i \times \boldsymbol{M}_i^{receive} \rightarrow M_i$

A set of observations

A set of actions

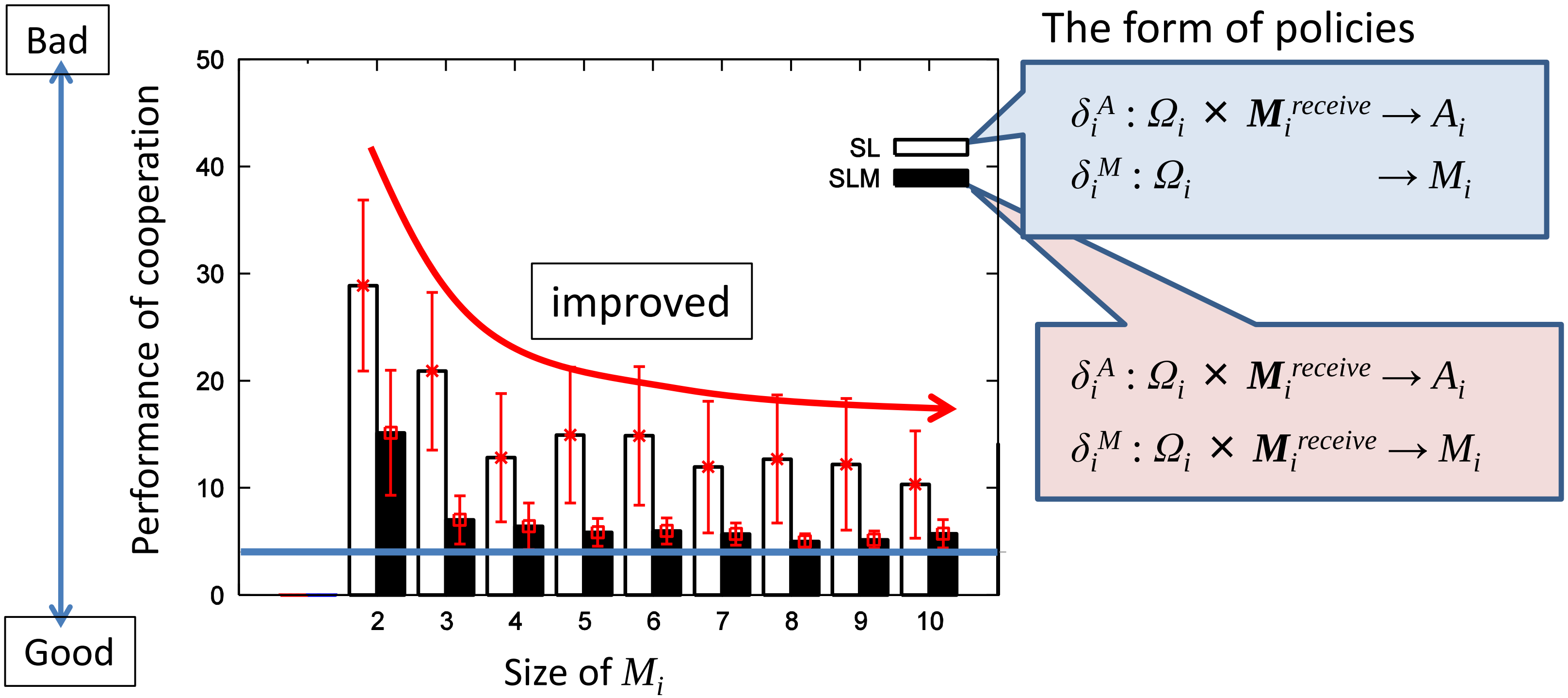
A set of received messages

A set of messages to send other agents

(SL and SLM are based on Multi-Agent Reinforcement Learning framework)

Motivation

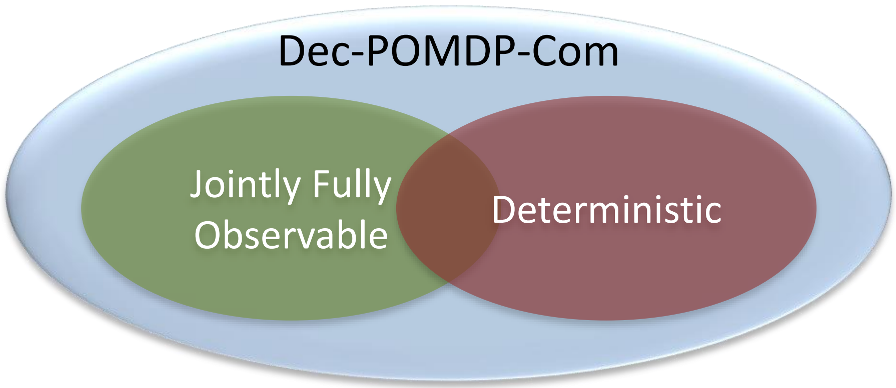
- Actual learning results of SL and SLM [Kasai+ AAMAS09]
  - The performance of cooperation when the size of M_i is increases.



- We have an interest about how much size of M_i for constructing the optimal policy ?

Scheme of talk

- We show *minimum required sizes* $|M_i|$ for achieving the optimal policy for
  - Signal Learning on *Jointly Fully Observable* *Dec-POMDP-Com*
  - Signal Learning with Messages on *Deterministic* *Dec-POMDP-Com*



Outline

- ☑ Background
- ☑ Scheme of talk
- ☐ Review : Dec-POMDP-Com [Goldman+ 04]
- ☐ Constrained model
  - Jointly Fully Observable Dec-POMDP-Com [Goldman+ 04]
  - Deterministic Dec-POMDP-Com (we define)
- ☐ Theoretical analysis
- ☐ Conclusion

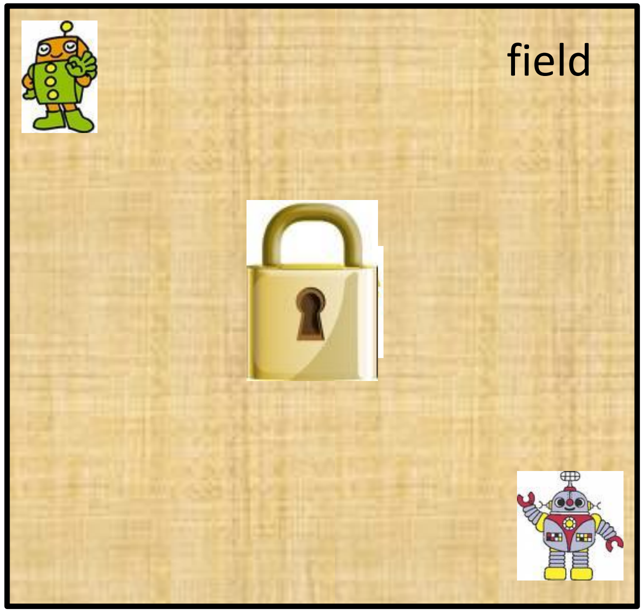
Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.

- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

Example of model



Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$ ← Formulation

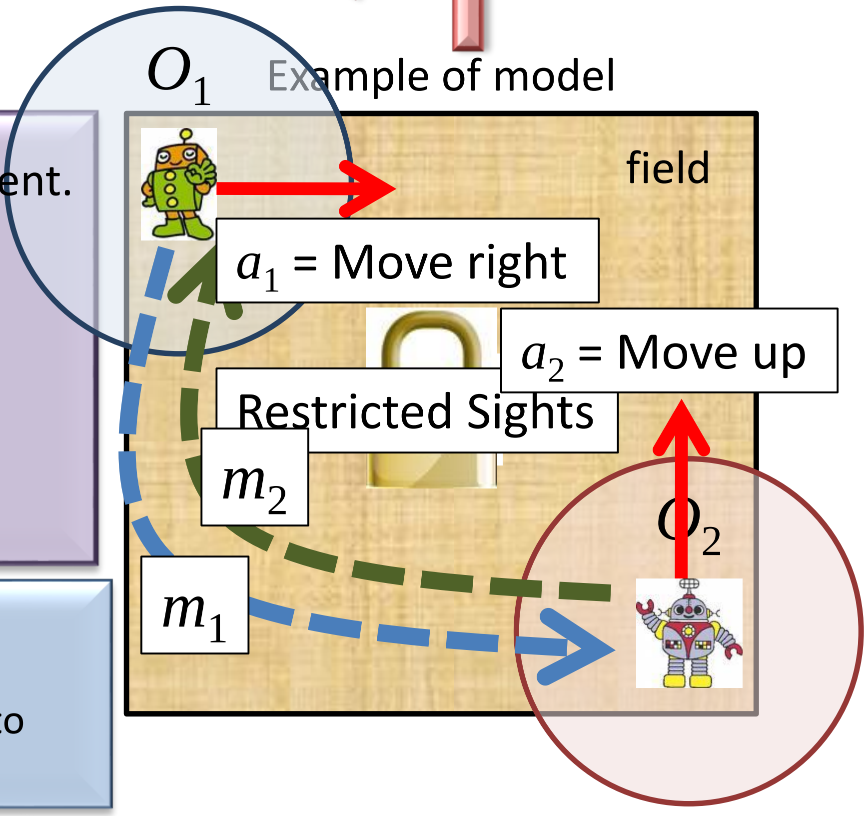
1step for agent i on Dec-POMDP-Com

1. Receive an observation O_i from the environment.
2. Send a message m_i to the other agents.
3. Perform an action a_i in the environment.

Repeat until both agent arrive at the treasure.

- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

Example of model



Dec-POMDP-Com [Goldman+ 04]

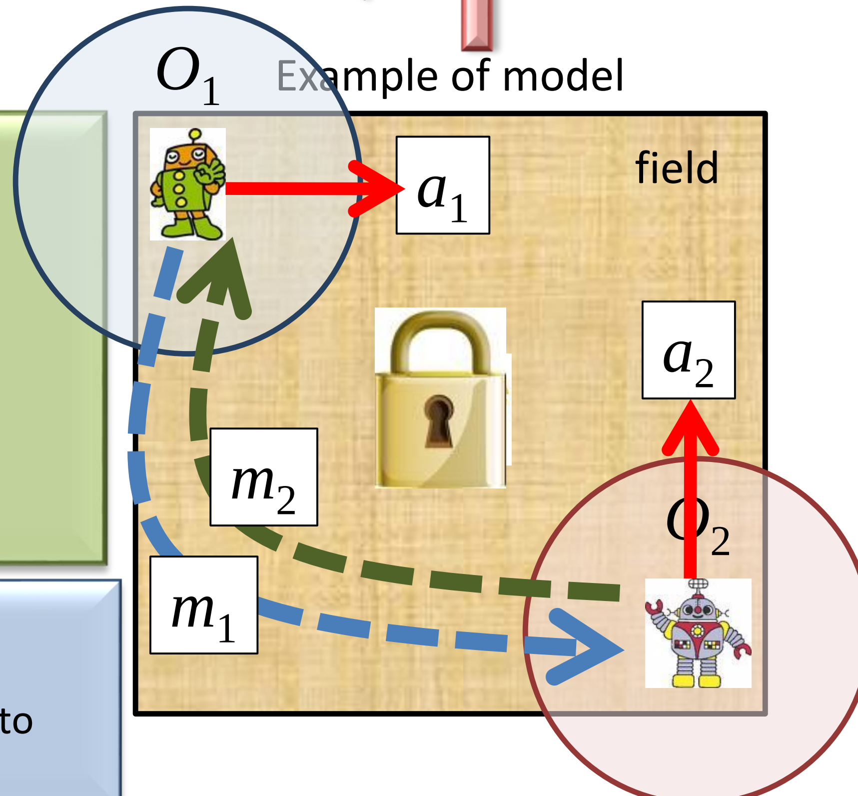
(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \Omega, A, M, C, P, O, R, T >$

Formulation

- A set of agents' indices
- e.g., $I = \{1, 2\}$

Example of model



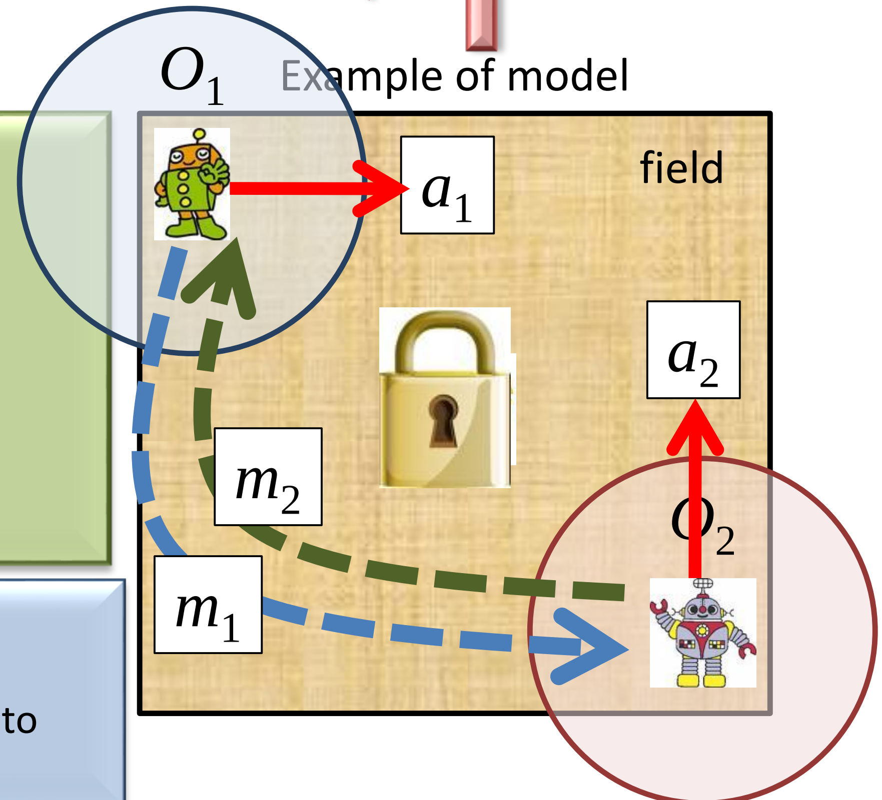
- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \Omega, A, M, C, P, O, R, T >$ ← Formulation

- A set of global states
- e.g., s = (position of agent 1,
  position of agent 2,
  position of treasure)
  , $s \in S$



- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

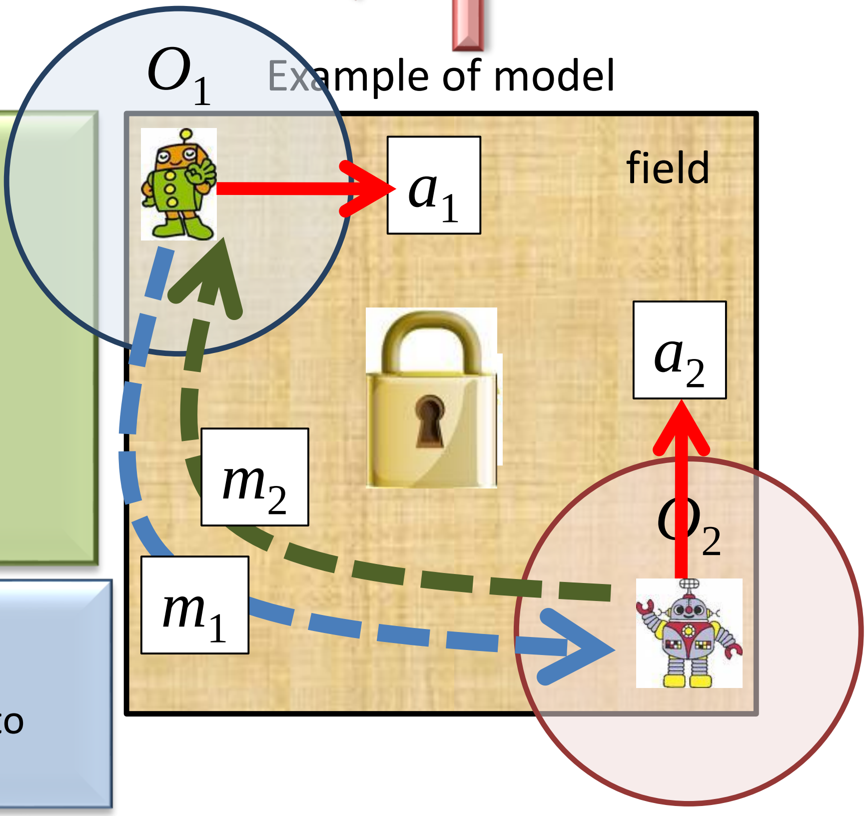
Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, M, C, P, O, R, T >$ ← Formulation

- $\boldsymbol{\Omega}$: a set of joint observations
- $\boldsymbol{\Omega} = \Omega_1 \times \Omega_2$,
  where Ω_i is a set of observations for agent i
- $\boldsymbol{A}$: a set of joint actions
- $\boldsymbol{A} = A_1 \times A_2$

Example of model



- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

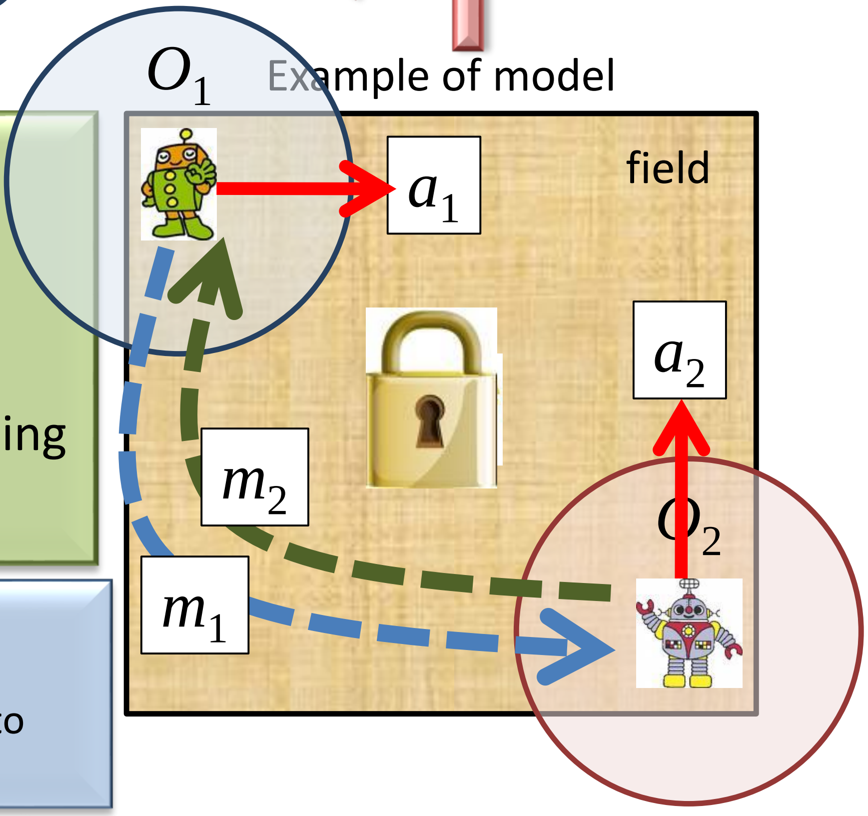
Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$ ⇐ Formulation

- $\boldsymbol{M}$: a set of joint messages
- $\boldsymbol{M} = M_1 \times M_2$
- $C : \boldsymbol{M} \rightarrow \mathfrak{R}$ is a cost function
- $C(\boldsymbol{m})$ represent the total cost of transmitting the messages sent by all agents.

Example of model



- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

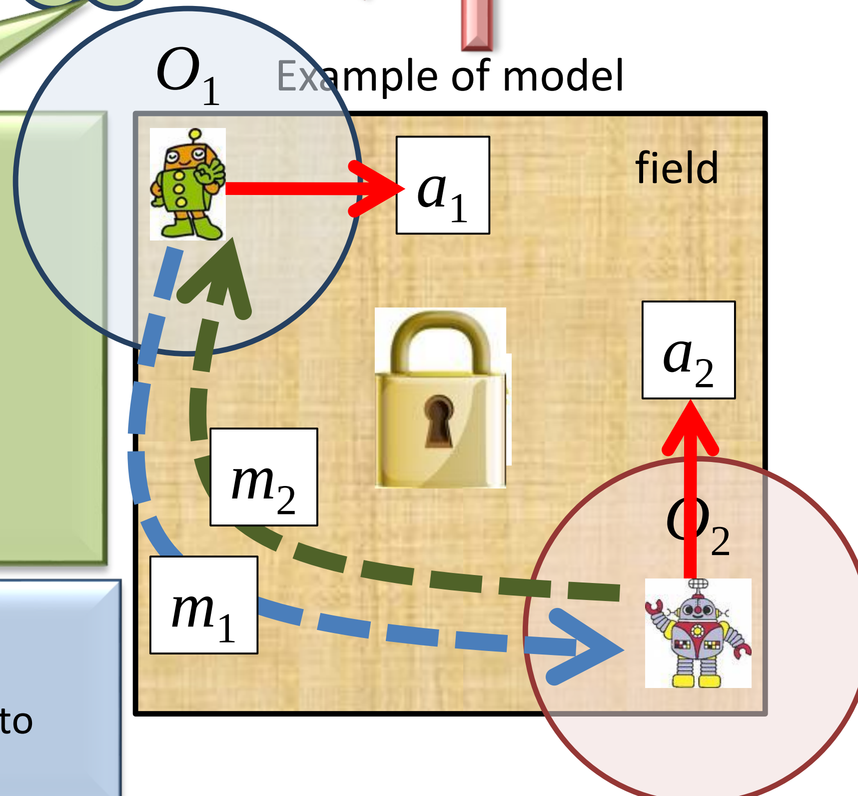
Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$

Formulation

- P : a transition probability function
- O : an observation probability function



- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

Dec-POMDP-Com [Goldman+ 04]

(Decentralized Partially Observable Markov Decision Process with Communication)

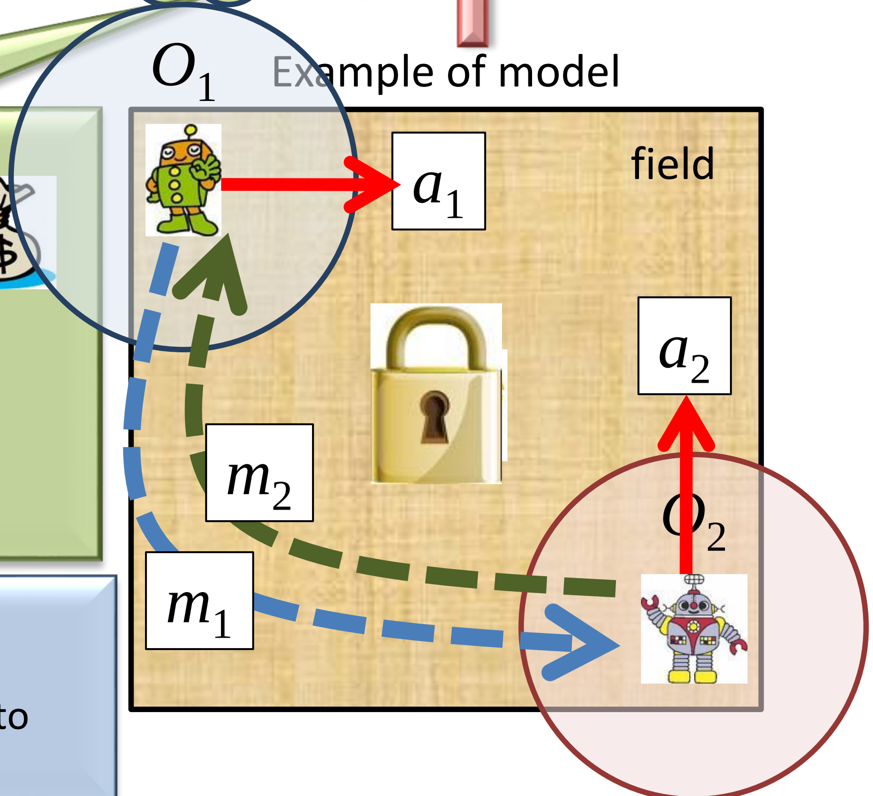
- A decentralized multi-agent system, where agents can communicate with each other and only observe the restricted information.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$

Formulation

- R : a reward function
- e.g., the treasure obtained by agents

- T : a time horizon

Example of model



- Two agents get a treasure cooperatively.
- The treasure is locked.
- Both agents must reach the treasure at the same time to open the lock.

Outline

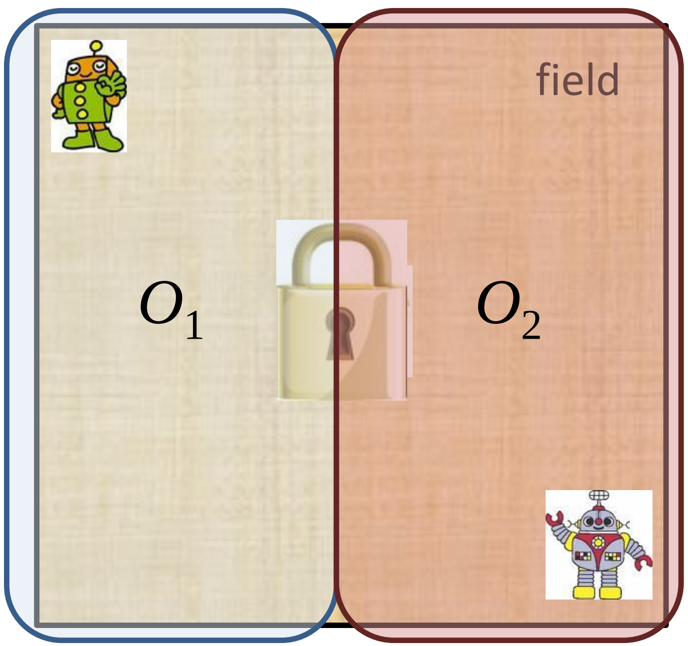
- ☑ Background
- ☑ Scheme of talk
- ☑ Review : Dec-POMDP-Com [Goldman+ 04]
- ☐ Constrained model
  - Jointly Fully Observable Dec-POMDP-Com [Goldman+ 04]
  - Deterministic Dec-POMDP-Com (we define)
- ☐ Theoretical analysis
- ☐ Conclusion

Jointly Fully Observable Dec-POMDP-Com

[Goldman+ 04]

- The Dec-POMDP-Com such that the combination of the agents' observations leads to the global state.

Jointly fully Observable Dec-POMDP-Com



$o_1 + o_2 =$ global state
(That is Jointly fully observable)

Deterministic Dec-POMDP-Com

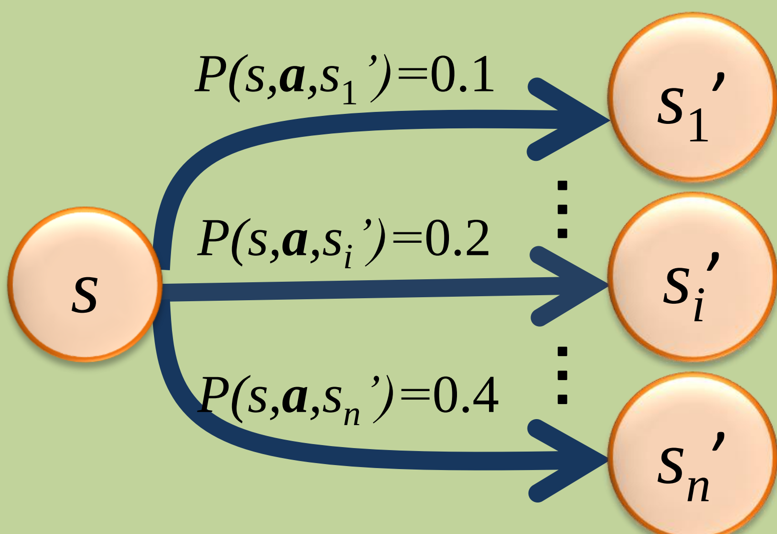
- The model where P and O on the definition are constrained.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$

Deterministic Dec-POMDP-Com

- The model where P and O on the definition are constrained.
- Dec-POMDP-Com := $\langle I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T \rangle$

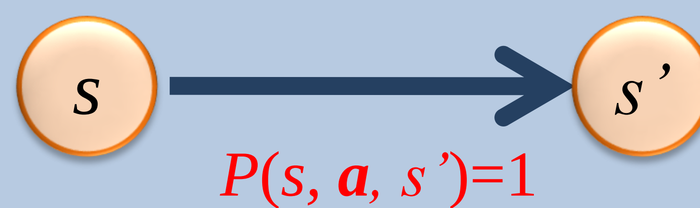
Restriction 1 : Deterministic transitions

- P is a transition probability function



$n = |S|$

- For any state $s \in S$ and any joint action $\boldsymbol{a} \in \boldsymbol{A}$, there exists a state $s' \in S$ such that $P(s, \boldsymbol{a}, s') = 1$.



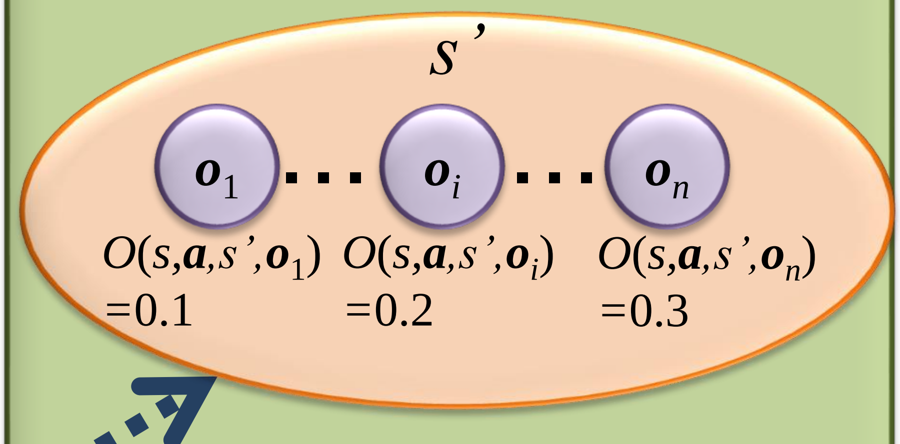
The next global state is decided uniquely.

Deterministic Dec-POMDP-Com

- The model where P and O on the definition are constrained.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$

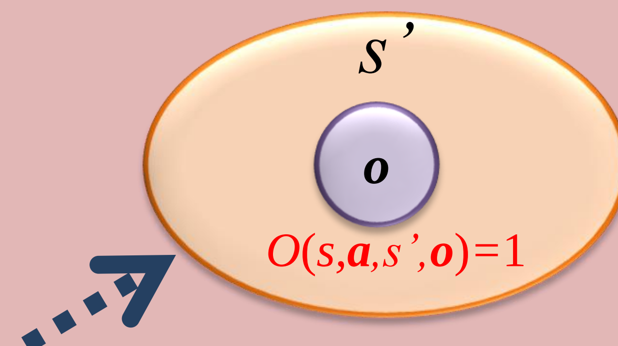
Restriction 2 : Deterministic observable

- O is a observation probability function



$n = |\boldsymbol{\Omega}|$

- For any state $s, s' \in S$ and any joint action $\boldsymbol{a} \in \boldsymbol{A}$, there exists a joint observation $\boldsymbol{o} \in \boldsymbol{\Omega}$ such that $O(s, \boldsymbol{a}, s', \boldsymbol{o}) = 1$.



The current observation is decided uniquely.

Deterministic Dec-POMDP-Com

- The model where P and O on the definition are constrained.
- Dec-POMDP-Com := $< I, S, \boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{M}, C, P, O, R, T >$

Restriction 1 : Deterministic transitions	The next global state is decided uniquely.
Restriction 2 : Deterministic observable	The current observation is decided uniquely.

- When Dec-POMDP-Com has Restriction 1 and 2, it is called Deterministic Dec-POMDP-Com

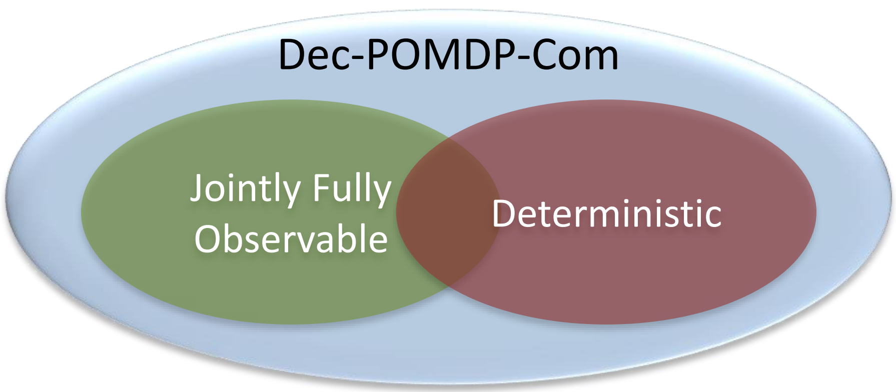
Outline

- ☑ Background
- ☑ Scheme of talk
- ☑ Review : Dec-POMDP-Com [Goldman+ 04]
- ☑ Constrained model
  - Jointly Fully Observable Dec-POMDP-Com [Goldman+ 04]
  - Deterministic Dec-POMDP-Com (we define)
- ☐ Theoretical analysis
- ☐ Conclusion

Main results

- Corollary 1 :
  Minimum required sizes $|M_i|$ for Signal Learning on *Jointly Fully Observable Dec-POMDP-Com*

- Theorem 2 :
  Minimum required sizes $|M_i|$ for Signal Learning with Messages on *Deterministic Dec-POMDP-Com*



Theorem 1 [Goldman 04]

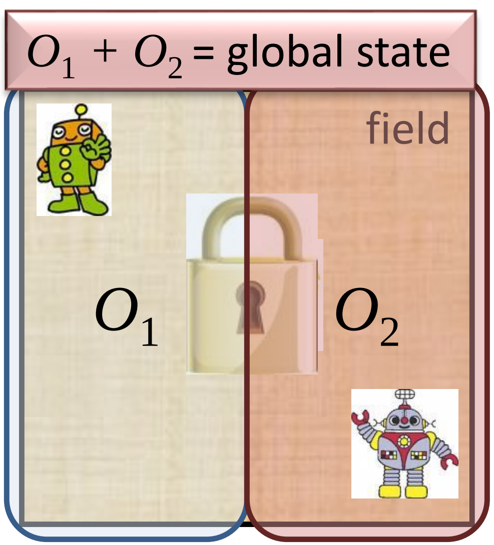
Theorem 1

For any jointly fully observable Dec-POMDP-Com, the following equation holds.

$$\forall \boldsymbol{M},\ \max_{\boldsymbol{\delta} \in \boldsymbol{D}} V^{T}_{\boldsymbol{\delta},\boldsymbol{M}}(s_0) \leq \max_{\boldsymbol{\delta} \in \boldsymbol{D}} V^{T}_{\boldsymbol{\delta},\boldsymbol{M}'}(s_0)$$

*the value of the optimal joint policy with respect to any joint message set **M***

*the value of the optimal joint policy with respect to the joint message set **M' := Ω**.*



This theorem means that the optimal communication policy of each agent is to send its own observation in jointly fully observable Dec-POMDP-Com (i.e. for agent i, $m_i := o_i$).

Each agent can always know the current global state by own observation and received message.
(e.g., for agent 1, $o_1 + m_2 = o_1 + o_2$ = global state)
Therefore, each agent always perform the optimal actions.

Corollary 1

Corollary 1

For any jointly fully observable Dec-POMDP-Com, if the size $|M_i|$ of the message set of each agent i satisfies the condition,

$$|M_i| \geqq |\Omega_i|$$

then the following equation holds:

$$\max_{\delta \in D^{SL}} V_{\delta}^{T}(s_0) = \max_{\delta' \in D} V_{\delta'}^{T}(s_0)$$

the value of the optimal joint policy on SL

the value of the optimal joint policy with history

From theorem 1 by Goldman,
the optimal communication policy of each agent is
to send its own observation in jointly fully observable Dec-POMDP-Com.

Therefore, If each agent has $|M_i|$ that is larger than $|\Omega_i|$,
it is possible to constructing the optimal policy such that each agent can send its own observation.

Theorem 2

Theorem 2

For any deterministic Dec-POMDP-Com, if the size $|M_i|$ of the message set of each agent i satisfies the condition,

$$|M_i| \geqq \max_{j \in I} \max_{o \in \Omega_j} |S_j^{obs}(o)|$$

then the following equation holds:

$$\max_{\boldsymbol{\delta} \in \boldsymbol{D}^{SLM}} V_{\boldsymbol{\delta}}^T(s_0) = \max_{\boldsymbol{\delta}' \in \boldsymbol{D}} V_{\boldsymbol{\delta}'}^T(s_0)$$

the value of the optimal joint policy on SLM

the value of the optimal joint policy with history

First, I explain a function S_j^{obs}.

S_j^{obs}

- Deterministic Dec-POMDP-Com has the following properties.

Deterministic transitions ⇒ The next global state is decided uniquely.

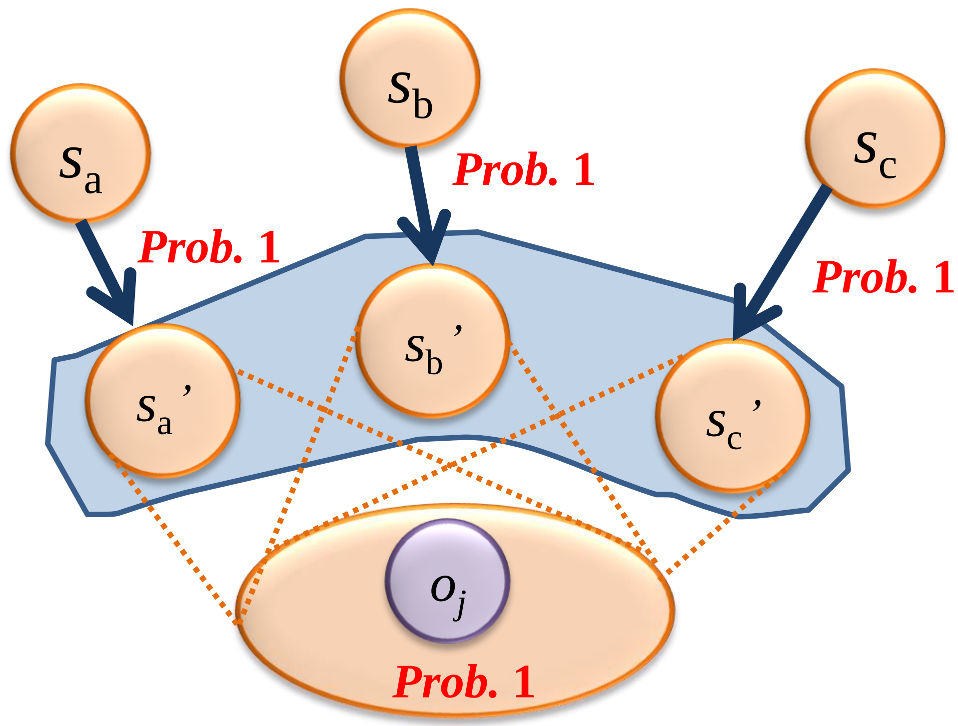
Deterministic observable ⇒ The observation is decided uniquely.

There exists some transitions such that agent *j* observes the same observation.

From the properties on deterministic Dec-POMDP-Com, we can compute the following function.

$$S_j^{obs}(o_j) = \{s_a', s_b', s_c'\}$$

S_j^{obs} returns the set of all states where agent *j* observes o_j.



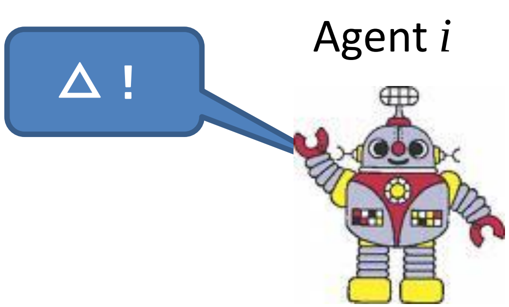
Proof sketch of Theorem2

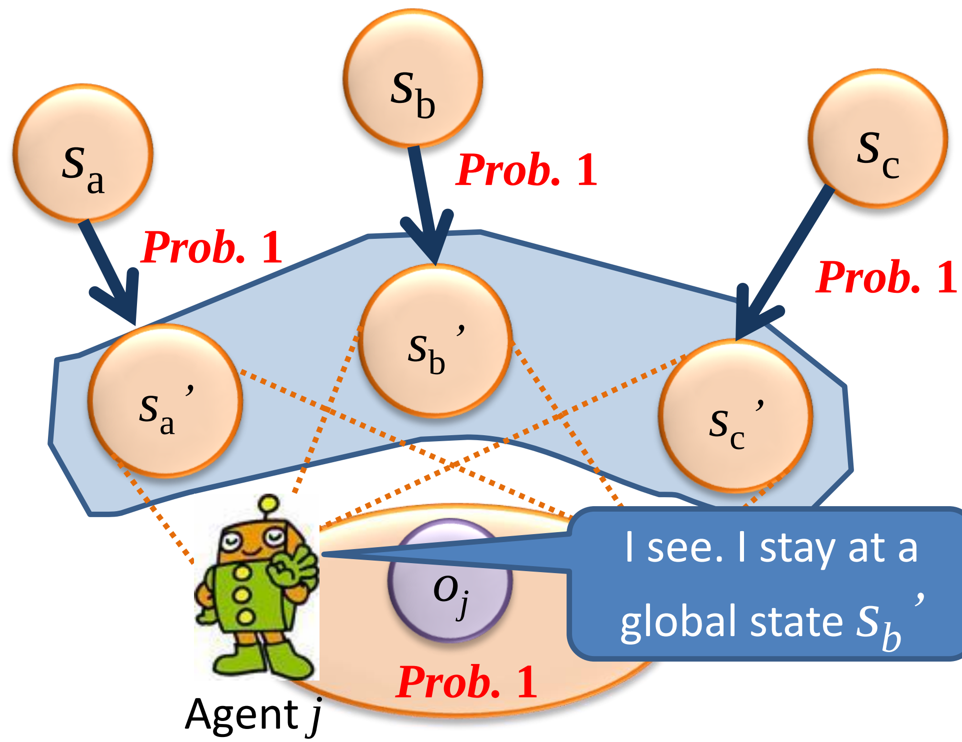
The condition of theorem 2 is $|M_i| \geqq \max_{j \in I} \max_{o \in \Omega_j} |S_j^{obs}(o)|$

The condition shows that
agent j can know the global state based on the message received from agent i by setting the set of message which have the maximum size of $S_j^{obs}(o_j)$.

$S_j^{obs}(o_j) = \{s_a', s_b', s_c'\}$

$M_i = \{□, △, ○\}$





Outline

- ☑ Background
- ☑ Scheme of talk
- ☑ Review : Dec-POMDP-Com [Goldman+ 04]
- ☑ Constrained model
  - Jointly Fully Observable Dec-POMDP-Com [Goldman+ 04]
  - Deterministic Dec-POMDP-Com (we define)
- ☑ Theoretical analysis
- ☐ Conclusion

Conclusion

- We defined deterministic Dec-POMDP-Com for theoretical analysis

Restriction 1 : Deterministic transitions	The next global state s' is decided uniquely.
Restriction 2 : Deterministic observable	The observation $\boldsymbol{o}$ is decided uniquely.

- We showed *Minimum required sizes* $|M_i|$ for
  - Signal Learning on *Jointly Fully Observable* *Dec-POMDP-Com* at corollary 1
  - Signal Learning with Messages on *Deterministic* *Dec-POMDP-Com* at Theorem 2