Inducing Desirable Behaviour through an Incentives Infrastructure

Roberto Centeno Holger Billhardt Sascha Ossowski

Centre for Intelligent Information Technologies (CETINIA)
University Rey Juan Carlos
Madrid, Spain

roberto.centeno@urjc.es

MATES'10 - September, 27th 2010

Outline

Outline

Introduction I

Open MAS

- designed with a general purpose in mind
- agents may join/leave the system
- at design time the population might be unknown
- agents could be heterogeneous, self-interested, built by a third party, etc.

PROBLEM

will agents behave according to the preferences of the system?

Introduction II

Solution 1: Organisational Structures

- *Soft Norms*
  - detection mechanisms + penalties/rewards
  - designed before knowing the population
  - Problem: what happens if the current population is not sensitive to these penalties/rewards?
- *Hard Norms*
  - agents cannot violate these norms
  - mechanisms to avoid such violations
  - Problem: in some domains it is almost impossible to take into account all possible exceptions due to their complexity and size

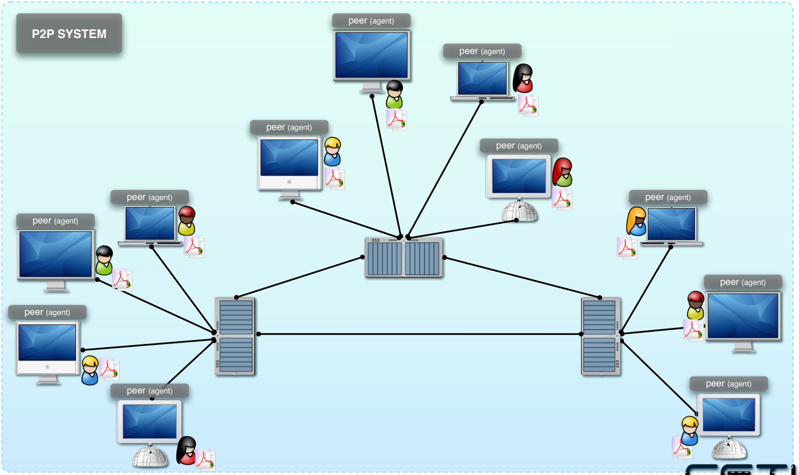
CETINIA

Introduction III

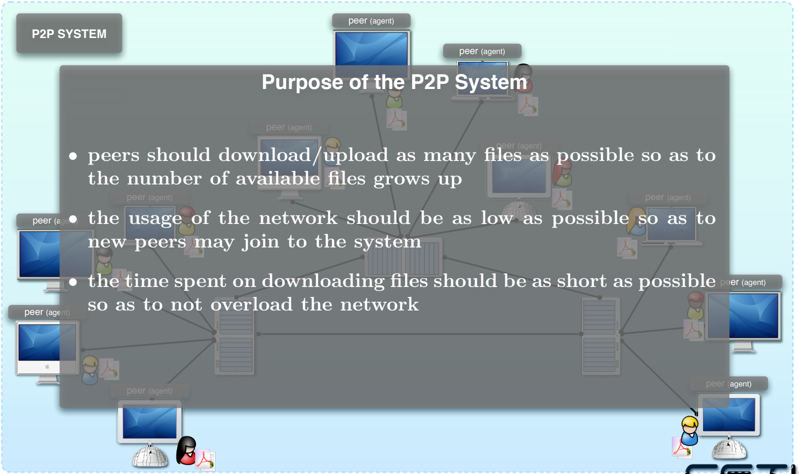
Solution 2: Incentives Infrastructure

- based on *"Economic Analysis of Law"* (R.A. Posner, 1977)
- to analyse and check how normative systems avoid the waste of resources and increase the efficiency
- it assumes rational agents, so they will violate norms if that action maximises their preferences
- agents have no reason to obey the law without sanctions/rewards
- it focuses on the effects of norms on outcomes (the effects of the norm on the behaviour of individuals)
- since this theory we propose a personalised Incentives System

Motivation: An Example



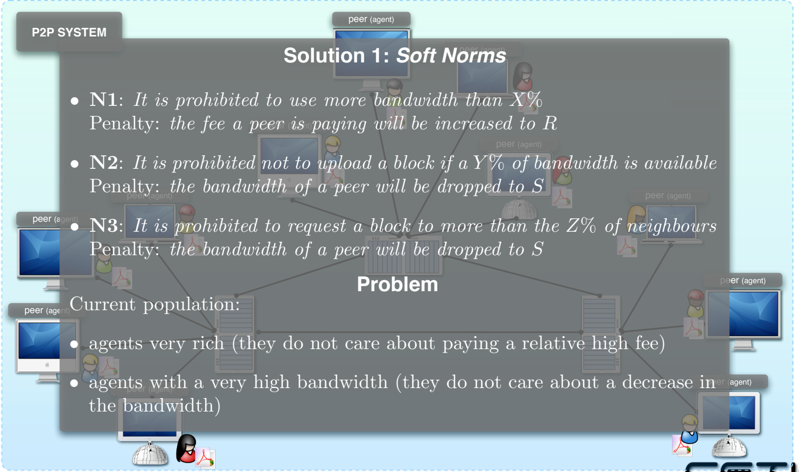
Motivation: An Example



Purpose of the P2P System

- peers should download/upload as many files as possible so as to the number of available files grows up
- the usage of the network should be as low as possible so as to new peers may join to the system
- the time spent on downloading files should be as short as possible so as to not overload the network

Motivation: An Example



Solution 1: *Soft Norms*

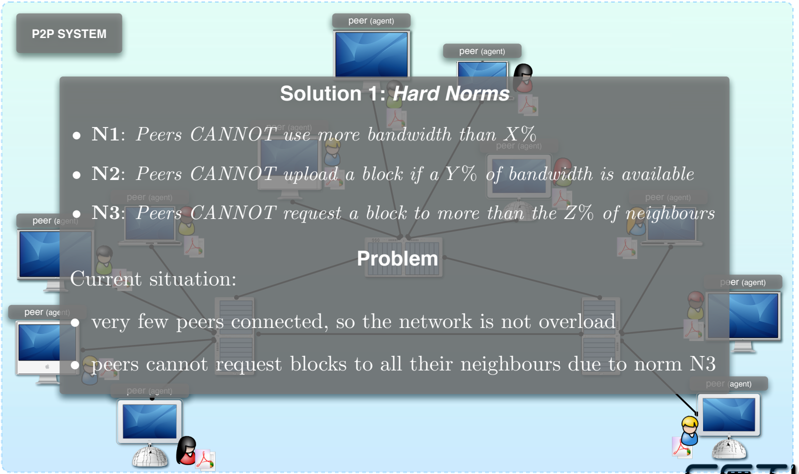
- **N1**: *It is prohibited to use more bandwidth than $X\%$*
  Penalty: *the fee a peer is paying will be increased to R*
- **N2**: *It is prohibited not to upload a block if a $Y\%$ of bandwidth is available*
  Penalty: *the bandwidth of a peer will be dropped to S*
- **N3**: *It is prohibited to request a block to more than the $Z\%$ of neighbours*
  Penalty: *the bandwidth of a peer will be dropped to S*

Problem

Current population:

- agents very rich (they do not care about paying a relative high fee)
- agents with a very high bandwidth (they do not care about a decrease in the bandwidth)

Motivation: An Example



Solution 1: *Hard Norms*

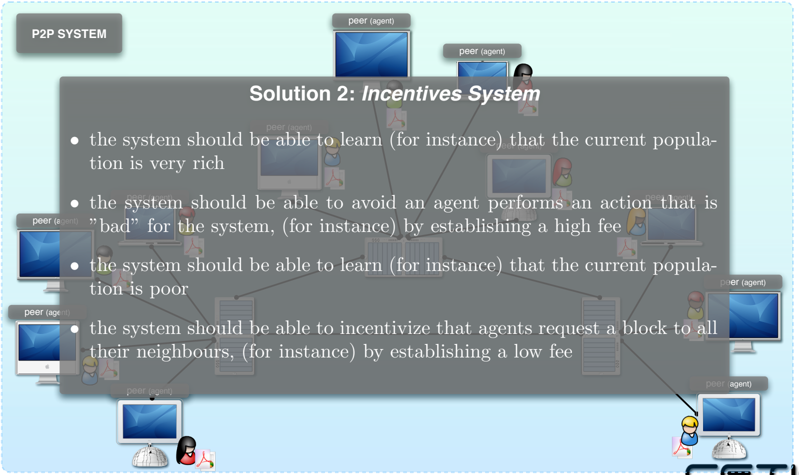
- **N1**: *Peers CANNOT use more bandwidth than X%*
- **N2**: *Peers CANNOT upload a block if a Y% of bandwidth is available*
- **N3**: *Peers CANNOT request a block to more than the Z% of neighbours*

Problem

Current situation:

- very few peers connected, so the network is not overload
- peers cannot request blocks to all their neighbours due to norm N3

Motivation: An Example



Solution 2: *Incentives System*

- the system should be able to learn (for instance) that the current population is very rich
- the system should be able to avoid an agent performs an action that is "bad" for the system, (for instance) by establishing a high fee
- the system should be able to learn (for instance) that the current population is poor
- the system should be able to incentivize that agents request a block to all their neighbours, (for instance) by establishing a low fee

Outline

Previous Work

- R.Centeno; H.Billhard; R.Hermoso; S.Ossowski, SAC'09: *"Organising MAS: A Formal Model Based on Organisational Mechanisms"*
- general formal framework for organising multiagent systems whose participants are rational
- light weight organisational model based on the idea of *organisational mechanisms*

Organisational Mechanisms I

How can we influence agents' behaviour if they are autonomous and independent?

manipulating the parameters which influence in their making decision process

$$t(s) = argmax_{a \in \mathcal{A}/\phi(a)=1} \sum_{s' \in \mathcal{S}} \mathcal{U}(s') \cdot \overline{P}(s'|s,a)$$

Organisational Mechanisms I

How can we influence agents' behaviour if they are autonomous and independent?

manipulating the parameters which influence in their making decision process

$$t(s) = argmax_{a \in \mathcal{A}/\phi(a)=1} \sum_{s' \in \mathcal{S}} \mathcal{U}(s') \cdot \overline{P}(s'|s,a)$$

changing agents' capability function
COERCIVE MECHANISM

Organisational Mechanisms I

How can we influence agents' behaviour if they are autonomous and independent?

manipulating the parameters which influence in their making decision process

$$t(s) = argmax_{a \in \mathcal{A}/\phi(a)=1} \sum_{s' \in \mathcal{S}} \mathcal{U}(s') \cdot \overline{P}(s'|s,a)$$

changing agents' capability function
COERCIVE MECHANISM

changing transition probability distribution
INCENTIVE MECHANISM

CETINIA

Organisational Mechanisms I

How can we influence agents' behaviour if they are autonomous and independent?

manipulating the parameters which influence in their making decision process

$$t(s) = argmax_{a \in \mathcal{A}/\phi(a)=1} \sum_{s' \in \mathcal{S}} \mathcal{U}(s') \cdot \overline{P}(s'|s,a)$$

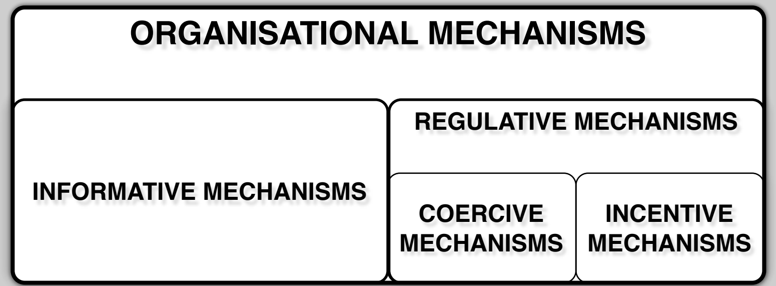
changing agents' capability function
COERCIVE MECHANISM

changing transition probability distribution
INCENTIVE MECHANISM

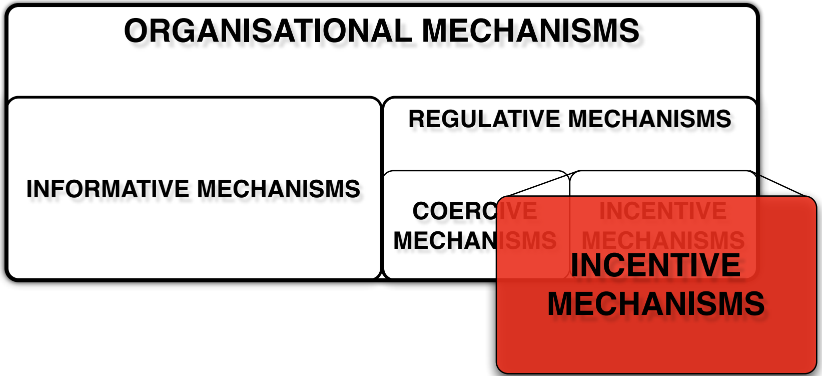
providing information to better estimation
INFORMATIVE MECHANISM

Organisational Mechanisms II

two kinds of organisational mechanisms



Organisational Mechanisms II



Outline

Incentive Organisational Mechanism

Definition

Let *MAS* be a multiagent system $MAS = \langle \mathcal{A}g, \mathcal{A}, \mathcal{X}, \Phi, \varphi, \mathcal{U}, x_0 \rangle$

- An incentive mechanism, Υ_{inc}, is a function that given a possibly partial description of an environmental state of *MAS* produces changes in the transition probability distribution of *MAS*

$$\Upsilon_{inc} : \mathcal{X}' \to \Phi$$
$$\Upsilon_{inc} : \mathcal{X}' \to [\mathcal{X} \times \mathcal{A}^{|\mathcal{A}g|} \times \mathcal{X} \to [0..1]]$$

  - $|\mathcal{A}g|$ is the number of agents;
  - $\mathcal{A}$ is the action space;
  - $\mathcal{X}$ is the environmental state space;
  - Φ is the MAS transition probability distribution;
  - φ is the agents' capability function (*physical* restrictions);
  - $\mathcal{U}$ is the global utility function;
  - $x_0 \in \mathcal{X}$ stands for the initial state;
  - $\mathcal{X}'$ is the set of partial descriptions of environmental states.

Assumptions

A1. the action space is finite

A2. agents are utility maximizers. (utility functions capture the utility at a long term)

A3. an environmental state $x_i \in \mathcal{X}$ can be modelled as a set of tuples $x_i = \langle attribute, value \rangle$

A4. the utility of an environmental state is the output of a multi-attribute utility function

A5. the attributes are additively independent $\mathcal{U}(x_i) = \sum_{j=1}^{n} w_j \cdot u_{i,j}$

A6. all participants in the system share the same ontology

The Problem

designing an incentive mechanism requires

- to learn which attributes should be modified, so as to make the consequences of a particular action more or less attractive for an agent
- to estimate agents' preferences
- to decide how the consequences of an action should be changed in order to incentivize it.

Outline

Incentives Infrastructure

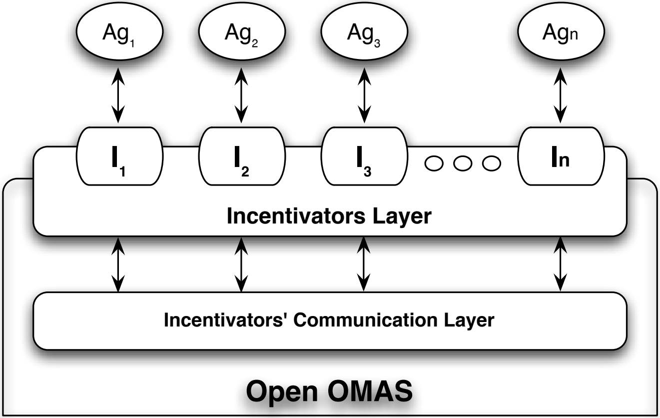
- Objective
  - to discover agents' preferences
  - to select the appropriate incentive (modification of the consequences of an action)
- Idea
  - similar to EIs (AMELI) with institutional agents
  - governors mediate external agents interactions in the institution

Incentives Infrastructure

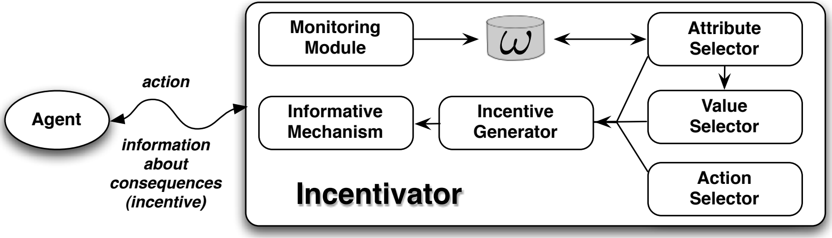
- Objective
  - to discover agents' preferences
  - to select the appropriate incentive (modification of the consequences of an action)
- Idea
  - similar to EIs (AMELI) with institutional agents
  - governors mediate external agents interactions in the institution

INCENTIVATORS

Incentives Infrastructure: Architecture



Incentivator Architecture



Discovering Agents' Preferences

learning an agent's preferences from monitoring its behaviour in response to given incentives

- Objective: to estimate which attributes (and their values) affects to the agent's decision
- Modules: Attribute Selector (selects the attribute to modify) and Value Selector (selects the value of such an attribute)
- How: Q-learning with immediate rewards and ϵ-greedy action selection
  - attribute selector action space: $Z_i \subseteq \{X_1, \ldots, X_n\}$
  - value selector action space: $Y_i = \{value_j \in [value_{X_i}^{min}, value_{X_i}^{max}]\}$
  - update action-value f.: $\mathcal{Q}_{t+1}(z_j) = \mathcal{Q}_t(z_j) + \alpha \cdot [\mathcal{R}_t(z_j) - \mathcal{Q}_t(z_j)]$
  - reward: $\mathcal{R}_t(z_j) = \begin{cases} +1 & \text{if agent performed the action} \\ -1 & \text{i.o.c.} \end{cases}$
- Output: $x_{i,j}^{*} = \langle attribute, value \rangle \mid attribute = z_i \wedge value = y_i$

Selecting the Action to Incentivize

the incentivator wants the agent to perform the action that would lead to the state with the best utility for the system

- Objective: trying to induce the action that gives the highest utility for the system
- Modules: Action Selector
- How:
    - to estimate the result of each possible action the agent is able to perform (domain-dependent)
    - to calculate the utility of the system in each resulting state
    - to rank the actions by the expected utility of the system
- Output: $\bigtriangledown_{x_j}^{ag_i} = \langle a_1, \ldots, a_n \rangle$

Testing the Proposed Incentive

it could be necessary to assure that such an incentive is not damaging the objective of the system

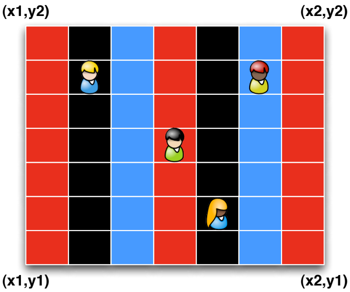
- Objective: to evaluate whether or not the new consequences of the action are still the *best option* for the system
- Modules: Incentive Generator
- How: to find an action such that..
  - if the agent performs it with the new consequences the expected utility of the system is greater or equal than if the agent performs the same action without the new consequences
  - the expected utility of the system with the new consequences is greater or equal than if the agent performs the following best action, w.r.t. the utility of the system
- Output: the best action to incentivize by using the incentive proposed or no action (not apply the incentive)

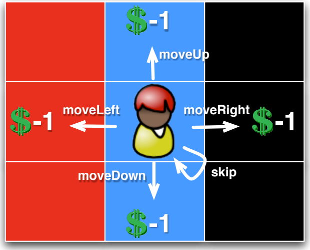
Monitoring and Informing the Agent

- Monitoring Module:
  - monitories the action actually performed by the agent
  - informs the attribute and value selectors (for updating q-values)
  - no way to distinguish if an agent performs an action because of its own interest instead of the incentive (exploration/exploitation process in the Q-learning algorithm will detect such situations)
- Informative Mechanism:
  - informs about the new consequences (attribute modification) of the selected action allowing agents to reason about it
  - agents can query the mechanism by providing an action

Outline

A “toy” Example





- $\mathcal{A} = \{moveRight, moveLeft, moveUp, moveDown, skip\}$
- $\mathcal{X} = \{agent_1 Position, \ldots, agent_1 Money, \ldots, gridSize, squareColour_{1,1}, \ldots, systemMoney\}$
- deterministic environment

A “toy” Example: the System

Objective of the System

1. *the system wants agents to be as close as possible to the middle point*
2. *the system wants to get as much money as possible*

$$\mathcal{U}(x_j) = \mathcal{U}_{systemMoney}(x_j) \cdot w_0 + \sum_{k=1}^{|\mathcal{A}g|} \mathcal{U}_{agent_k Position}(x_j) \cdot w_k$$

- $\mathcal{U}_{agent_k Position}(x_j)$: how far they are from the central point (Manhattan distance)
- $\mathcal{U}_{systemMoney}(x_j)$: the more money the system gets, the more utility it obtains

A “toy” Example: the Agents

Objective of an Agent

1. *to reach a corner*
2. *to remain in a particular colour*
3. *to save as much money as possible*

$$\mathcal{U}_{a_k}(x_j) = \mathcal{U}_{agent_k Position}(x_j) \cdot w_1 + \mathcal{U}_{squaresColours_{agent_k Position}}(x_j) \cdot w_2 + \mathcal{U}_{agent_k Money}(x_j) \cdot w_3$$

- $\mathcal{U}_{agent_k Position}(x_j)$: how far they are from their corner (Manhattan distance)
- $\mathcal{U}_{squaresColours_{agent_k Position}}(x_j)$: 1 when they are on their preferred colour (0 i.o.c.)
- $\mathcal{U}_{agent_k Money}(x_j)$: the more money the agent gets, the more utility it obtains

Regulating the System..

Normative System

- **Norm**

  "it is prohibited to go beyond an established area from the central point"

- **Punishment/Reward**

$$consequences = \{\omega^*\} = \{agent_i Money^*\}$$

- **Detection Mechanism**

  infallible mechanism: *100% times detects when an agent crosses the area*

- **Informative Mechanism**

  informs about the norms and their punishment/reward

Regulating the System..

Incentives System (*Incentivators*)

- **Action Selector**
  able to simulate the result of an action (deterministic environment)
- **Attribute Selector**
  action space: $\mathcal{Z}_i = \{agent_i Money, square_{agent_i Position} Colour\}$
- **Value Selector**
  action space: $\mathcal{Y}_i = \{value^{min}_{agent_i Money} \ldots value^{max}_{agent_i Money}\}$
  action space: $\mathcal{Y}_j = \{black, red, blue\}$
- **Informative Mechanism**
  informs about the consequences (new values of the attributes $agent_i Money$ or $square_{i,j} Colour$) of the selected action

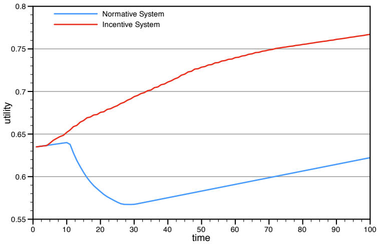
CETINIA

System Setup

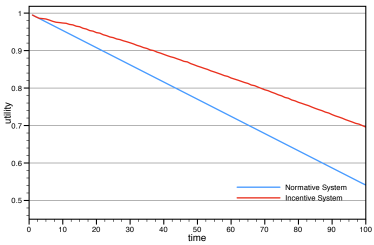
	Exp1	**Exp2**
Grid/Agents/Steps	$200/30/100$	$200/30/100$
$\mathcal{U}_{Ag}$	$Random(w_1, w_2, w_3)$	$w_1, w_2 = 0.45 \; w_3 = 0.1$
$\mathcal{U}_{MAS}$	$Random(w_0, w_k)$	$Random(w_0, w_k)$
Norm Limit/Penalty	$10/-50$	$10/-50$
$agent_i Money_0 / agent_i Money^*$	$Random(1.000)/\pm 5\%$	$Random(1.000)/\pm 5\%$
$agent_i Position_0$/corner to reach	$(100, 100)/Random(4)$	$(100, 100)/Random(4)$

CETINIA

Results I

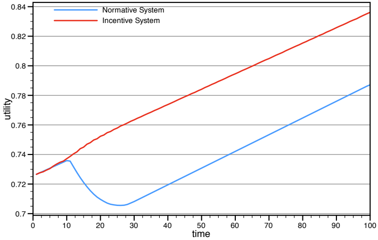


(a) Agents' Utility: Exp1

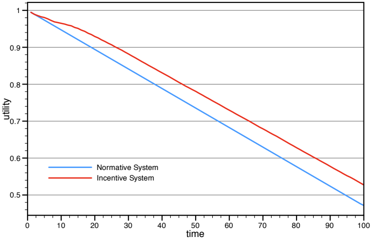


(b) System's Utility: Exp1

Results II



(c) Agents' Utility: Exp2



(d) System's Utility: Exp2

Outline

Conclusions I

1. to regulate an open MAS by using an incentive system
2. incentives infrastructure is able to:
   - discover agents' preferences (modifications in the environment that affect agents)
   - provide the suitable incentive by modifying the consequences of an action
3. incentives infrastructure architecture
   - institutional agents: *incentivators*
   - an incentivator is in charge of an external agent
   - to learn the best incentivation policy by using learning techniques
4. it seems to be a promising mechanism for:
   - micro level: agents
   - macro level: system

Conclusions II

❌ **Incentives Vs Norms** ✅ **Incentives + Norms**

Future Work

- current work: incentivizing just an agent (actions performed by an agent do not influence other agents)
- future work: agents do influence each other (multiagent learning techniques to coordinate incentivators)
- to apply the approach in real world domains: P2P systems

THAT'S ALL

Thank you for your attention!!