Identifying tractable decentralized control problems on the basis of information structure

Aditya Mahajan, Ashutosh Nayyar, and Demosthenis Teneketzis

Dept of EECS,

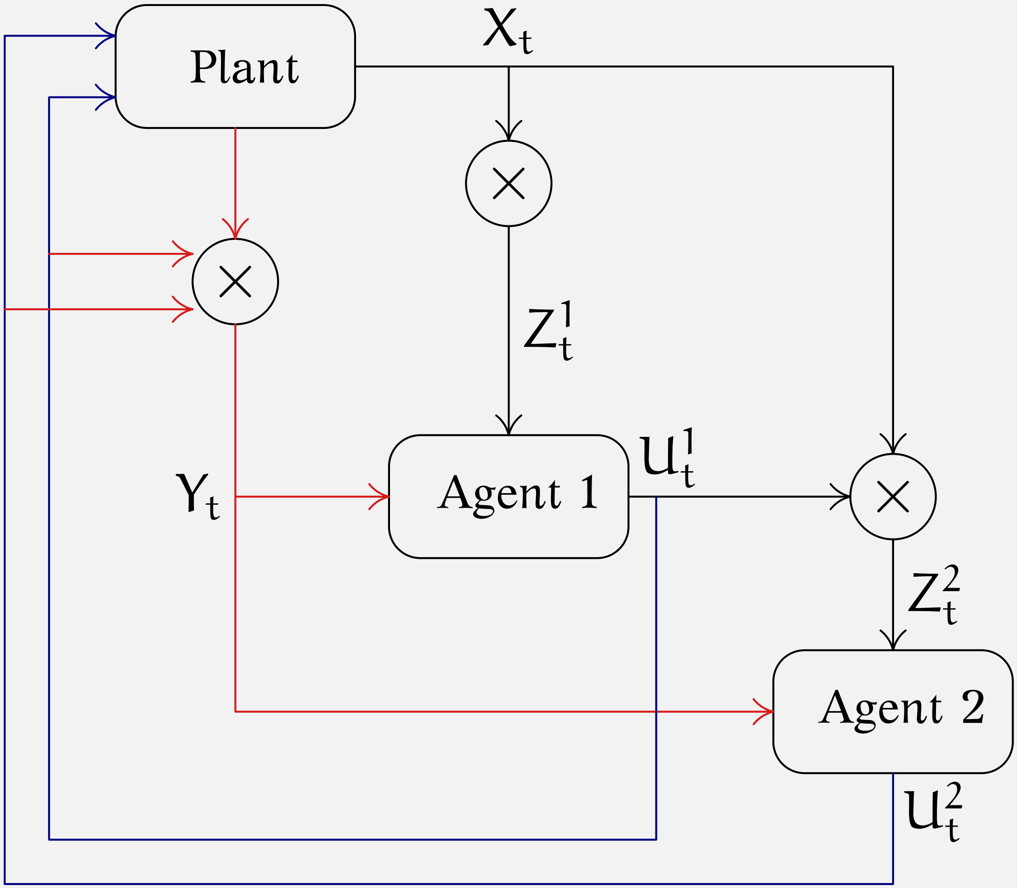
University of Michigan, Ann Arbor

September 26, 2008

Optimal design of decentralized systems with non-classical information structures

- **Difficulties**: Conceptual and computational
- **Results of this paper**: Consider two general models of decentralized systems and obtain a sequential decomposition for their finite and infinite horizon cases.
- Our models encompass
  - Standard form (Witsenhausen, 1973)
  - k-step delay sharing pattern (Walrand and Varaiya, 1978)
  - Generic team model of Witsenhausen (1988)
- **Main idea**: viewed appropriately, these models are equivalent to POMDPs with functions as control actions
- Numerical solution can be obtained using existing techiques for POMDPs

Model A for two agents



Model A for two agents

- **Plant**: $X_{t+1} = f_t(X_t, U_t^1, U_t^2, W_t)$

- **Observations**
  - Common message:

    $$Y_t = c_t(X_t, U_{t-1}^1, U_{t-1}^2, Q_t)$$

  - Private message: $Z_t^1 = h_t^1(X_t, N_t^1)$

    $$Z_t^2 = h_t^2(X_t, U_t^1, N_t^2)$$

- **Agent** k
  - Control: $U_t^k = g_t^k(Y^t, Z_t^k, M_{t-1}^k)$
  - Memory update: $M_t^k = l_t^k(Y^t, Z_t^k, M_{t-1}^k)$

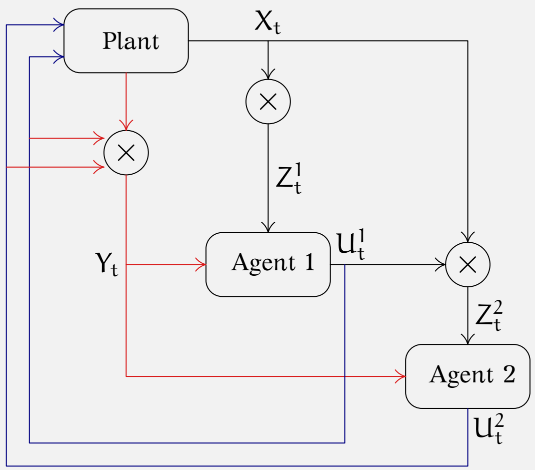
- **Design** $\equiv$ all control and memory update functions of both agents

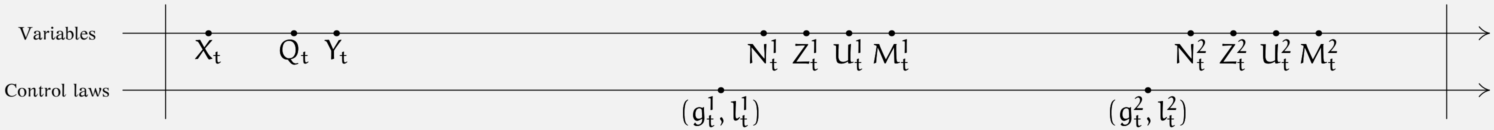
- **Cost at time** t: $\rho_t(X_t, U_t^1, U_t^2)$. **Cost of a design**: $E\left\{ \sum_{t=1}^{T} \rho_t(X_t, U_t^1, U_t^2) \,\middle|\, \text{Design} \right\}$

- **Objective**: Determine an optimal design

Model A for two agents

- Salient features
  - Non-classical information structures
  - Sequential system





Consider the model from the point of view of a fictitious common agent

Common Agent

Common agent observes all common messages

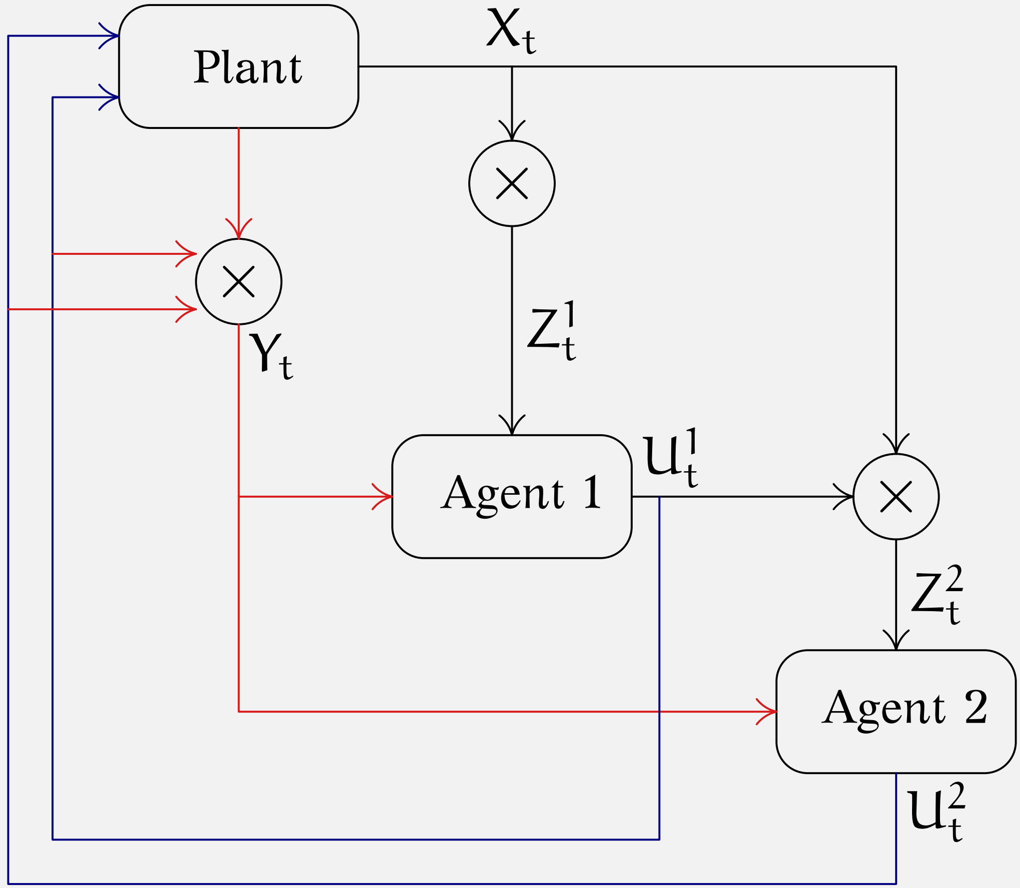
- Think of control and memory update functions in two steps

$$U_t^k = g_t^k(Y^t, Z_t^k, M_{t-1}^k)$$
$$= \hat{g}_t^k(Z_t^k, M_{t-1}^k), \quad \text{where } \hat{g}_t^k = \gamma_t^k(Y^t)$$

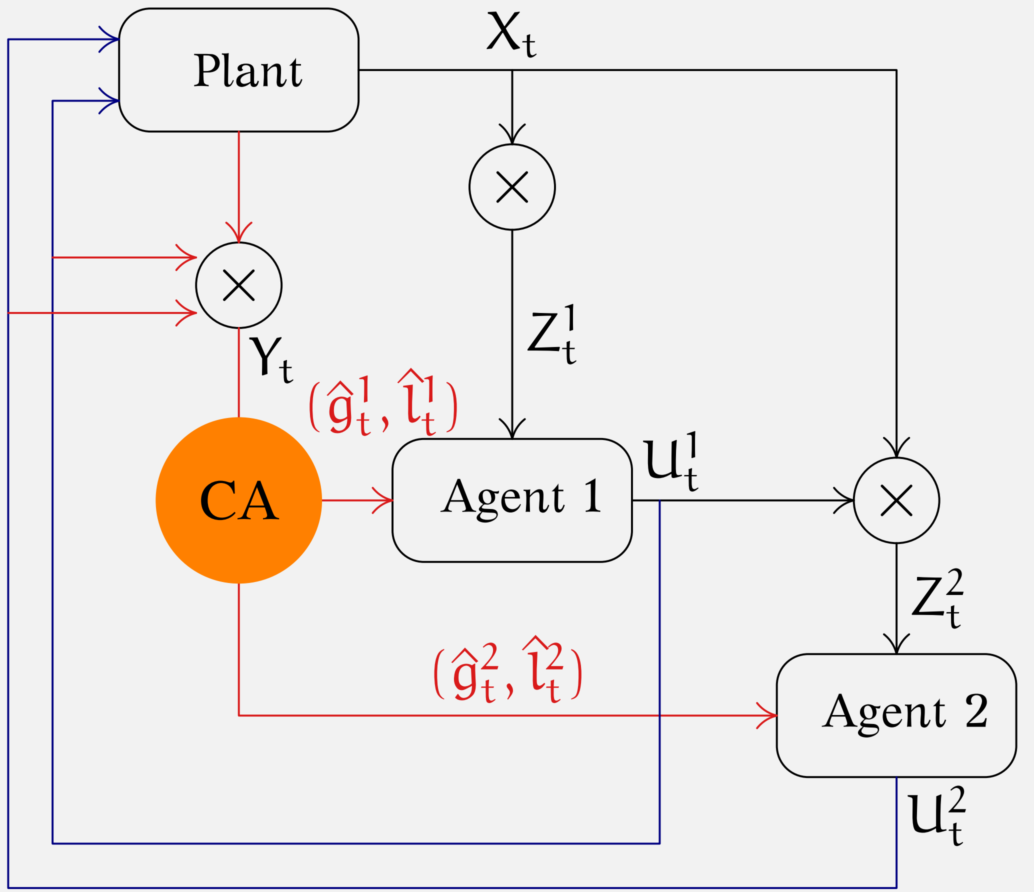
Similarly,

$$M_t^k = l_t^k(Y^t, Z_t^k, M_{t-1}^k)$$
$$= \hat{l}_t^k(Z_t^k, M_{t-1}^k), \quad \text{where } \hat{l}_t^k = \lambda_t^k(Y^t)$$

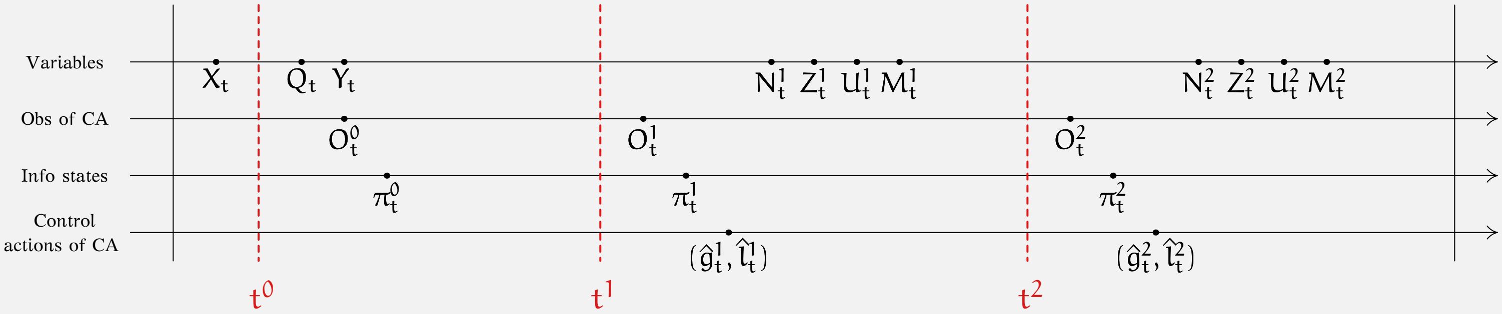
Common Agent's viewpoint



Common Agent's viewpoint



Common Agent's viewpoint



- Consider three time steps t^0, t^1, and t^2 in time interval t

$$S_t^0 = (X_t, M_{t-1}^1, M_{t-1}^2, U_{t-1}^1, U_{t-1}^2), \qquad O_t^0 = Y_t$$

$$S_t^1 = (X_t, M_{t-1}^1, M_{t-1}^2), \qquad O_t^1 = -$$

$$S_t^2 = (X_t, M_t^1, M_{t-1}^2, U_t^1), \qquad O_t^2 = -$$

- POMDP with: ▷ **State**: S_t^i, ▷ **Obs**: O_t^i, ▷ **Control actions**: $(\hat{g}_t^k, \hat{l}_t^k)$

From the common agent's viewpoint $\{S_t^0, S_t^1, S_t^2, t = 1, \ldots, T\}$ is a POMDP (partially observable Markov decision process)

Sequential decomposition

- Information states

$$\pi_t^0 = \Pr\left(S_t^0 \,\middle|\, Y^t, \hat{g}^{1,t-1}, \hat{l}^{1,t-1}, \hat{g}^{2,t-1}, \hat{l}^{t-1}\right)$$

$$\pi_t^1 = \Pr\left(S_t^1 \,\middle|\, Y^t, \hat{g}^{1,t-1}, \hat{l}^{1,t-1}, \hat{g}^{2,t-1}, \hat{l}^{t-1}\right)$$

$$\pi_t^2 = \Pr\left(S_t^2 \,\middle|\, Y^t, \hat{g}^{1,t}, \hat{l}^{1,t}, \hat{g}^{2,t-1}, \hat{l}^{t-1}\right)$$

- Optimality equations

$$V_{T+1}^0(\pi_{T+1}^0) \equiv 0,$$

for $t = 1, \ldots, T$

$$V_t^0(\pi_t^0) = E\left\{V_t^1(\pi_t^1) \,\middle|\, \pi_t^0\right\},$$

$$V_t^1(\pi_t^1) = \min_{\theta_t^1}\left\{ E\left\{V_t^2(\pi_t^2) \,\middle|\, \pi_t^1, \theta_t^1\right\} \right\},$$

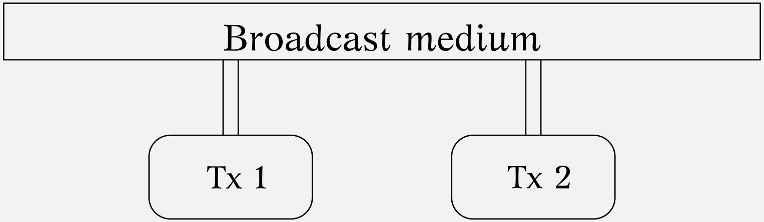
$$V_t^2(\pi_t^2) = \min_{\theta_t^2}\left\{ E\left\{\rho_t(X_t, U_t^1, U_t^2) + V_{t+1}^0(\pi_{t+1}^0) \,\middle|\, \pi_t^2, \theta_t^2\right\} \right\},$$

where $\theta_t^k = (\hat{g}_t^k, \hat{l}_t^k)$

Models considered in the paper

- Model A
  - n-agent version of what was presented here
- Model B
  - Model A with no common messages
- Also consider infinite horizon problems

Example – multiaccess broadcast



- MAB Channel
  - Single user transmits $\Longrightarrow$ successful transmission
  - Both users transmit $\Longrightarrow$ packet collision
- Transmitters
  - Queues with buffer size 1
  - Packet held in queue until successful transmission
  - Packet arrival is independent Bernoulli process

Example – multiaccess broadcast

- Channel feedback

  Both transmitters know if there was no transmission, successful transmission, or a collision

- Policy of transmitters

  If packet is available, decide whether or not to transmit based on all past channel feedback

- Objective: Maximize throughput
  - Avoid collisions
  - Avoid idle

History of multiaccess broadcast

- Hluchyj and Gallager,

  "Multiaccess of a slotted channel by finitely many users", NTC 81.

  - Considered symmetric arrival rates
  - Restricted attention to "window protocols"

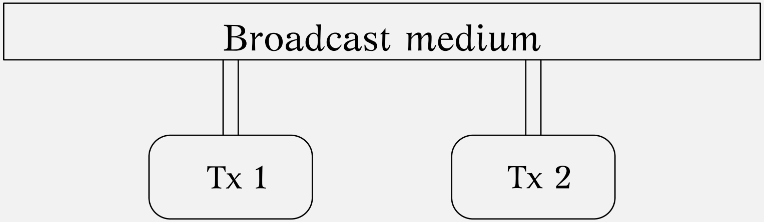
- Ooi and Wornell,

  "Decentralized control of multiple access broadcast channels", CDC 96.

  - Considered a relaxation of the problem
  - Numerically find optimal performance of the relaxed problem
  - Hluchyj and Gallager's scheme meets this upper bound

- AI Literature
  - Consider the case of asymmetric arrival rates
  - Approximate heuristic solutions for small horizons

Multi-access broadcast is equivalent to Model A



Tx 1	$\equiv$	Agent 1
Tx 2	$\equiv$	Agent 2
Channel feedback	$\equiv$	Common message
Number of packets in each buffer	$\equiv$	Private messages

- **Information state**: $\pi_t = \Pr\left(Z_t^1, Z_t^2 \,\middle|\, \text{feedback}\right), \quad Z_t^k = \{0, 1\}$
- **Action Space**: $\hat{g}_t^k : \{0, 1\} \to \{\text{Tx, Don't Tx}\}$

Equivalent to a POMDP with finite state and action spaces

Tractability

- Finite horizon problem
  - All system variables are **finite valued**
- Infinite horizon
  - All system variables take values in a **time-invariant** space
  - The system is **time-homogeneous**

Conclusions

- Sequential decomposition of two general models of decentralized systems
- Equivalent to POMDPs (sometimes to POMDPs with finite state and action spaces)
- Harder to solve than POMDPs due to expansion of state and action spaces.

Thank you